

Ruth Henquin

EPIDEMIOLOGÍA y ESTADÍSTICA

para principiantes



**EPIDEMIOLOGÍA Y ESTADÍSTICA
PARA PRINCIPIANTES**

Ruth P. Henquin

**EPIDEMIOLOGÍA Y ESTADÍSTICA
PARA PRINCIPIANTES**



La presente es una publicación de:



www.corpuslibros.com

Henquin, Ruth P.

Epidemiología y estadística para principiantes. - 1a ed. - Ciudad Autónoma de Buenos Aires: Corpus Libros Médicos y Científicos, 2013.

264 p.; 24x16 cm.

ISBN 978-987-1860-16-6

1. Epidemiología. 2. Estadísticas. I. Título
CDD 614.4

DERECHOS RESERVADOS

©2013 Corpus Editorial y Distribuidora

editorial@corpuslibros.com

www.corpuslibros.com

Tucumán 2180 - Tel/Fax: (011) 4371-8635 / 4373-5128
(C1050AAR) Ciudad Autónoma de Buenos Aires - Argentina

Editor: CORPUS Editorial y Distribuidora S.A.

Tirada: 1000 ejemplares

Se terminó de imprimir en Abril de 2013

Buenos Aires - Argentina

ISBN: 978-987-1860-16-6



No está permitida la reproducción total o parcial de esta obra, ni su tratamiento o transmisión por cualquier medio o método, sin autorización escrita de la Editorial.

NOTA

La Medicina es una ciencia en constante desarrollo. Conforme surjan nuevos conocimientos, se requerirán cambios de la terapéutica. El autor y los editores se han esforzado para que los cuadros de dosificación medicamentosa sean precisos y acordes con los establecidos en la fecha de publicación. Sin embargo, ante los posibles errores humanos y cambios en la medicina, ni los editores, ni cualquier otra persona que haya participado en la preparación de la obra garantizan que la información contenida en ella sea precisa o completa. Convendría recurrir a otras fuentes de datos, por ejemplo, y de manera particular, habrá que consultar la hoja de información que se adjunta con cada medicamento, para tener certeza de que la información de esta obra es precisa y no se han introducido cambios en la dosis recomendada o en las contraindicaciones para su administración. Esto es de particular importancia con respecto a fármacos nuevos o de uso no frecuente. También deberá consultarse a los organismos de control de medicamentos de cada país para obtener información sobre los valores normales y medicamentos permitidos o recomendados.

A Marta y Arnoldo, sin ellos no se habría escrito este libro.

A Miriam, no sólo por Lucas y Sofía.

A AB y GB por su amistad incondicional.

A Patry por compartir nuestras aventuras intelectuales.

CAPÍTULO 1. INTRODUCCIÓN.....	17
¿POR QUÉ EMPEZAR POR EPIDEMIOLOGÍA?	17
¿QUÉ ES LA EPIDEMIOLOGÍA?	17
OBJETIVOS DE LA EPIDEMIOLOGÍA	17
FACTOR DE EXPOSICIÓN Y EVENTO	18
 CAPÍTULO 2 CLASIFICACIÓN DE LOS ESTUDIOS DE INVESTIGACIÓN	 21
DISEÑO EXPERIMENTAL Y DISEÑO OBSERVACIONAL	21
DISEÑO PROSPECTIVO Y DISEÑO RETROSPECTIVO.....	22
DISEÑO LONGITUDINAL Y DISEÑO TRANSVERSAL	22
DISEÑO DESCRIPTIVO Y DISEÑO ANALÍTICO	23
 CAPÍTULO 3 DISEÑO DE ESTUDIOS EPIDEMIOLÓGICOS.....	 25
INTRODUCCIÓN.....	25
ESTUDIOS TRANSVERSALES	25
ESTUDIOS TRANSVERSALES DESCRIPTIVOS Y ANALÍTICOS	26
ESTUDIOS TRANSVERSALES RETROSPECTIVOS Y PROSPECTIVOS.....	28
UTILIDAD DE LOS ESTUDIOS TRANSVERSALES.....	29
POBLACIÓN DE ESTUDIO	30
PARTICIPANTES DEL ESTUDIO	30
RECOLECCIÓN DE DATOS.....	31
ANÁLISIS DE DATOS.....	31
INTERPRETACIÓN DE LOS RESULTADOS	31
VENTAJAS Y DESVENTAJAS.....	31
ESTUDIOS DE COHORTES.....	32
ESTUDIOS DE COHORTES DESCRIPTIVOS	32
ESTUDIOS DE COHORTES ANALÍTICOS	33
ESTUDIOS DE COHORTES PROSPECTIVOS Y RETROSPECTIVOS	33
UTILIDAD DE LOS ESTUDIOS DE COHORTES	34
SELECCIÓN DE LA POBLACIÓN DE ESTUDIO.....	34
CLASIFICACIÓN DE LA POBLACIÓN EN EXPUESTOS Y NO EXPUESTOS.....	35
SEGUIMIENTO DE LA POBLACIÓN DE ESTUDIO	35
DETERMINACIÓN DEL EVENTO.....	35
ANÁLISIS DE DATOS	36
INTERPRETACIÓN DE LOS RESULTADOS.....	36
VENTAJAS Y DESVENTAJAS.....	36
ESTUDIOS CASO—CONTROL.....	37
UTILIDAD DE LOS ESTUDIOS CASO-CONTROL.....	38
SELECCIÓN DE CASOS	38
SELECCIÓN DE CONTROLES	39
IDENTIFICACIÓN DEL FACTOR DE EXPOSICIÓN	40
ANÁLISIS DE DATOS.....	41

INTERPRETACIÓN DE LOS RESULTADOS	41
VENTAJAS Y DESVENTAJAS	42
ESTUDIOS CASO-CONTROL ANIDADOS.....	42
ESTUDIOS EXPERIMENTALES	42
OBJETIVO DEL ESTUDIO	43
DEFINICIÓN DE LA POBLACIÓN DE INTERÉS	44
DETERMINACIÓN DEL/LOS EVENTO/S DE INTERÉS	44
¿EFICACIA O EFECTIVIDAD?	45
CÁLCULO DEL TAMAÑO DE LA MUESTRA	46
ASIGNACIÓN DE LA INTERVENCIÓN. ALEATORIZACIÓN.....	46
DIFERENTES TIPOS DE ALEATORIZACIÓN.....	46
ALEATORIZACIÓN SIMPLE	46
ALEATORIZACIÓN POR BLOQUES	47
ALEATORIZACIÓN ESTRATIFICADA	47
DISEÑO FACTORIAL	48
DISEÑO CRUZADO (<i>CROSS-OVER</i>)	48
¿PLACEBO O CONTROL?.....	49
SEGUIMIENTO.....	49
ANÁLISIS	49
INTERPRETACIÓN DE LOS RESULTADOS	50
VENTAJAS Y DESVENTAJAS.....	50
ESTUDIOS ECOLÓGICOS	51
TIPOS DE ESTUDIOS ECOLÓGICOS	52
ESTUDIOS MULTIGRUPO	52
ESTUDIOS DE TENDENCIA EN EL TIEMPO.....	53
DESCRIPTIVOS.....	53
ANALÍTICOS.....	53
ESTUDIOS EXPERIMENTALES ECOLÓGICOS.....	53
¿CUÁNDO ES NECESARIO REALIZAR UN ESTUDIO ECOLÓGICO?	53
DIFERENTES GRUPOS.....	53
GRUPOS MÁS ADECUADOS QUE INDIVIDUOS	54
¿PARA QUÉ SON ÚTILES LOS ESTUDIOS ECOLÓGICOS?	54
¿CÓMO SE RECLUTAN LOS GRUPOS DE ESTUDIO?	55
MEDICIÓN DEL NIVEL DE EXPOSICIÓN Y EVENTOS A NIVEL GRUPAL	55
ANÁLISIS DE DATOS	56
INTERPRETACIÓN DE LOS RESULTADOS.....	56
INFERENCIAS DE LOS RESULTADOS DE UN ESTUDIO ECOLÓGICO	57
VENTAJAS Y DESVENTAJAS.....	58
 CAPÍTULO 4 Sesgos	 59
CLASIFICACIÓN DE SESGOS	60
SESGOS DE SELECCIÓN	60
ESTUDIOS DESCRIPTIVOS	60
ESTUDIOS ANALÍTICOS	61
ESTUDIOS CASO-CONTROL.....	61
SELECCIÓN DE CASOS.....	61
SELECCIÓN DE LOS CONTROLES.....	62
ESTUDIOS DE COHORTES.....	63
ELECCIÓN EQUIVOCADA DE POBLACIÓN NO EXPUESTA	63
SEGUIMIENTO DIFERENCIAL ENTRE EXPUESTOS Y NO EXPUESTOS	64
SESGOS DE INFORMACIÓN.....	64
SESGOS DE REPORTE	65
SESGOS DE OBSERVACIÓN	65
VEAMOS ALGUNOS EJEMPLOS.....	65

CAPÍTULO 5 CONFUNDIDORES E INTERACCIÓN	67
CONCEPTO.....	67
CONDICIONES PARA QUE UN FACTOR DE EXPOSICIÓN O VARIABLE SEA CONSIDERADO COMO POSIBLE CONFUNDIDOR	68
MÉTODOS PARA CONTROLAR CONFUNDIDORES EN LA ETAPA DE DISEÑO	70
MÉTODOS PARA CONTROLAR CONFUNDIDORES EN LA ETAPA DE ANÁLISIS	71
INTERACCIÓN	71
CAPÍTULO 6 REPORTE DE FRECUENCIA DE FACTOR DE EXPOSICIÓN Y EVENTO.....	73
ESTIMACIONES	73
PREVALENCIA E INCIDENCIA	74
PREVALENCIA	74
INCIDENCIA	75
INCIDENCIA RIESGO	75
INCIDENCIA TASA	75
RELACIÓN ENTRE INCIDENCIA Y PREVALENCIA.....	76
CAPÍTULO 7 CASO Y POBLACIÓN EN RIESGO.....	79
DEFINICIÓN DE CASO.....	79
CRITERIOS PARA IDENTIFICAR CASOS.....	80
LÍMITE DE CASO	80
UNIDAD DE ANÁLISIS	80
DESARROLLO DE LA DEFINICIÓN DE CASO.....	81
LA PREGUNTA QUE DESEAMOS RESPONDER	82
RECURSOS DISPONIBLES	82
DEFINICIÓN DE LA POBLACIÓN DE ESTUDIO	83
POBLACIÓN EN RIESGO	83
CAPÍTULO 8 RECOLECCIÓN DE DATOS.....	85
CLASIFICACIÓN DE VARIABLES.....	85
CLASIFICACIÓN DE DATOS NUMÉRICOS.....	85
CLASIFICACIÓN DE DATOS CATEGÓRICOS	85
CAPÍTULO 9 PRINCIPIOS DE LA ESTADÍSTICA	89
ÁZAR Y PROBABILIDAD.....	90
LÓGICA DE LAS PRUEBAS ESTADÍSTICAS.....	91
CONCEPTO DE INTERVALO DE CONFIANZA.....	92
CAPÍTULO 10 DATOS CUANTITATIVOS. MEDIDAS DE POSICIÓN Y DISPERSIÓN	95
MEDIDAS DE POSICIÓN	95
MEDIDAS DE DISPERSIÓN	97
MÉTODOS GRÁFICOS PARA VALORAR LA VARIABILIDAD	97
MÉTODOS NUMÉRICOS PARA EVALUAR LA VARIABILIDAD	101
RANGOS.....	101
PERCENTILES.....	101
DESVÍO ESTÁNDAR.....	104
CÁLCULO DE VARIANZA Y DESVÍO ESTÁNDAR.....	105
CÁLCULO DEL INTERVALO DE CONFIANZA	107
CAPÍTULO 11 DATOS CUANTITATIVOS. DISTRIBUCIÓN NORMAL.....	111
CAPÍTULO 12 ANÁLISIS DE DATOS CUANTITATIVOS. COMPARACIÓN DE MEDIAS	121
ERROR ESTÁNDAR.....	121
INTERVALO DE CONFIANZA DE LA MEDIA.....	122

CAPÍTULO 13 ANÁLISIS DE DATOS CUANTITATIVOS. PRUEBAS DE SIGNIFICACIÓN PARA MEDIAS.....	127
PRUEBA DE T PARA UNA MUESTRA.....	127
PRUEBA DE T PARA MUESTRAS APAREADAS.....	130
PRUEBA DE T PARA DOS MUESTRAS INDEPENDIENTES.....	132
PRUEBAS NO PARAMÉTRICAS.....	135
TEST DE T PARA MUESTRAS INDEPENDIENTES.....	135
TEST DE T NO VÁLIDO. ÚTILICE TEST DE MANN-WHITNEY.....	136
TEST DE MANN-WHITNEY.....	136
COMPARACIÓN DE MÁS DE DOS MEDIAS.....	137
CAPÍTULO 14 ERRORES ALFA Y BETA	139
PODER ESTADÍSTICO	139
ELECCIÓN DEL ERROR ALFA O NIVEL DE SIGNIFICACIÓN	142
CAPÍTULO 15 CÁLCULO DEL TAMAÑO DE LA MUESTRA	145
DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA POR PRECISIÓN.....	145
DETERMINACIÓN DE LA MUESTRA POR PODER	146
AJUSTE POR PÉRDIDA DE SUJETOS.....	148
CAPÍTULO 16 CORRELACIÓN Y REGRESIÓN LINEAL	149
CORRELACIÓN	149
REGRESIÓN LINEAL	150
CAPÍTULO 17 DATOS CUALITATIVOS. MEDIDAS DE RESUMEN	155
MEDIDA DE RESUMEN NUMÉRICA.....	155
RESUMEN POR MÉTODOS GRÁFICOS.....	155
VARIABILIDAD DE DATOS CATEGÓRICOS.....	156
CAPÍTULO 18 DATOS CUALITATIVOS. DISTRIBUCIÓN BINOMIAL	159
CAPÍTULO 19 DATOS CUALITATIVOS. PRUEBAS DE SIGNIFICACIÓN ESTADÍSTICA PARA PROPORCIONES	163
PRUEBA PARA UNA PROPORCIÓN.....	163
PRUEBA PARA DOS PROPORCIONES.....	164
CAPÍTULO 20 DATOS CUALITATIVOS. TABLAS DE CONTINGENCIA	167
CAPÍTULO 21 PRUEBA DE CHI CUADRADO	169
PRUEBA DE CHI CUADRADO PARA DOS PROPORCIONES	169
PRUEBA DE CHI CUADRADO PARA MUESTRAS PEQUEÑAS	172
PRUEBA DE CHI CUADRADO PARA DATOS APAREADOS.....	173
PRUEBA DE CHI CUADRADO PARA ANÁLISIS DE TABLAS R X C.....	175
PRUEBA DE CHI CUADRADO MÚLTIPLE	175
PRUEBA DE CHI CUADRADO TREND.....	177
CAPÍTULO 22 MEDIDAS DE EFECTO.....	181
RIESGO ABSOLUTO.....	181
RIESGO RELATIVO	181
REDUCCIÓN DEL RIESGO RELATIVO	183
NÚMERO DE PACIENTES NECESARIOS A TRATAR. NNT	185
NÚMERO DE PACIENTES NECESARIOS A DAÑAR. NNH	186
ODDS RATIO	186
ODDS RATIO PARA ESTUDIOS CASO-CONTROL	187

<i>ODDS RATIO</i> PARA ESTUDIOS CASO-CONTROL APAREADOS	189
<i>ODDS RATIO</i> PARA ESTUDIOS DE COHORTES.....	190
CAPÍTULO 23 MEDIDAS DE IMPACTO	193
RIESGO ATRIBUIBLE PARA SUJETOS EXPUESTOS	193
RIESGO ATRIBUIBLE POBLACIONAL.....	195
COMPARACIÓN ENTRE RIESGO RELATIVO Y RIESGO ATRIBUIBLE	196
CAPÍTULO 24 ANÁLISIS DE SOBREVIDA.....	199
TIEMPO DE SOBREVIDA.....	199
TABLAS DE SOBREVIDA.....	200
CURVAS DE SOBREVIDA O KAPLAN MEIER	202
COMPARACIÓN DE CURVAS DE KAPLAN-MEIER . PRUEBA LOG RANK TEST	204
HAZARD RATIO	205
CAPÍTULO 25 TAMIZAJE Y DIAGNÓSTICO.....	207
DEFINICIÓN	207
CRITERIOS PARA INTRODUCIR UN PROGRAMA DE TAMIZAJE	208
SENSIBILIDAD Y ESPECIFICIDAD.....	209
ESTUDIOS CON RESULTADOS DICOTÓMICOS (POSITIVO O NEGATIVO).....	209
ESTUDIOS CON RESULTADOS CONTINUOS	212
CURVAS ROC	213
CRITERIOS DE CONFIABILIDAD.....	214
VALOR PREDICTIVO POSITIVO Y NEGATIVO.....	214
CRITERIOS DE REPETIBILIDAD.....	218
VARIABILIDAD INTRASUJETO	218
VARIABILIDAD INTEROBSERVADOR	218
PRUEBA DE KAPPA	218
CAPÍTULO 26 EVALUACIÓN PROGRAMAS DE SCREENING.....	221
ACEPTABILIDAD	221
SESGOS EN ESTUDIOS DE EVALUACIÓN DE PROGRAMAS DE TAMIZAJE.....	221
SESGO DE SELECCIÓN	221
EXISTEN TRES SESGOS PARTICULARES DE LOS ESTUDIOS DE SCREENING	222
SESGO POR ADELANTO O <i>LEAD TIME BIAS</i>	222
SESGO DE DURACIÓN O <i>LENGHT TIME BIAS</i>	223
SESGO POR SOBREDIAGNÓSTICO.....	223
ASPECTOS A TENER EN CUENTA EN UN PROGRAMA DE SCREENING	223
CAPÍTULO 27 METANÁLISIS.....	225
OBJETIVOS DEL METANÁLISIS.....	226
METODOLOGÍA.....	226
MÉTODOS ESTADÍSTICOS	227
MÉTODO DE EFECTO FIJO.....	227
MÉTODO DE EFECTO AL AZAR	228
SESGOS EN METANÁLISIS	229
SESGOS DE PUBLICACIÓN	229
SESGO POR PUBLICACIÓN MÚLTIPLE	230
SESGOS DE SELECCIÓN.....	230
EVALUACIÓN DE SESGOS	231
GRÁFICO CONO INVERTIDO O <i>FUNNEL PLOT</i>	231

CAPÍTULO 28 INTRODUCCIÓN AL ANÁLISIS MULTIVARIABLE.....	233
PRUEBA DE MANTEL-HAENZEL. INTERACCIÓN	233
CONTROL DE CONFUNDIDORES.....	236
DETERMINACIÓN DE PREDICTORES INDEPENDIENTES	237
DESARROLLO DE PUNTAJES PRONÓSTICOS.....	237
IGUALACIÓN DE CARACTERÍSTICAS BASALES.....	238
TIPOS DE MÉTODOS ESTADÍSTICOS PARA ANÁLISIS MULTIVARIABLE	239
CAPÍTULO 29 REGRESIÓN LOGÍSTICA MÚLTIPLE	241
CAPÍTULO 30 ANÁLISIS MULTIVARIABLE. REGRESION DE COX	249
DESCRIPCIÓN DEL MÉTODO.....	249
MODELO MATEMÁTICO DE COX.....	250
HAZARD RATIO.....	250
CRITERIO DE PROPORCIONALIDAD	251
REGRESIÓN DE COX SIMPLE.....	253
REGRESIÓN DE COX MÚLTIPLE.....	254
CAPÍTULO 31 CRITERIOS DE CAUSALIDAD.....	257
TIPOS DE RELACIÓN CAUSAL	257
CRITERIOS DE BRADFORD HILL.....	259
ÍNDICE ANALÍTICO	263



PRÓLOGO

En los primeros años de mi formación como médica, no existía una educación formal en Estadística y Epidemiología.

En mi caso personal tenía una gran inquietud por estos temas, pero además la verdadera necesidad de aprender, porque estaba involucrada en la realización de una de las primeras encuestas epidemiológicas en Cardiología que se realizaban en nuestro país.

Comencé a estudiar apoyándome en colegas que tenían alguna experiencia en este campo, buscando libros de texto en inglés y en castellano, concurriendo a cursos de posgrado dictados en general por especialistas en Estadística y compartiendo experiencias con colegas en etapa de formación que tenían mis mismas inquietudes.

Este camino no fue de los más fáciles. El diálogo con mis colegas no siempre era sencillo, y el que podía entablar con los profesionales de la Estadística parecía el de personas con idiomas completamente diferentes.

Leer libros extranjeros no implicaba adquirir un total entendimiento de su contenido y los textos traducidos al español eran tediosos, incomprensibles y no daban ejemplos claros que ayudaran a entender un tema.

No obstante, a pesar de las dificultades, pude lograr una educación formal que me permitió comenzar mi tarea educativa dirigida a otros colegas.

Esta actividad fue de gran ayuda para mí. Las inquietudes de los alumnos y sus dificultades para entender algunos conceptos representaron un reto que debía enfrentar.

El objetivo de mis clases fue entonces transmitir conocimientos en un lenguaje claro y ameno para los médicos, y el método elegido fue el que sentía que yo misma habría necesitado en aquella etapa de formación, haciendo las veces de traductor entre el idioma de la Estadística y el del médico, ofreciendo ejemplos simples para que el tema de la clase se comprendiera y luego se recordara y, no menos importante, estimulando a los alumnos a participar.

De esta participación de los alumnos es que pude generar este lenguaje que, estoy convencida, los acerca de una forma más placentera al mundo de las Matemáticas y Epidemiología que estamos obligados a aprender.

Ya había alcanzado dos metas: concluir mi Maestría de Estadística y Epidemiología en la Universidad de Londres y dictar cursos donde los alumnos se sintieran interesados y motivados.

Pero algo seguía faltando, al final de cada clase formulaban la misma pregunta: “¿Puede recomendarme un libro fácil de entender?”

Con la experiencia ganada como alumna y docente, me sentí en condiciones de enfrentar el desafío de escribir ese libro que reclamaban los alumnos.

Ese fue otro problema. ¿Cómo escribir un libro que tuviera todos los requisitos que para mí (y según ellos mismos, los alumnos) debería tener?

Tuve la idea de poner por escrito las clases que dicto habitualmente en los cursos.

Este libro se escribió como si el lector estuviera presente en la clase, ya sea un especialista, un alumno de pregrado o un estudiante de alguna carrera afín a la medicina.

Los temas elegidos son los que los médicos debemos saber para entender las publicaciones médicas que leemos a diario y para realizar investigaciones en nuestros propios ámbitos de trabajo.

Evité deliberadamente escribir un libro de Estadística. Ella es sólo una parte de un gran proceso de planeamiento, diseño, implementación y conducción de los estudios de investigación que, sin el conocimiento sobre Epidemiología, no es posible realizar.

La terminología específica está explicada con palabras sencillas, los ejemplos fueron obtenidos de la práctica diaria y los razonamientos matemáticos están desarrollados de forma simple y sólo con fines didácticos.

Mi intención, creo que ya clara, es la de ayudar a incorporar estos conocimientos de forma sencilla y amena, para que el lector adquiera la habilidad de comprender y criticar publicaciones médicas, además de estimularlo a iniciar sus propios proyectos de investigación.

Es mi deseo que este libro derribe las barreras existentes en la enseñanza de la Epidemiología y sea un punto a partir del cual se sientan estimulados a seguir aprendiendo acerca de este fascinante y tan necesario aspecto de nuestra profesión.

LA AUTORA

¿POR QUÉ EMPEZAR POR EPIDEMIOLOGÍA?

Cuando se dice que se enseña *investigación clínica* en general se hace referencia nada más que a los métodos estadísticos utilizados para analizar los datos obtenidos en un estudio. Esto es sólo una parte del proceso de investigación, que de hecho exige cumplir etapas y utilizar una metodología propia.

La Epidemiología es la ciencia que nos proporciona la metodología apropiada para *investigar*.

¿QUÉ ES LA EPIDEMIOLOGÍA?

Es el estudio de la distribución y de los determinantes de estados relacionados con la salud en una población, en un punto específico de tiempo, y la aplicación de estrategias para su control.

Hablamos de estados relacionados con la salud debido a que la Epidemiología intenta evaluar todas las posibles situaciones que modifiquen el estado de salud, ya sean beneficiosas o perjudiciales. Por ejemplo, es importante conocer la proporción de la población que tiene acceso a agua potable, ya que de esto depende la aparición de algunas enfermedades infecciosas gastrointestinales, también es importante conocer el porcentaje de sujetos que usan cinturón de seguridad al conducir o bien el porcentaje de la población de niños vacunados contra el sarampión.

Conocer estos datos permite implementar, por ejemplo, una campaña para uso de cinturón de seguridad con el propósito de reducir las lesiones o muerte por accidente de tránsito, o bien decidir que no es necesario introducir una nueva campaña de vacunación contra el sarampión, si la proporción vacunada de niños es adecuada para controlar esta enfermedad.

OBJETIVOS DE LA EPIDEMIOLOGÍA

- 1) Identificar la / las causas de una enfermedad o de estados relacionados con la salud y sus factores de riesgo.
- 2) Determinar la extensión de la enfermedad en la población.
- 3) Estudiar la historia natural de la enfermedad y su pronóstico.
- 4) Evaluar medidas preventivas y terapéuticas.
- 5) Proporcionar fundamento para el desarrollo de políticas de salud.

Basta recordar la historia del HIV/SIDA en los últimos 20 años para observar cómo, a través del cumplimiento de los objetivos de la Epidemiología, se pudo identificar el agente

etiológico, desarrollar métodos para su diagnóstico, establecer nuevas alternativas terapéuticas y políticas de Salud Pública, con lo cual se logró alterar la historia natural de la enfermedad de una manera drástica.

Los objetivos de la Epidemiología se reúnen en lo que hoy llamamos Investigación Clínica.

La mayoría de las publicaciones médicas actuales aportan información acerca de alguno de estos objetivos. En Cardiología y Oncología, la investigación sobre las causas genéticas está en pleno desarrollo, pero no abandona la búsqueda de factores de riesgo *ambientales*, medidas preventivas y terapéuticas que modifiquen la actual historia natural de estas enfermedades.

A la hora de leer un estudio de investigación hay que recordar que, de hecho, se trata de Epidemiología y que, por lo tanto, para poder evaluar la validez de dicho estudio hay que conocer la metodología epidemiológica empleada.

La mayoría de las guías de “apreciación crítica de la literatura médica” sólo ofrecen un listado de condiciones que debe poseer el estudio de investigación para ser considerado *válido*, con poco énfasis en el diseño del estudio y el método de recolección de los datos y mucho énfasis en su análisis. Esto es práctico sólo si se tiene el conocimiento epidemiológico y estadístico previo.

Son muchas las causas por las cuales los resultados de un estudio de investigación pueden no ser válidos; la primera es un mal diseño del estudio.

Cuando hablamos del diseño de un estudio de investigación nos referimos al modo en que se seleccionó la muestra de sujetos a estudiar, cómo fueron definidos los factores de riesgo y eventos a estudiar, cómo serán obtenidos los datos, etc.

Todos estos pasos son fundamentales, porque, por más que la **p** resulte significativa, si la elección de la población no fue correcta o el modo en que se recabaron los datos fue erróneo, además de considerar la presencia de sesgos y confundidores, el estudio no tendrá validez.

En los siguientes capítulos se desarrollarán en profundidad los diferentes diseños epidemiológicos empleados para la conducción de estudios de investigación, posibles fuentes de errores como la presencia de sesgos y la evaluación de confundidores.

FACTOR DE EXPOSICIÓN Y EVENTO

La Epidemiología *trabaja* con factores de exposición y eventos, que son los estados relacionados a la salud.

Estos son los dos elementos principales que se registran en los estudios epidemiológicos.

El factor de exposición (*exposure* en inglés) es el factor de riesgo que estamos investigando, que puede o no ser la causa de una enfermedad o evento (situación relacionada con la salud).

Un factor de exposición puede ser cualquier agente que pueda influenciar la aparición o disminución del evento.

Por ejemplo, los factores de exposición que pueden ser relevantes en el desarrollo de cáncer de pulmón podrían ser: tabaquismo, exposición a asbesto, consejo médico sobre abandono del hábito, características culturales de la población, clase socioeconómica, legislación sobre publicidad de tabaco o prohibición de fumar en lugares públicos, precio de los cigarrillos, entre muchos otros.

Aquí hemos mencionado factores de exposición que incrementan o reducen el desarrollo de cáncer de pulmón.

Un factor de exposición no necesariamente es un factor que incrementa la frecuencia de un evento, también puede reducir su aparición, y por lo tanto puede ser beneficioso. Por este motivo es mejor hablar de factor de exposición y no de riesgo para no malinterpretar este concepto.

Por ejemplo, podemos investigar la aplicación de la ley antibacno en lugares públicos (factor de exposición *beneficioso*) y reducción de la incidencia de infarto de miocardio (evento beneficioso).

El evento (*outcome* en inglés) es lo que comúnmente entendemos como enfermedad (o estado relacionado) con la salud en el cual estamos interesados.

Es importante destacar que el evento no siempre es un hecho perjudicial para la salud, un evento puede ser mejoría de alguna enfermedad, uso de casco al conducir una motocicleta, acceso a agua potable.

El evento puede ser cualquier estado relacionado con la salud y también un factor de exposición.

Un determinado evento en un estudio de investigación puede ser un factor de exposición en otro estudio.

En un estudio sobre el efecto del cigarrillo sobre el cáncer de pulmón, el factor de exposición es el hábito de fumar y el evento el cáncer de pulmón.

En un estudio sobre el efecto de la publicidad de cigarrillos sobre el hábito de fumar, el factor de exposición es la publicidad de cigarrillos y el evento el hábito de fumar.

Lo que la Epidemiología trata de establecer, a través del diseño adecuado de un estudio de investigación y su posterior análisis estadístico, es si existe una *verdadera* asociación entre un factor de riesgo o factor de exposición y un evento. ¿Existe asociación entre fumar (factor de riesgo) y cáncer de pulmón (evento)? ¿Existe asociación entre el uso de cinturón de seguridad (factor de exposición) y reducción de la mortalidad por accidente de tránsito?

Determinar asociación entre un factor de exposición y un evento, sólo significa que cuando uno de ellos está presente, la probabilidad de que el otro también esté presente es alta. Esto no quiere decir que exista relación etiológica entre el factor de exposición y el evento.

Si observamos asociación entre gastroenteritis y falta de agua potable, significa que cuando se presenta un caso de gastroenteritis debemos preguntar acerca del tipo de agua que bebe el sujeto. La causa o etiología será algún agente infeccioso, que es lo que trataremos para curar al paciente.

La determinación de una asociación definitiva entre un factor de exposición y un evento es lo que nos permite en la práctica diaria de la medicina, a través del interrogatorio del paciente, determinar con cierta probabilidad la existencia de una enfermedad.

Ante un paciente que se presente con una hemorragia digestiva alta, una de las primeras preguntas que le haremos es si consumió altas dosis de antiinflamatorios. Si la respuesta es afirmativa, podremos pensar que el paciente tiene una hemorragia digestiva erosiva con alta probabilidad de que así sea.

Otro ejemplo es el diagnóstico clínico de angina de pecho, donde además de las características del dolor, la presencia de factores de exposición, en este caso denominados factores de riesgo cardiovasculares, como hipertensión arterial, aumento de colesterol, tabaquismo, etc., nos permiten hacer el diagnóstico e iniciar el tratamiento de la enfermedad.

La identificación de factores de exposición asociados a eventos, por ejemplo el caso de falta de agua potable o la presencia de factores de riesgo cardiovasculares, permitirá implementar políticas de salud destinadas a corregir estos factores de exposición para reducir la frecuencia de aparición de las enfermedades. Dicho en otras palabras, a través de

la identificación de factores de exposición asociados a un evento es posible prevenir la enfermedad, lo que es mucho más eficiente que sólo tratar los casos que presentan el evento.

Además de recabar información sobre el factor de exposición de interés y del evento, en todo estudio de investigación se recolecta información de muchas otras características de los sujetos incluidos en el estudio de investigación; a estas características las denominamos variables o covariables para diferenciarlas del factor de exposición y evento, que también son variables.

Las variables de un estudio de investigación son fundamentales para un buen análisis estadístico del mismo, como veremos en próximos capítulos.

CLASIFICACIÓN DE LOS ESTUDIOS DE INVESTIGACIÓN



En investigación médica, los objetivos más frecuentemente perseguidos son dos:

1. Determinar causas y factores de exposición asociados a una enfermedad.
2. Evaluar un tratamiento farmacológico u otra intervención para determinar su impacto en la historia natural de la enfermedad.

Cada uno de estos objetivos puede llevarse a cabo aplicando diseños de investigación apropiados.

En general los estudios que buscan establecer la etiología y/o los factores de exposición utilizan diseños epidemiológicos llamados observacionales, como estudios transversales, de cohortes y los caso-control.

Aquellos estudios que buscan evaluar una intervención emplean los conocidos estudios aleatorizados (conocidos comúnmente como estudios randomizados) o estudios de experimentación.

Los estudios de investigación tienen diseños preespecificados, pero algunas de sus características deben mencionarse, debido a que dependen de diferentes aspectos del estudio.

Por lo general las características de un estudio se clasifican de acuerdo con las siguientes consideraciones:

1. Si se interviene sobre las características de los sujetos de investigación
2. El modo de recolección de datos
3. Si la población es seguida en el tiempo
4. El modo de analizar la información

DISEÑO EXPERIMENTAL Y DISEÑO OBSERVACIONAL

Si el investigador realiza alguna intervención para que los sujetos adquieran alguna característica en particular, el diseño es experimental. En este tipo de estudios, la población seleccionada generalmente se divide en dos grupos: a uno se le asigna una droga o tratamiento nuevo y al otro una droga o tratamiento estándar, o bien (si no existe droga estándar) un placebo. A este grupo se lo denomina grupo control, ya que la nueva intervención se comparará con él.

Cuando se habla de *diseño experimental* se hace referencia sólo al hecho de haber ejercido una acción sobre los sujetos.

En un estudio con *diseño observacional*, sólo se observan características que la población ha adquirido naturalmente. Aquí los investigadores no realizan ninguna intervención sobre los sujetos. La población se selecciona en base a alguna característica que ella posea y que sea de interés estudiar.

Por ejemplo, podemos tomar un grupo de pacientes con infarto agudo de miocardio tratados con terapia fibrinolítica y determinar la mortalidad. Luego comparamos dicha mortalidad con la de un grupo que no haya recibido este tratamiento. En este estudio, la asignación a tratamiento fibrinolítico no la realiza el investigador: la determinación de tratar o no a los pacientes ya fue realizada por el médico tratante y lo único que hace el investigador es *observar* la mortalidad, o cualquier otro evento, en este grupo de pacientes.

DISEÑO PROSPECTIVO Y DISEÑO RETROSPECTIVO

Otra forma de clasificar los estudios de investigación es con relación a cómo se generan o producen los datos del estudio, en otras palabras, si los datos son nuevos o se utilizan datos existentes.

En un estudio con *diseño prospectivo* los datos se generan al iniciarlo.

Por ejemplo, en un estudio cuyo interés es conocer las complicaciones que padecen pacientes diabéticos que asisten a un centro especializado en esta enfermedad, si para incluir al sujeto en el estudio se realiza una determinación de glucemia, el dato sobre el estatus de diabético del sujeto es nuevo o prospectivo.

Al mismo tiempo se puede evaluar la presencia de complicaciones a través de diferentes estudios que confirmen, por ejemplo, la existencia de deterioro de la función renal, enfermedad vascular periférica, retinopatía o enfermedad coronaria. Estos datos también son prospectivos o nuevos.

En un estudio con *diseño retrospectivo* los datos se obtienen a través del interrogatorio o anamnesis del sujeto, su historia clínica, u otros registros existentes. En este caso, los datos ya han sido recabados o ya existen.

Siguiendo el ejemplo anterior, se puede realizar el mismo estudio de pacientes diabéticos y sus complicaciones, recurriendo a los datos disponibles en sus historias clínicas.

A veces surgen dudas acerca de si un estudio es prospectivo o retrospectivo, porque muchas veces parte de la información necesaria para llevar a cabo el estudio ya ha sido recabada.

Se puede determinar, en base a datos de la historia clínica, si los pacientes son diabéticos y cuál es la tasa de complicaciones que padecen al ingresar al estudio realizando los estudios correspondientes, o bien recabar las complicaciones que padecerán durante cierto período de seguimiento. En este caso el diagnóstico de diabetes es retrospectivo (porque ya se hizo previamente y no fue reconfirmado al inicio del estudio) y la determinación de las complicaciones es prospectiva (se hace en el momento del ingreso al estudio o bien a medida que ellas surjan en el tiempo).

Cuando la forma de obtener los datos es combinada, lo mejor es preguntarse: ¿Cuál es el objetivo del estudio? o ¿los datos que van a responder el objetivo ya existen o van a recabarse?

En el ejemplo anterior el objetivo es determinar la tasa de complicaciones. Como la recolección de los datos se realizará a partir del momento de iniciado el estudio, este podría definirse como un estudio prospectivo, aclarando que el diagnóstico de diabetes se realizó de manera retrospectiva.

DISEÑO LONGITUDINAL Y DISEÑO TRANSVERSAL

Un estudio con *diseño longitudinal* es aquel en el que los sujetos son seguidos en el tiempo para observar cuál es su evolución. En general los sujetos son observados varias veces

durante un período de tiempo determinado. (Por ejemplo, pacientes diagnosticados en el pasado como diabéticos son seguidos a partir del ingreso al estudio cada 6 meses, para determinar la tasa de complicaciones vasculares que desarrollarán en los próximos 5 años.)

Cuando las características y los eventos de una población se determinan al mismo tiempo, durante un período específico de tiempo, el diseño se llama transversal.

Un estudio con diseño transversal incluiría, durante un período de 3 meses, pacientes con diagnóstico previo o reciente de diabetes (característica de los pacientes o factor de exposición) para determinar qué complicaciones vasculares presentan en ese momento, pudiendo obtener los datos sobre las complicaciones de las historias clínicas, interrogatorio o bien realizando estudios destinados a identificar las complicaciones de interés.

Generalmente este tipo de estudio aclara, en su apartado de materiales y métodos, que se recabó información entre los meses x y x del año x, de pacientes atendidos en un determinado centro, con determinadas características y para evaluar determinados eventos.

DISEÑO DESCRIPTIVO Y DISEÑO ANALÍTICO

Los estudios observacionales pueden clasificarse como *diseño descriptivo* si sólo se reportan los datos observados obtenidos de toda la población estudiada (por ejemplo, porcentaje de hipertensos en adultos mayores de 65 años y su mortalidad).

Si la población se subclasifica de acuerdo con alguna característica relevante (factor de exposición) y se comparan los datos de los eventos entre las subpoblaciones, el estudio se denomina *diseño analítico*. (La población mayor de 65 años se subclasifica en hipertensos vs. no hipertensos, se compara entre estos dos grupos la mortalidad.)

Está claro que un determinado diseño puede entonces describirse de diversas formas, de acuerdo con los criterios mencionados.

Un diseño observacional puede ser transversal y prospectivo.

El caso de los pacientes con diagnóstico reciente de diabetes, que son estudiados para determinar las complicaciones vasculares que presentan en ese momento mediante estudios específicos, es un estudio *observacional*, porque la condición de diabético no es impuesta por el investigador; es *transversal*, porque tanto el factor de exposición (diabetes) como su consecuencia o evento (complicaciones vasculares) son recolectados al mismo tiempo (los pacientes no son seguidos en el tiempo para determinar la tasa de complicaciones vasculares) y es *prospectivo* porque el diagnóstico de diabetes se realiza al ingreso al estudio, lo que implica adoptar una definición de diabetes previa para que todos los sujetos se clasifiquen de igual manera. Lo mismo ocurre con las complicaciones vasculares: se definen antes del inicio del estudio y todos los sujetos se clasifican de igual manera y son sometidos a estudios para determinar la presencia o ausencia de las complicaciones. Los datos acerca del factor de exposición y sus consecuencias o eventos se recolectarán desde el inicio del estudio en adelante; no se utilizan datos previos del paciente.

Fácilmente este estudio podría ser diseñado como observacional, retrospectivo y transversal si incluimos pacientes con historia de diabetes (sin necesidad de reconfirmar el diagnóstico) y buscamos datos acerca de complicaciones vasculares pasadas o presentes, sin realizar estudios diagnósticos. Sigue siendo un estudio observacional porque el investigador no modifica el factor de exposición (diabetes).

DISEÑO DE ESTUDIOS EPIDEMIOLÓGICOS



INTRODUCCIÓN

Ya conocen las características que definen un estudio de investigación.

De la combinación de estas características surgen los diseños preestablecidos de los estudios de investigación, estudios transversales, de cohorte, caso-control y experimental.

Cada uno de estos diseños tiene sus propios procesos para seleccionar la población de estudio, para determinar la secuencia de identificación de factor de exposición y evento, para evitar incurrir en errores y para analizar los resultados.

Los estudios de investigación se centran en la asociación entre uno o varios factores de exposición y uno o varios eventos. Dependiendo del objetivo del estudio y del tipo de factor de exposición y evento que se desea estudiar, el investigador debe elegir, entre los diferentes diseños, el más apropiado para lograr resultados válidos.

Para seleccionar el diseño más adecuado para un estudio de investigación es imprescindible conocer detalladamente cada diseño.

Comencemos a conocer cada uno de estos diseños.

ESTUDIOS TRANSVERSALES

En los estudios transversales (*cross sectional* en inglés) se determina la frecuencia de uno o varios factores de exposición y/o uno o varios eventos de interés al mismo tiempo, en una población definida, en un punto particular de tiempo.

Un estudio transversal recolecta datos sobre factores de exposición y/o eventos ya presentes en la población.

La frecuencia de un factor de exposición o evento ya presente en la población, independientemente del tiempo en que los sujetos presenten el factor de exposición o evento, se denomina prevalencia.

$$\text{Prevalencia} = \frac{\text{N}^\circ \text{ de casos existentes en un punto específico de tiempo}}{\text{Población total en un punto específico de tiempo}}$$

Caso es un sujeto con el factor de exposición o evento de interés.

El factor de exposición y/o el evento ya existentes en la población pueden ser recientemente diagnosticados o bien existir desde hace mucho tiempo. Lo importante es que ya están presentes en la población.

Por este motivo a estos estudios se los llama de prevalencia. También son conocidos como encuestas.

En este tipo de estudio los factores de exposición y/o eventos a estudiar se recaban al mismo tiempo, *no hay período de seguimiento de los sujetos*, por eso son transversales, y debido a que los factores de exposición y/o eventos a estudiar surgen espontáneamente de la población estos estudios son por definición observacionales.

Por lo tanto, estos estudios *siempre son observacionales* y como su nombre indica, transversales.

Solo dos características deben ser definidas: cómo van a reportarse y cómo van a recabarse los datos.

ESTUDIOS TRANSVERSALES DESCRIPTIVOS Y ANALÍTICOS

En un estudio descriptivo transversal sólo se describe la frecuencia encontrada de uno o varios factores de exposición y/o eventos de interés.

Por ejemplo, podemos tomar como población los pacientes que asisten a servicios de clínica médica en hospitales públicos de la ciudad de Buenos Aires (población definida) y realizar una encuesta durante una semana (período particular de tiempo) para obtener datos acerca de obesidad y niveles plasmáticos de glucemia en 100 pacientes. Para este estudio el factor de exposición es obesidad y el evento de interés, alteración de los niveles de glucemia. Previamente se ha establecido cómo se va a definir obesidad y alteración de la glucemia. Si el estudio reporta sólo el porcentaje o el número de sujetos observados con el factor de exposición y evento en la población de estudio decimos que es descriptivo. (Figura 3.1)

La forma de reportar estos datos es mediante una tabla. (Tabla 3.1)

En esta tabla se muestra el porcentaje de sujetos con obesidad y alteración de la glucemia, hallados en el total de la población. De un total de 100 pacientes, 60 eran obesos y 60 tenían alteración de la glucemia, muchos de los cuales tenían las dos características.

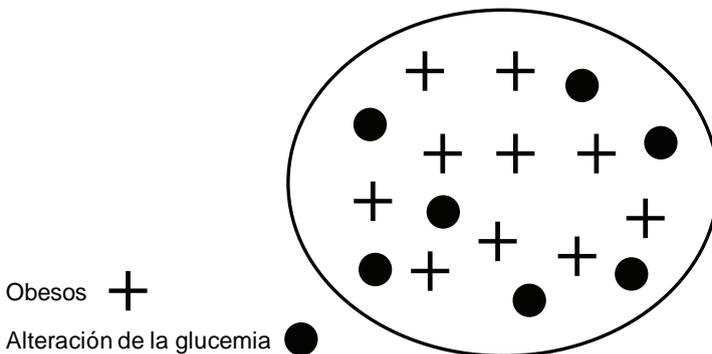


Figura 3.1: Estudio descriptivo

Variable	Porcentaje (n° de sujetos)
Obesidad (IMC \geq 25)	60% (60)
Alteración de la glucemia (\geq 110 mg/dL ayunas)	60% (60)

Tabla 3.1: Descripción de los porcentajes de obesidad y alteración de la glucemia

En un estudio transversal analítico, se obtiene simultáneamente información sobre factores de exposición y eventos; se compara la prevalencia del evento entre aquellos sujetos expuestos y no expuestos al factor. Como la información sobre la frecuencia de factores de exposición y evento se recaba al mismo tiempo, también es posible comparar la frecuencia de un factor de exposición entre sujetos con y sin el evento de interés. (Figura 3.2)

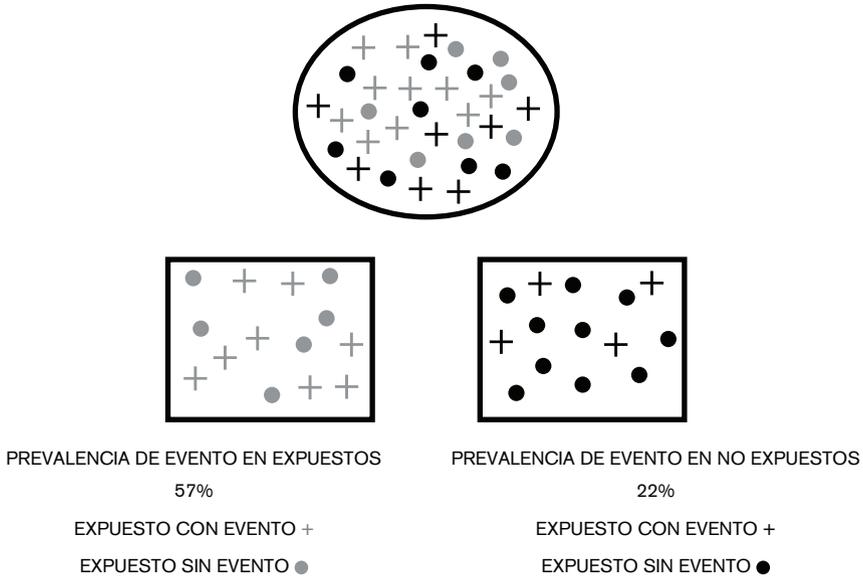


Figura 3.2: Diseño de estudio transversal analítico

El análisis de los datos de este tipo de estudio se realiza con métodos estadísticos apropiados, empezando por la construcción de la llamada tabla de contingencia.

En estas tablas se coloca en las columnas el número y/o porcentaje de sujetos con y sin el evento de interés y en las filas el número y/o porcentaje de sujetos con y sin el factor de exposición. Cuando sólo tenemos dos categorías de eventos y dos de factor de exposición decimos que estamos trabajando con una tabla 2 x 2, en la que hay 4 categorías posibles para clasificar a los sujetos del estudio. Estas celdas por convención se denominan con letras. (Tabla 3.2)

Evento Factor de exposición	Evento Sí (n) %	Evento NO (n) %	Total
Sí % (n)	a	b	a + b
No % (n)	c	d	c + d
Total	a + c	b + d	a + b + c + d

Tabla 3.2: Tabla de contingencia 2 x 2. Notación general

Evento Factor de exposición	Alteración de la glucemia No (n) %	Alteración de la glucemia Sí (n) %	Total
Obesidad			
Sí	25	35	60
No	15	25	40
Total	40	60	100

Tabla 3.3: Porcentaje de obesidad de acuerdo con alteración de la glucemia

En la Tabla 3.3 observamos que los sujetos con alteración de la glucemia 35% también eran obesos; asimismo observamos que en el grupo de pacientes con alteración de la glucemia hay más prevalencia de sujetos obesos, comparado con el grupo de pacientes sin alteración de la glucemia (35 vs. 25%).

ESTUDIOS TRANSVERSALES RETROSPECTIVOS Y PROSPECTIVOS

La información acerca de factores de exposición y/o eventos en los estudios transversales pueden provenir de fuentes de datos existentes, como historias clínicas, registros de ingresos y egresos hospitalarios, registros de datos gubernamentales (certificados de defunción, etc.) y por supuesto también del interrogatorio a los sujetos. Por lo tanto estos datos, como ya existen, serán retrospectivos.

En el caso de que la información sobre factores de exposición y/o eventos se obtenga a partir de datos generados al momento de iniciar el estudio, este es prospectivo. Por ejemplo, se puede solicitar a laboratorio para determinar el valor de la glucemia y medir y pesar al paciente a fin de calcular el índice de masa corporal, para determinar si el paciente es obeso durante la visita en la que el paciente es incluido en el estudio.

En resumen, en el estudio propuesto como ejemplo se desea conocer la prevalencia de alteración de la glucemia (evento) en una población de hombres y mujeres ≥ 50 años, proveniente de todo el país, durante el primer trimestre del año 2011. Como factor de exposición posiblemente asociado a alteración de la glucemia se obtendrán datos acerca de presencia de obesidad, y como variables necesarias para caracterizar la población se relevarán datos acerca de sexo, edad, tipo de dieta, enfermedades concomitantes, lugar de residencia, nivel socioeconómico, presencia de sedentarismo, tabaquismo, entre otros.

En este estudio se recolectarán datos acerca de la presencia o ausencia del evento (alteración de la glucemia) y de la presencia o ausencia del posible factor de exposición asociado (obesidad) y de las variables en el mismo momento. Por eso es *transversal*.

Como el factor de exposición, evento y variables surgen espontáneamente de la población, el estudio es *observacional*.

Si la información sobre la presencia o ausencia del evento y factor de exposición se obtienen a partir del interrogatorio de los sujetos, estos datos serán *retrospectivos*.

Si a los sujetos se les realiza medición de la glucemia para determinar si la tienen alterada y se les realiza medición del índice de masa corporal, los datos sobre factor de exposición y evento serán *prospectivos*.

La información acerca de las variables que caracterizan a la población y/o que podrían *interferir* en la posible asociación entre el evento y el factor de exposición, casi siempre se obtienen a partir del interrogatorio de los sujetos.

La clasificación de prospectivo o retrospectivo se realiza sobre la base de cómo se obtuvieron los datos sobre el factor de exposición y evento. Es imposible obtener información sobre todas las variables de forma prospectiva.

Una vez obtenida la información de todos los sujetos, se procede a su análisis.

Si sólo reportamos datos acerca de la prevalencia, alteración de la glucemia, obesidad y frecuencia de las variables, el estudio será descriptivo. Se podría informar que la población estuvo compuesta por un 70% de hombres, la edad media de hombres y mujeres, que 60% de los pacientes presentaron alteración de la glucemia y 50% eran obesos, además de reportar la frecuencia de las otras variables. Este tipo de reporte es *descriptivo*.

Como el objetivo del estudio es determinar si la alteración de la glucemia está asociada a obesidad, es necesario comparar la prevalencia de alteración de la glucemia entre los sujetos con y sin obesidad, aplicando los métodos estadísticos apropiados. En este caso el estudio es *analítico*.

Además de presentar los datos sobre prevalencia de alteración de la glucemia en sujetos con y sin obesidad, se deben informar las características de cada grupo comparado, debido a que si los grupos son diferentes en alguna característica, podría ser necesario investigar si esas diferencias no están afectando la asociación encontrada (a esto se lo denomina *controlar por factores de confusión o ajustar*, tema que desarrollaremos más adelante).

UTILIDAD DE LOS ESTUDIOS TRANSVERSALES

Los estudios transversales se utilizan en dos circunstancias:

1. *Planeamiento de estrategias en salud pública*

La información obtenida de un estudio transversal es muy útil para conocer la situación de una enfermedad o condición específica cuando se necesita planear una estrategia preventiva o de tratamiento.

Los estudios transversales también recolectan información acerca de la utilización de medidas preventivas y curativas en una población.

En muchos países se llevan a cabo encuestas periódicas para obtener datos relacionados con la salud en muestras poblacionales recolectadas al azar.

Los datos que se obtienen, características demográficas, prevalencia de enfermedades agudas y crónicas, necesidades y utilización de ciertos servicios de salud, son de ayuda para planear la atención sanitaria de la población.

Otro tipo de información que se recolecta con estos estudios es la relacionada al conocimiento, actitud y hábitos de una población en relación con un particular evento vinculado a la salud.

Esta información puede ser utilizada para desarrollar campañas de educación.

2. *Investigación causal de las enfermedades*

Los estudios transversales no proporcionan evidencia fuerte acerca de relación causal entre factor de exposición y evento. Para establecer la relación causal es importante determinar la secuencia temporal entre factor de exposición y evento: si el factor de exposición se presenta primero y luego de cierto tiempo aparece la enfermedad, esto constituye uno de los criterios requeridos para establecer causalidad.

En los estudios transversales no es posible determinar relación temporal entre factor de exposición y evento. Más aún, puede erróneamente aceptarse que un factor de exposi-

ción produce un evento, cuando en realidad es al revés (causalidad reversa). Por ejemplo, como recabamos al mismo tiempo prevalencia de obesidad y alteración de la glucemia, podemos suponer que la alteración de la glucemia produce obesidad y en realidad es la obesidad la que produce alteración de la glucemia.

Sin embargo, estos estudios proporcionan información para generar hipótesis acerca de posibles factores de exposición asociados y/o causa/s de una enfermedad, que es el primer paso en la investigación etiológica.

POBLACIÓN DE ESTUDIO

La población de estudio es aquella a la cual se extrapolarán o aplicarán los resultados.

Como no es posible encuestar a toda la población, se debe obtener una muestra representativa de esta, lo que significa que la muestra debe tener todas las características de la población total.

Si la muestra no representa a la población de estudio, introduciremos lo que se denomina *sesgo*, error de diseño de un estudio de investigación por selección equivocada de la población (*bías* en inglés).

La mejor manera de seleccionar la muestra es eligiendo los sujetos al azar (muestreo al azar) por medio de un método conocido como aleatorización (conocido comúnmente como randomización, que proviene de la palabra inglesa *random*, que significa azar).

En el ejemplo del estudio de prevalencia de alteración de la glucemia, sería imposible estudiar a todos los hombres y mujeres ≥ 50 años de todo el país. Por lo tanto, lo que se hace es a partir, por ejemplo, de los padrones electorales, seleccionar al azar un porcentaje de hombres y mujeres ≥ 50 años de la población total de cada población del país. Este porcentaje elegido al azar es la muestra seleccionada con la que se realizará el estudio. La muestra debe ser proporcional a la distribución de la población de todo el país. Por ejemplo, si en Capital Federal y Gran Buenos Aires se encuentra el 50% de los sujetos con las características para ingresar al estudio de investigación en todo el país, en la muestra seleccionada el 50% de los sujetos debe provenir de estas áreas geográficas.

PARTICIPANTES DEL ESTUDIO

Una vez que se han elegido los sujetos que participarán en el estudio, se deben realizar todos los esfuerzos posibles para obtener TODA la información deseada de TODOS los sujetos elegidos, de otra manera introduciremos un sesgo.

Los sujetos que no quieren participar pueden tener características muy diferentes de los que participan, con lo cual la población de estudio ya no será igual que la población de la muestra seleccionada y por lo tanto los resultados obtenidos podrían no ser los correctos, además de que no pueden ser aplicados a la población originalmente seleccionada como población de interés.

Por ejemplo, si muchos sujetos obesos no desean participar, podría no ser posible establecer una relación entre este factor de exposición y alteración de la glucemia, y los datos obtenidos serían sólo aplicables a sujetos no obesos.

De los sujetos que no desean participar hay que obtener la mayor cantidad de información posible (generalmente sexo, edad, estatus socioeconómico) para determinar si es una población muy diferente a la que está participando del estudio.

Si la población que rechaza participar en el estudio es diferente a la que está participando, se dice que la población incluida está sesgada, específicamente se introdujo un sesgo de selección.

Si esta situación no es detectada, estaremos cometiendo un error de diseño al inicio del estudio, que no puede corregirse una vez finalizado este.

RECOLECCIÓN DE DATOS

Antes de comenzar el estudio hay que definir claramente el/los factores de exposición y evento/s de interés. Esto es fundamental para que todos los investigadores recaben la información de forma homogénea. Si, por ejemplo, no se define claramente el criterio de obesidad, algunos investigadores podrían clasificar sujetos no obesos como obesos y viceversa. Si esto ocurre se introduce un sesgo u error.

Los métodos específicos que se usarán para recolectar los datos dependerán de los factores de exposición y de los eventos a relevar, del tipo de estudio que se desee realizar, prospectivo o retrospectivo, así como de los recursos económicos y humanos.

Por ejemplo, se pueden utilizar entrevistas personales, exámenes físicos, estudios complementarios, cuestionarios autoadministrados, historias clínicas, etc.

En todo estudio de investigación se recaba información de muchas otras variables para caracterizar a la población y determinar, en el caso en que se desee hacer una comparación de grupos, si los grupos son similares o diferentes en relación a sus características. Estas variables también podrían estar influenciando la asociación entre el factor de exposición y evento, lo que se denomina factor de confusión.

ANÁLISIS DE DATOS

Los resultados principales se reportan como porcentaje de factor de exposición y/o eventos referidos como prevalencia, lo adecuado es informarla por sexo y grupo de edad.

Si lo que se desea es reportar asociación entre factor de exposición y evento, la medida estadística que se utiliza es la razón de chance u *odds ratio*.

INTERPRETACIÓN DE LOS RESULTADOS

Al observar los resultados de un estudio hay que preguntarse qué errores pudieron haberse cometido al diseñar, recolectar o analizar los datos: siempre hay que pensar en los posibles sesgos. También hay que evaluar la posible existencia de confundidores y el tipo de análisis estadístico empleado.

Además de estos errores, en los estudios transversales hay que tener en cuenta que lo que se está midiendo es prevalencia, que como veremos más adelante depende de otros factores característicos de cada población, por lo que la comparación de prevalencias entre diferentes poblaciones no siempre es válida.

VENTAJAS Y DESVENTAJAS

Ventajas:

- 1) Son fáciles de realizar, económicos y relativamente rápidos.
- 2) Proveen importante información acerca de la distribución y amenaza para la salud pública que los factores de exposición y eventos pueden representar para la población.
- 3) Son el primer paso en los estudios etiológicos.

Desventajas:

- 1) Miden sólo prevalencia, por lo cual tienen valor limitado para la evaluación de asociación entre factor de exposición y evento.
- 2) No es posible establecer una secuencia temporal entre el factor de exposición y el evento.
- 3) No son útiles para factores de riesgo y/o eventos poco frecuentes.

ESTUDIOS DE COHORTES

Los estudios de cohortes son estudios observacionales en los que el punto de inicio es la exposición a un factor de riesgo y luego se efectúa el seguimiento en el tiempo del grupo/s observado/s de sujetos para recolectar información acerca de la aparición del/los evento/s de interés.

Por definición, *los estudios de cohortes son entonces observacionales y longitudinales.*

Al igual que los estudios transversales, los de cohortes se clasifican en descriptivos o analíticos y prospectivos o retrospectivos.

ESTUDIOS DE COHORTES DESCRIPTIVOS

En este tipo de estudios se selecciona una población expuesta a un factor de exposición de interés, que no presente el/los evento/s de interés al momento de ingresar al estudio (a la que se denomina población en riesgo, en riesgo de sufrir el/los evento/s) y se la sigue en el tiempo para determinar la incidencia de uno o varios eventos.

Como el estudio es longitudinal, al valorar la frecuencia de eventos no hablamos de Prevalencia, sino de Incidencia, puesto que el objetivo de estos estudios es detectar sujetos que desarrollan el o los eventos de interés.

$$\text{Incidencia} = \frac{\text{N}^\circ \text{ de nuevos casos en un período de tiempo}}{\text{N}^\circ \text{ de personas en riesgo de sufrir el evento al inicio del período de tiempo de seguimiento}}$$

La medida que acabamos de definir como incidencia es también llamada *incidencia acumulativa*.

Si estos estudios sólo describen lo que le sucede a la población en términos de incidencia de eventos, se los denomina estudios de historia natural de la enfermedad.

Por ejemplo, podemos seleccionar pacientes con insuficiencia renal moderada y seguirlos en el tiempo para determinar la incidencia de insuficiencia renal severa y/o necesidad de diálisis a 5 años de seguimiento. En este caso, el cálculo de la incidencia de insuficiencia renal severa nos permitirá estimar cuál es el riesgo de sufrir el evento o condición. Entonces, la principal utilidad de la incidencia es poder estimar el riesgo absoluto o riesgo de la población.

En este ejemplo queda claro que el estudio es observacional y longitudinal.

La característica distintiva de un estudio de cohorte es la relación temporal entre factor de exposición y evento. Siempre el factor de exposición está presente al inicio del estudio, en una población sin el/los eventos, los cuales van ocurriendo en el transcurso del tiempo de seguimiento.

Si lo que se desea es establecer una asociación entre un factor de exposición y un evento, entonces debemos diseñar un estudio de tipo analítico.

ESTUDIOS DE COHORTES ANALÍTICOS

En un estudio de cohortes analítico el investigador selecciona un grupo de sujetos expuestos a un factor de exposición y otro grupo no expuesto, ambos grupos sin el evento de interés, y sigue a ambos grupos en el tiempo para comparar la incidencia de eventos de interés al final del seguimiento.

Factor de exposición Evento	Obesidad Sí (n)	Obesidad No (n)	Total
Alteración de la glucemia			
Sí	a 25	b 25	50
No	c 5	d 45	50
Total	30	70	100

Tabla 3.4: Tabla de contingencias 2 x 2. Obesidad como factor de exposición y alteración de la glucemia como evento

Los resultados obtenidos se analizan utilizando una tabla de contingencias.

Supongamos que seleccionamos una población de obesos y no obesos y los seguimos 5 años para determinar la incidencia de diabetes.

El cálculo de la incidencia de diabetes entre los expuestos y los no expuestos surge de los datos de la tabla:

$$\text{Incidencia de alteración de la glucemia en expuestos} = \frac{a}{a+c} = \frac{25}{30} = 0,83 \times 100 = 83\%$$

$$\text{Incidencia de alteración de la glucemia en no expuestos} = \frac{b}{b+d} = \frac{25}{70} = 0,36 \times 100 = 36\%$$

La incidencia de diabetes en el grupo expuesto es de 83% y en la de no expuestos 36%. Esta diferencia en la incidencia de diabetes entre los grupos será analizada luego con métodos estadísticos para determinar si existe asociación entre obesidad y alteración de la glucemia.

ESTUDIOS DE COHORTES PROSPECTIVOS Y RETROSPECTIVOS

Utilizando el ejemplo de un estudio que tenga como evento alteración de la glucemia y factor de exposición obesidad, si los datos de obesidad se recolectan a partir de información generada al inicio del estudio, mediante el cálculo de índice de masa corporal, pesando y midiendo altura de los sujetos, estos datos serán prospectivos y el estudio una cohorte prospectiva.

Como la información acerca de los eventos se irá recolectando a medida que estos aparezcan en el tiempo, estos datos serán siempre prospectivos.

Algunas veces la información sobre el factor de exposición, dependiendo de la naturaleza del mismo, podría ser retrospectivo, pero el estudio sigue siendo prospectivo debido a que el objetivo de este tipo de estudios es determinar la incidencia de evento/s.

¿Cómo se entiende entonces qué es un estudio de cohorte retrospectivo?

Supongamos que el estudio prospectivo sobre incidencia de alteración de la glucemia y asociación con obesidad ha concluido. En este estudio se han relevado datos de muchas variables, por ejemplo edad, sexo, sedentarismo, antecedentes familiares de obesidad, hipertensión arterial, tipo de dieta, tabaquismo, entre muchos otros.

Pensemos que, un tiempo después, algún investigador involucrado en el estudio inicial tiene interés en investigar si el tipo de dieta está asociado al desarrollo de alteración de la glucemia. Como dispone de los datos del estudio original, los utiliza para investigar este

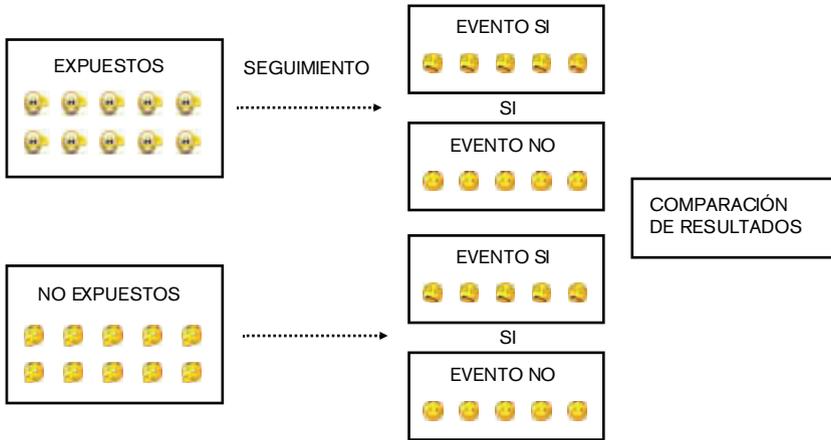


Figura 3.3: Diseño de un estudio de cohortes analítico

nuevo interrogante. Los datos ya han sido recabados y por lo tanto ya existen, es decir que para este nuevo estudio los datos son retrospectivos. Como el factor de exposición ha sido relevado tiempo antes del desarrollo del evento, el estudio sigue siendo un estudio de cohorte, en este caso retrospectivo.

UTILIDAD DE LOS ESTUDIOS DE COHORTES

Este tipo de estudios son particularmente útiles cuando:

- 1) Se desea investigar la exposición a un factor de exposición raro o de poca frecuencia.
- 2) Cuando hay interés en evaluar más de un evento.
- 3) Cuando se quiere establecer temporalidad entre la exposición al factor de exposición y la aparición del evento. Esto refuerza la asociación en términos de causalidad.

SELECCIÓN DE LA POBLACIÓN DE ESTUDIO

En un estudio de cohorte descriptivo la selección de la población se realiza obteniendo al azar una muestra de la población total expuesta al factor de exposición, de manera que la muestra sea representativa de la población a la cual extrapolaremos los datos.

Si el estudio de cohortes es analítico y el factor de exposición es común en la población general, se selecciona al azar una muestra de sujetos expuestos y no expuestos de una misma población, quienes deben ser lo más parecidos posible en sus características, salvo la presencia del factor de exposición.

En este caso se dice que usamos una *comparación interna*. Estamos comparando expuestos vs. no expuestos dentro de una misma población.

En el caso de que el factor de exposición no sea frecuente, debemos primero seleccionar la población expuesta y luego elegir una población no expuesta adecuada.

En estudios que analizan contaminantes ambientales como factor de exposición asociado a una enfermedad, se elige como población expuesta a los trabajadores de una determinada industria y sería posible elegir de una misma fábrica un grupo de trabajadores que no estén expuestos directamente al factor de exposición, por ejemplo trabajadores de un reactor nuclear comparado con personal administrativo de la misma planta.

Con eso logramos tener un grupo de comparación interna y por lo tanto ambos grupos serán comparables, porque provienen de una misma población.

Pero si el grupo de comparación interna tiene alguna probabilidad de estar expuesto al factor de exposición, se debería buscar un grupo de no expuestos de una población diferente (comparación externa).

En este caso la dificultad es encontrar un grupo que sea comparable al de expuestos en todas sus características, a excepción de la exposición al factor de riesgo. Se puede utilizar como grupo de comparación a trabajadores de otro tipo de industria en la que no exista el factor de exposición a estudiar.

Si utilizamos una muestra obtenida de la población general podemos introducir el sesgo del *trabajador sano*. En general los trabajadores son más sanos que la población general, por lo que el resultado final sería que el factor de exposición no incrementa la incidencia del evento cuando la comparamos con la población general. Para evitar este sesgo se comparan grupos con diferentes niveles de exposición dentro de la misma población.

CLASIFICACIÓN DE LA POBLACIÓN EN EXPUESTOS Y NO EXPUESTOS

La definición del factor de exposición y la correcta clasificación de los sujetos en expuestos o no expuestos, de acuerdo a una única definición mediante la cual los investigadores clasifiquen de la misma forma a expuestos y no expuestos, es muy importante debido a que una clasificación equivocada puede resultar en una falsa estimación de la asociación entre el factor de exposición y el evento.

Si los investigadores no siguen la clasificación de los sujetos de acuerdo con la definición adoptada por el estudio, se introduce un sesgo. Algunos sujetos con la exposición podrían ser clasificados como no expuestos o viceversa.

Para determinar la situación de exposición de los sujetos, con frecuencia se recurre a diferentes fuentes de información: entrevistas, registros médicos, registros del lugar de trabajo, determinación del factor de exposición en sangre, examen físico, otros estudios, etc.

En los estudios de cohortes es frecuente que el factor de exposición cambie en el tiempo, por ejemplo un fumador puede abandonar el hábito durante el estudio. Es por esto que muchos factores de exposición se evalúan varias veces durante el transcurso del estudio. Para el análisis de estas situaciones hay técnicas que permiten controlar este problema.

SEGUIMIENTO DE LA POBLACIÓN DE ESTUDIO

En general, una vez clasificada correctamente la población como expuestos y no expuestos, y obtenidos los datos de todas las variables que caracterizan ambos grupos y que podrían ser confundidoras, se comienza el período de seguimiento. Este, dependiendo del evento de interés, puede durar varios años, lo que resulta en pérdida de sujetos y costos elevados.

Una opción es realizar un estudio de cohorte retrospectivo.

Si bien esto ahorra mucho tiempo, la veracidad de los datos puede ponerse en duda porque no se determinó previamente la definición de factor de exposición y podrían no haberse recabado datos de variables posiblemente confundidoras para la nueva asociación a estudiar.

DETERMINACIÓN DEL EVENTO

Durante el seguimiento es necesario hacer todos los esfuerzos posibles para garantizar que el seguimiento de los grupos expuesto y no expuesto sea completo y asegurarse de que sea comparable, y así evitar el sesgo de seguimiento.

Es por eso que el seguimiento debe realizarse mediante contactos frecuentes con los sujetos, ya sea con entrevistas, consultas, contactos telefónicos, etc., para no perder información acerca de la aparición del evento de interés.

El evento debe ser asignado usando una definición estandarizada previamente al inicio del estudio y el investigador que lo asigna debe estar *ciego* respecto de la situación de expuesto o no expuesto del sujeto para evitar sesgos en la asignación del evento.

Como en todo estudio también se recaban datos de muchas otras variables para describir a la población o a los grupos a comparar y para tener en cuenta la posibilidad de presencia de factores de confusión.

ANÁLISIS DE DATOS

La frecuencia de eventos se mide en términos de Incidencia (número de nuevos casos) en relación con un número de sujetos. Por ejemplo, incidencia de cáncer de mama a un año podría ser de 2% o, lo que es lo mismo, 2 por cada 100 sujetos seguidos. A esto se lo denomina *riesgo*.

Como en los estudios de cohortes no todos los sujetos participan el mismo tiempo durante el seguimiento, se puede calcular la incidencia con relación al número de sujetos que participaron *en promedio* por cada año de seguimiento. A esto se lo denomina *tasa*. En este caso, si todos los participantes fueron seguidos por un año, la tasa de cáncer de mama sería de 2 casos cada 100 personas/año.

La estimación del riesgo es útil cuando todos los sujetos completaron el mismo tiempo de seguimiento; de lo contrario se utiliza como estimación de eventos el cálculo de tasa de eventos.

En los estudios de cohortes analíticos se utiliza como medida de asociación o efecto la razón de riesgo o riesgo relativo (*risk ratio-relative risk*) o la razón de tasa (*rate ratio*) que desarrollaremos más adelante.

INTERPRETACIÓN DE LOS RESULTADOS

Como en todo estudio epidemiológico, antes de aceptar un resultado como verdadero hay que evaluar los posibles sesgos del estudio, el ajuste por confundidores y el método estadístico utilizado.

VENTAJAS Y DESVENTAJAS

Ventajas:

- 1) Elimina sesgos de información, debido a que el factor de exposición y el evento son definidos previamente a la realización del estudio.
- 2) Provee información sobre temporalidad, factor de exposición/evento, lo que refuerza la relación causal entre el factor de exposición y el evento.
- 3) Permite evaluar más de un evento.
- 4) Pueden estudiarse factores de exposición infrecuentes.

Desventajas:

- 1) Requieren mucho tiempo hasta obtener los resultados.
- 2) En los retrospectivos, los datos pueden no ser los correctos.
- 3) El factor de exposición puede cambiar en el tiempo de seguimiento.

- 4) Pueden perderse sujetos durante el seguimiento.
- 5) En aquellos con un seguimiento muy largo, el criterio diagnóstico del evento puede cambiar.

ESTUDIOS CASO-CONTROL

Los estudios caso-control son siempre observacionales y analíticos.

El estudio se inicia con la selección de sujetos enfermos (casos) que luego se compararán con sujetos sanos (controles) con relación al porcentaje de expuestos a un factor de exposición en el pasado que se observe en cada grupo.

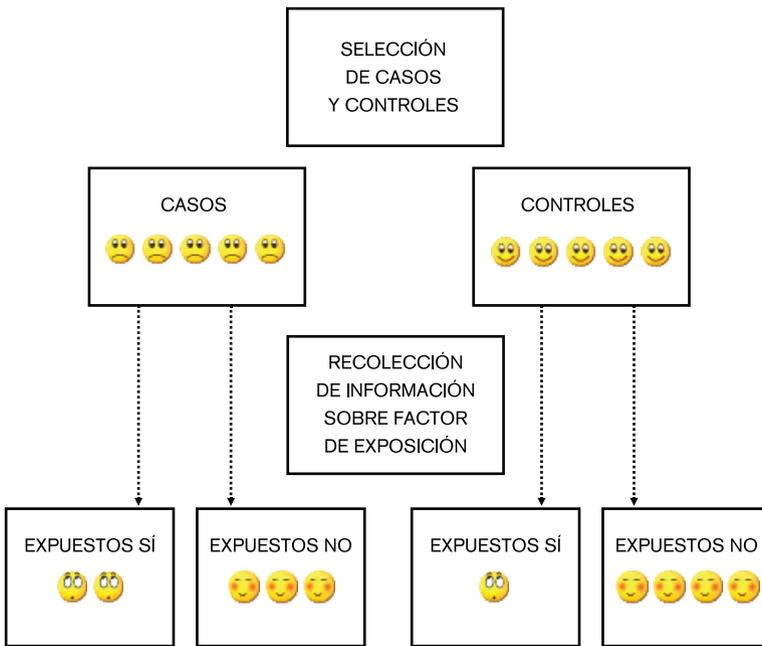


Figura 3.4: Diseño estudio caso-control

Como ejemplo podemos seleccionar 100 mujeres con cáncer de mama y 100 sin cáncer de mama y determinamos quiénes tomaron anticonceptivos orales. Los datos podrían analizarse como en la Tabla 3.6.

Vemos que, entre los casos de cáncer de mama, el 70% tomó anticonceptivos, mientras que en el grupo control sólo un 30% los utilizó.

En este caso no hablamos ni de prevalencia ni de incidencia.

La prevalencia de empleo de anticonceptivos orales no se puede determinar, porque no hemos elegido una población general: sólo tenemos una población de pacientes con y sin cáncer de mama, que no representa la población general.

Tampoco determinamos la incidencia de cáncer de mama, debido a que trabajamos sobre un grupo que ya presentaba esta enfermedad.

Evento Factor de exposición	Casos n	Controles n
Expuestos		
Sí	a	b
No	c	d
Total	a + c	b + d
Proporción de expuestos	$\frac{a}{a + c}$	$\frac{b}{b + d}$

Tabla 3.5: Tabla de contingencia 2x2 para estudio caso-control con los datos obtenidos

Evento Factor de exposición	Casos con cáncer de mama	Controles sin cáncer de mama
Expuestos		
Anticonceptivos No	30	70
Anticonceptivos Sí	70	30112
Total	100	100
Proporción de expuestos	70%	30%

Tabla 3.6: Estudio caso-control. Cáncer de mama y anticonceptivos

Por lo tanto, lo que hacemos es comparar la proporción de expuestos al factor de exposición entre casos y controles y calculamos (como medida de asociación) la razón de chances, conocida en inglés como *Odds Ratio* o por sus siglas OR.

UTILIDAD DE LOS ESTUDIOS CASO-CONTROL

Esta clase de estudio es útil para:

- 1) Investigar causas de nuevas enfermedades
- 2) Establecer la eficacia de las vacunas
- 3) Investigar enfermedades raras
- 4) Evaluar múltiples factores de riesgo

SELECCIÓN DE CASOS

En primer término, la definición de caso debe ser claramente especificada. Esta puede determinarse por la clínica o por los resultados de un estudio (invasivo, no invasivo, histológico o serológico).

Un sujeto puede clasificarse como portador de cáncer de próstata de acuerdo con diferentes definiciones: por exploración física, ecografía, toma de biopsia o marcadores serológicos.

El origen de la muestra de la que se obtendrán los casos debe definirse bien, ya que los controles deberán provenir de la misma población.

Los casos podrán ser todos los pacientes que cumplan con la definición de caso de una población específica que tengan la enfermedad. Deben ser representativos de todos los casos de una población.

Es importante especificar si los casos representan a todos los de la población de referencia (población a la cual serán extrapolados los datos) o si constituyen un subgrupo de casos (selección por asistencia a hospital). Para algunas enfermedades con diferentes grados de severidad, los pacientes que asisten a hospitales especializados pueden representar sólo los casos más severos; los pacientes con estadios menos avanzados podrían estar tratándose en otro tipo de centro de salud o seleccionarse de la población general.

Los casos pueden ser pacientes con diagnóstico reciente de la enfermedad y, a partir de este diagnóstico, ser incluidos en el estudio, o bien ser individuos a los que ya se les hizo el diagnóstico anteriormente.

En la primera situación se dice que el estudio incluye sólo casos nuevos, o *casos incidentes*, mientras que la elección de casos preexistentes se denomina selección de *casos prevalentes*.

SELECCIÓN DE CONTROLES

Esta es la parte más difícil de este tipo de estudios.

El punto crítico es que los controles deben ser representativos de la población que produjo los casos. Dicho en otras palabras, los controles deben ser iguales a los casos en todas sus características (distribución de sexo, edad, situación socioeconómica, nivel educacional, etc.) salvo por no padecer la enfermedad en estudio.

La población de donde se obtienen los controles debe ser la misma que la de los casos. Si los casos provienen de una muestra de la población general, los controles deben seleccionarse al azar de la misma población.

Existen distintas fuentes de información para la selección de controles que representen a la población de referencia: lista de pacientes registrados en centros de atención primaria, listas sujetos obtenidas de municipios o censos.

Todas estas fuentes de datos tienen limitaciones; es por eso que en general se usan todas las fuentes posibles de datos.

Cuando los casos son derivados de centros especializados, muchos estudios utilizan como controles a pacientes que asisten a estos centros y que no padecen la enfermedad en estudio, debido a que son más fáciles de reclutar.

Sin embargo, estos controles pueden no estar representando a la población que ha producido los casos en relación a uno o más factores de exposición, con lo cual se introduce un sesgo.

Si tomamos niños con leucemia de un hospital especializado de Buenos Aires al cual se derivan casos de todo el país, los controles que asisten a este centro no provienen de la misma población de los casos, probablemente sus lugares de residencia habitual y situación económica sean muy diferentes. Esto trae como consecuencia que, si estamos investigando un agente químico como causa de leucemia, no hallemos una relación entre el factor de exposición y la enfermedad.

Como muchas veces no es posible identificar una población de controles adecuada, se elige más de un grupo control: uno hospitalario y otro seleccionado de la población general, por ejemplo del área de influencia del hospital de donde se seleccionan los casos.

Una buena alternativa es seleccionar como control alguna persona amiga del caso del mismo sexo y edad similar (± 5 años). Al ser amigo/a comparten nivel socioeconómico, educación y lugar de residencia. Todas estas variables caracterizan muy bien a la población de casos y controles.

Lo que no se debe hacer es elegir un familiar, porque comparten muchas costumbres, lugar de residencia y otras numerosas características que lo hacen muy parecido al caso, lo que estaría muy bien, si no fuera porque pueden compartir el factor de exposición. Si esto ocurre, no sería posible encontrar diferencias en la frecuencia del factor de exposición entre casos y controles.

Otro problema es cuántos controles elegir por cada caso.

Al incluir más controles que casos, se aumenta la población y mejora la precisión estadística. La desventaja radica en que es difícil identificar controles adecuados.

Cuando se elige un control con sexo y/o edad similar a los casos, lo que se desea es hacer que las dos poblaciones sean más parecidas, sobre todo en las variables que podrían ser confundidoras. A este procedimiento se lo denomina *aparear*.

El aparear también tiene sus desventajas. Requiere de un análisis estadístico específico. Con esta técnica no es posible determinar el efecto de las variables elegidas para aparear sobre la relación factor de exposición y evento.

IDENTIFICACIÓN DEL FACTOR DE EXPOSICIÓN

Luego de haber incluido los casos y sus respectivos controles, hay que determinar en ambos grupos quiénes estuvieron o no expuestos al factor de riesgo. Hay diferentes formas de obtener esta información. Se puede realizar por medio de entrevistas, cuestionarios telefónicos o enviados por correo (esto es muy común en países con servicios de estadística de salud pública muy eficientes), entrevista a familiares, historias clínicas, datos del lugar de trabajo o bien de muestras biológicas.

Es muy importante determinar qué período de posible exposición es relevante. Si fumó hasta hace 20 años o hasta hace 5, desde cuándo está medicado por colesterol alto o bien si en los últimos 5 años recibió tal tratamiento, etc.

Esto es importante, sobre todo para enfermedades con período de latencia largo.

Independientemente del método utilizado para recabar los datos, lo importante es disminuir al mínimo la posibilidad de sesgos, lo que veremos en profundidad más adelante.

Si se adjudica mal el factor de exposición, la asociación entre factor de exposición y evento será falsa.

Para evitar estos sesgos hay que utilizar fuentes objetivas de datos de las cuales se haya recabado la información antes de que la enfermedad apareciera y tratar de que la persona que está recabando la información no sepa si el sujeto es un caso o un control.

Además de recabar los datos sobre la exposición, se deben obtener datos que describan ambas poblaciones (casos y controles) para poder compararlas y determinar cuán similares son. De existir diferencias en algunas variables, estas pueden estar influenciando o *confundiendo* la asociación entre el factor de exposición y el evento, a estas variables se las denomina *confundidoras*.

Supongamos que se desea conocer los factores de exposición asociados a cáncer de vejiga.

Se elige un número determinado de pacientes con diagnóstico reciente de cáncer de vejiga, definido por biopsia postoperatoria (casos incidentes) en hombres ≥ 60 años. Los enfermos (casos) serán seleccionados de centros especializados en el tratamiento de cáncer, así como de hospitales públicos de todo el país, para que la muestra sea representativa de la mayor cantidad de casos de cáncer de vejiga del país.

Como controles se elegirán obviamente hombres de edad similar a los casos, amigos, que no padezcan cáncer de vejiga (controles apareados por edad).

Como factores de exposición posiblemente asociados a cáncer de vejiga, se relevarán datos de antecedente de tabaquismo, trabajo relacionado a químicos, como anilinas, tinturas, hidrocarburos, etc. Para corroborar estos datos se interrogará a los sujetos y se tratará de obtener algún tipo de evidencia objetiva acerca de la exposición. Esto es importante para clasificar correctamente expuestos y no expuestos en la muestra de casos y controles.

Como variables posiblemente confundidoras de la asociación entre cáncer de vejiga y estos factores de exposición se relevarán datos acerca de tipo de centro de atención, tipo de tratamiento, nivel educacional, estatus socioeconómico, tanto en casos como en controles. De estas variables también se tratará de obtener evidencia objetiva.

El estatus de expuesto debe ser definido al inicio del estudio, por ejemplo cantidad de cigarrillos fumados, duración del hábito, fecha de abandono (si aplica); en el caso de exposición a químicos, inicio y duración de la exposición, podría realizarse algún tipo de análisis para detectar residuos de estos químicos, si el sujeto aún trabaja o dejó de hacerlo hace poco tiempo antes del diagnóstico de cáncer de vejiga. En todos los casos se definirá un tiempo mínimo de exposición.

Una vez obtenida la muestra de casos y de controles, se procede a obtener los datos sobre los factores de exposición y posibles variables confundidoras. Luego se cuantifica la frecuencia de factores de exposición en casos y controles y se realizan pruebas estadísticas adecuadas para establecer si existe asociación entre los factores de exposición investigados y cáncer de vejiga.

ANÁLISIS DE DATOS

Como ya hemos mencionado, los estudios de caso-control son analíticos, por lo que se comienza con la construcción de una tabla de contingencia 2 x 2, que ya hemos presentado.

Como en este tipo de estudio no recabamos datos de prevalencia del factor de riesgo ni incidencia de evento, se utiliza como medida de asociación el *odds ratio* (razón de chances).

Desde ya, si el estudio es apareado, lo correcto es construir una tabla de contingencia 2 x 2 para datos apareados, realizar el cálculo del *odds ratio* y las pruebas estadísticas con las técnicas correctas para datos apareados.

INTERPRETACIÓN DE LOS RESULTADOS

Ya hemos mencionado que es posible cometer errores en el diseño (sesgos) y en el análisis de los resultados, por no tener en cuenta las variables confundidoras y por no aplicar pruebas estadísticas adecuadas al diseño del estudio y tipo de datos que se están analizando.

Particularmente para los estudios caso-control, los sesgos son muy frecuentes si no se han tomado medidas necesarias para evitarlos.

Como el factor de exposición se mide después de haber aparecido el evento (o alguno de sus síntomas) es posible que el factor de exposición se haya visto influido por el evento.

Por ejemplo, si investigamos la dieta actual como factor de riesgo para infarto de miocardio, lo más probable es que la dieta haya cambiado como consecuencia de la enfermedad. A este hecho se lo denomina causalidad reversa, debido a que el evento influyó el factor de exposición. Si se recaban datos de la dieta actual, no estaremos midiendo el factor de exposición adecuado.

Es por esto que es tan importante determinar el período de tiempo en el que el sujeto estuvo expuesto previamente a la aparición del evento.

VENTAJAS Y DESVENTAJAS

Ventajas:

- 1) Son estudios que se realizan rápidamente.
- 2) Resultan útiles para enfermedades raras y de largo período de latencia.
- 3) Permiten estudiar más de un factor de exposición.

Desventajas:

- 1) El control de los sesgos suele ser difícil.
- 2) No pueden determinar la secuencia de eventos entre exposición al factor de exposición y evento (causalidad reversa).
- 3) No son útiles para factores de exposición infrecuentes.
- 4) No permiten determinar prevalencia ni incidencia.

ESTUDIOS CASO-CONTROL ANIDADOS

Este tipo de estudio suele resolver muchos de los problemas de sesgos que tienen los estudios caso-control.

Se realizan mientras se está llevando a cabo un estudio de cohorte.

Durante la fase de seguimiento, algunos sujetos irán presentando eventos y serán catalogados como casos; a los sujetos que hasta ese momento no tengan eventos se los clasificará como controles.

Como la información sobre el factor de exposición de interés y otros factores que pueden ser confundidores se recabaron al inicio del estudio, es posible determinar cuáles son los sujetos que, dentro de los casos y controles, estuvieron expuestos a un factor de exposición de interés determinado para el estudio caso-control anidado.

La gran ventaja de este tipo de estudio es que la información sobre los factores de riesgo se obtuvo antes de que apareciera el evento, con lo cual se evitan sesgos y causalidad reversa.

Como los casos y controles derivan de la misma población, el sesgo de selección se reduce sustancialmente.

Por ejemplo, si estamos siguiendo niños desde su nacimiento y tenemos datos de peso al nacer, anomalías en el parto, tipo y tiempo de lactancia, enfermedades del primer año, rendimiento escolar, etc., es posible realizar un estudio caso-control anidado a los 7 años de seguimiento para estudiar causas de bajo rendimiento escolar.

Seleccionamos aquellos niños con bajo rendimiento como casos y a los de rendimiento normal como controles, y calculamos la proporción de bajo peso al nacer en cada grupo como factor de exposición.

Como hemos observado, llevar a cabo un estudio caso-control es difícil, dada la dificultad de encontrar controles adecuados a la cantidad de sesgos que pueden cometerse y a la dificultad de determinar con exactitud la presencia o ausencia del factor de exposición en el pasado.

ESTUDIOS EXPERIMENTALES

Son similares a los estudios de cohortes analíticos. En lugar de observar el estatus de expuesto, el investigador asigna el factor de exposición de interés a un grupo y al otro ningún factor o algún otro que sirva para realizar la comparación. Por lo tanto, los estudios

experimentales son estudios de cohortes donde la situación de expuesto o no expuesto es determinada por la investigación.

Ambos grupos son seguidos en el tiempo para observar si desarrollan uno o varios eventos de interés.

El patrón oro o método de referencia (*gold standard*) de los estudios de intervención son los llamados estudios controlados aleatorizados doble ciego o doble enmascaramiento, paralelos, conocidos comúnmente como estudios randomizados (*random*, del inglés azar).

Controlados significa que un grupo recibe la intervención de interés (factor de exposición) mientras que el otro grupo no; este último es el denominado *grupo control*. Este grupo es necesario para que la incidencia de eventos en el grupo de intervención pueda ser comparada.

Por *aleatorización* entendemos la asignación al azar de la intervención de interés entre los sujetos elegidos para participar del estudio. La principal ventaja de aleatorizar es que asegura que ambos grupos sean comparables. Elimina los sesgos de selección, igualando las variables confundidoras, ya sean conocidas o no. También elimina el sesgo de asignación. Todos los sujetos tienen la misma probabilidad de pertenecer al grupo intervención o al grupo control. Si la aleatorización se realizó correctamente, ambos grupos serán iguales en todas sus características, salvo en la intervención.

El término *doble ciego* se refiere a que ni el sujeto ni el investigador conocen si están recibiendo la intervención o no. Esto es para eliminar sesgos, si el sujeto sabe que recibe la intervención puede, por ejemplo, reportar más efectos adversos, o bien si el investigador sabe si el sujeto está tratado con la intervención y cree en ella, podría no reportar ciertos eventos, con lo cual la incidencia de eventos en el grupo intervención podrían ser inferiores a la real.

Esto tendría como consecuencia mayor frecuencia de efectos adversos y menor incidencia de eventos en el grupo tratado con la intervención. Lo mismo ocurriría en el grupo control, podrían reportar menos efectos adversos y el investigador, mayor cantidad de eventos.

Los resultados del estudio, si ocurre esto, no serían los *verdaderos*, estarían sesgados.

El término estudio paralelo se refiere a que los sujetos asignados a una determinada intervención continúan hasta el final del estudio con la misma intervención.

Los estudios experimentales cumplen con todas las características de los estudios de cohorte, salvo por la introducción del factor de exposición por parte del investigador.

Veamos ahora aspectos especiales de los estudios experimentales.

OBJETIVO DEL ESTUDIO

El primer paso para realizar un estudio de intervención es definir la pregunta que deseamos que el estudio conteste.

Por ejemplo, ¿afecta la intervención la incidencia de eventos (betabloqueantes en el infarto de miocardio), modifica la historia natural de la enfermedad (antirretrovirales en HIV)?

Por lo general la pregunta debe ser más específica.

En un estudio experimental normalmente hay varias preguntas que se desea responder, pero es fundamental determinar cuál de ellas es la más importante o factible de responder con certeza estadística.

A la pregunta principal del estudio se la llama *objetivo primario* del estudio, y a las otras se las denomina *objetivos secundarios*.

La importancia de determinar el objetivo primario con claridad radica en que este determina qué eventos van evaluarse durante el seguimiento. A su vez, la tasa de eventos va a determinar el número de sujetos que se deben incluir en el estudio.

A los eventos que serán evaluados para responder el objetivo primario se los denomina punto final primario.

El objetivo primario de un estudio podría ser determinar si una droga reduce la hospitalización de una determinada patología (punto final primario) y el objetivo secundario determinar si existe algún impacto en la mortalidad (punto final secundario).

La tasa de hospitalizaciones de una enfermedad suele ser mayor a su mortalidad, por lo que la evaluación de este punto es más factible y seguramente necesitará menos sujetos.

Una vez definidos los objetivos primarios y secundarios del estudio debemos definir la población a incluir en el estudio o población de interés, determinar cuáles son los eventos de interés y determinar si deseamos realizar un estudio de eficacia o efectividad.

DEFINICIÓN DE LA POBLACIÓN DE INTERÉS

Es la población en la que se desea evaluar una intervención, para determinar si existe beneficio con la intervención y qué grado de perjuicio puede acarrear la implementación de la misma, llamada población objetivo o población blanco (*target population*), que es la población a la cual los datos serán extrapolados.

¿Reducen los betabloqueantes la incidencia de muerte posinfarto? ¿En quiénes? ¿En la población general de infartos (todos los infartos), en los sujetos de alto riesgo (infartos con insuficiencia cardíaca) o en los ancianos?

Los efectos adversos de los betabloqueantes postinfarto de miocardio ¿son los mismos en la población general de infartos, en pacientes con insuficiencia cardíaca o en ancianos?

Una vez determinada la población, se debe seleccionar la población de estudio. Esta es una muestra representativa de la población de interés, por ejemplo pacientes postinfarto mayores de 70 años con determinadas características.

Los criterios de inclusión y los criterios de exclusión de un estudio son los que en última instancia definen la población de estudio.

DETERMINACIÓN DEL/LOS EVENTO/S DE INTERÉS

La definición del evento de interés principal o punto final primario del estudio es muy importante, ya que a través de él se responderá la pregunta del objetivo primario y se determinará el tamaño de la muestra del estudio.

Si el punto final es determinar la eficacia de un antibiótico en el tratamiento de la neumonía intrahospitalaria, esta pregunta podría responderse de varias maneras: por una reducción de la mortalidad, del tiempo de internación, de la necesidad de asistencia respiratoria mecánica, etc.

La elección del punto final primario o evento de interés depende de muchos factores.

Si la tasa de mortalidad es muy baja y no tiene un impacto importante en la salud pública o la población afectada es escasa, no tiene mucho sentido tomar este evento como punto final primario. En cambio, si el impacto de esta enfermedad en el tiempo de internación o en la tasa de necesidad de asistencia respiratoria es muy alto, parece adecuado tomar alguno de estos eventos como punto final primario.

Muchos estudios tienen como objetivo o punto final determinar si la intervención reduce la morbilidad, lo que implica determinar la incidencia de varios eventos.

Es muy frecuente en la actualidad observar que los estudios clínicos aleatorizados evalúan tasa de rehospitalización, progresión a estadios más severos de la enfermedad, incidencia de intervenciones como cirugía, o bien aparición de enfermedades que son parte de la historia natural de una enfermedad.

En un estudio aleatorizado en el que se explora el beneficio de una droga hipoglucemiante se puede determinar como objetivo primario la evaluación conjunta de deterioro de la función renal, hospitalizaciones por descompensación (hiper o hipoglucemia), incidencia de infarto de miocardio o desarrollo de angina de pecho, etc.

Otros estudios agregan a la morbilidad la evaluación del efecto de la intervención sobre la mortalidad de la enfermedad.

Cuando el objetivo primario de un estudio evalúa más de un evento se denomina *punto final primario combinado*.

Las razones para elegir este tipo de punto final generalmente responden a cuestiones estadísticas relacionadas con el tamaño de la muestra.

En algunas situaciones, como por ejemplo una tasa de eventos baja y/o un seguimiento demasiado extenso, es posible elegir un evento intermedio. Este es un evento que se encuentra en la cadena fisiopatológica de la enfermedad antes de la aparición de eventos más severos. También se los denomina *puntos finales surrogantes*.

Después de un infarto de miocardio la función ventricular izquierda se modifica. Si se deteriora lo suficiente puede llevar a insuficiencia cardíaca y esta a la muerte. Si deseamos determinar si la intervención previene el desarrollo de insuficiencia cardíaca, lo cual puede llevar años, podemos elegir como punto final la evaluación de la función ventricular izquierda como surrogante de insuficiencia cardíaca.

Para utilizar surrogantes hay que estar seguros de que el evento intermedio sea una determinación válida del evento final, en otras palabras, si aparece el surrogante invariablemente aparecerá el evento más serio.

Esto tiene sus desventajas, porque podríamos no detectar un incremento en la morbilidad o mortalidad debido a un efecto adverso de la intervención.

Si evaluamos un tratamiento antirretroviral para el tratamiento del VIH a través del número de CD 4 circulantes, no estamos valorando la morbilidad que estas drogas pueden causar, que en última instancia podrían provocar la muerte del paciente.

¿EFICACIA O EFECTIVIDAD?

Se define *eficacia* (por ejemplo de una droga) como el efecto bajo condiciones establecidas por el estudio clínico.

La eficacia se evalúa cuando se desea determinar el máximo beneficio que puede alcanzarse con una intervención. Para lograr este objetivo usualmente se exige a los sujetos la máxima adherencia al tratamiento asignado en el estudio. Estos sujetos son vistos frecuentemente para reforzar la adherencia al tratamiento y para determinar si ocurrió algún evento de interés o efecto adverso.

La evaluación de efectividad determina el efecto de una intervención bajo circunstancias habituales en la práctica médica. En este caso no se exige la máxima adherencia, sólo se obtienen datos del porcentaje de adherencia y se evalúa el efecto de la intervención en un contexto más *real*.

En la práctica diaria es común observar que los pacientes se olvidan de tomar la medicación o la abandonan. En estas circunstancias el efecto de la intervención será menor, es por esto que los estudios de efectividad reflejan el efecto de una intervención más cercana a lo que ocurre en la práctica diaria.

La decisión de evaluar eficacia o efectividad depende del objetivo del estudio y de la factibilidad de exigir una adherencia del 100%.

CÁLCULO DEL TAMAÑO DE LA MUESTRA

Una vez definida la población de interés y los eventos a relevar, debemos calcular cuántos sujetos será necesario incluir en cada grupo con el propósito de observar una diferencia de eventos lo suficientemente importante para determinar, al final del estudio sin errores, el grado de beneficio, si lo hay, que produce la intervención.

Este aspecto será desarrollado en un capítulo destinado específicamente a este procedimiento.

ASIGNACIÓN DE LA INTERVENCIÓN. ALEATORIZACIÓN

La mayor ventaja cuando la asignación de tratamientos es mediante aleatorización es que los grupos serán similares con respecto a todas las variables confundidoras, excepto por la intervención.

Esto es lo que muestra la Tabla 3.7 cuando se publican los estudios aleatorizados. Ninguna de las características entre los grupos que se quieren comparar mostrará diferencias significativas.

Esto hace que los grupos sean completamente comparables, aun para variables desconocidas, por ejemplo clase social, lugar de residencia (ciudad o campo), etc.

Por medio de la aleatorización evitamos sesgos de selección. El investigador no tiene control sobre qué paciente recibirá una u otra intervención y por lo tanto los eventos en el seguimiento no estarán influenciados por la preferencia del investigador.

Aleatorizar significa que cada sujeto tiene la misma probabilidad de ser asignado a uno u otro tratamiento, y nadie, ni los investigadores ni los sujetos, puede predecir cuál será la asignación.

DIFERENTES TIPOS DE ALEATORIZACIÓN

ALEATORIZACIÓN SIMPLE

La asignación de tratamiento se realiza de acuerdo con una lista generada al azar por una computadora.

En general la lista se basa en números. Si los números son pares se asigna un tratamiento y si son impares el otro. Los pacientes 2, 4, 6, 8, etc. serán asignados a un tratamiento, los sujetos 1, 3, 5, 7 etc. a otra intervención.

Otra forma es generar una secuencia de asignación completamente al azar, es asignar al azar la intervención, como vemos en el ejemplo.

A medida que los sujetos son incorporados al estudio, la asignación del factor de exposición se realiza en base a la tabla de aleatorización, donde se observa el número de paciente incorporado y la asignación del factor de exposición. El primero, segundo sexto u octavo sujeto incorporado serán asignados a la intervención. Los sujetos incorporados en tercero, cuarto y quinto lugar se asignan al grupo control.

Como observamos, la asignación de la intervención es completamente aleatoria, pero al final, con la inclusión del paciente número 50, tendremos la misma cantidad de sujetos en el grupo intervención y en el grupo control.

INTERVENCIÓN 1	CONTROL 3
INTERVENCIÓN 2	CONTROL 4
INTERVENCIÓN 6	CONTROL 5
INTERVENCIÓN 8	CONTROL 7
INTERVENCIÓN 9	CONTROL 10
INTERVENCIÓN 11	CONTROL 12
INTERVENCIÓN 15	CONTROL 13
INTERVENCIÓN 16	CONTROL 14
INTERVENCIÓN 17	CONTROL 18
INTERVENCIÓN 19	CONTROL 20
INTERVENCIÓN 24	CONTROL 21
INTERVENCIÓN 25	CONTROL 22
INTERVENCIÓN 26	CONTROL 23
INTERVENCIÓN 27	CONTROL 30
INTERVENCIÓN 28	CONTROL 35
INTERVENCIÓN 29	CONTROL 36
INTERVENCIÓN 31	CONTROL 38
INTERVENCIÓN 32	CONTROL 41
INTERVENCIÓN 33	CONTROL 42
INTERVENCIÓN 34	CONTROL 43
INTERVENCIÓN 37	CONTROL 44
INTERVENCIÓN 39	CONTROL 45
INTERVENCIÓN 40	CONTROL 46
INTERVENCIÓN 47	CONTROL 49
INTERVENCIÓN 48	CONTROL 50

Tabla 3.7: Aleatorización simple

ALEATORIZACIÓN POR BLOQUES

Es una forma de asegurar que cada determinado número de sujetos asignados habrá casi la misma cantidad de sujetos en un grupo y el otro.

En una aleatorización de 4 sujetos por bloque, cada 4 sujetos asignados habrá 2 en la rama intervención y la misma cantidad en la rama control. Esto no implica saber la secuencia de asignación, puesto que, por cada 4 sujetos, esta secuencia varía siempre.

Si, por ejemplo, detenemos el estudio luego de incluir el paciente número 22, veremos que en el grupo intervención se incluyeron 10 sujetos y en el grupo control 12 sujetos. Si la aleatorización fuera simple podría haber ocurrido que, al suspender el estudio, por azar tuviéramos 5 sujetos en la intervención y 17 en el grupo control, con lo cual la comparación de estos grupos sería poco válida.

La aleatorización por bloques permite, en caso de detener el estudio precozmente, tener una cantidad similar de sujetos en cada grupo y poder hacer las comparaciones necesarias.

ALEATORIZACIÓN ESTRATIFICADA

Este tipo de aleatorización es útil cuando alguna variable tiene una gran influencia sobre el evento de interés.

Supongamos que deseamos determinar si una droga reduce la tasa de infarto de miocardio pero sabemos que el hecho de ser diabético eleva mucho el riesgo de infarto y la

evolución de este evento es peor en sujetos diabéticos. Para asegurarnos de que diabéticos y no diabéticos tendrán las mismas características y serán asignados en la misma proporción a intervención o control, procedemos a estratificar.

Lo que se hace es generar una lista de aleatorización para diabéticos y otra para no diabéticos. Primero se determina si el sujeto es diabético o no y luego se elige la lista correspondiente. Esto sucede en general cuando la muestra de sujetos es pequeña.

Además del estudio aleatorizado paralelo, existen otras modalidades de estudios experimentales aleatorizados.

DISEÑO FACTORIAL

Una clase de estudio aleatorizado muy utilizado es el diseño factorial. Permite explorar la eficacia de 2 intervenciones al mismo tiempo con un solo grupo de pacientes. A estos estudios se los denomina factorial 2 x 2 (2 grupos intervenciones y 2 grupos control).

En la Tabla 3.8 se observa que algunos sujetos (25%) serán asignados a ambas intervenciones A y B, otro 25% a la intervención A y placebo de la intervención B, etc.

Al final tendremos un 50% de la muestra en intervención A y un 50% en su rama placebo y al mismo tiempo un 50% en intervención B y un 50% en su rama placebo.

Este diseño permite evaluar cada intervención por separado y también la eficacia de ambas intervenciones juntas (denominada interacción).

DISEÑO CRUZADO (CROSS-OVER)

En la mayoría de los estudios de intervención se utiliza el diseño paralelo, un grupo recibe durante todo el seguimiento la intervención y el otro el control.

En el caso del diseño cruzado los sujetos son asignados primero a intervención o control y luego se invierte la asignación.

Por lo tanto, durante el seguimiento ambos grupos han sido asignados a intervención y a control.

Quiénes serán asignados primero a una u otra rama se determina por aleatorización. En el primer sujeto puede asignarse primero a control y luego a intervención y en el segundo sujeto al revés, y así sucesivamente.

Cuando se cambia de rama de tratamiento, en general debe considerarse la necesidad de un período de *lavado* si la droga previamente asignada tiene un efecto temporal más o menos prolongado, para evitar el efecto conjunto de las drogas.

	Intervención A Nº pacientes	Placebo A Nº pacientes	Total
Intervención B Nº pacientes	25	25	50
Placebo B Nº pacientes	25	25	50
Total	50	50	100

Tabla 3.8: Diseño factorial 2 x 2

Nº de paciente	Secuencia de asignación	Nº de paciente	Secuencia de asignación
1	Control - Intervención	9	Intervención - Control
2	Intervención - Control	10	Intervención - Control
3	Intervención - Control	11	Control - Intervención
4	Intervención - Control	12	Intervención - Control
5	Control - Intervención	13	Control - Intervención
6	Control - Intervención	14	Control - Intervención
7	Control - Intervención	15	Control - Intervención
8	Control - Intervención	16	Control - Intervención

Tabla 3.9: Asignación cruzada de la intervención (*cross-over*)

¿PLACEBO O CONTROL?

La decisión de asignar al grupo control un placebo o un tratamiento estándar se establece de la siguiente forma:

1) El placebo es apropiado cuando no hay evidencia clara de si la intervención es mejor al tratamiento estándar o a ningún tratamiento.

2) El tratamiento estándar de valor establecido para la patología a estudiar debe indicarse en lugar de placebo cuando se utiliza en la práctica habitual de forma rutinaria.

Por supuesto, siempre hay discusión al respecto; lo que es éticamente aceptable es que el grupo control reciba el *mejor* tratamiento disponible demostrado al momento de realizar el estudio.

SEGUIMIENTO

Como en los estudios de cohortes, el seguimiento debe ser lo más completo posible en ambas ramas del estudio para evitar sesgos.

Se debe seguir a los sujetos aunque estos hayan discontinuado el tratamiento asignado. Si sólo se sigue a aquellos que continúan tomando la droga/placebo o tratamiento control, introduciremos un sesgo. Es muy posible que los que hayan abandonado por efectos adversos sean los asignados a la droga activa o los más graves; por lo tanto, si analizamos sólo el grupo que continúa recibiendo la intervención asignada, estaremos analizando a un subgrupo de pacientes que puede estar desbalanceado en las ramas de tratamiento.

Los eventos deben definirse antes de iniciar el estudio, para que todos los sujetos puedan ser clasificados de igual manera. Es muy importante determinar si un evento se ajusta a la definición adoptada por el estudio para no equivocarse en la asignación del mismo (sesgo de información).

La determinación de la aparición de un evento deben realizarla investigadores ciegos con respecto al tratamiento recibido por el paciente.

ANÁLISIS

La medida de efecto que se utiliza en este tipo de estudio es el riesgo relativo o razón de tasas. Dependiendo del tiempo de seguimiento, se pueden utilizar las llamadas curvas de

sobrevida o tiempo a la ocurrencia del evento (Kaplan-Meier), que se comparan entre los grupos mediante una prueba estadística denominada *long rank*.

La medida de efecto Hazard Ratio (HR), que es utilizada en los estudios con seguimiento extenso, es el equivalente del riesgo relativo.

Cuando se realiza un estudio donde el tiempo de seguimiento es prolongado y constituye una variable a tener en cuenta, cuanto más tiempo un sujeto esté en un estudio tiene más probabilidad de tener eventos comparado con un sujeto que esté menos tiempo en el estudio, es por este motivo que el tiempo de permanencia o seguimiento es un factor que influye en la incidencia de eventos y debe ser tenido en cuenta.

Con independencia de la medida de efecto elegida, lo importante es analizar los grupos de acuerdo con el criterio de *intención de tratar* (*Intention To Treat o ITT*).

Lo importante es comparar la tasa de eventos entre todos los sujetos asignados a intervención vs. todos los sujetos asignados a control, independientemente de que algunos hayan abandonado el tratamiento asignado o se hayan perdido en el seguimiento. El análisis se realiza de acuerdo con la asignación original, siendo esta la forma de análisis más confiable.

Si analizamos sólo aquellos que continúan en tratamiento estamos introduciendo un sesgo.

Supongamos que en el grupo de intervención hay más abandono, porque los pacientes más graves no toleran la droga. Estos pacientes seguirán en el estudio con el tratamiento estándar. Si analizamos el estudio de acuerdo con el tratamiento recibido y no con el asignado, podríamos encontrar que en el grupo estándar hay más eventos, sólo porque los pacientes más graves tuvieron que abandonar el tratamiento. De esta manera estaríamos subestimando el efecto de la intervención.

Si, en cambio, analizamos por intención de tratar, veremos que en ambos grupos hay la misma cantidad de eventos o tal vez menos.

INTERPRETACIÓN DE LOS RESULTADOS

Si la asignación de la intervención fue aleatoria y la determinación de eventos y efectos adversos fue ciega en relación con el factor de exposición asignado para los sujetos y para los investigadores, y si el seguimiento fue completo en ambos grupos, no hay posibilidad de sesgos. También las variables que caracterizan los grupos y los posibles factores de confusión estarán equilibrados en ambos grupos y no será necesario realizar ajuste por estos factores.

De esta manera, si el análisis estadístico fue el correcto, los resultados del estudio serán confiables.

VENTAJAS Y DESVENTAJAS

Ventajas:

La aleatorización y el doble ciego permiten realizar un estudio sin sesgos. Debido a esto, proporcionan una fuerte evidencia acerca de la relación causal entre intervención y eventos.

Es posible determinar el impacto de la intervención en múltiples eventos al mismo tiempo. En general se observa el efecto de la droga en la incidencia de muerte, progresión de enfermedad, hospitalizaciones, etc.

Desventajas:

Este tipo de estudios requiere de una estructura muy importante y, en general, de una gran muestra de pacientes con un seguimiento largo. Por lo tanto, los recursos necesarios son muchos y su costo muy elevado.

En muchas circunstancias, problemas éticos no permiten realizar este tipo de estudios.

Por ejemplo, no sería ético realizar un estudio aleatorizado para demostrar que el uso de profiláctico disminuye la probabilidad de contraer HIV, o aleatorizar a un grupo a fumar y a otro a no fumar para determinar la tasa de eventos producida por el tabaquismo.

Estos son algunos ejemplos donde los estudios observacionales han aportado una información invaluable, que de otra manera no sería posible obtener.

ESTUDIOS ECOLÓGICOS

En algunas circunstancias no es posible determinar el impacto de un factor de exposición a nivel individual, por lo que se recurre a los estudios que comparan el impacto del factor de exposición entre grupos de individuos (familias, escuelas, ciudades o países), los que en inglés se denominan *clusters*.

En los estudios ecológicos se compara el factor de exposición y los eventos de cada grupo. La medición tanto del factor de exposición como de los eventos se evalúa para el grupo total estudiado.

No se obtienen datos de los individuos sino del grupo en su totalidad.

Por ejemplo, porcentaje de fumadores de las ciudades A, B y C y la mortalidad de causa cardiovascular en estas ciudades. No sabemos la tasa de mortalidad de causa cardiovascular entre fumadores y no fumadores de cada ciudad.

Lo que se determina es la asociación entre factor de exposición y evento a nivel grupal. Se explora si el evento es más frecuente en el grupo donde el factor de exposición es más frecuente. (Figura 3.5)

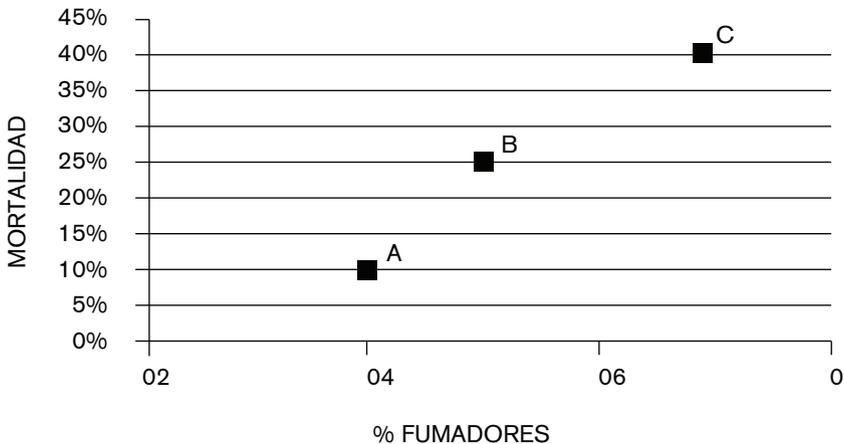


Figura 3.5: Estudio ecológico. Asociación entre tabaquismo y mortalidad en 3 ciudades

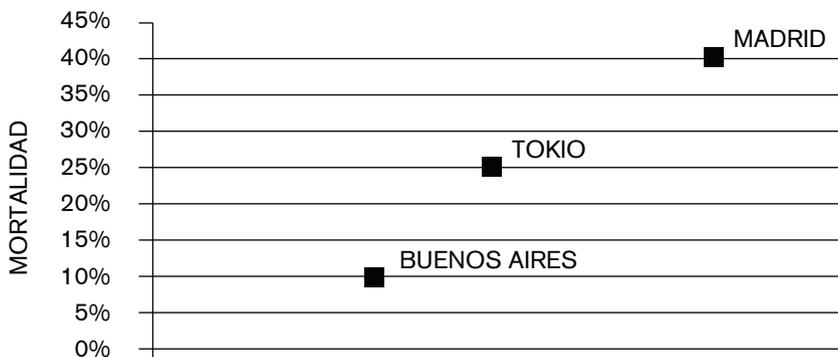


Figura 3.6: Mortalidad cardiovascular en 3 ciudades en el año 2004

TIPOS DE ESTUDIOS ECOLÓGICOS

ESTUDIOS MULTIGRUPO

En este tipo de estudios se compara en un tiempo específico la frecuencia del o los factores de exposición y del o los eventos en diferentes grupos. Es como un estudio de corte transversal.

Estos estudios pueden ser descriptivos, analíticos o experimentales.

En los estudios descriptivos se muestra la tasa del evento de varios grupos durante el mismo período de tiempo. No se postula ninguna hipótesis. Lo que se busca son patrones de distribución de eventos entre diferentes grupos para generar una posible explicación a la variación observada.

Por ejemplo, podemos tomar la tasa de mortalidad cardiovascular del año 2004 de diferentes países y observar la tendencia. (Figura 3.6)

En los estudios analíticos, se explora una hipótesis específica que explique las diferencias entre eventos. Por ejemplo, la mortalidad de causa cardiovascular es más elevada en Madrid que en Tokio y Buenos Aires debido al tipo de alimentación de la región.

Lo que se busca es determinar si el factor de exposición (alimentación) está asociado al evento (mortalidad).

Este tipo de estudio ecológico es el que se observa con más frecuencia.

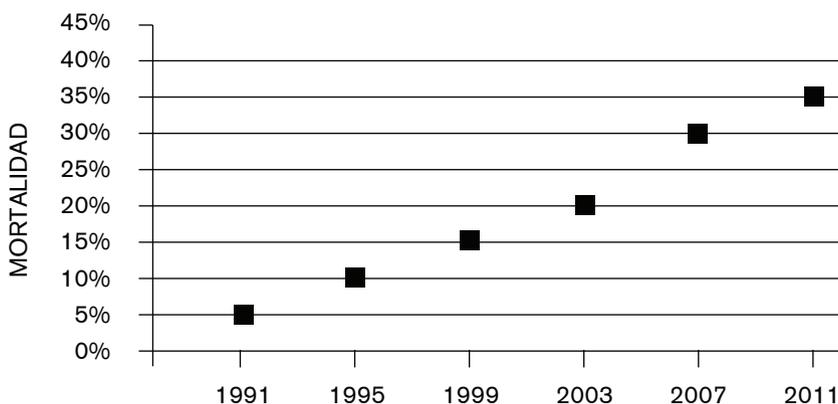


Figura 3.7: Variación de la mortalidad cardiovascular en Argentina en los últimos 20 años

ESTUDIOS DE TENDENCIA EN EL TIEMPO

DESCRIPTIVOS

En este caso lo que se compara es la tendencia en el tiempo de la frecuencia de un evento en un solo grupo de sujetos o región. Por ejemplo, la variación de la mortalidad cardiovascular en los últimos 20 años en Argentina. (Figura 3.7)

Lo que se busca son patentes temporales que puedan sugerir una hipótesis etiológica más específica. Es muy parecido a los estudios ecológicos multigrupo, pero en este tipo de estudio se sigue a un solo grupo durante un periodo de tiempo y se compara dentro de este grupo la tendencia del evento. No se comparan diferentes grupos, es la misma población en la cual se describe la variación en el tiempo de un evento.

ANALÍTICOS

En los estudios analíticos de tendencia de tiempo se busca explorar una hipótesis determinada. Lo mismo que con los estudios descriptivos analíticos multigrupo, se explora la asociación entre la variación de un factor de riesgo en el tiempo y la variación del evento de interés, siempre en una sola población.

Podríamos, por ejemplo, explorar si la disminución del uso compartido de jeringas se acompañó de un descenso en la tasa de infección por HIV en un grupo drogadependiente endovenoso en un período determinado de tiempo.

ESTUDIOS EXPERIMENTALES ECOLÓGICOS

Este tipo de estudios es similar a los estudios experimentales, pero la asignación de la intervención se hace en los grupos y no en individuos.

Se podría tomar 2 o más ciudades y asignar de forma aleatoria una intervención; por ejemplo, implementar un test de screening para cáncer de mama y a otras ciudades asignar la conducta estándar. Luego se debería comparar la tasa de mortalidad por cáncer de mama entre las ciudades que implementaron la intervención vs. las que no lo hicieron.

De los datos obtenidos de estudios ecológicos es que hoy se sabe del beneficio o no de ciertas intervenciones a nivel social. Por ejemplo, los países con uso obligatorio de casco o de cinturón de seguridad tienen menos muertes por accidentes de tránsito.

¿CUÁNDO ES NECESARIO REALIZAR UN ESTUDIO ECOLÓGICO?

Hay tres circunstancias donde se aplican este tipo de estudios.

- 1) Para investigar diferentes grupos.
- 2) Para investigar el efecto de una intervención a nivel grupal.
- 3) Cuando los grupos son más adecuados para estudiar que los individuos.

DIFERENTES GRUPOS

En salud pública muy a menudo hay interés en determinar por qué un grupo de sujetos tiene una mayor incidencia o prevalencia de un evento en particular en comparación con otros grupos.

El factor de exposición responsable de la diferencia de eventos entre grupos debe ser diferente de los factores de exposición responsables del evento a nivel individual. Esto es porque la exposición al factor de exposición no varía mucho entre individuos, pero sí entre grupos.

La única forma de investigar esta diferencia es a través del estudio de la distinta prevalencia del factor de exposición entre los grupos.

Por ejemplo, el bocio y la deficiencia de yodo. A nivel individual, todos los sujetos de una misma población consumen una alimentación baja en yodo, aunque no todos desarrollan bocio.

Si exploramos la posible asociación entre dieta baja en yodo y desarrollo de bocio en esta población, no encontraremos asociación, porque la causa de bocio está más ligada a la predisposición genética de cada individuo que a la dieta, que es muy similar entre todos los individuos. En este caso la dieta baja en yodo no es la causa de bocio para los individuos de esta población.

Sin embargo, si comparamos la prevalencia de bocio entre diferentes regiones con dietas altas, moderadas y bajas en yodo, veremos que la tasa de bocio sí se relaciona con la dieta.

Efecto a nivel grupal

Algunos factores de exposición no actúan a nivel individual. Estos factores de riesgo son compartidos por todos los miembros del grupo. Por ejemplo, sistema de salud, normas de tránsito, leyes como prohibición de alcohol a menores, etc.

Hay otros factores de riesgo que no sólo actúan a nivel individual sino cuyo efecto a nivel grupal es mayor.

La exposición del grupo tiene un efecto contextual sobre el evento. Ser pobre tiene sus efectos sobre la salud a nivel individual, pero el efecto de vivir en condiciones de pobreza puede aumentar sus efectos sobre la salud del grupo.

No sólo los factores de exposición tienen un efecto contextual, algunos eventos también. Esto es común en las enfermedades infecciosas, donde la prevalencia de la infección y la prevalencia de inmunidad influyen el riesgo de infección en el grupo.

GRUPOS MÁS ADECUADOS QUE INDIVIDUOS

Algunos factores ambientales pueden medirse sólo a nivel grupal y es casi imposible hacerlo a nivel individual debido al tipo de factor de exposición o a la gran variabilidad interpersonal.

La polución ambiental, la radiación solar o el consumo de alcohol son algunos ejemplos.

La determinación de la exposición a la radiación ultravioleta a nivel individual es muy difícil de medir, por lo que es más adecuado medir el nivel de radiación en una región y compararla con el nivel de otra y luego determinar la tasa de cáncer de piel en cada región.

Determinar el consumo de alcohol de un individuo puede ser difícil, porque muchas veces es una respuesta sesgada por el estigma social; por lo tanto la medición promedio de consumo de alcohol de una población comparada con otra podría explicar las diferencias en la incidencia de algunos eventos como cáncer de estómago, accidentes, etc.

¿PARA QUÉ SON ÚTILES LOS ESTUDIOS ECOLÓGICOS?

Estos estudios se llevan a cabo como primer paso en la investigación de una posible asociación. Una de las ventajas de estos estudios es que toman datos que son relevados

frecuentemente por los organismos estatales u otra clase de organizaciones (a estos datos se los denomina datos rutinarios).

Ejemplo de estos son: tasa de natalidad, tasa de mortalidad específica, prevalencia de enfermedades infecciosas y no infecciosas, como cáncer o cardiovasculares, aplicación de leyes, etc.

¿CÓMO SE RECLUTAN LOS GRUPOS DE ESTUDIO?

Los grupos a estudiar deben poseer las siguientes características:

- 1) Se debe elegir grupos que cubran un amplio rango de niveles de exposición y eventos. En el caso del bocio y la dieta baja en yodo, se debe tratar de incluir algunos grupos con muy alta prevalencia de bocio y otros con mediana y baja prevalencia; lo mismo debe hacerse con la dieta.
- 2) Los grupos deben ser comparables excepto por la prevalencia de factor de exposición y eventos. Esto reduce la posibilidad de sesgos. Por ejemplo, al estudiar la asociación entre exposición solar y cáncer de piel, los grupos deben tener la misma composición social, ya que si comparamos trabajadores de la construcción con un grupo que se exponga al sol por motivos estéticos, seguramente veremos más incidencia de cáncer de piel en el grupo trabajador, ya que el otro grupo muy frecuentemente pertenece a una clase social que accede a controles preventivos de cáncer de piel.
- 3) Si el factor de exposición está altamente correlacionado con un potente confundidor, se debe incluir algún grupo donde esto no suceda; de otra manera no será posible discriminar si la asociación se debe al factor de riesgo o al factor confundidor. El efecto de las leyes que obligan al uso de casco y su asociación con accidentes de tránsito puede estar *confundida* por la ingesta de alcohol, la cual está muy asociada con el hábito de usar casco y los accidentes de tránsito. Por lo tanto, hay que elegir grupos con diferentes niveles de consumo de alcohol.
- 4) Si se usan datos rutinarios hay que asegurarse de que tanto el factor de exposición como el evento estén definidos de la misma manera en todas las fuentes de información. De lo contrario, estaremos comparando diferentes eventos y factores de exposición. Por ejemplo, para algunos países la incidencia de sarampión es de notificación obligatoria, mientras que en otros sólo se informan los casos graves o las muertes. También el nivel de inmunidad puede ser determinado de distintas maneras (certificado de vacunación, detección de anticuerpos, etc.).
- 5) No se debe incluir grupos donde la asociación a estudiar ya esté establecida. Si la inclusión del grupo se basa en la asociación establecida, se introduce un sesgo de selección.

MEDICIÓN DEL NIVEL DE EXPOSICIÓN Y EVENTOS A NIVEL GRUPAL

La evaluación o estimación de eventos de un grupo se realiza mediante datos globales, tales como mortalidad por causa específica, incidencia o prevalencia de otros eventos. En general se expresan como tasas o porcentajes.

Para evaluar o medir el factor de exposición también se usan medidas globales, como consumo promedio de alcohol, porcentaje de población bajo la línea de pobreza, promedio de contaminación ambiental de una ciudad, promedio mensual de radiación ultravioleta, etc.

Otros factores de riesgo se miden por las llamadas medidas integrales, que son características de toda la población y no tienen equivalencia a nivel individual. Por ejemplo, la aplicación de las leyes.

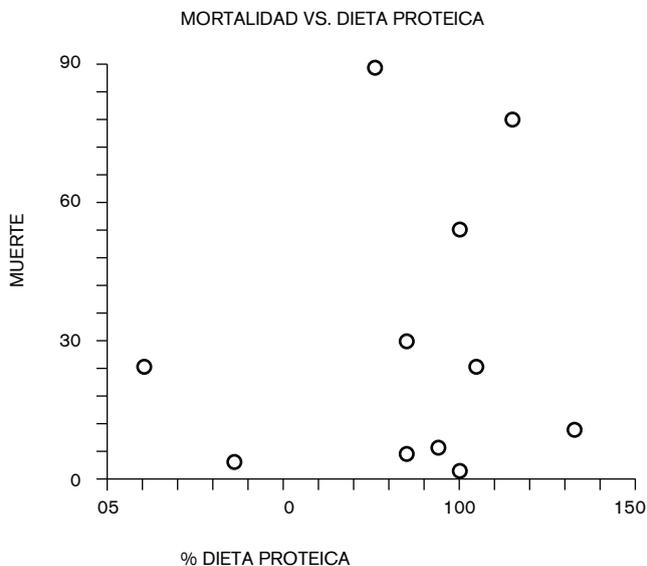


Figura 3.8: *Scatter plot*

ANÁLISIS DE DATOS

En los estudios ecológicos las mediciones del nivel de exposición y de eventos en general son variables continuas. Por ejemplo, al valorar el nivel de exposición podemos expresar el porcentaje de casas sin agua potable, consumo per cápita de cigarrillos o mediana de ingresos de una familia.

En el caso del evento, se miden incidencia de diarrea infantil, tasa de mortalidad de causa cardiovascular, etc.

Debido a que estas variables se transforman en continuas, su asociación no se puede expresar con el OR o riesgo relativo, que se usa en estudios a nivel individual. En este caso lo que se hace es analizar la relación entre 2 variables continuas, utilizando un gráfico *scatter plot*, donde en el eje **x** se introducen los datos del evento y en el eje **y** los del factor de exposición. (Figura 3.8)

Como puede verse, en la Figura 3-8 parece no existir una relación entre mortalidad y dieta alta en proteínas.

Para cuantificar si esta relación es estadísticamente significativa se utiliza la prueba de correlación o de regresión lineal.

Para este ejemplo, el Coeficiente de Correlación es del 9% con $p = 0,78$. Por lo tanto, no podemos sostener que exista una asociación entre mortalidad y dieta rica en proteínas.

INTERPRETACIÓN DE LOS RESULTADOS

Como en todo estudio, una posibilidad es que los resultados sean falsos en el sentido estadístico, es decir que los resultados hayan sido producto del azar. En este tipo de estudios el tamaño de la muestra es pequeño, con lo cual la certeza estadística es baja.

Otro factor muy importante es la dificultad de identificar y controlar la influencia de factores de confusión.

Muchos factores están más fuertemente asociados con otros a nivel grupal que a nivel individual. Ya mencionamos la *interferencia* del consumo de alcohol en el efecto de las leyes de tránsito con accidentes. Otros ejemplos son el mayor uso de protección solar en regiones cálidas que puede interferir en el nivel de radiación y cáncer de piel, la pobreza, las regiones con dieta alta en grasas que se asocia a baja ingesta de fibras y que puede llevar a concluir que la dieta alta en grasa es la responsable de cáncer de colon en lugar de la baja ingesta de fibras.

En los estudios ecológicos existe la posibilidad de que haya factores que a nivel individual no sean factores de confusión pero que sí lo sean a nivel grupal. Esta es una de las principales debilidades de estos estudios.

Hay que tener en cuenta que los grupos son diferentes en muchos otros factores, por lo tanto su comparación es relativa.

También los sesgos son un problema en esta clase de estudios. Los sesgos de selección surgen cuando los grupos se eligen en base a una asociación ya establecida.

El mayor problema en estos estudios son los sesgos de información, debido a que los datos provienen de relevamientos globales, que muchas veces no contienen la totalidad de los datos individuales o bien no comparten la misma definición.

Como en los estudios transversales, existe la posibilidad de causalidad reversa, porque tanto el factor de exposición como el evento se relevan al mismo tiempo, no puede verificarse relación temporal entre ambos. Por ejemplo, la presencia de cáncer de estómago lleva a consumir menos carnes, ¿o el consumo bajo de carnes produce cáncer de estómago?

INFERENCIAS DE LOS RESULTADOS DE UN ESTUDIO ECOLÓGICO

Es muy difícil asegurar que una asociación observada en estos estudios sea válida.

Sobre la base de un estudio ecológico no se puede concluir que, dentro de un grupo, los individuos estén en riesgo elevado de sufrir un evento, debido a que los datos no provienen de individuos sino del grupo total. Es más probable que otros factores dentro del grupo determinen el riesgo individual de sufrir un evento.

Los estudios ecológicos sólo permiten hacer inferencias a nivel grupal. El efecto de un factor de exposición a nivel grupal no necesariamente es el mismo a nivel individual.

A esto se lo denomina *falacia ecológica*.

Hay tres razones por las cuales no se puede asumir que la causa de un evento a nivel grupal sea la misma que a nivel individual.

- Una causa particular puede sólo ser responsable de las diferencias de eventos entre grupos, mientras que dentro del grupo puede estar uniformemente distribuida.
- Algunos factores de exposición tienen un efecto contextual en el grupo, como la pobreza.
- Algunos eventos tienen un efecto contextual a nivel grupal, como la inmunidad.

Por ejemplo, se puede encontrar en un grupo una asociación entre susceptibilidad genética y bocio. Sería erróneo concluir para otros grupos que la causa de bocio es genética. Por el contrario, cuando estudiamos otros grupos encontramos que la causa es la deficiencia de yodo.

Por lo tanto, las conclusiones obtenidas de un estudio ecológico no deben ser extrapoladas a nivel individual y viceversa.

VENTAJAS Y DESVENTAJAS

Ventajas:

- 1) Permiten investigar diferencias entre grupos, lo cual es importante a nivel de Salud Pública.
- 2) También son útiles para factores de exposición difíciles de evaluar a nivel individual.
- 3) Se pueden realizar de forma rápida y barata, debido a que utilizan datos existentes.
- 4) Constituyen el primer paso en la investigación de una posible asociación causal entre factor de exposición y evento.

Desventajas:

- 1) Los resultados deben ser considerados con cuidado, por la dificultad de controlar confundidores y la introducción de sesgos.
- 2) Las conclusiones no pueden extrapolarse a nivel individual.

Hemos mencionado que la Epidemiología trata de establecer si existe una verdadera asociación entre un factor de exposición y un evento.

Suponiendo que los resultados de un estudio concluyan que existe dicha asociación, siempre debemos recordar que pudo cometerse algún error; por lo tanto los resultados deben aceptarse como válidos si hemos descartado tales errores.

Los errores que siempre debemos tener en cuenta son:

1. Error por azar

El resultado del estudio se obtuvo por azar y no es posible replicarlo en otros estudios. Veremos más adelante cómo, a través de las pruebas estadísticas, podemos considerar si el resultado de un estudio fue por azar o no.

2. Factores de confusión o confundidores

Los factores de confusión son factores de exposición diferentes al factor de exposición que se está estudiando. Estos otros factores hacen que la asociación descubierta entre el factor de exposición de interés y el evento estudiado no sea la que verdaderamente es.

Estos factores de confusión pueden ser una explicación alternativa de la asociación. Veremos más adelante, con mayor detalle, este tema.

3. Sesgos

A la hora de diseñar un estudio epidemiológico, de acuerdo con el tipo de estudio seleccionado, debemos seguir ciertas reglas en lo que se refiere a la selección de los sujetos que participarán del estudio y al modo en que los factores de exposición y/o eventos de interés serán clasificados.

Si seleccionamos mal la población es muy probable que, al comparar dos grupos, estos no sean iguales, por lo tanto no serán comparables y las conclusiones del estudio no serán válidas.

Se llama sesgo (*bias* en inglés) a un error en el diseño o de la conducción del estudio que resulta en una conclusión que no es cierta.

Un estudio con sesgos (sesgado) no ofrece una verdadera representación de la situación o de la asociación que queremos estudiar.

Las potenciales fuentes de sesgos deben determinarse en el momento de diseñar el estudio, porque si nos damos cuenta durante o al final del estudio no habrá forma de corregir el error.

A diferencia de lo que ocurre con los confundidores o el error por azar, no hay métodos estadísticos que podamos emplear para corregir sesgos.

Podemos pensar a los sesgos desde dos situaciones:

1. En estudios descriptivos: los sesgos aparecen cuando la población en estudio no representa a la población que queremos describir (población de referencia). Si queremos describir la mortalidad de la diabetes en la población general y elegimos sólo a los pacientes hospitalizados por esta causa podríamos pensar que la mortalidad de la enfermedad es alta, pero no estaríamos teniendo en cuenta que en general se interna únicamente a los pacientes más graves; o también podríamos llegar a la conclusión de que la mortalidad es baja, si los pacientes más graves no llegaran a hospitalizarse porque mueren antes de consultar. Los datos de mortalidad de este estudio serían entonces los de los pacientes con diabetes hospitalizados y no los de la población general con diabetes.
2. En los estudios analíticos: los sesgos surgen cuando las poblaciones a estudiar no son comparables. Supongamos que deseamos determinar si la exposición a cierto agente químico es causa de infertilidad femenina. Para esto buscamos en clínicas de infertilidad mujeres con este diagnóstico y las comparamos con mujeres que hayan tenido al menos un hijo en hospitales públicos. Cuando comparamos ambos grupos de mujeres observaremos que muchas de las características son diferentes, por ejemplo la clase social a la que pertenecen. Las mujeres que concurren a clínicas de infertilidad suelen pertenecer a una clase social diferente a la de aquellas que asisten a un hospital público. Este factor individual ya implica diferencias en muchos otros factores, como dieta, estado de salud y el lugar de residencia, este último puede estar directamente relacionado con la exposición al agente químico de interés.

Los sesgos se clasifican de acuerdo a que el error esté en la selección de la población o en la forma de reportar los eventos.

CLASIFICACIÓN DE SESGOS

SESGOS DE SELECCIÓN

ESTUDIOS DESCRIPTIVOS

El sesgo de selección aparece si la población seleccionada no es representativa de la población de referencia (aquella a la cual queremos extrapolar los datos).

Por ejemplo, en una encuesta sobre la ingesta de alcohol en adolescentes, los sujetos bebedores de altas cantidades probablemente participarán menos. La consecuencia de esto es que los sujetos bebedores estarán subrepresentados y la ingesta de alcohol promedio obtenida en esta encuesta estará subvalorada respecto al verdadero nivel de ingesta alcohólica en adolescentes. A este tipo de sesgos de selección se los denomina sesgos de respuesta.

Al tratar de estimar la prevalencia de HIV de una población a través de un estudio serológico, probablemente tendremos una alta tasa de rechazo entre aquellos individuos de riesgo elevado. Este grupo puede incluir una proporción elevada de HIV+, con lo cual el grupo que acepta realizarse el test probablemente será de bajo riesgo y por lo tanto la prevalencia reportada será baja y estará sesgada por la alta tasa de rechazo (alta proporción de sesgo de respuesta).

En cualquier estudio descriptivo es muy importante obtener algunos datos del grupo que se niega a participar. Los más frecuentemente recabados son: edad, sexo, ciertos hábitos y clase social. Estos datos son muy útiles para poder hacer una comparación con el grupo que acepta participar: si existen diferencias importantes entre ambos grupos diremos que el grupo analizado no es representativo de la población que queremos estudiar, o que está sesgado.

En la mayoría de los estudios es de buena práctica reportar qué porcentaje de la población a la que se invitó a participar aceptó hacerlo. De esta forma reportamos la probabilidad de que la población esté o no sesgada.

Si en un estudio el porcentaje de rechazo fue del 10%, probablemente no se haya introducido un sesgo significativo; pero si el porcentaje de rechazo es del 50%, seguramente la población que aceptó participar será muy diferente a la que rechazó el estudio.

ESTUDIOS ANALÍTICOS

La fuente de sesgos de selección en los estudios analíticos depende de qué tipo de estudio se trate.

ESTUDIOS CASO-CONTROL

Existe sesgo si los casos (sujetos enfermos) no son representativos de todos los que existen dentro de la población a estudiar y/o si los controles (sujetos sanos) no son representativos de la población sana.

SELECCIÓN DE CASOS

Se introduce un sesgo al elegir los casos (sujetos con el evento), cuando los casos expuestos al factor de exposición de interés tienen una mayor o menor probabilidad de ser seleccionados, en comparación con los casos no expuestos al factor de exposición.

Por ejemplo, si estudiamos la asociación entre hipertensión arterial (caso) y uso de anticonceptivos orales (factor de exposición) y decidimos elegir los casos de una población de mujeres hipertensas diagnosticadas en servicios de ginecología, veremos que estas mujeres toman anticonceptivos en mayor proporción que otras mujeres hipertensas provenientes de otro servicio.

Por lo tanto, las mujeres hipertensas expuestas al factor de riesgo (por provenir de un servicio de ginecología) tienen más probabilidad de ser seleccionadas que otras mujeres hipertensas no expuestas al factor de exposición, por no concurrir a un servicio de ginecología.

En otras palabras, la probabilidad de seleccionar casos expuestos es mayor debido a la elección de la población de mujeres de donde se obtendrán los casos.

El resultado más probable será que la mayoría de los casos tendrán el antecedente de tratamiento anticonceptivo y se inferirá erróneamente que existe una asociación entre uso de anticonceptivos e hipertensión arterial, lo cual podría no ser cierto, debido a que la población está *sesgada*.

A este tipo de sesgos se los denomina *sesgos de vigilancia*.

Para evitar este sesgo se podría tomar la tensión arterial a mujeres que concurren a diferentes supermercados y confirmar el diagnóstico de hipertensión arterial y preguntar acerca del uso de anticonceptivos. Estos sujetos provienen de la población general.

SELECCIÓN DE LOS CONTROLES

Los controles deben ser sujetos cuya única diferencia con los casos es la ausencia de la enfermedad o evento de interés.

Los controles deben provenir de la misma población que produjo los casos.

En el ejemplo de la asociación entre anticonceptivos e hipertensión, la población que produce mujeres hipertensas es la población general y no una muestra hospitalaria; por lo tanto los controles (mujeres no hipertensas) deben seleccionarse de la población general.

Si seleccionamos controles más similares a los casos con respecto al factor de riesgo de la población que produjo los casos introducimos un sesgo. Esto resulta en una subestimación del efecto del factor de riesgo.

Por ejemplo, las mujeres no hipertensas que concurren a un servicio de ginecología probablemente usen más anticoncepción oral que la población general, y es posible que el porcentaje de uso de anticoncepción oral sea muy similar al de las mujeres hipertensas (casos). Si los casos y los controles tienen la misma proporción de exposición al factor de riesgo, no encontraremos asociación válida, pero esto se debe al sesgo en la selección de la población.

Es más fácil seleccionar casos y controles buscando en los registros hospitalarios, pero en ambos casos esta población no es representativa de la población de casos ni de la población general. Los sujetos que asisten regularmente a un centro de salud suelen ser más cuidadosos con su salud y por lo tanto tienen hábitos diferentes a los de la población general.

Cuando se seleccionan controles hospitalizados, se corre el riesgo de que el sujeto padezca de una enfermedad relacionada al factor de exposición que está siendo investigado.

Si se está investigando la asociación entre cáncer de vejiga y tabaquismo, los casos serán sujetos con cáncer de vejiga. Si seleccionamos sujetos hospitalizados por enfermedad coronaria como controles, es muy probable que estos pacientes tengan exposición a tabaquismo y por lo tanto casos y controles tengan similar frecuencia de exposición a tabaquismo y no se pueda determinar que existe asociación entre tabaquismo y cáncer de vejiga.

El *sesgo de Berkson* o *sesgo de admisión* es un ejemplo especial de sesgo de selección para estudios caso-control.

Cuando casos y controles son seleccionados entre pacientes hospitalizados, las características de estos grupos con relación al factor de exposición estarán influenciadas por su tasa de admisión hospitalaria.

Esto ocurre cuando la combinación de factor de exposición y evento en estudio aumenta la probabilidad de hospitalización, produciendo una mayor tasa de exposición entre los casos hospitalizados comparados con los controles hospitalizados.

Veamos algunos ejemplos:

- Si una mujer tiene sospecha de padecer trombosis venosa profunda y toma anticonceptivos orales, es mucho más probable que sea hospitalizada que una mujer con los mismos síntomas pero que no tome anticonceptivos orales. Si seleccionamos como controles a mujeres hospitalizadas por otra causa, por ejemplo enfermedad cardíaca, veremos que estas mujeres tienen baja tasa de uso de anticonceptivos.
- Ante uso de antiinflamatorios no esteroides y síntomas de úlcera péptica, habrá mayor tendencia a hospitalizar a estos pacientes que a sujetos con síntomas de úlcera pero que no tomen antiinflamatorios. Si seleccionamos controles en el mismo servicio de gastroenterología, sólo por el hecho de estar admitidos a un servicio de gastroenterología no recibirán antiinflamatorios.

En ambas situaciones, si seleccionamos casos hospitalizados veremos que tienen una alta proporción de exposición, comparada con la proporción de exposición de la población general.

Si también seleccionamos controles hospitalizados, estos sujetos tendrán una tasa de exposición mucho menor a la que presenta la población general, sólo por el hecho de estar hospitalizados.

Como los casos tienen alta tasa de exposición y los controles baja tasa, la asociación entre factor de exposición y evento aparecerá como muy evidente.

Esto es una falacia, debido a que seleccionamos poblaciones muy sesgadas.

En cualquier circunstancia, seleccionar controles hospitalizados introducirá un sesgo.

Dependiendo del evento de interés, como enfermedades que necesitan hospitalización o son poco frecuentes en la práctica, se suele seleccionar los casos en los centros de salud y se toman como controles personas cercanas al caso, por ejemplo amigos (que representan a la población general), no siendo elegibles parientes, debido a que podrían estar tan expuestos al factor de exposición como el caso.

También se utiliza para la selección de controles personas que vivan en el área de influencia del centro hospitalario o del área de residencia del caso, que provienen de la población general, ya que el hecho de vivir en la misma área hace que sean más similares en sus características socioeconómicas, lo que implica poseer otras características también similares entre casos y controles, sin el riesgo de igualar la exposición al factor de riesgo.

Para eventos comunes o que no necesiten hospitalización se pueden obtener los casos y los controles de la población general. Para el estudio sobre la asociación entre hipertensión arterial y uso de anticonceptivos, se podría incluir mujeres hipertensas (casos) que concurren a diferentes supermercados y preguntar sobre el uso de anticonceptivos, y los controles podrían ser las mujeres no hipertensas (controles) que concurren a los mismos supermercados y preguntarles si utilizan anticonceptivos. Ambos grupos provienen de la población general y por el hecho de asistir a ciertos supermercados igualan las características de los grupos. Las mujeres que asisten a ciertos supermercados comparten características socioeconómicas, lo que tiene influencia en su cuidado sobre su salud.

ESTUDIOS DE COHORTES

La fuente de sesgos surge cuando los sujetos expuestos y no expuestos a un factor de riesgo no son comparables, ya sea por una mala selección de los no expuestos o bien por un seguimiento diferencial entre los expuestos y los no expuestos.

ELECCIÓN EQUIVOCADA DE POBLACIÓN NO EXPUESTA

Es importante que los sujetos no expuestos sean similares a los expuestos en todas sus características, menos en la exposición al factor de riesgo.

Si estamos comparando la morbilidad de trabajadores de una planta nuclear (expuestos) con la de la población general, veremos que los trabajadores de la planta nuclear tienen menos morbilidad. Esto se debe a que la población general incluye a sujetos sanos y sujetos enfermos. Por definición, los trabajadores no deben estar enfermos para poder trabajar.

Por lo tanto una población trabajadora tiene una alta probabilidad de presentar menos morbilidad que la población general.

A esto se lo denomina *efecto del trabajador sano*.

Entonces, para comparar la morbimortalidad de los trabajadores de una planta nuclear, debemos elegir trabajadores de otra industria no expuestos a la radiación.

SEGUIMIENTO DIFERENCIAL ENTRE EXPUESTOS Y NO EXPUESTOS

Si en el grupo expuesto se pierden en el seguimiento más sujetos que en el grupo no expuesto, no podremos detectar los eventos que se produzcan (debido a esta pérdida de seguimiento), lo que traerá como consecuencia poca diferencia de eventos entre expuestos y no expuestos y no se podrá detectar asociación, si es que la hay, por este sesgo.

Esta pérdida diferencial puede asociarse al hecho de estar expuesto o no al factor de interés.

Por ejemplo, si estamos siguiendo sujetos expuestos a un agente carcinógeno y sujetos no expuestos a este agente, seguramente se enfermarán y morirán más sujetos en el grupo expuesto. Es posible que por estar expuestos y enfermarse más, estos sujetos deban trasladarse a otro sitio para su atención. Si no tomamos las medidas necesarias para asegurar el contacto con estos sujetos o sus familiares, el seguimiento de este grupo sufrirá una pérdida mayor a la del grupo no expuesto y los eventos producidos en los expuestos no podrán recabarse. El resultado será una diferencia de eventos menor a la esperada, y existirá la posibilidad de no detectar la asociación supuesta entre factor de riesgo y eventos.

A esto se lo denomina *pérdida de seguimiento diferencial*: es la pérdida que está relacionada con la condición de estar expuesto.

En todo estudio de cohortes, el porcentaje de seguimiento mínimo aceptable para cada grupo es de 80% y siempre debe reportarse.

En resumen: en los estudios descriptivos los sesgos de selección se evitan llevando la tasa de aceptación o respuesta al máximo posible. Si esto no es posible, se debe recabar toda la información posible de los sujetos no respondedores.

En los estudios analíticos caso-control estos sujetos deben ser representativos de sus respectivas poblaciones y en los estudios de cohortes se debe evitar la pérdida de seguimiento diferencial.

SESGOS DE INFORMACIÓN

Cualquier error en la medición o en la detección del factor de exposición o del evento que resulte en una diferencia sistemática relacionada a la situación del sujeto (caso o control, expuesto o no expuesto), en la precisión de la información recolectada entre los grupos a comparar, introducirá un sesgo de información.

En un estudio caso-control, el objetivo es determinar cuántos casos y cuántos controles estuvieron expuestos a un determinado factor de riesgo.

En el caso de los estudios de cohortes el objetivo es determinar la tasa de eventos que sufren los individuos expuestos y los no expuestos en el tiempo.

Si nos equivocamos al determinar si un sujeto estuvo expuesto o no por pertenecer al grupo caso o control, o bien si sufrió un evento o no en el tiempo por pertenecer al grupo expuesto o no expuesto al factor de exposición, el resultado será que clasificaremos mal la condición de haber estado expuesto en los estudios caso-control o en los estudios de cohortes clasificaremos mal al sujeto que sufrió un evento. A esto se lo denomina *falla de clasificación* (*misclassification*).

Los sesgos de información se clasifican según de dónde provenga el error, si es de los sujetos al reportar una condición o de los investigadores al observarla.

SESGOS DE REPORTE

En un estudio caso-control, los sesgos de reporte surgen cuando los sujetos enfermos (casos) reportan un antecedente de exposición al factor de exposición de forma sistemáticamente diferente al grupo sano (controles) y en un estudio de cohortes, cuando los sujetos expuestos reportan eventos de forma sistemáticamente diferente al grupo no expuesto.

En un estudio caso-control sobre factores de riesgo para leucemia, los sujetos con la enfermedad tendrán más probabilidad de reportar exposición a químicos que los sanos. Las personas que están enfermas piensan más acerca de las posibles causas de su enfermedad y consideran mucho más la posibilidad de haber estado expuestos a determinados factores de riesgo que la población sana.

A este tipo de sesgo se lo llama *sesgo de recuerdo*.

En un estudio de cohortes que investiga la tolerancia y el efecto de cierta droga, los pacientes que reciben la droga reportarán más eventos adversos comparados con los individuos que no la reciben.

En un estudio transversal sobre hábitos sexuales, llevado a cabo mediante una encuesta, los sujetos pueden sobrerreportar hábitos considerados socialmente correctos o bien subreportar hábitos socialmente considerados indeseables (un ejemplo es el uso de preservativo).

Esto, claro, depende en gran medida de la sociedad en la que estemos trabajando.

En una sociedad donde se acepte que los hombres pueden tener muchos contactos sexuales en su vida y las mujeres no, obviamente los hombres reportarán mayor número de contactos sexuales que las mujeres.

Los sesgos de reporte son un problema particular cuando los sujetos están advertidos sobre la asociación que se está estudiando. Esto puede resultar en sub o sobreestimación de la asociación que se está estudiando.

La mejor forma de evitar el sesgo de recuerdo es obtener los datos de fuentes objetivas que se hayan registrado con anterioridad a la aparición de la enfermedad. Los sesgos de reporte pueden minimizarse usando datos que provengan de fuentes objetivas y evitando que los sujetos de estudio sepan cuál es la asociación que se está estudiando. Cuando los sujetos no conocen, por ejemplo, qué droga están tomando se dice que están *ciegos* al factor de exposición.

SESGOS DE OBSERVACIÓN

En un estudio transversal o en uno de tipo caso-control, el investigador introduce el error al asignar mal el estatus de expuesto o no expuesto de forma diferente, dependiendo de que los sujetos sean considerados enfermos o sanos.

En un estudio de cohortes, el error surge cuando la asignación de un evento depende de que el sujeto haya estado expuesto o no.

Los sesgos de observación pueden ser un problema en los estudios analíticos. Es más probable que ocurran si la clasificación del factor de exposición o evento se basa en un criterio subjetivo del investigador.

VEAMOS ALGUNOS EJEMPLOS

El serotelioma es una forma rara de cáncer y es difícil de diagnosticar histológicamente. Su factor de riesgo más importante es la exposición al asbesto, y muchas veces este antecedente es el que permite hacer el diagnóstico.

Si un patólogo conoce este antecedente estará más propenso a diagnosticar serotelio-
ma que cualquier otra forma de cáncer difícil de diagnosticar.

En un estudio caso-control se investiga la exposición a vacuna BCG y desarrollo de
tuberculosis. La determinación de si el sujeto fue vacunado se realiza por la presencia de
cicatriz residual confirmada por el investigador. Si este conoce el estatus de caso o control
de los pacientes (y cree que la vacuna es efectiva) determinará que las cicatrices en los
pacientes con TBC se deben a otra causa y no a la vacunación BCG. Esto obviamente
alterará la asociación: los controles estarán más vacunados que los casos y la eficacia de
la vacuna parecerá superior a su eficacia real. Este dato está sesgado por el investigador.

La mejor forma de evitar sesgos de observación es mantener ciego al investigador acer-
ca del estatus de los sujetos.

En un estudio de cohortes, quien asigne los eventos no debe saber si el sujeto está
expuesto o no; en un estudio caso-control, el investigador que asigna el antecedente de
expuesto o no expuesto entre casos y controles no debe conocer este dato.

Cuando el sujeto de estudio no conoce la asociación que se investiga, decimos que
está ciego; cuando el investigador no conoce el estatus del paciente también decimos que
está ciego. Cuando ambos, sujetos e investigadores, son ciegos, decimos que el estudio
es doble ciego.

Los sesgos de información se evitan con un diseño doble ciego y con una adecuada
recolección de datos.

Se deben buscar datos sobre exposición pasada de manera objetiva y no confiar en la
memoria del paciente. Si esto no es posible, debe considerarse el dato más cercano en el
tiempo a la aparición de la enfermedad.

También es preciso estandarizar los cuestionarios y entrenar a los investigadores para
que todos pregunten de la misma manera y eviten interferir en las respuestas de los sujetos.

Es necesario tener definiciones claras de los eventos a recabar durante un estudio de
cohortes, del estatus de caso o control (quiénes se consideran enfermos y quiénes sanos),
así como de quiénes deben clasificarse como expuestos y quiénes no.

CONFUNDIDORES E INTERACCIÓN

CAPÍTULO

5

CONCEPTO

Supongamos que estamos explorando la asociación entre ingesta de café (factor de exposición de interés) y cáncer de páncreas (evento). Llevamos a cabo un estudio y encontramos que los sujetos que beben café tienen una tasa mayor de cáncer de páncreas. El tomar café duplica el riesgo de cáncer de páncreas. (Tabla 5.1)

¿Es *verdadero* este hallazgo? ¿Omitimos considerar algo importante?

Lo que omitimos fue tener en cuenta la presencia de otras variables posiblemente asociadas a cáncer de páncreas como el tabaquismo. Si analizamos la población, veremos que los sujetos que toman café fuman más que los que no toman café. (Tabla 5.2)

Si ahora calculamos la tasa de cáncer de páncreas en los bebedores y no bebedores de café, dividiendo a la población en fumadores y no fumadores, veremos que los bebedores de café comparados con los sujetos no bebedores de café en ambos grupos (fumadores y no fumadores) tienen casi la misma tasa de cáncer de páncreas, no fumadores 20 vs. 15% y en los fumadores 40 vs. 35% y que el grupo fumador tiene más cáncer de páncreas, 40% en los fumadores que beben café vs. 20% para los no fumadores que beben café. (Tabla 5.3)

	Cáncer de páncreas
Bebedores de café	30%
No bebedores de café	15%

Tabla 5.1: Porcentaje de cáncer de páncreas en la población de bebedores y no bebedores de café

	Bebedores de café	No bebedores de café
Fumadores	80%	30%
No fumadores	20%	70%

Tabla 5.2: Porcentaje de fumadores entre población bebedora y no bebedora de café

	No fumadores Cáncer de páncreas	Fumadores Cáncer de páncreas
Bebedores de café	20%	40%
No bebedores de café	15%	35%

Tabla 5.3: Porcentaje de cáncer de páncreas en población bebedora y no bebedora de café clasificados de acuerdo con el hábito de fumar

	Cáncer de páncreas
Fumadores	30%
No fumadores	10%

Tabla 5.4: Porcentaje de cáncer de páncreas en población fumadora y no fumadora

Por lo tanto, el factor que parece aumentar el riesgo de cáncer de páncreas es el tabaquismo y no el beber café. (Tabla 5.4)

Cuando un factor de exposición está asociado con otro (tomar café y fumar) hay que determinar si la asociación observada entre el factor de exposición de interés y el evento no se ve afectada o confundida por la presencia de otra variable. A este tipo de variables se las denomina *confundidores* o *variables confundidoras*.

En este caso fumar es un confundidor y afecta claramente la asociación entre beber café y cáncer de páncreas. En realidad, la verdadera variable asociada a cáncer de páncreas es el tabaquismo y, por lo tanto, es el factor de exposición asociado al evento.

CONDICIONES PARA QUE UN FACTOR DE EXPOSICIÓN O VARIABLE SEA CONSIDERADO COMO POSIBLE CONFUNDIDOR

1. Debe estar asociado al factor de exposición de interés. En el ejemplo, los sujetos que toman café fuman más que los que no toman café. Si empleamos una prueba estadística veremos que la proporción de fumadores es estadísticamente mayor en el grupo que toma café.
2. Debe ser un factor de riesgo para el evento de interés, tanto en la población expuesta como no expuesta. Claramente, el hábito de fumar es un factor de riesgo para cáncer de páncreas, tanto en los sujetos que beben café como en los que no lo beben. Podemos determinar estadísticamente que los sujetos con cáncer de páncreas tienen una proporción de tabaquismo más alta que los sujetos sin cáncer de páncreas y en ambos grupos, bebedores y no bebedores de café.
3. No debe estar en la vía causal entre el factor de exposición a estudiar y el evento de interés. La vía por la cual tomar café produciría cáncer de páncreas es completamente distinta a la vía causal o fisiopatológica por la cual el tabaquismo produce esta enfermedad.

Supongamos que consideramos como factor de exposición principal el tabaquismo, como posible variable confundidora el antecedente de padecer bronquitis crónica y como evento el cáncer de pulmón. En este caso el hábito de fumar produce bronquitis crónica y esta constituye una de las etapas fisiopatológicas hacia el cáncer de pulmón; por lo tanto en este caso no debe considerarse a la bronquitis como un confundidor.

Cualquier variable puede ser un factor de exposición en un estudio y un confundidor en otro.

Por ejemplo, en un estudio que explora la asociación entre dislipemia y enfermedad coronaria, además de considerar la dislipemia debemos tener en cuenta otros factores de exposición o variables como edad y sexo. Estas variables deberán considerarse confundidoras porque cumplen con las tres condiciones necesarias: (Figura 5.1)

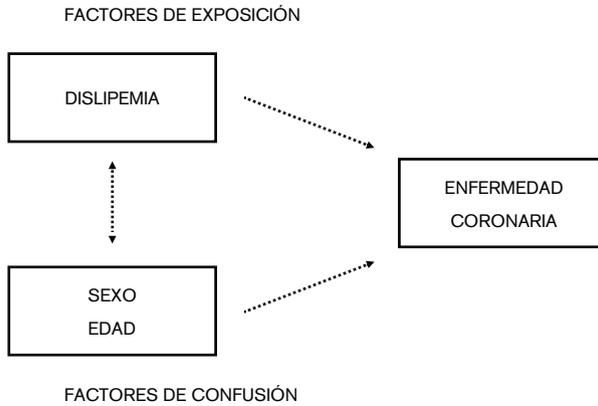


Figura 5.1: Variables confundidoras

1. Se asocian con dislipemia (los sujetos dislipémicos son más añosos y generalmente hombres).
2. A mayor edad, mayor probabilidad de padecer enfermedad coronaria, sobre todo en el sexo masculino.
3. La edad y el sexo no participan del mismo proceso fisiopatológico que la dislipemia en la etiología de la enfermedad coronaria. (Figura 5.1)

En otro estudio se explora la asociación entre un agente químico (anilina) como factor de exposición y cáncer de vejiga como evento. Habrá que considerar si el hábito de fumar no es un posible factor de confusión. Nos preguntamos entonces si las personas que trabajan con anilinas fuman más que las que no están expuestas a anilinas. En otras palabras ¿existe asociación entre exposición a anilinas y tabaquismo? (asociación entre factor de exposición y variable posiblemente confundidora). Si encontramos igual tasa de fumadores entre expuestos y no expuesto a anilinas, no podemos considerar al tabaquismo como confundidor.

Bajo estas condiciones, supongamos que encontramos que los fumadores tienen una tasa de cáncer de vejiga más alta que los no fumadores, independientemente de estar expuesto a anilinas. Entonces en este estudio consideraremos al tabaquismo como otro factor de exposición asociado a cáncer de vejiga, pero no como un confundidor. (Figura 5.2)

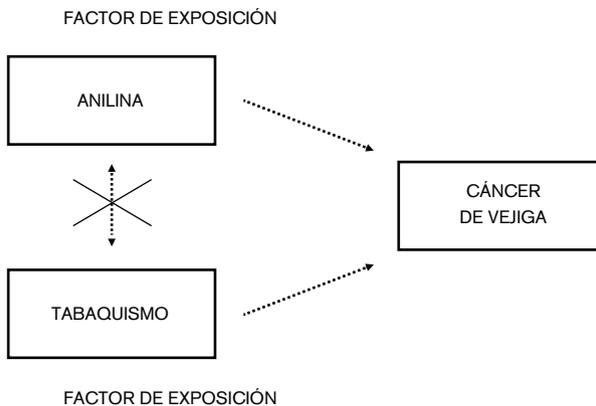


Figura 5.2: Factores de exposición

En vista de lo observado, cuando se planea un estudio de investigación, es de suma importancia determinar a priori qué variables podrían ser potenciales confundidores y recabar estos datos. Si esto no se hace, se corre el riesgo de no poder controlar la asociación investigada por variables confundidoras que podrían afectar esta asociación de forma importante y, por ende, nuestros resultados podrían no ser válidos.

Cuando al valorar una asociación se tienen en cuenta uno o varios confundidores, se dice que la asociación está ajustada por esos factores de exposición.

Muchas veces vemos que un estudio reporta sus resultados aclarando que se realizó un *ajuste* por sexo y edad u otras variables.

Sexo y edad casi siempre cumplen con las condiciones para ser confundidores.

En el ejemplo de beber café y cáncer de páncreas diríamos que “una vez ajustada por la variable tabaquismo”, no existe asociación entre tomar café y cáncer de páncreas. También podríamos verificar la asociación luego de ajustar por sexo, edad.

Ahora que ya sabemos que existen factores que pueden afectar la verdadera asociación entre un factor de exposición y un evento, debemos aprender cómo controlarlos o manejarlos.

Podemos evitar la influencia de los factores de confusión en dos etapas diferentes del desarrollo de un estudio epidemiológico: en el momento del diseño o en el momento de analizar estadísticamente los datos.

MÉTODOS PARA CONTROLAR CONFUNDIDORES EN LA ETAPA DE DISEÑO

1. *Restricción*: se puede seleccionar la población incluyendo en el estudio sólo sujetos que pertenezcan a un mismo nivel o categoría del posible confundidor. Por ejemplo, podríamos haber elegido sólo a no fumadores para testear la asociación entre beber café y cáncer de páncreas.

Este es un método simple y permite evitar el ajuste posterior. Su desventaja es que, al seleccionar sólo un grupo de sujetos, los resultados no son generalizables. Además, el número de sujetos incluidos es menor, con lo cual podríamos tener problemas con el poder estadístico de la muestra.

2. *Aparear la muestra*: si el estudio tiene un diseño caso-control, los sujetos controles deben seleccionarse de tal manera que tengan las mismas características que los casos en relación con el posible factor de confusión; por ejemplo, si los factores de confusión que tenemos que controlar son edad y sexo, por cada caso se seleccionará un control del mismo sexo y de la misma edad.

De esta manera, como cada par caso-control tiene el mismo nivel o categoría del confundidor, este ya no afecta la asociación.

Cuando apareamos por factores de confusión, no es posible examinar el efecto de las variables confundidoras sobre el evento.

Si apareamos por edad, no podremos examinar si la edad afecta el evento de interés, por ejemplo la mortalidad.

No obstante, aparear hace que el estudio sea más eficiente, dado que identifica otros factores de exposición y permite controlar variables difíciles de medir como nivel socioeconómico. Si apareamos por la variable lugar de residencia, se puede indirectamente aparear por nivel socioeconómico porque en general son variables asociadas.

Si utilizamos la técnica de apareamiento, los datos deben analizarse con pruebas estadísticas específicas para estos diseños. De otra manera el factor de confusión no estará controlado.

La dificultad que representa aparear radica en el reclutamiento de los sujetos controles, debido a que estos deben tener los mismos factores de confusión que los casos.

3. *Aleatorización*: esta es la técnica más utilizada en los diseños experimentales. Es más conocida como randomización (random en inglés significa azar).

Su gran ventaja es que permite igualar todos los confundidores posibles, conocidos y no conocidos, entre los grupos a comparar. Cuando seleccionamos la población por aleatorización o, lo que es lo mismo, asignamos un tratamiento a los sujetos por azar, los grupos a comparar son iguales respecto de todas las variables recabadas y las no recabadas. Las desventajas de la aleatorización son: que el método no se puede utilizar en los diseños observacionales, que requiere una gran cantidad de sujetos y que muchas veces no es ético asignar un tratamiento al azar a un determinado grupo de pacientes.

MÉTODOS PARA CONTROLAR CONFUNDIDORES EN LA ETAPA DE ANÁLISIS

1. *Estratificación*: consiste en analizar la asociación por separado en sujetos que pertenecen a niveles o categorías diferentes del confundidor. Como vimos antes, podemos verificar la asociación entre beber café y cáncer de páncreas en los fumadores y no fumadores.

Cuando estratificamos asumimos que, dentro de cada categoría o estrato, el efecto del confundidor fue removido. Esto es cierto pero depende de cuán amplio sea el estrato. Por ejemplo, si el factor de confusión es la edad, podemos dividir la población en intervalos de 5 años y verificar la asociación en cada estrato. Pero si los intervalos son de 20 años, se puede tener un impacto diferente: podemos comparar los sujetos dentro del grupo etario de 40 a 45 años pero comparar un estrato que incluya sujetos entre 50 y 70 años no parece muy adecuado. En este estrato hay una diferencia de edad tan grande que la asociación buscada puede verse influenciada, por lo que se debe considerar que el factor de confusión no ha sido removido.

Otro problema es que el factor de confusión haya sido adecuadamente asignado en todos los sujetos del estudio. Si un sujeto fuma y lo niega, será clasificado erróneamente en el grupo no fumador y el factor de confusión seguirá afectando la asociación. En cualquiera de estos dos casos tendremos lo que se denomina confusión residual.

2. *Análisis multivariable*: este tipo de análisis es el que se emplea con mayor frecuencia.

Es una técnica estadística que permite valorar la asociación en estudio teniendo en cuenta al mismo tiempo todos los posibles factores de confusión. Los análisis multivariados más frecuentemente usados son la *regresión logística* y la *regresión de Cox*. Cada uno se aplica a diferente tipo de diseño de estudios y serán estudiados en profundidad en otros capítulos.

Es importante destacar que no existe un criterio claro y *preestablecido* para definir una variable como confundidora. Determinar que una variable es un confundidor depende de los resultados obtenidos en las pruebas estadísticas necesarias a aplicar y del criterio del investigador.

INTERACCIÓN

Cuando analizáramos una asociación (por ejemplo, beber café y cáncer de páncreas) de acuerdo con estratos de una tercera variable (por ejemplo, edad) podríamos encontrar que, a medida que el estrato se incrementa, la asociación en estudio puede variar.

Edad	Cáncer de páncreas en bebedores de café	Cáncer de páncreas en no bebedores de café
20-29 años	1%	1%
30-39	1,5%	1,2%
40-49	6%	6,3%
50-59	9%	10%
60-69	9,5%	10,2%
70-79	10%	10,5%
> 79	11%	10,9%

Tabla 5.5: Datos hipotéticos de incidencia de cáncer de páncreas en bebedores de café y no bebedores estratificada por edad

Dicho de otro modo, la relación entre factor de exposición y evento varía en cada nivel de una tercera variable (confundidora o no) o, lo que es lo mismo, la asociación no tiene la misma fuerza cuando la analizamos en los diferentes estratos de una tercera variable. (Tabla 5.5)

Como vemos en la Tabla 5.5, a medida que se incrementa la edad aumenta la incidencia de cáncer de páncreas, tanto en el grupo bebedor de café como en el no bebedor. Si tomamos como referencia el grupo de 20-29 años, vemos que la incidencia de cáncer de páncreas es 6 veces mayor en el grupo 40-49 años, 9 veces mayor en el grupo siguiente y 10 veces mayor en el grupo de 70-79 años.

Si se verifica, con una prueba estadística que este incremento de la incidencia está relacionado con el estrato de edad (tercera variable), podremos decir que existe una interacción entre el factor de exposición y la tercera variable, porque la combinación de ambos tiene un impacto diferente en la tasa de eventos que el que tiene cada variable por separado sobre el evento.

Si un paciente bebe café y tiene 40 años tendrá una incidencia de cáncer de páncreas menor que aquel sujeto que bebe café y tiene 70 años. Al mismo tiempo, el riesgo de cáncer de páncreas no es el mismo sólo por beber café o sólo por tener 70 años; si se tienen los dos factores de exposición el riesgo es mayor. Este riesgo puede ser la suma de los riesgos independientes o excederla.

Cuando encontramos interacción, es muy importante reportar los datos por cada estrato. Es una evidencia muy fuerte a favor de un supuesto mecanismo fisiopatológico que queremos demostrar.

Es muy importante resaltar que cualquier factor de exposición puede ser sólo un factor de exposición, sólo un confundidor, sólo un factor de interacción, ninguno o más de uno.

Siempre deben evaluarse todos los factores de exposición que se analizan en un estudio a fin de determinar qué tipo de factores son.

Fumar es un factor de exposición o de riesgo y un confundidor en un estudio que examine la asociación entre hipertensión y enfermedad coronaria.

Para determinar qué tipo de variable es un factor de exposición se deben emplear los conceptos teóricos aquí desarrollados, pruebas estadísticas y aplicar criterio epidemiológico.

REPORTE DE FRECUENCIA DE FACTOR DE EXPOSICIÓN Y EVENTO



ESTIMACIONES

La Epidemiología tiene entre sus objetivos describir la distribución de eventos o enfermedades e identificar sus factores de exposición asociados.

Para lograr este objetivo, lo primero es determinar la frecuencia del evento de interés y de otros factores relacionados.

Esto permite luego la comparación de las frecuencias de la enfermedad entre diferentes poblaciones así como entre individuos con y sin determinados factores de exposición o características en la misma población.

La forma más simple de medir la frecuencia de un evento y/o factor de exposición es sencillamente contar la cantidad de casos, lo que no es de mucha utilidad en Epidemiología.

Para investigar la distribución y los determinantes de un evento o factor de exposición es necesario conocer el tamaño de la población de donde provienen estos casos y el período de tiempo durante el cual esta información fue recolectada.

Con esta información es posible comparar directamente la frecuencia del factor de exposición/evento entre diferentes poblaciones o entre diferentes grupos de individuos.

En la Tabla 6.1 puede observarse el número de casos de infarto de miocardio en dos ciudades diferentes, el período de tiempo en que se reportaron los casos y el tamaño de la población.

En la ciudad A se reportaron 58 nuevos casos de infarto, mientras que en la ciudad B hubo 35 casos en 2 años. Esta información podría servirle a los servicios de salud pública de cada ciudad con el fin de determinar la necesidad de recursos para tratar esta patología.

Sin embargo, la observación cruda de estos datos podría llevar a conclusiones erróneas con respecto a qué ciudad tiene más casos.

Para evitar este error y poder comparar la frecuencia del infarto entre las dos ciudades, hay que tener en cuenta el tamaño de la población y el período de tiempo de reporte.

En la ciudad A se reportaron en un año 58 casos, en una población de 25 000 personas. Por lo tanto la tasa (*rate*) anual de infarto para esta ciudad es de 58 por 25 000 habitantes. En la ciudad B se reportaron en dos años 35 casos en una población de 7 000 habitantes; la tasa anual es de 17,5 por 7 000 (35 en dos años, 17,5 en un año).

Hasta aquí se han igualado los períodos de tiempo, pero el tamaño de las poblaciones sigue siendo diferente. Por lo que todavía no es posible hacer una comparación directa.

Ciudad	Nuevos casos de infarto	Período	Población
A	58	1986	25 000
B	35	1984-1985	7 000

Tabla 6.1: Frecuencia de infarto de miocardio en dos ciudades

Para hacer más fácil la comparación de las tasas se debe usar la misma unidad de población, es decir, emplear el mismo numerador en ambas tasas.

En general las tasas se reportan como número de eventos por cada 100 personas observadas en un año, o sus múltiplos 1 000, 10 000, 100 000.

En el caso de la ciudad A, tendríamos 232 casos por 100 000 personas/año (calculado por regla de tres simple) mientras que en la ciudad B habría 250 casos por 100 000 personas/año.

Una vez igualados el tamaño de la población y el período de tiempo, la comparación se hace de manera directa.

Ahora podemos observar que, a pesar de haberse reportado menos casos de infarto en la ciudad B, en realidad esta tiene una tasa de infartos más alta comparada con la ciudad A.

En general el término *tasa* se utiliza en Epidemiología para referirse a la frecuencia de una enfermedad reportada como: número de casos/personas/año. Esta medida habitualmente se emplea en estudios de cohortes, y como factor distintivo tiene incluido el factor tiempo en el denominador.

Pero a veces el término *tasa* también se refiere a proporciones o razones (*ratio*). La forma más simple de medición de la frecuencia de una enfermedad es la razón o *ratio*, que se obtiene dividiendo una cantidad por otra sin que estas tengan una relación específica: por ejemplo número de prematuros por 1 000 nacidos vivos.

Una proporción es un tipo de razón en la que los casos del numerador están incluidos en el denominador: por ejemplo, porcentaje de mujeres mayores a 50 años con histerectomía. (El denominador incluye a todas las mujeres mayores a 50 años y el numerador sólo a aquellas con histerectomía.)

Dado que el término *tasa* puede referirse tanto a verdaderas tasas, proporciones o razones, debe aclararse qué tipo de medida de frecuencia de enfermedad se está reportando.

PREVALENCIA E INCIDENCIA

PREVALENCIA

La cuantificación de la frecuencia de un factor de exposición o evento está contenida dentro de dos amplias categorías: la prevalencia y la incidencia.

La prevalencia cuantifica la proporción de individuos que poseen un factor de exposición o evento (independientemente de cuándo la hayan adquirido) en una población y en un punto específico de tiempo. Nos proporciona una estimación de la situación que presenta una población en relación con el factor de exposición o evento en ese período de tiempo.

$$\text{Prevalencia} = \frac{\text{Nº de casos existentes en un punto específico de tiempo}}{\text{Población total}}$$

Por ejemplo, durante una encuesta realizada en individuos de entre 50 y 80 años se encontraron 310 casos de miopía de un total de 2 700 encuestados durante un mes. En consecuencia, la prevalencia de la miopía en esta población es de 11,5% durante el período de tiempo en que se realizó la encuesta.

Este punto específico de tiempo puede ser un período de tiempo calendario o un período fijo en el curso de los eventos, y podría variar en tiempo real de persona a persona, como por ejemplo el comienzo de la menopausia (prevalencia de histerectomía en pacientes menopáusicas) o la pubertad (prevalencia del acné en la pubertad). Otro ejemplo es la prevalencia de la fibrilación auricular en el tercer día postoperatorio.

Esta medición de la cantidad de sujetos con factor de exposición o evento (casos) es la forma que se utiliza para reportar factores de exposición y/o eventos en los estudios transversales.

INCIDENCIA

En contraste con la prevalencia, la incidencia cuantifica el número de nuevos casos con un evento en una población de individuos en riesgo (sujetos que no tienen el evento) durante un período específico de tiempo.

La distinción entre prevalencia e incidencia es que en la segunda se incluyen sólo los nuevos casos como numerador y sólo la población en riesgo en el denominador. Esto significa que los sujetos que ya tienen el evento de interés quedan excluidos del denominador.

INCIDENCIA RIESGO

La incidencia puede calcularse como riesgo o como tasa.

La incidencia calculada como riesgo (*incidence risk*) es la probabilidad que tiene una persona libre de enfermedad de desarrollarla en algún período de tiempo y generalmente se la expresa como porcentaje.

$$\text{Incidencia riesgo} = \frac{\text{N}^\circ \text{ de nuevos casos en un período específico de tiempo}}{\text{N}^\circ \text{ de personas en riesgo de contraer la enfermedad al inicio del período de tiempo}}$$

Por ejemplo, en una población de 1 000 hipertensos y dislipémicos seguidos durante 5 años se observaron 100 infartos de miocardio: la incidencia de infarto, calculada como riesgo es del 10%.

Es posible reportar la incidencia como riesgo cuando todos los individuos se siguen durante la misma cantidad de tiempo.

INCIDENCIA TASA

Cuando la incidencia se calcula como tasa (*incidence rate*) el número de casos nuevos no está relacionado con el total de individuos en riesgo al comienzo del período de observación, sino con el número de personas/años en riesgo (*person/year at risk*) observadas durante el período de estudio.

Supongamos que se inicia la observación de 1 000 individuos hipertensos y dislipémicos, sin infarto previo. No todos estos individuos van a contribuir con la misma cantidad de tiempo al estudio. Algunos serán observados desde el inicio del estudio hasta que tengan un infarto, otros empezarán a ser observados luego de iniciado el estudio y otros empezarán y terminarán el estudio completando el tiempo total de seguimiento sin desarrollar eventos.

En la Figura 6.1 se puede observar un ejemplo con cinco pacientes.

El paciente 1 fue observado 2 años, el 2 durante 3 años, el 3 durante 5 años, etc.

Al tiempo de seguimiento se lo denomina tiempo total en riesgo.

Para el cálculo de personas/años en riesgo se suman los períodos de tiempo que cada paciente fue observado, en este caso 16,5.

Sólo dos pacientes desarrollaron infarto, por lo que la incidencia calculada como tasa es de 2 casos/16,5 personas/año.

	ENE 1976	JUL 1976	ENE 1977	JUL 1977	ENE 1978	JUL 1978	ENE 1979	JUL 1979	ENE 1980	JUL 1980	ENE 1981	TTR
A		■	■	■	■	■						2
B	■							X				3
C	■	■	■	■	■	■	■	■	■	■	■	5
D		■	■	■	■	■	■	■	■	■		4
E					■	■	■	■	■	X		2,5
TAR												16,5

Figura 6.1: Cálculo tiempo / personas

TTR: Tiempo total en riesgo

x = eventos

TAR: Total años en riesgo

■ = tiempo de seguimiento

Como las tasas se reportan con un denominador basado en el 100 o sus múltiplos, esta tasa puede reportarse como 12,1/100 personas/año.

Decimos entonces que la incidencia de infarto es de 12 casos/100 personas/año en riesgo.

Este tipo de reporte de la incidencia es el que debe emplearse en los estudios de cohortes.

En el caso de una enfermedad rara o de baja ocurrencia, las incidencias calculadas como riesgo y como tasa serán muy parecidas, ya que el número de sujetos en riesgo será aproximadamente igual a la población total.

Para una enfermedad común la incidencia será diferente dependiendo de cómo se calcule.

Por ejemplo, si consideramos la incidencia de varicela durante el segundo año de vida en una población con bajo índice de vacunación, siguiendo a 1 000 niños libres de esta enfermedad desde su primer año de vida y por un año, podríamos observar que 140 niños tuvieron varicela, dejando 860 todavía en riesgo de contraerla al final de ese año de seguimiento.

La incidencia calculada como riesgo es $140/1\ 000 = 14\%$. Esta es la probabilidad de que un niño contraiga varicela durante el segundo año de vida.

Si se deseara calcular la incidencia como tasa y el tiempo de seguimiento fuera de un año para los 1 000 niños, el número de niños en riesgo/año sería el promedio del número de niños al comienzo + número de niños aún en riesgo al final, es decir $= (1\ 000 + 860)/2 = 930$.

La incidencia como tasa entonces es $140/930 = 15/100$ niños/año en riesgo.

La tasa esperada de riesgo de varicela por cada 100 niños no vacunados durante el segundo año de vida es de 15 casos.

RELACIÓN ENTRE INCIDENCIA Y PREVALENCIA

La prevalencia incluye todos los casos existentes en una población. Por lo tanto la cantidad de estos casos se verá afectada por la incidencia de la enfermedad (número de nuevo casos) y por la duración de la misma. (Figura 6.2)

Si una enfermedad tiene un incidencia baja y su mortalidad es alta (tiempo de duración corto) la prevalencia en dicha población será baja. Por el contrario, si la incidencia es alta y el tratamiento permite prolongar la enfermedad por mucho tiempo, la prevalencia será mucho más alta. (Figura 6.2)

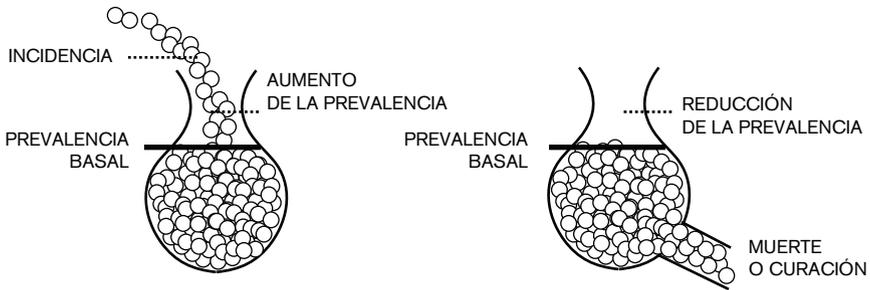


Figura 6.2: Relación entre incidencia y prevalencia

Es por esto que la comparación de prevalencias entre diferentes poblaciones, por ejemplo entre individuos del mismo lugar pero de diferentes clases sociales, no es muy útil y debe interpretarse con cautela.

La prevalencia está muy afectada por todos los factores que condicionan la incidencia y la duración de la enfermedad.

Supongamos que la prevalencia de diabetes en el país A es del 7 vs. el 3% en el país B. Si no tomáramos en cuenta los factores que influyen la prevalencia, concluiríamos que el país A tiene más diabetes. Si en cambio descubriéramos que en el país B el diagnóstico de diabetes es bajo, se buscan menos nuevos casos (menor incidencia), y que además el acceso al tratamiento es menor (mueren más rápido) podríamos explicar la diferencia de prevalencia de diabetes entre ambos países.

Consideren el siguiente ejemplo: en un país de 100 000 habitantes, 2 000 sujetos tienen asma. Diez años más tarde, 300 personas murieron (25 de las cuales tenían asma), nacieron 250 bebés y 500 personas desarrollaron asma.

La prevalencia de asma basal en la población fue de 2% (2 000 casos existentes de asma/100 000 población total).

La incidencia a 10 años, considerando la población basal, fue de 0,51% (500 nuevos casos/98 000 habitantes totales en riesgo).

La prevalencia a 10 años fue de 2,5% (2 475 casos existentes/99 950 población total a 10 años). De los 2 000 casos iniciales se agregaron, a 10 años, 500, pero se murieron 25 casos, por lo tanto quedaron en la población 2 475 casos de asma.

De la población inicial de 100 000 habitantes, en 10 años murieron 300 personas pero nacieron 250, por lo que la población se redujo sólo en 50 personas.

La prevalencia a 10 años aumentó debido a que en este período murieron menos casos de asma de los que se desarrollaron (25 casos muertos y 500 nuevos casos).

Supongamos ahora que en los 10 años murieron 200 casos y se desarrollaron sólo 20 casos.

La incidencia a 10 años sería de 0,0002% y la prevalencia de 1,8% (1 820 casos existentes/99 950 población total).

Como vemos, prevalencia e incidencia están íntimamente relacionadas, dependiendo de la dinámica de la enfermedad y de la población.

CASO Y POBLACIÓN EN RIESGO



La base de la investigación epidemiológica es cuantificar la ocurrencia de estados relacionados con la salud (factores de exposición y/o eventos) en una población específica.

Dependiendo del motivo de investigación, se puede cuantificar factores de exposición y/o eventos, entonces un sujeto será un caso si posee la condición que se está estudiando.

Antes de cuantificar cualquier factor de exposición o evento de interés, debemos claramente definir que queremos decir con caso y la población de la cual estos casos provenirán (población en riesgo).

DEFINICIÓN DE CASO

Un caso será un sujeto que presente el factor de exposición y/o evento de interés.

En un estudio de cohortes en el cual se estudie la relación entre uso de anticonceptivos y cáncer de mama, caso será toda mujer que, según la definición adoptada para el factor de exposición, haya usado anticonceptivos.

Toda mujer que durante el desarrollo del estudio presente cáncer de mama, según la definición adoptada para el evento, será un caso.

Tomemos dos poblaciones X e Y. En la población X, se reportó que 50 de 100 000 hombres desarrollaron enfermedad coronaria y en la población Y, 70 de 100 000 en el mismo año. (Figura 7.1)

¿Podemos decir que en la población Y hubo más casos de enfermedad coronaria que en la población X?

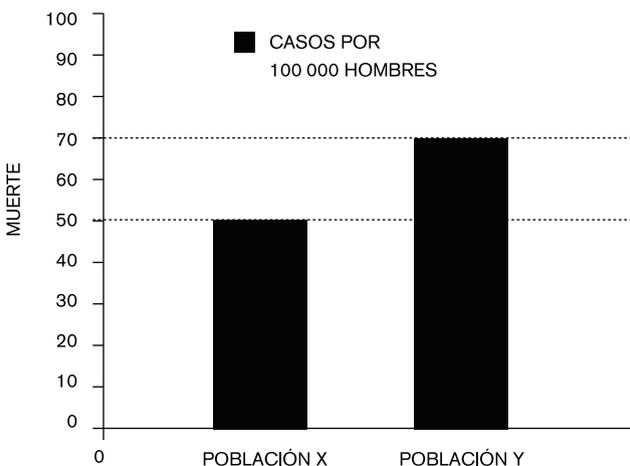


Figura 7.1: Prevalencia de enfermedad coronaria

No necesariamente. Hay muchas maneras de definir casos de enfermedad coronaria. Solo podemos comparar datos de las dos poblaciones si ambas han utilizado la misma definición de caso.

Una definición de caso es un conjunto de criterios estandarizados usados para identificarlos.

Una buena definición de caso debe permitir identificar claramente si una persona podría ser definida como caso o no.

La definición de caso debe incluir:

1. El/los métodos (criterios) usados para identificar casos.
2. Los límites de un caso.
3. La unidad de análisis.

CRITERIOS PARA IDENTIFICAR CASOS

Es importante conocer los criterios que han sido utilizados para identificar los casos.

En el ejemplo anterior, ¿cómo fueron diagnosticados los pacientes con enfermedad coronaria? ¿Por antecedentes clínicos (historia de dolor de pecho, infarto o síndrome coronario agudo) mediante estudios de provocación de isquemia miocárdica o mediante angiografía coronaria?

LÍMITE DE CASO

Debe dejarse claro dónde está la línea divisoria entre caso y no caso.

Por ejemplo, la definición de enfermedad coronaria ¿incluye casos subclínicos (diagnóstico por test de provocación de isquemia) o sólo casos sintomáticos?

Si el factor de exposición o evento de interés es un número continuo, debemos establecer un punto de corte para separar casos de no casos.

Por ejemplo en un estudio se podría definir hipertensión como tensión arterial diastólica ≥ 90 mmHg y en otro estudio como ≥ 80 mmHg. Dependiendo del límite que se establezca, un sujeto con tensión arterial ≥ 80 mmHg podría ser considerado como hipertenso en un estudio y en el otro no.

Años atrás el diagnóstico de diabetes requería de dos determinaciones consecutivas de glucemia en ayunas ≥ 140 mg/dL. En años recientes esta definición cambió a dos determinaciones consecutivas de glucemia en ayunas ≥ 126 mg/dL.

Cualquier comparación de estudios llevados a cabo en cada uno de estos períodos de tiempo no será válida, debido a que el límite para identificar casos fue diferente.

UNIDAD DE ANÁLISIS

Un caso podría ser desde una escuela, una familia, una persona con cierta condición, un episodio de un evento, una enfermedad y aun un órgano o tejido.

La unidad de análisis utilizada para definir el caso afectará la estimación de su ocurrencia.

Un caso de tumor renal podría referirse a hombres con tumor renal (sólo un caso posible por persona), a riñones con tumor (dos casos posibles por persona) o a tumores malignos renales (varios casos por persona).

Por ejemplo, una comparación de dos registros separados mostró que los pacientes de la población X asistieron a la sala de emergencia del hospital comunitario mucho más frecuentemente que en la población Y.

Un análisis de los datos mostró que en la población Y sólo se contó la primera consulta por una determinada condición a la sala de emergencia como caso, mientras que en la población X se contaron todas las veces que el paciente asistió por la misma condición como casos separados.

Veamos un ejemplo con los tres criterios para definir caso.

En un estudio donde se desea determinar la frecuencia de insuficiencia cardíaca, se definió caso como “Un episodio que presente un paciente con antecedentes de insuficiencia cardíaca que incluya síntomas como: disnea de reposo, edema periférico, hepatomegalia y tercer ruido cardíaco de duración mayor a 24 horas”.

Según esta definición la unidad de análisis es *episodio* de insuficiencia cardíaca.

El método o criterio que se utilizó fue: “síntomas como disnea de reposo, edema periférico, hepatomegalia y tercer ruido cardíaco” y el límite fue la duración de los síntomas > 24 horas.

Supongamos que la definición de caso cambia a:

Paciente con antecedentes de insuficiencia cardíaca con síntomas que incluyan disnea de reposo, edema periférico, hepatomegalia y tercer ruido cardíaco de duración mayor a 24 horas.

¿Cómo esperaría que cambie la frecuencia de insuficiencia cardíaca descompensada?

Disminuiría la frecuencia de insuficiencia cardíaca descompensada debido a que la unidad de análisis es ahora el paciente y no como antes cada episodio.

Puede haber varios episodios por paciente y obviamente menos pacientes que episodios.

Otro cambio de la definición de caso:

Paciente con antecedentes de insuficiencia cardíaca con síntomas que incluyan disnea de reposo, edema periférico, hepatomegalia y tercer ruido cardíaco de duración mayor a 24 horas, causado por abandono de la medicación.

¿Cómo cambiaría la frecuencia de insuficiencia cardíaca descompensada?

La frecuencia de insuficiencia cardíaca descompensada disminuiría debido a que ahora es el método para identificar casos el que cambió.

Hay menos pacientes con insuficiencia cardíaca descompensada por abandono de la medicación que pacientes con insuficiencia cardíaca descompensada de cualquier causa.

Ahora la definición cambia a:

Paciente con antecedentes de insuficiencia cardíaca con síntomas que incluyan disnea de reposo, edema periférico, hepatomegalia y tercer ruido cardíaco.

En este caso la frecuencia de insuficiencia cardíaca descompensada aumentaría. Su límite ha cambiado. De hecho no se ha establecido un límite.

Hay más pacientes con insuficiencia cardíaca descompensada con los síntomas requeridos que no tengan límite de tiempo, que pacientes con insuficiencia cardíaca descompensada cuyos síntomas duren más de 24 horas. Con esta definición podemos incluir cualquier paciente que padezca estos síntomas por un período menor o mayor a 24 horas.

Como vemos en los ejemplos, si se cambia alguno de los elementos de la definición de caso, la frecuencia del evento cambiará.

DESARROLLO DE LA DEFINICIÓN DE CASO

Con tantas posibilidades para definir un caso ¿cómo establecemos nuestra propia definición?

No hay una sola respuesta a esta pregunta, pero existen diferentes factores que debemos tener en cuenta a la hora de desarrollar una definición.

1. La pregunta que deseamos responder en nuestra investigación
2. Los recursos disponibles
3. Aspectos éticos

LA PREGUNTA QUE DESEAMOS RESPONDER

El objetivo de nuestra investigación influirá en la definición de caso que desarrollemos. Consideraremos tres objetivos típicos.

1) Si queremos comparar nuestros resultados con los provenientes de otros estudios, debemos utilizar la misma definición de caso que fue empleada en aquellos.

Para comparar la ocurrencia de trastornos de ansiedad debemos definir esta patología según los criterios del DSM IV o según los criterios del ICD 10. Estas dos clasificaciones son diferentes. Si queremos comparar la frecuencia de trastornos de ansiedad entre dos o más poblaciones, debemos asegurarnos de que se ha utilizado la misma definición en cada estudio reportado.

2) Testear una hipótesis

Si queremos explorar una hipótesis acerca de determinantes de una enfermedad, debemos usar una definición de caso muy precisa. Esta debe incluir el máximo de verdaderos casos de la condición en estudio y excluir el máximo número de casos de otras condiciones que sean similares pero diferentes.

En un estudio sobre factores de riesgo para sarampión en una población de niños que viven en un área donde las enfermedades eruptivas de la infancia son frecuentes, debemos exigir confirmación de laboratorio, debido a que el diagnóstico clínico podría hacernos incluir casos que sean parecidos pero que no se deban a infección por sarampión.

3) Investigación de epidemias

En este caso se puede usar diferentes definiciones de caso (caso sospechoso, probable o confirmado) dependiendo de cuán certero es el diagnóstico y de si se realizó o no laboratorio confirmatorio de la enfermedad.

En una investigación de epidemia de fiebre amarilla, un caso probable es aquel que muestra síntomas típicos de la enfermedad, y un caso confirmado es el que tiene serología positiva para esta enfermedad.

RECURSOS DISPONIBLES

Si tenemos recursos limitados, no podremos utilizar instrumentos o pruebas diagnósticas caras para identificar casos.

Podemos definir nuestros casos basados en criterios diagnósticos menos rigurosos, sin embargo esta circunstancia no nos exime de ser bien explícitos en nuestra definición.

Si deseamos investigar frecuencia de menopausia precoz, podría ser muy caro realizar todas las determinaciones de laboratorio para confirmar el diagnóstico. No obstante, podemos adoptar una definición clínica estricta que incluya a aquellas mujeres menores de 40 años, con amenorrea consecutiva de más de un año de evolución, excluyendo aquellas < 40 años que no menstrúen desde hace un año pero que estén recibiendo algún

tratamiento que cause amenorrea, que se les haya realizado histerectomía o que tengan cualquier otro factor secundario que produzca falta de menstruación.

DEFINICIÓN DE LA POBLACIÓN DE ESTUDIO

Una vez desarrollada claramente la definición de caso, debemos definir la población que debemos estudiar.

La población de estudio es el conjunto de sujetos, quienes potencialmente podrían ser casos.

La definición de caso influencia entonces la definición de la población de estudio.

En un estudio de sobrevivida a largo plazo de cáncer de ovario, llevado a cabo en hospitales públicos de regiones del norte de Argentina, se definió caso como aquellas mujeres que han sobrevivido más de 5 años luego del diagnóstico de cáncer de ovario, confirmado por biopsia.

La población de estudio serán todas las mujeres con cáncer de ovario confirmado por biopsia, con sobrevivida mayor a 5 años y que asisten a hospitales públicos de regiones del norte de Argentina.

POBLACIÓN EN RIESGO

La población de estudio generalmente está limitada a una locación, podemos estudiar todos los sujetos que podrían ser casos en un hospital, pueblo, colegio, provincia o país.

La población de estudio también es denominada población en riesgo, debido a que representa a la población que está en riesgo de desarrollar el evento de interés.

En un estudio sobre trombosis venosa profunda poscirugía ortopédica, llevado a cabo en hospitales de alta complejidad de la ciudad de Santiago de Chile en el año 2007, trombosis venosa profunda fue definida como la aparición de signos y síntomas característicos dentro de las 72 horas poscirugía.

La población en riesgo son todos los pacientes con cirugía ortopédica operados en hospitales de alta complejidad en Santiago de Chile en el año 2007. Cada sujeto es parte de la población en riesgo sólo durante las primeras 72 horas poscirugía.

Supongamos que se desea determinar la prevalencia en hombres de infarto de miocardio en todo el país. Se considera que los hombres < 30 años no están en riesgo de sufrir un infarto.

La población total del país incluye niños, mujeres y hombres adultos, por supuesto también hombres < 30 años y hombres con infarto.

La población en riesgo incluye sólo a hombres mayores de 30 años sin infarto.

Los casos serán aquellos hombres con infarto.

Como vemos, la cantidad de sujetos en cada población va disminuyendo.

Para determinar la prevalencia de infarto en hombres de todo el país, elegiremos como población sólo a hombres > 30 años de todo el país.

En la práctica puede ser difícil excluir de la población de estudio a sujetos que no están en riesgo de ser casos. Sin embargo, si somos precisos sobre quiénes han sido incluidos y quiénes han sido excluidos de la población de estudio, nuestros resultados podrán ser interpretados claramente.

Supongamos que a partir de estadísticas estatales se quiere determinar la ocurrencia de cáncer de riñón en hombres en Argentina, se reporta la ocurrencia de cáncer de riñón como el porcentaje de la población total de hombres residentes en Argentina que lo padecen.

Aunque los hombres con nefrectomía bilateral en diálisis no están en riesgo de desarrollar cáncer de riñón, como las estadísticas estatales los contabilizan como hombres, estos serán incluidos dentro de la población en riesgo.

Debemos entonces aclarar que este grupo de sujetos fue incluido dentro de la población en riesgo.

Esto es importante, debido a que la población en riesgo tiene mayor número de sujetos y por lo tanto el porcentaje de cáncer de riñón será menor en comparación a un estudio que haya excluido a los sujetos con nefrectomía bilateral en diálisis.

Como vemos, dependiendo de la definición de caso y de la población en riesgo, el resultado de la cuantificación de factores de riesgo y/o eventos cambiará.

CLASIFICACIÓN DE VARIABLES

Llamamos variables a los datos recolectados para un estudio de investigación.

Tales datos o variables pueden ser inicialmente números, como edad, peso, altura, resultados de laboratorio, etc., o bien descripciones de varios aspectos de los pacientes, como antecedentes patológicos (insuficiencia renal, EPOC, cirugías previas, estadios de tumores), descripciones de estudios diagnósticos (función ventricular izquierda, resultados de endoscopias digestivas, clasificación de PAP), procedimiento realizado a los sujetos (tipo de cirugía), etc.

La distinción entre ambos tipos de variables es fácil: unos son números y los otros son palabras que luego tendremos que *traducir* a un lenguaje que la estadística entienda.

A los datos que en su origen son números se los denomina cuantitativos o numéricos, y a aquellos que describen características (las palabras) se los denomina datos cualitativos o categóricos.

¿Por qué es importante esta clasificación? Porque la elección de la prueba estadística a aplicar dependerá del tipo de dato recogido.

CLASIFICACIÓN DE DATOS NUMÉRICOS

La clasificación de los datos numéricos no tiene implicancias sobre el tipo de prueba estadística que vamos a emplear. A dichos datos simplemente se los divide en dos grupos, de acuerdo a si la variable emplea números con decimales o sólo enteros, como por ejemplo cantidad de visitas al médico, número de hijos, vasos coronarios enfermos, etc.

A estos datos numéricos se los denomina *variables discretas*.

Aquellos datos numéricos cuyo número entero está seguido por decimales se llaman *variables continuas*: la tensión arterial puede ser de 140,5 mmHg, el potasio plasmático de 4,5 mEq/L, la altura de 1,23 cm, etc.

Esta clasificación también resulta útil para distinguir los datos discretos de aquellos de tipo categóricos que veremos a continuación.

CLASIFICACIÓN DE DATOS CATEGÓRICOS

Los datos cualitativos pueden dividirse en categorías; de ahí su nombre.

El tipo de variable categórica con la que estemos trabajando determinará la prueba estadística que emplearemos, de ahí la importancia de conocer y saber clasificar este tipo de variables.

Aquellas variables que sólo tienen 2 categorías o resultados posibles se denominan *variables dicotómicas*.

Por ejemplo: sexo, muerte Sí/No, diabético/no diabético, cirugía cardíaca previa Sí/No, fumador/no fumador, hipertenso/no hipertenso, etc.

Si la variable no posee naturalmente sólo 2 categorías, la decisión de que una variable se clasifique en 2 categorías la toma el investigador.

Cuando hay más de 2 categorías posibles, hay también dos posibilidades: que exista un *orden* natural de categorías (por ejemplo estadios de tumores, clasificación de disnea, valoración de la función ventricular izquierda, etc.), o que no exista dicho orden.

El primer caso es el de las variables que tienen un orden natural. A este tipo se las denomina *variables ordinales*, porque a medida que las categorías cambian implican en general peores estadios. Un estadio I de cáncer de mama no es lo mismo que un estadio IV; la función ventricular izquierda severa no es lo mismo que la función normal.

Algunas variables que han sido determinadas como dicotómicas por el investigador pueden también clasificarse como variables ordinales. Por ejemplo, podríamos clasificar el tabaquismo como nunca fumó/ex fumador/fumador actual, o la hipertensión en leve, moderada o severa, categorías que tienen diferentes implicancias para el paciente.

Otras variables no tienen un orden natural. Es lo mismo (en términos de gravedad) ser soltero, divorciado, casado o viudo, tener grupo sanguíneo A, B, o cero; estas situaciones no implican diferentes estados de gravedad. A estos datos se los llama *variables nominales*: sólo describen una característica del sujeto que no implica un peor estadio con respecto a otra categoría. De nuevo el investigador puede decidir clasificar una variable nominal como dicotómica o al revés. Por ejemplo, casado/no casado vs. soltero, divorciado, casado o viudo, diabético/no diabético vs. diabetes tipo 1, tipo 2, intolerancia a la glucosa.

Lo más recomendable siempre es recolectar la información como se presenta naturalmente y luego convertirla en el tipo de variable que sea más apropiado reportar en el estudio.

Si la variable es hipertensión, lo más apropiado es recabar el dato de la presión arterial; luego, según los puntos de corte que se elijan para los valores, estos pueden clasificarse en leves, moderados o severos.

Variable original	Cambia a
Numérica discreta	Catégorica ordinal
Edad en años (promedio)	0-17 años % niños
	18-30 años % jóvenes
	31-50 años % adultos jóvenes
	51-69 años % adultos mayores
	≥ 70 años % ancianos
Albuminuria mg/24 hs. (promedio)	< 30 Normal
	30-299 microalbuminuria
	≥ 300 albuminuria clínica
Catégorica nominal	Catégorica dicotómica
Infarto previo: no/inferior/anterior/lateral/ otros (porcentaje de cada categoría)	Infarto previo SÍ/NO (porcentaje)

Tabla 8.1: Conversión de variables

En lugar de preguntar por tabaquismo SÍ/NO, lo mejor es recabar lo más extensamente posible el dato: nunca fumó, ex fumador o fumador actual.

El número de vasos coronarios enfermos podría ser una variable numérica discreta (sin decimales): 1, 2 o 3, y se reportaría el promedio de vasos enfermos de la población, o bien podría ser una variable categórica ordinal (aunque 1, 2 o 3 no son palabras, podríamos asumirlas como tales); en este caso reportaríamos porcentaje de pacientes con 1, 2 y 3 vasos enfermos.

De acuerdo a cómo tomemos o consideremos el dato será la forma en que lo analizaremos y lo reportaremos.

Los datos cuantitativos se reportan, en general, como promedios, y los categóricos como porcentajes.

Como podemos deducir, las variables de tipo numérico, ya sean discretas o continuas, pueden entonces convertirse en variables categóricas, generalmente de tipo ordinal. En la Tabla 8.1 se muestran algunos ejemplos.

PRINCIPIOS DE LA ESTADÍSTICA

CAPÍTULO

9

Una vez recolectados los datos, ¿qué hacemos ahora? ¿Cómo traducimos los datos clínicos individuales en valores numéricos que resuman los datos de la población estudiada? ¿Cómo comparamos datos de una manera válida? ¿Qué podemos concluir con esta comparación?

La Estadística es la herramienta fundamental que nos permite traducir la experiencia de un grupo de individuos en datos numéricos para hacer inferencias acerca de la población estudiada y derivar conclusiones generales.

La Estadística, como toda ciencia, tiene su propio lenguaje, razonamiento y metodología. Por lo tanto, es fundamental entender cómo *razona* la Estadística y su lenguaje antes de comenzar a estudiar sus métodos.

Supongamos que deseamos saber cuál es la mortalidad del infarto agudo de miocardio en su etapa intrahospitalaria.

Para determinar este valor deberíamos tomar a todos los pacientes internados por infarto de miocardio del mundo. Como esto, obviamente, no es posible, se obtiene lo que se llama una *muestra* de pacientes, que representan a todos los infartos internados. Este es un grupo de sujetos que, teniendo las mismas características, representan a la población de interés (universo entero de todos los casos que están siendo estudiados).

La manera correcta de elegir una muestra para que sea representativa de la población es seleccionar los sujetos al azar, mediante un procedimiento conocido como aleatorización.

Si tomamos 10 000 pacientes provenientes de diferentes países y calculamos su mortalidad, podríamos concluir que la mortalidad del infarto de miocardio es, por ejemplo, del 8%, pero ¿es este valor único e inamovible?

Obviamente no es así, porque si tomamos otra muestra de 10 000 pacientes, casi seguro la mortalidad que observaremos no será exactamente 8%; quizás esta vez sea de un 10%.

De este ejemplo podemos concluir que, cada vez que realicemos un mismo experimento repetidas veces, nos encontraremos con distintos resultados.

Los datos obtenidos a través de métodos estadísticos no nos informan un valor absoluto e inamovible; lo que hace la estadística es *inferir* o deducir un valor que se acerque lo más posible al *supuesto valor real* (que nunca conoceremos) y asumir este valor como el *verdadero valor*.

La estadística entonces no mide valores exactos sino que los *estima*. Cuando hablamos de estimación nos referimos al cálculo más cercano que podamos hacer del verdadero valor y cuál es el grado de precisión con el cual este valor ha sido calculado.

El ejemplo más común son las encuestas que se realizan antes de un acto electoral: se toma una muestra suficientemente grande de sujetos al azar para estimar el porcentaje de votos de cada candidato en la población de interés (todas las personas mayores de 18 años del país). Una vez realizadas las elecciones se pueden comparar las estimaciones hechas a partir de las muestras con los resultados obtenidos.

Pensemos que en una elección no todas las personas mayores de 18 años concurren a votar, por lo que el acto electoral podría ser asumido como la muestra más grande posible de obtener, debido a que la única forma de saber el *verdadero porcentaje* de cada candidato sería que nadie falte a votar. Es por eso que decimos que el *verdadero valor* nunca lo conoceremos, sólo lo podemos estimar.

ÁZAR Y PROBABILIDAD

¿Por qué si tomamos la misma cantidad de pacientes en varias oportunidades nunca obtenemos el mismo valor, por ejemplo de mortalidad?

Más allá de todos los factores que puedan afectar la evolución de un infarto, sin ninguna duda y en última instancia, es el azar el que determina estas variaciones.

Pensemos por ejemplo en una maternidad y en qué porcentaje de varones nacen cada día durante un mes. Seguramente algunos días no nacerá ningún varón y otro día serán todos varones, lo cual intuitivamente sabemos ocurrirá muy pocas veces. Con mayor frecuencia observaremos que el porcentaje de varones está entre un 40 y un 60%. Esto se debe a que la *verdadera* proporción de varones y mujeres es del 50% y que sólo existen estas dos posibilidades.

Estas variaciones son debidas al azar, pero debemos estar seguros. La estadística nos ayuda a determinar cuál es la variación que puede presentar un valor debido solo al azar o, dicho de otro modo, cuál es el rango de valores que probablemente obtendremos por azar al medir la ocurrencia de un evento.

Si los eventos no variaran al azar serían enteramente predecibles y no necesitaríamos de la estadística.

A través del cálculo de probabilidades es posible determinar cuál es la probabilidad que tiene un evento de ocurrir bajo determinadas condiciones y cuál es la variación que puede presentar ese mismo valor debido al azar.

Por ejemplo, la probabilidad de ser varón o mujer es del 50%, debido a que sólo existen dos resultados posibles o condiciones.

La posibilidad de obtener el número 1 cuando tiramos un dado es 1 en 6, porque sólo existen 6 posibilidades o condiciones.

Probabilidad es simplemente una forma de describir cuán posible es que un evento ocurra dadas ciertas condiciones y por azar. La probabilidad se expresa con números decimales. La probabilidad de obtener un número dado por azar tirando un dado se expresa como 0,166 (1 dividido 6).

Su interpretación es fácil: si un evento tiene una probabilidad muy baja de ocurrencia, por ejemplo 0,001 (probabilidad 1 en 1 000 de que ocurra), se concluye que es muy improbable que ocurra sólo por azar, a diferencia de un evento con probabilidad 0,9 (probabilidad 9 en 10) el cual es altamente probable que ocurra sólo por azar.

Si un evento puede ocurrir o no, entonces la probabilidad toma valores entre 1 y 0: 1 significa que un evento ocurrirá y 0 significa que el evento nunca ocurrirá.

Entre la posibilidad absoluta de que un evento ocurra o no hay diferentes posibilidades intermedias: un evento puede ser más o menos probable. Por eso los valores intermedios entre 1 y 0 nos transmiten esta probabilidad. Por ejemplo, la probabilidad de que una persona muera es 1, ya que todos en algún momento moriremos, pero la probabilidad de que muera mañana podría ser 1 en 100 000, o 0,000001. Si bien es cercana a cero y muy improbable, puede existir la posibilidad de que sufra un accidente de tránsito.

El cálculo de probabilidad es el eje de las pruebas estadísticas.

El término p o valor de p se refiere a probabilidad y es lo que siempre se ve expresado como $p = x$ (por ejemplo, $p = 0,003$, que indica que un evento tiene una probabilidad o chance de ocurrir por azar de 3 en 1 000 bajo ciertas condiciones).

LÓGICA DE LAS PRUEBAS ESTADÍSTICAS

Las pruebas estadísticas asumen entonces que todos los eventos (en principio) ocurren por azar y calculan cuál es la probabilidad de que un determinado evento se haya producido por azar para confirmar esta asunción.

A las asunciones en que se basa la Estadística se las llama supuestos.

Entonces, *el supuesto de la estadística es que, hasta que se pruebe lo contrario, todos los eventos ocurren por azar, dadas ciertas condiciones*, y el deber de las pruebas estadísticas es probarlo a través del cálculo de la probabilidad de ocurrencia de un evento o valor p .

Supongamos que en un estudio que compara una nueva droga se observa que la tasa de mortalidad es del 5% y en el grupo tratado con droga estándar (grupo control) es del 8%. Si las dos muestras tienen la misma cantidad de pacientes con las mismas características y fueron seleccionados al azar, ¿cuál es la probabilidad de haber obtenido 5% de mortalidad con la nueva droga sólo por azar, asumiendo que ambos valores no son diferentes desde el punto de vista estadístico?

Las pruebas estadísticas asumen en principio que estos valores no son diferentes. Sólo por azar en un grupo se obtuvo 5 y en el otro 8%: a esto se lo denomina *hipótesis nula*.

Cuando en términos estadísticos decimos que no hay diferencias estadísticamente significativas, queremos decir que el haber obtenido un 5% de mortalidad fue una de las tantas variaciones posibles de encontrar este resultado sólo por azar cuando la hipótesis nula es cierta.

¿Qué significa cuando la hipótesis nula es cierta? Quiere decir que asumimos como verdadero que no hay diferencias entre estos dos porcentajes.

Cuando asumimos como hipótesis nula que 5% y 8% son iguales, estamos haciendo justamente eso: “asumir como cierto que no hay diferencia”. Debido a que no conocemos el verdadero valor de mortalidad, algo hay que fijar como cierto: ambos porcentajes son iguales.

Es muy común que se malinterprete el valor de p cuando su resultado nos dice que dos valores no son estadísticamente diferentes. En general se lo entiende como la probabilidad de obtener un valor sólo por azar y que el efecto observado no es real: 5% de mortalidad no es real, sólo fue por azar. Esto no es así. Sólo podemos decir que no hemos podido probar esta vez que fuera un efecto real.

Es como el principio de inocencia: toda persona es inocente hasta que se demuestre lo contrario. Pues bien, esto es lo que hacen las pruebas estadísticas: tratar de encontrar la evidencia para demostrar *culpabilidad*.

La *culpabilidad* en términos estadísticos es lo opuesto a la hipótesis nula, o sea dos valores son *verdaderamente distintos*: no fue el azar el responsable de esta diferencia. A esta asunción se la denomina *hipótesis alternativa*.

Es como si lleváramos a cabo un juicio donde el abogado defensor dice: “mi cliente es inocente” (las diferencias observadas son al azar, hipótesis nula) vs. el fiscal que dice “el acusado es culpable” (las diferencias observadas no fueron al azar, algo las provocó, hipótesis alternativa). Entonces el juez pide pruebas para demostrar la culpabilidad o, lo que es lo mismo, para demostrar que la hipótesis alternativa es la correcta.

La forma de demostrar una u otra hipótesis es aportando *evidencia o pruebas*. La *evidencia* de la que dispone la estadística es el cálculo de probabilidades.

En Estadística, cuando el valor p es grande (alta probabilidad de que un evento ocurra por azar) no decimos que no hay diferencias estadísticas; decimos que no tenemos evidencia para demostrar lo contrario y, hasta que no la tengamos, debemos asumir que ambos valores fueron obtenidos por azar. En términos jurídicos, no hemos podido probar culpabilidad; por lo tanto, aunque la persona sea culpable, debemos asumir, hasta encontrar más pruebas, que es inocente.

Si vamos a decidir que los eventos no son por azar en base a un valor de p debemos, por convención, fijar un valor a partir del cual estemos todos de acuerdo en que tenemos la *evidencia suficiente* para probarlo. Este valor es el que se reporta en los estudios como $p < 0,05$, que significa que, cuando la probabilidad es de 5 en 100 (5%) o menor, asumimos que es suficiente evidencia para rechazar la hipótesis nula, y aceptar que el evento no ocurrió por azar.

De lo que se trata es de probar o no que la hipótesis nula es cierta.

Por lo tanto, cuando p es, por ejemplo, 0,005 decimos que la probabilidad de obtener 5% de mortalidad (siendo la *verdad* que 5 y 8% son iguales) es de 5 en 1 000, muy poco probable, y la conclusión que sacamos es que la hipótesis nula no es cierta (los porcentajes no son iguales) y, traducido a un estudio clínico, concluimos que un tratamiento fue mejor que otro.

Si el valor p fuera de 0,09 diríamos que la probabilidad de obtener 5% de mortalidad bajo el supuesto de no diferencia es de 9 en 100, bastante probable. Entonces diríamos que no podemos rechazar la hipótesis nula. No rechazar no es lo mismo que decir que la hipótesis nula sea cierta, sino asumir que *todavía* no hay evidencia para aseverarlo y que, por lo tanto, hay que aceptar que el supuesto desde donde partimos sigue siendo cierto.

Traducido al estudio clínico, concluiríamos que ambos tratamientos son iguales, porque, si partimos del supuesto de no diferencia, no tenemos forma de probar lo contrario. Esta vez no pudimos probar que una droga fue mejor que otra.

CONCEPTO DE INTERVALO DE CONFIANZA

El cálculo del intervalo de confianza es una forma alternativa de evaluar el rol del azar y es más informativo que el solo cálculo del valor p , ya que aporta el grado de certidumbre con el cual hemos estimado un valor.

Volviendo al ejemplo anterior, en el que la nueva droga tuvo una mortalidad de 5% mientras que la droga estándar tuvo una mortalidad de 8%, lo que el intervalo de confianza nos informa son los valores que probablemente obtendremos por variaciones al azar, si realizamos el experimento varias veces.

En el caso de usar la nueva droga, por ejemplo, en 100 grupos de 1 000 pacientes, cada uno con las mismas características, veremos que no siempre obtendremos los mismos valores de mortalidad. Obtendremos un rango de valores, por ejemplo entre 2 y 7%.

La Estadística es capaz de *estimar* o inferir cuáles serían esos valores a partir de una sola muestra. Si obtuvimos una mortalidad del 5%, el intervalo de confianza nos informa entre qué valores pudo haber variado este 5% sólo por azar (en este caso entre un 2 a 7%). Este rango de 2 a 7% también nos habla del grado de certeza que tuvimos al estimar el valor de mortalidad o cuán cerca estamos del supuesto verdadero valor. Si hubiéramos obtenido un intervalo de confianza entre 2 y 28%, el verdadero valor estaría entre 2 y 28. Una mortalidad de 28% es muy diferente a una del 2%, por lo que este intervalo nos dice que la estimación que hicimos no es muy precisa.

Expresado de otra manera, el intervalo de confianza nos dice que de cada 100 veces que realicemos un experimento, en 95 oportunidades obtendremos valores entre x y x , por

ejemplo entre 2 y 7%, y que dentro de este rango está el *supuesto verdadero valor* que queremos estimar.

A este rango de posibles valores es al que llamados *intervalo de confianza del 95%*. En general se lo reporta como IC 95%.

Para el ejemplo anterior, si la mortalidad de la nueva droga fue de 5% en nuestro experimento y sabemos, por el cálculo del intervalo de confianza, que por azar es posible que en el 95% de los casos obtengamos valores entre 2 y 7%, entonces el verdadero valor de la mortalidad de este grupo estará entre 2 y 7%.

Entonces 8%, la mortalidad del grupo control, es un valor que no esperamos encontrar por azar.

Si la hipótesis nula es cierta, no diferencia entre valores, ambos porcentajes deberían estar dentro del IC 95%.

Si alguno de estos valores no está dentro del IC 95%, entonces queda dentro del 5% de probabilidad que utilizamos para decir que un evento no fue al azar. Lo que es igual a obtener un valor de $p < 0,05$.

Como sabemos que el verdadero valor de la mortalidad de la nueva droga está entre 2 y 7%, obviamente 8% no es un valor posible de hallar por azar.

Veamos otro ejemplo. Supongamos que la droga A disminuye la tensión arterial en 15 mmHg y la droga B en 5 mmHg. La diferencia entre ambas es de 10 mmHg, ahora sabemos que estos valores pueden variar sólo por azar, por lo que esta diferencia podría incluir valores calculando el IC 95% de entre 3 y 12 mmHg, con lo cual podemos concluir que, aun interviniendo el azar en el 95% de los casos, la tensión arterial será diferente entre ambas drogas, como mínimo 3 mmHg y como máximo 12 mmHg. Podemos estar seguros de que en 95 veces de 100 no vamos a encontrar que ambas drogas reduzcan la tensión arterial con la misma magnitud. Siempre la droga A reduce más la tensión arterial que la droga B aunque a veces esta reducción podría ser mínima (3 mmHg) y otras veces máxima (12 mmHg). El verdadero valor de la reducción de la tensión arterial por la droga A está entre 3 y 12.

Si el IC 95% contiene al cero, por ejemplo IC 95% -4 a 20 mmHg, en algún experimento que volvamos a realizar con estas dos drogas ambas reducirán la tensión arterial de igual forma (la diferencia sería cero) o, lo que es lo mismo, de 100 veces que realicemos el experimento puede ocurrir que, en algunos casos, no haya diferencias en la reducción de la tensión arterial entre ambas drogas. Entonces decimos que, por azar, la diferencia de la tensión arterial puede variar entre tres posibilidades: que la droga B disminuya más la tensión arterial que la droga A (-4 mmHg), que ambas drogas sean iguales (diferencia cero) o que la droga A disminuya más la tensión arterial que la droga B (20 mmHg).

La conclusión obvia con este intervalo de confianza es que la droga A no siempre muestra ser superior a la droga B, por lo que no hay evidencia para rechazar la hipótesis nula. En este caso ambas drogas reducen de igual forma la tensión arterial y la diferencia observada se debió sólo al azar. El verdadero valor de la reducción de la tensión arterial es algún valor entre -4 y 20 mmHg

En resumen, la Estadística basa sus mediciones en la *estimación*, lo más cercana posible al verdadero valor, a partir de una muestra representativa de sujetos.

Estas estimaciones son sólo una aproximación al verdadero valor, debido a que por azar los valores cambian de un experimento a otro.

El grado de certeza con respecto a cuán cerca o lejos estamos del verdadero valor nos lo da el IC 95% y la evidencia para afirmar que los valores observados no se deben a variaciones por azar cuando asumimos que son iguales, la obtenemos con el cálculo de p .

DATOS CUANTITATIVOS. MEDIDAS DE POSICIÓN Y DISPERSIÓN



En los estudios descriptivos sólo nos remitimos a reportar lo observado, y en los estudios analíticos comparamos los resultados entre dos o más grupos de sujetos.

En ambos casos, lo primero que necesitamos hacer es *resumir* la información obtenida.

En el caso de variables cuantitativas, si tenemos 100 valores de creatinina, no podemos mostrar una tabla con las 100 observaciones; esto no sería de mucha ayuda a la hora de reportar la información. Lo que necesitamos es un valor único que represente a estas 100 observaciones de creatinina.

Además, como ya sabemos, estos valores varían entre los sujetos, ya sea debido al azar o por patologías que puedan padecer los sujetos: los valores de creatinina de pacientes con insuficiencia renal serán superiores a los de sujetos sin esta patología. Y aun dentro de cada grupo de pacientes, los valores de creatinina también serán diferentes entre los sujetos.

Podemos entonces dividir la variabilidad de los datos en de causas conocidas o de causa no explicable. Esta última es la que, decimos, se debe al azar y se denomina variabilidad aleatoria.

Consecuentemente, lo que vamos a reportar son dos tipos de valores: un único valor que resuma a todos los valores y otro que nos muestre el grado de variabilidad de las observaciones que presenta la muestra.

A las estimaciones que resumen todas las observaciones se las denomina *medidas de posición* y a aquellas que resumen la variabilidad se las llama *medidas de dispersión*.

MEDIDAS DE POSICIÓN

El primer paso para describir un conjunto de observaciones numéricas es calcular el promedio. En Estadística el promedio se conoce como *media* y su notación matemática es:

$$\bar{X} \text{ o } \mu$$

\bar{X} se usa para denominar la media de una muestra y μ para referirse a la media poblacional (aquella que queremos estimar).

Su cálculo es muy sencillo, es la suma de todos los valores dividida por el número de sujetos de la muestra.

Otra medida que resume los valores es la *mediana*. Esta se define como el valor que divide a la muestra en dos partes iguales o 50% cuando los datos están ordenados.

En la Tabla 10.1 se observan valores de edad y superficie corporal de 40 pacientes.

La media de la edad es igual a $2422/40 = 60,1$ años

La mediana es el valor que ocupa la posición que divide a la muestra en dos partes iguales. En este caso, como tenemos 40 observaciones, la posición de la mediana está entre el sujeto 20 y el 21. Lo que hacemos aquí es tomar los valores de la posición 20 y 21 y promediarlos: $60+61/2 = 60,5$ años.

SUJETO	EDAD	SUP. CORP.
1	20	1,32
2	43	1,40
3	43	1,70
4	48	1,70
5	50	1,72
6	51	1,74
7	53	1,75
8	55	1,75
9	56	1,76
10	56	1,77
11	56	1,81
12	56	1,81
13	57	1,84
14	57	1,84
15	58	1,89
16	60	1,91
17	60	1,91
18	60	1,94
19	60	1,95
20	60	1,97
21	61	1,98
22	61	1,99
23	61	2,00
24	62	2,01
25	62	2,02
26	63	2,02
27	64	2,03
28	65	2,05
29	66	2,08
30	66	2,10
31	68	2,10
32	68	2,12
33	69	2,13
34	71	2,14
35	71	2,14
36	71	2,16
37	73	2,16
38	73	2,19
39	75	2,19
40	76	2,28

Tabla 10.1: Datos de edad y superficie corporal

Si tuviéramos 39 valores, la mediana ocuparía la posición 20 y sería 60 años.

La superficie corporal para la muestra de 40 sujetos tiene una media de $77,4/40 = 1,93$ m² y una mediana de $1,97+1,98 = 1,975$ m².

Supongamos que la edad en la posición de los pacientes 39 y 40 ha cambiado a 90 años; la media es ahora 61,2 años y la mediana sigue siendo la misma, 60,5 años. Si los valores en la posición 39 y 40 de superficie corporal cambian a 2,40, la media es 1,94 m². Y la mediana continúa siendo la misma.

Como vemos, la media es muy sensible a valores extremos, por lo que en muestras cuya variabilidad esté desviada hacia algún lado de la media (muestras asimétricas o no normales) esta medida de posición no refleja el valor central de la muestra sino un valor que está desviado hacia el lado donde hay valores más extremos.

Por lo tanto, dependiendo de la variabilidad que presente la muestra, puede ser más apropiado presentar la mediana, que no se ve influenciada por valores extremos.

El último indicador estadístico de centralidad es el *modo*, que es el valor más frecuentemente observado. El modo de edad para la muestra de 40 pacientes es 60 años: se lo ha observado en 5 sujetos. Ningún otro valor ha sido observado con más frecuencia en esta muestra.

El modo rara vez se reporta como medida de posición central de una muestra.

MEDIDAS DE DISPERSIÓN

El segundo aspecto en la descripción de un conjunto de variables numéricas es evaluar la variabilidad de las observaciones.

Cualquier conjunto de datos contendrá datos de diferente magnitud, como los que vimos en la Tabla 10.1. El valor absoluto de los datos determina la variabilidad de las observaciones. Si el menor valor de glucemia hallado en una población es de 60 mg% y el mayor es de 300 mg%, entre estos valores se encuentran muchos otros posibles, hay mucha variabilidad. Si, por el contrario, estos valores están entre 70 y 120 mg%, hay menos valores posibles y por lo tanto menor variabilidad.

En lo que estamos interesados ahora es en determinar la variabilidad de los datos y cómo están dispersos. ¿Son similares entre sí o varían mucho? ¿Hay mayor concentración de ciertos valores a un lado u otro de la media o de la mediana?

Hay dos formas de evaluar la variabilidad: por métodos gráficos y por métodos numéricos.

MÉTODOS GRÁFICOS PARA VALORAR LA VARIABILIDAD

Una forma simple de observar la variabilidad de los datos es construyendo un histograma.

En este tipo de gráfico, el número de observaciones o el porcentaje de sujetos queda graficado en función de los diferentes valores o grupos de valores.

En la Tabla 10.2 se observan los datos de edad de 3 800 sujetos. Podemos observar que las edades varían entre 17 y 90 años, por el número de sujetos (frecuencia) o la frecuencia relativa, vemos que la mayoría de los pacientes están entre 56 y 76 años. Si bien por esta tabla podemos inferir que los datos tienen una amplia dispersión, que la mayoría de los sujetos tienen entre 56 y 76 años y que por debajo o por encima de estos valores hay pocos sujetos, no es una manera muy amigable ni resumida de verlo.

El histograma es una forma mucho más resumida y fácil de interpretar.

Para construir un histograma, lo más adecuado es agrupar los datos en categorías iguales; por ejemplo creando grupos de 5 años de edad y calculando cuántos sujetos están dentro de esa categoría.

EDAD	FRECUENCIA	FRECUENCIA RELATIVA %
17	1	0,026316
18	3	0,078947
19	1	0,026316
20	3	0,078947
22	2	0,052632
23	6	0,157895
24	2	0,052632
25	2	0,052632
26	1	0,026316
27	1	0,026316
28	2	0,052632
29	4	0,105263
30	8	0,210526
31	2	0,052632
32	5	0,131579
33	8	0,210526
34	4	0,105263
35	7	0,184211
36	5	0,131579
37	9	0,236842
38	8	0,210526
39	18	0,473684
40	13	0,342105
41	13	0,342105
42	28	0,736842
43	30	0,789474
44	36	0,947368
45	31	0,815789
46	28	0,736842
47	39	1,026316
48	56	1,473684
49	63	1,657895
50	62	1,631579
51	71	1,868421
52	64	1,684211
53	92	2,421053
54	87	2,289474
55	86	2,263158
56	116	3,052632
57	104	2,736842
58	114	3
59	104	2,736842

EDAD	FRECUENCIA	FRECUENCIA RELATIVA %
60	113	2,973684
61	115	3,026316
62	135	3,552632
63	117	3,078947
64	140	3,684211
65	132	3,473684
66	116	3,052632
67	122	3,210526
68	145	3,815789
69	126	3,315789
70	140	3,684211
71	133	3,5
72	135	3,552632
73	103	2,710526
74	126	3,315789
75	117	3,078947
76	110	2,894737
77	63	1,657895
78	74	1,947368
79	60	1,578947
80	43	1,131579
81	35	0,921053
82	21	0,552632
83	10	0,263158
84	10	0,263158
85	6	0,157895
86	9	0,236842
87	2	0,052632
88	2	0,052632
90	1	0,026316
TOTAL	3800	100

Tabla 10.2: Edad de 3 800 sujetos

En la Figura 10.1 se observa el histograma correspondiente a la Tabla 10.2.

Como se puede observar, se han creado grupos de 5 años, la altura de cada barra representa el número de sujetos que cada intervalo de edad contiene. El número que se observa en el eje x es el punto medio de edad de cada intervalo. Así, el intervalo 40 está construido con valores que van desde 37,5 a 42,5, siendo el punto medio de este intervalo los 40 años.

La elección de cómo construir cada intervalo es decisión del investigador, pero en general se recomienda no tener más de 15 grupos.

Analizando el histograma, vemos fácilmente que la mayor concentración de datos está en los grupos 55 a 75 años, que el intervalo con mayor número de sujetos es el de 70 años (67,5 a 72,5 años) con 680 pacientes y que el número de sujetos a la derecha del intervalo

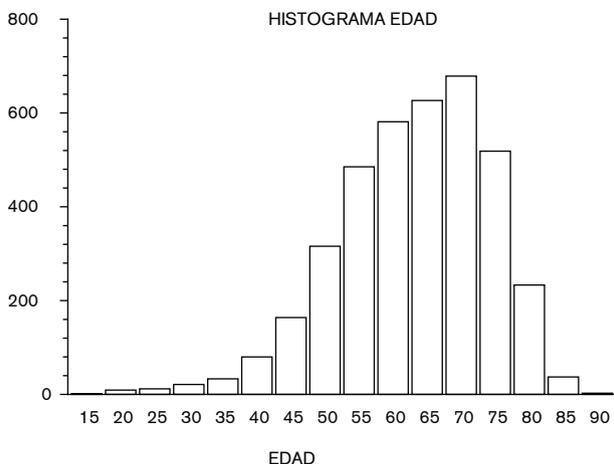


Figura 10.1: Histograma. Edad de 3 800 sujetos

70 es mayor que el número de sujetos a la izquierda. Se dice entonces que los datos tienen distribución asimétrica.

Observemos ahora el histograma construido con los datos de superficie corporal de 3 800 pacientes. (Figura 10.2)

Aquí los intervalos fueron construidos con categorías que van de 0,10 m², adecuándose a los valores de superficie corporal.

Lo que podemos ver claramente es que la mayoría de los pacientes (aproximadamente 840 sujetos) tuvieron una superficie corporal de entre 1,85 y 1,95 m².

Además, podemos distinguir que a la derecha y a la izquierda, si bien hay menos pacientes, la cantidad a uno y otro lado es similar.

Estos datos tienen distribución simétrica.

Por lo tanto, a través del histograma, podemos obtener información acerca de la dispersión de los datos (valor mínimo y máximo), sobre qué grupos de valores son más frecuentes y cómo se distribuyen alrededor de la media muestral (simétricos o asimétricos).

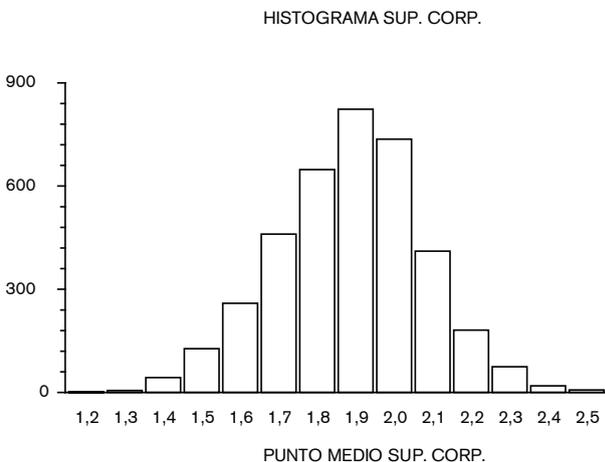


Figura 10.2: Histograma de superficie corporal de 3 800 sujetos.

MÉTODOS NUMÉRICOS PARA EVALUAR LA VARIABILIDAD

Además de los métodos gráficos, es necesario contar con evaluaciones numéricas que resuman la variabilidad de los datos.

Estas, junto con las medidas de posición, proveen un resumen muy informativo de un grupo de observaciones.

Existen tres formas numéricas de cuantificar variabilidad: rango, centilos y desvío estándar.

RANGOS

Es la forma más simple de describir la dispersión de los datos. Lo que se reporta son los valores mínimo y máximo de una muestra de datos cuantitativos.

Para el ejemplo anterior de edad, el rango de la muestra de 3 800 sujetos es 17 a 90 años y para la superficie corporal de 1,23 a 2,53 m².

La desventaja del rango es que sólo tiene en cuenta los valores extremos, que probablemente sean los más atípicos o menos frecuentes, sin describir cómo se distribuyen los valores intermedios. Por esta razón, los rangos no son muy utilizados como medidas de resumen de la variabilidad de los datos.

PERCENTILOS

Si tomamos una muestra y calculamos progresivamente cuántos pacientes hay desde el valor mínimo hasta un valor determinado y los dividimos por el número de pacientes totales obtenemos lo que se llama frecuencia relativa. En la Tabla 10.3 se muestran los valores de edad de 40 pacientes con sus frecuencias absolutas, relativas, acumulativas absolutas (suma del número de sujetos) y acumulativas relativas (suma de los porcentajes relativos).

Con 50 años de edad hay un único paciente, que por sí solo representa el 2,5% de la muestra (1/40).

Entre 20 y 50 años hay 5 pacientes (frecuencia acumulativa absoluta) los cuales representan el 12,5% de la muestra total (frecuencia acumulativa relativa).

A través de la frecuencia acumulativa relativa podemos saber qué porcentaje de la muestra está por debajo de este valor y a esta frecuencia acumulativa relativa la llamamos *percentilo*.

El percentilo 50 de la Tabla 10.3 corresponde a 60 años de edad y, como puede verse, contiene 20 pacientes, que representan el 50% de la muestra. A este percentilo también lo llamamos mediana: la mitad de las observaciones son menores o mayores a este valor.

Los percentilos más frecuentemente reportados son los 25 y 75, llamados cuartiles, y nos permiten dividir a la muestra en 4 partes iguales, cada una de las cuales contiene un 25% de la muestra total. La diferencia numérica entre el cuartilo 25 y el 75 es el rango intercuartilo. En este caso la distancia intercuartilos 25-75 sería de 55,5 a 66 años.

El cuartilo 25 no se observa en la tabla, estaría entre el percentilo 20 y 30, cuyos valores corresponden a 55 y 56 años, por lo tanto, el valor correspondiente al percentilo 25 es el promedio entre los valores de los percentilos 20 y 30.

En la Figura 10.3 se muestra un gráfico llamado *box plot*, en el que se puede observar cómo está distribuida la muestra de 40 pacientes de acuerdo con los cuartiles 25, 50 y 75.

La línea vertical contenida dentro del rectángulo es la mediana o percentilo 50. Los lados izquierdo y derecho del rectángulo son los percentilos 25 y 75 respectivamente.

Los valores extremos o rangos están representados por los corchetes unidos al rectángulo.

Edad	Frecuencia absoluta	Frecuencia relativa %	Frecuencia acumulativa absoluta	Frecuencia acumulativa relativa %
20	1	2,5	1	2,5
43	2	5	3	7,5
48	1	2,5	4	10
50	1	2,5	5	12,5
51	1	2,5	6	15
53	1	2,5	7	17,5
55	1	2,5	8	20
56	4	10	12	30
57	2	5	14	35
58	1	2,5	15	37,5
60	5	12,5	20	50
61	3	7,5	23	57,5
62	2	5	25	62,5
63	1	2,5	26	65
64	1	2,5	27	67,5
65	1	2,5	28	70
66	2	5	30	75
68	2	5	32	80
69	1	2,5	33	82,5
71	3	7,5	36	90
73	2	5	38	95
90	2	5	40	100

Tabla 10.3: Edad de 40 pacientes

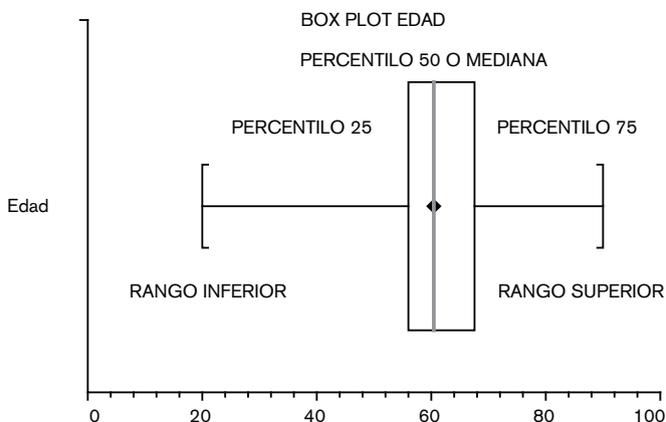


Figura 10.3: Box plot edad

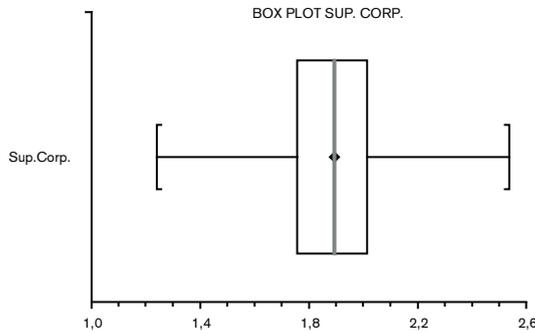


Figura 10.4: *Blox plot* superficie corporal

¿Cómo interpretamos este gráfico?

- El valor que corresponde a la mediana o percentilo 50 es 60 años. Por ser este el percentilo 50, ya sabemos que a un lado y otro del mismo hay igual cantidad de sujetos (50% a cada lado).
- Entre el percentilo 25 y 50 los sujetos tienen una edad entre 54 y 60 años, por lo tanto su variabilidad es escasa.
- Por debajo del percentilo 25 los sujetos tienen entre 17 (valor mínimo) y 54 años, mostrando una amplia variabilidad.
- Entre la mediana y el percentilo 75 los sujetos tienen edades de entre 60 y 68 años, y más allá del percentilo 75, entre 68 y el rango máximo de 90 años.
- En conclusión, vemos que los cuartos cuartiles tienen distancias diferentes y por lo tanto la muestra es asimétrica.

De esta manera tenemos una mejor idea de cómo se distribuyen los valores en la población.

En la Figura 10.3 vemos que la distribución de los valores alrededor de la mediana es asimétrica: los valores están más concentrados a la derecha de la mediana.

Comparemos el *box plot* de edad con el de superficie corporal. (Figura 10.4) ¿Qué diferencias encontramos?

Al observar este gráfico, rápidamente nos damos cuenta de que la distribución de los datos es mucho más simétrica. Las dos mitades del rectángulo central son casi del mismo tamaño, al igual que las líneas que llegan hasta los rangos.

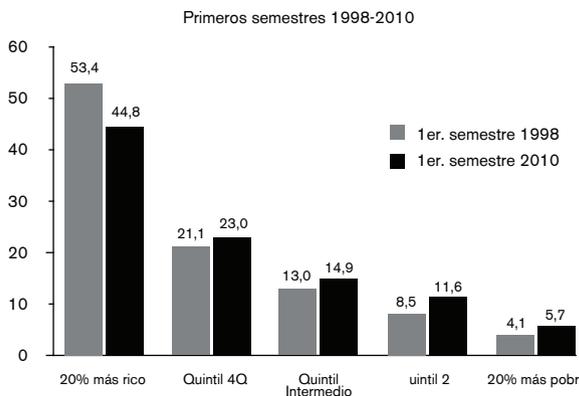


Figura 10.5: Ingreso en hogares por quintiles

Otro percentilo muy comúnmente reportado en los estudios científicos son los quintilos, que no es nada más que dividir la muestra en 5 grupos que contienen el 20% de la población cada uno.

Por ejemplo, podemos reportar el ingreso anual de una población dividido en 5 grupos, de acuerdo con los valores de sus quintilos. (Figura 10.5)

El mejor ejemplo de la utilización de percentilos en medicina es el control del peso y la altura en niños. (Figura 10.6)

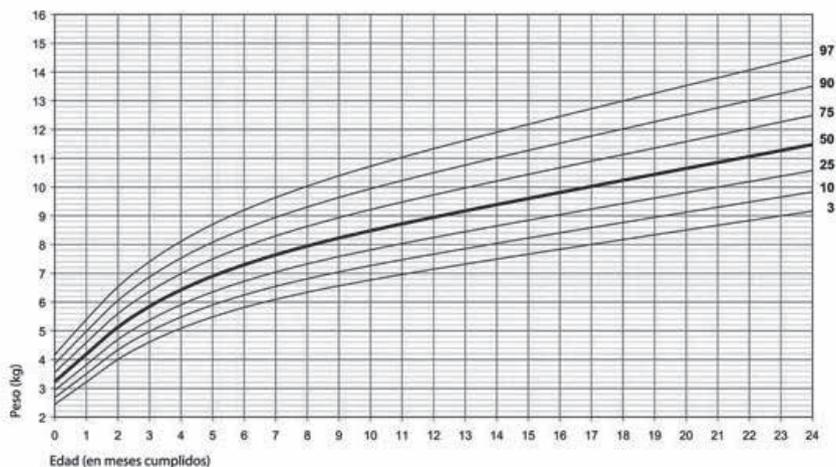


Figura 10.6: Ejemplo de cálculo de percentilos de peso en niños

DESVÍO ESTÁNDAR

Esta forma de resumir la variabilidad de un conjunto de datos cuantitativos es la más frecuentemente reportada en la literatura y a su vez la que más dificultades de comprensión presenta.

La unidad de distancia que usamos en estadística se denomina desvío estándar. En lugar de medir la distancia en centímetros, metros o kilómetros, ahora hablamos de *desvío estándar (DS)*.

El desvío estándar es la distancia desde un valor x hasta su media, en una muestra de datos cuantitativos.

De la misma forma que al calcular la media de una muestra *estimamos* la media poblacional, cuando calculamos el DS también estamos estimando la variabilidad de los datos de una muestra.

Para un grupo de datos es posible calcular la distancia entre cada valor de la muestra y su media; al sumar cada distancia obtenemos una distancia promedio que representa el valor de un desvío estándar para esa muestra.

El valor absoluto que adquiere 1 DS es propio de cada muestra y depende de varios factores que veremos luego.

Cuando hablamos de 1 DS nos referimos a una distancia *fija* entre la media y el DS, que es común a cualquier muestra de datos cuantitativos, siempre y cuando la muestra sea simétrica.

De esta forma, la nueva medida de distancia nos independiza de los valores particulares de cada muestra y establece un lenguaje común para todos. Cuando decimos tal valor está a 1 DS, todos entendemos a qué distancia está. Por ahora la equivalencia a otras unidades del DS la dejaremos en suspenso.

CÁLCULO DE VARIANZA Y DESVÍO ESTÁNDAR

Si tomamos los valores de la Tabla 10.2, cuya media es 62,98 años, podríamos calcular para cada uno de los valores su distancia con respecto a la media. Por ejemplo para el 1º valor = 17.

$$\text{Distancia} = 17 - 62,98 = - 45,98 \text{ años}$$

Cada vez que calculemos la distancia de un valor menor a la media obtendremos valores negativos. Para evitar esto, la matemática siempre tiene una solución, que en este caso es elevar la resta al cuadrado. Por lo tanto ahora la diferencia es:

$$(17 - 62,98)^2 = 2\ 114 \text{ años}^2$$

Este valor es ahora positivo, pero está en unidades que no son las originales, ¡años al cuadrado! y además es extremadamente grande.

Por el momento dejémoslo así.

Ahora bien, dijimos que sumando todas las distancias de cada valor (ya sabemos por qué están elevadas al cuadrado) podemos obtener el *promedio* de las distancias. Si queremos obtener un promedio, lo que debemos hacer entonces es dividir la suma de las distancias por el número de sujetos de la muestra. Así, en notación matemática tendremos:

$$\text{Promedio de distancia} = DS = \sum \frac{(x - \bar{X})^2}{n - 1}$$

Donde:

\sum = sumatoria de las diferencias

x = cada valor individual de la muestra

\bar{X} = media

$n - 1$ = número total de la muestra - 1

Aplicando esta fórmula, lo que se obtiene es un valor de la distancia promedio elevada al cuadrado conocida como *varianza muestral*. Su resultado se expresa en unidades al cuadrado (en este caso años al cuadrado), lo que no tiene mucho sentido; por lo tanto, para volver a los valores a las unidades originales de la muestra y obtener valores más adecuados en cuanto a magnitud, lo que se hace es calcular la raíz cuadrada de la varianza muestral (notación matemática S^2 o σ^2) que no es otra cosa que el desvío estándar (notación matemática: DS o S o σ).

La notación S^2 se utiliza para hacer referencia a la varianza de la muestra y σ^2 para hacer referencia a la varianza poblacional.

Lo mismo para el desvío estándar, S y σ se refieren al DS muestral y poblacional respectivamente.

$$\text{Siendo } S^2 \text{ o varianza} = \sum \frac{(x - \bar{X})^2}{n - 1}$$

$$\text{El desvío estándar o } S \text{ es } = \sqrt{\sum \frac{(x - \bar{X})^2}{n-1}}$$

Tenemos entonces un valor con sus unidades originales que equivale a 1 DS. Aplicando esta fórmula a los datos de la Tabla 10.2 obtenemos:

Una varianza (S^2) = 122 años² y un DS (S) = $\sqrt{122} = 11,05$ años

Para esta muestra el valor de 1 DS es igual a 11,05 años.

Para medir la distancia de cualquier valor de la muestra y su media, usamos la unidad de distancia DS. Decimos entonces que un valor x está, por ejemplo, a 1,5 DS de la media.

Por suerte, en la actualidad contamos con programas que hacen cálculos estadísticos. Estos programas calculan el valor que equivale a un desvío estándar para la muestra con la que estamos trabajando.

Si sabemos a cuántas unidades equivale 1 DS para una muestra determinada, podemos calcular la distancia en DS a que se encuentra cualquier valor de su media.

Siguiendo con el mismo ejemplo de la Tabla 10.2 cuya media es 62,98 años, decimos que hay un valor a 1,5 DS ¿cuál es ese valor?

Si 1 DS es igual a 11,05 años:

1,5 DS es igual a 11,05 (valor de 1 DS) * 1,5 (nº de DS) = 16,6 años

El valor que está a 1,5 DS = 62,98 (media) + 16,6 (1,5 DS) = 79,5 años

Entonces 79,5 años está a 1,5 DS de la media.

También podemos calcular a la inversa: si un sujeto tiene 73 años y el valor del DS es de 11 años, ¿a cuántos DS se encuentra de la media?

El cálculo es muy fácil: si la distancia es 73 - 62,98 (media muestral) = 10,02 años.

Siendo el valor de 1 DS igual a 11 años y la diferencia de 10,02 años ¿a cuántos DS equivale?

Nº de desvíos estándar es igual a: 10,02/11 = 0,9 DS

Decimos entonces que el valor 73 años se encuentra a 0,9 DS de la media de esa muestra.

La idea de tomar al DS como una unidad de distancia permite adoptar una medida de distancia universal aplicable a cualquier muestra que se desee estudiar. Luego veremos que la unidad de 1 DS es derivada matemáticamente teniendo en cuenta sobre todo el tipo de distribución de datos que presente la muestra (simétrica o asimétrica).

Tomemos ahora la muestra con 3 800 datos de superficie corporal.

La media para esta muestra es de 1,88 m².

Su varianza es de 0,036 m⁴.

Su DS = 0,19 m² (para cada muestra el valor de 1 DS es propio de ella).

Para un valor de 1,4 m² la diferencia o distancia a la media es de:

1,4 - 1,88 = - 0,48

Si -0,48 es la distancia y el valor del DS es de 0,19 ¿a cuántos DS se encuentra este valor?

Nº DS = - 0,48/0,19 = -2,5 DS

Esto significa que este valor, al ser negativo, se encuentra a la izquierda de la media a 2,5 DS.

Si tomamos la media como punto central, podemos contar DS a la derecha y a la izquierda de la media.

La estadística nos dice que entre $\pm 1,96$ DS de la media (1,96 DS a la derecha y 1,96 DS a la izquierda) se encuentra el 95% de los valores esperados de obtener por variación al azar, siempre y cuando la distribución de los datos sea simétrica. Es lo que comúnmente se dice ± 2 DS debido a que 1,96 es casi 2.

Si la muestra no fuera simétrica a un lado y otro de la media habría diferente proporción de valores esperados y no tenemos entre $\pm 1,96$ DS el 95%.

Si observamos la Figura 10.7, vemos que el valor 1,4 m² se encuentra por fuera de este límite porque está a más de 1,96 DS (a la izquierda). Podríamos concluir que, como no es esperable obtenerlo por variación al azar, este valor puede deberse a otras causas.

Aquí no podemos decir que hubo variabilidad aleatoria. ¿Podría ser un niño que fue incluido por error? ¿Podría tener alguna enfermedad que se asocie a talla baja?

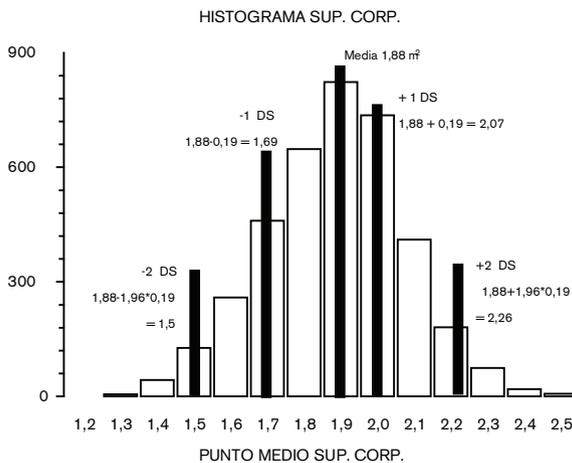


Figura 10.7: Superficie corporal. Desvío estándar

CÁLCULO DEL INTERVALO DE CONFIANZA

Como ya mencionamos antes, al rango de posibles valores esperables de obtener por azar se lo denomina intervalo de confianza del 95% (IC 95%).

Hablamos de rango puesto que el IC 95% nos informa un valor mínimo y un valor máximo dentro de los cuales encontramos el 95% de los valores esperados. Por lo tanto, tenemos lo que se llama límite inferior y límite superior del intervalo de confianza.

Cuando reportamos el IC 95% de nuestra muestra, estamos informando entre qué valores se encuentra el 95% de la población; esto nos da una idea del grado de dispersión que presentan nuestros datos.

Si entre $\pm 1,96$ DS de la media muestral se encuentra el 95% de los valores esperados por azar o intervalo de confianza 95% ¿cuáles serían estos valores?

Para el cálculo del IC 95% sólo necesitamos el valor de la media muestral y el del DS.

El IC 95% es igual a: $\text{media} \pm 1,96 * \text{DS}$

Si la media de superficie corporal es de 1,88 m² y el del DS 0,19 m².

El límite inferior del IC 95% es igual a: $1,88 - (1,96 * 0,19) = 1,88 - 0,38 = 1,5$

El límite superior es igual a: $1,88 + 1,96 * 0,19 = 1,88 + 0,38 = 2,26$

Por lo tanto, el IC 95% de los valores de superficie corporal de esta muestra están entre $1,5 \text{ m}^2$ y $2,26 \text{ m}^2$. (Figura 10.7)

¿Qué podemos concluir de este intervalo?

Pues, como explicamos en la sección azar y probabilidad, la estadística sólo infiere un valor que debe aproximarse al *verdadero valor buscado*. El cálculo del IC 95% nos dice que en el 95% de los casos que volvamos a hacer el experimento encontraremos valores que van de $1,5$ a $2,26 \text{ m}^2$, siempre y cuando usemos poblaciones con las mismas características.

También nos informa acerca de la variabilidad que presentan los valores de la muestra a través de sus límites inferior y superior y el grado de precisión con el que hemos estimado estos valores.

Para construir un IC 95% se utiliza el DS; por lo tanto, para que este intervalo sea válido, la muestra debe ser simétrica.

¿Cuál sería el IC 95% de la superficie corporal para la muestra de 40 pacientes de la Tabla 10.1?

Ya sabemos que para calcularlo necesitamos el valor de la media y del DS.

El valor de la media para esta muestra es de $1,93 \text{ m}^2$ y su DS $0,20 \text{ m}^2$

El límite inferior del IC 95% es: $1,93 - 1,96 * 0,20 = 1,93 - 0,392 = 1,54$

El límite superior del IC 95% es: $1,93 + 1,96 * 0,20 = 1,93 + 0,392 = 2,32$

El IC 95% está entonces entre $1,54$ y $2,32$ (IC 95% $1,54 - 2,32$ así se reporta el límite inferior y superior de un IC 95%)

Podemos entonces resumir los datos de una muestra como:

Media (IC 95%)

Aplicado al ejemplo: $1,93 \text{ m}^2$ (IC 95% $1,54 - 2,32$)

¿Qué diferencia encontramos con el IC 95% de la muestra con 3 800 pacientes?

El IC 95% de la muestra de 3 800 es $1,5 \text{ m}^2 - 2,26 \text{ m}^2$ y el IC 95% de la muestra con 40 pacientes $1,54 \text{ m}^2 - 2,32 \text{ m}^2$; sin dudas este último es más amplio que el de la muestra con 3 800 pacientes.

Si recordamos la fórmula para obtener el desvío estándar, las diferencias entre los valores y la media se dividían por el número de sujetos de la muestra; esto es lo que se llama en matemática una relación inversa.

Lo que quiere decir es que a menor cantidad de pacientes mayor será el valor del DS y por lo tanto más amplio el IC 95%.

El DS de la muestra de 40 pacientes es mayor al de la muestra de 3 800 ($0,20$ vs. $0,19$). Esta diferencia puede ser despreciable, pero a la hora de aplicar alguna fórmula estadística se amplifica, muchas veces de forma exponencial, conduciendo a diferencias muy importantes.

La importancia de esto radica en que a menor número de sujetos, mayor variabilidad de las observaciones y menor probabilidad de estimar con precisión el valor deseado. Este es uno de los factores que más influencia el valor del DS.

Usando el ejemplo anterior, sabemos que los valores de superficie corporal en la muestra de 3 800 pacientes está entre $1,5 \text{ m}^2 - 2,26 \text{ m}^2$, mientras que en la muestra de 40 pacientes estará entre algún valor que va desde $1,54 \text{ m}^2$ a $2,32 \text{ m}^2$. Este último IC 95% es menos preciso. Aquí se observa cómo una diferencia de $0,01$ en los valores de los DS se tradujo en una diferencia de $0,06$ en el límite superior del IC 95%.

Veamos un ejemplo práctico para resumir y entender la aplicabilidad del concepto de DS e IC 95%.

¿Cómo sabemos que los valores normales de glucemia son 70-110 mg%? (por ahora...). ¿Se midió la glucemia a todos los seres humanos del mundo en el mismo día?

La respuesta obvia es que no. Lo que se hizo fue seleccionar una muestra suficientemente grande de sujetos al azar supuestamente *normales* y se les midió la glucemia. Con todos los datos se calculó la media = 90 mg% y DS 10 mg%.

¿Cuál sería el IC 95%?

Como ya sabemos, es igual a: media \pm 1,96 * DS

Para nuestro ejemplo: $90 \pm 1,96 * 10 = 90 \pm 19,6 =$

$= 90 - 19,6 = 70,4 \text{ mg\%}$

$= 90 + 19,6 = 109,6 \text{ mg\%}$

IC 95% para glucemia = 70,4 – 109,6

Estos valores redondeados dan 70 – 110 mg%

Por eso decimos que los valores *normales* de glucemia están entre 70 y 110 mg%.

Es lo mismo que decir que el 95% de la población tendrá valores de glucemia considerados normales si están dentro de este rango.

Resumiendo, para reportar adecuadamente datos cuantitativos debemos informar al menos una medida de posición y una de dispersión.

Cuando los datos tienen distribución simétrica, las medidas que se reportan son la media y el DS, debido a que en este caso todas las medidas de posición coinciden y sólo es posible el cálculo del DS cuando la muestra es simétrica.

En el caso en que la muestra sea asimétrica, la medida de posición que se reporta es la mediana. Este es el valor que mejor representa al conjunto de los datos (recordemos que la media es muy sensible a valores extremos).

La medida de dispersión que se informa en general es el 2° y 3° cuartilo o alguna forma de percentilo, como los quintilos.

DATOS CUANTITATIVOS. DISTRIBUCIÓN NORMAL



Como hemos visto, el histograma nos muestra la *forma* que adquieren los datos debido a su variabilidad, siendo algunas formas simétricas y otras no.

¿Por qué es importante determinar la forma en que se distribuyen los datos?

Es importante porque mediante cálculos matemáticos, dependiendo de la forma que adquieran los datos, se pueden derivar probabilidades teóricas de ocurrencia de un evento.

Es decir que podemos calcular la probabilidad de que un evento ocurra por azar o no. Es el cálculo del valor p , al que nos referimos en capítulos anteriores, necesario para aceptar o no la hipótesis nula, que sostiene que todo ocurre por azar, o que los valores contenidos dentro del IC 95% son sólo variaciones al azar de un mismo valor.

A la forma que adquieren los datos se la conoce como distribución de probabilidades. Se podría decir que la distribución de probabilidades son las diferentes formas en que se distribuyen los datos y algunas de ellas son tomadas como *patrones* matemáticos para derivar probabilidades.

Hay varias formas de distribución, *patrones* o formas teóricas de distribución, la más conocida es la distribución *normal* para datos cuantitativos y la binomial para porcentajes. Otras distribuciones utilizadas en estadística son la de Poisson, gamma, la distribución F, etc., que presentan diferentes formas.

Muchos métodos estadísticos están basados en la asunción de que los datos observados tienen una distribución o forma teórica conocida y, siendo esta una distribución conocida, se pueden derivar probabilidades teóricas.

Todas las distribuciones de probabilidad son descritas o definidas por lo que se llaman parámetros, por ejemplo, la media y el DS son parámetros y definen a la distribución normal. Cuando decimos *definen* queremos decir que, a partir de sólo esos dos datos, podemos dibujar la forma en que se distribuyen los datos de una muestra.

Si nos reportan la media y el DS de dos muestras podemos saber cuál es el valor central de cada una y cómo es la distribución de los datos a los lados de la media y cuáles son los valores que corresponden a 1, 2 y 3 DS.

Como podemos observar en la Figura 11.1, dos muestras pueden tener la misma media pero diferente variabilidad en sus datos; una tiene datos que se alejan más de su media (11-1A); como consecuencia su DS será mayor, al igual que su IC 95%. Esta muestra, comparada con la que tiene menor variabilidad, tiene menos precisión.

En la Figura 11.1B, en las muestras se estimaron diferentes valores de las medias, pero con la misma variabilidad de sus datos (igual distancia desde el extremo de las curvas a la media en cada una de las dos muestras).

Cuando analizamos datos cuantitativos tenemos que determinar si pertenecen a alguna de las distribuciones conocidas. Si por error asumimos una distribución que no es la correcta, todas las pruebas estadísticas que apliquemos y que asuman esa distribución arrojarán resultados falsos.

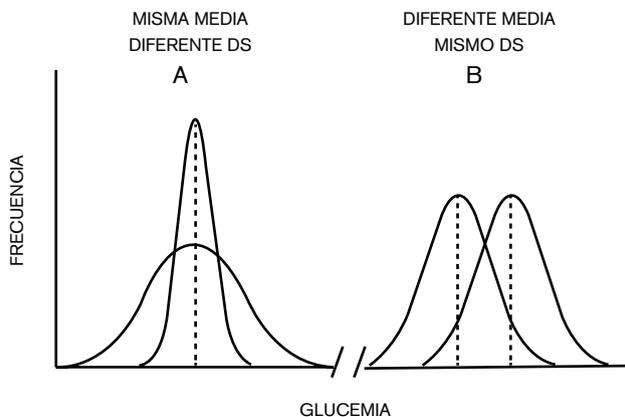


Figura 11.1: Comparación de muestras de acuerdo con su variabilidad

Algunas muestras no tienen una distribución conocida de sus datos y a estas se las denomina libres de distribución.

Otras muestras tienen una distribución simétrica y como vimos pueden definirse por sus parámetros, es por esto que las pruebas estadísticas que se basan en este tipo de distribución se llaman métodos paramétricos y a aquellos que no asumen ninguna distribución o tienen una distribución asimétrica se los denomina métodos no paramétricos.

Veamos ahora la distribución normal. Sin dudas es la más importante en Estadística, debido a que la mayoría de las pruebas estadísticas se basan en ella. También se la llama distribución gaussiana o paramétrica.

Se dice que para que un conjunto de datos tenga una distribución normal debe cumplir con los siguientes requisitos:

1. Su forma debe ser simétrica.
2. Las medidas de posición de la muestra deben coincidir (media, mediana y modo).
3. El 68% de la muestra debe estar contenido entre ± 1 DS, el 95% de los casos de la muestra debe estar contenido entre ± 2 DS y el 99% de la muestra entre ± 3 DS.

Si tomamos el área formada por la curva tenemos el 100% de la muestra.

Si queremos dividir la muestra en porcentajes, como por ejemplo 68%, necesitamos del DS.

Se define 1 DS como la distancia entre la media y el 34% del área bajo la curva que se halla a la derecha o izquierda de la misma (puede ser a la derecha o izquierda porque la muestra es simétrica).

Esta es la unidad equivalente de distancia a la que hacíamos referencia en el apartado de desvío estándar (Capítulo 10).

Así como existe un *patrón* internacional para definir un metro o un kilogramo y se tomarán como patrones para definir otras medidas de distancia o peso, en Estadística el patrón del DS es el porcentaje de área bajo la curva, siendo la equivalencia $1 \text{ DS} = 34\%$ del área bajo la curva a partir de la media muestral.

Por lo tanto, entre ± 1 DS se encuentra el 68% de la población.

Entre $\pm 1,96$ DS (o 2 DS como comúnmente se dice) se encuentra el 95% de la población debido a que el patrón para 2 DS es 47,5% del área bajo la curva a cada lado de la media muestral y entre $3 \pm$ DS (en realidad 2,57 DS) el 99%, el patrón para 3 DS es entonces 49,5% del área bajo la curva, a cada lado de la media muestral.

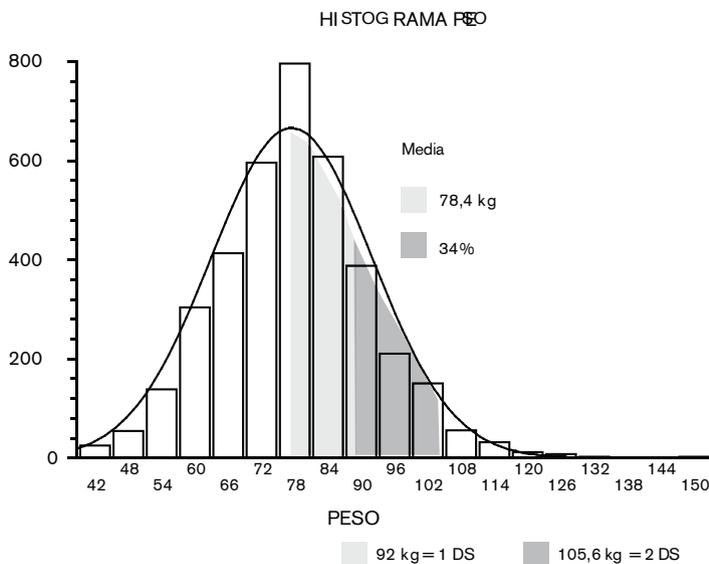


Figura 11.2: Histograma de peso de 3 800 pacientes

Si bien el porcentaje patrón para cada desvío estándar es diferente, en términos de distancia entre uno y otro son iguales.

Miremos el histograma de peso. (Figura 11.2)

La media de peso de esta población es de 78,4 kg, coincide perfectamente con el punto medio de la barra más alta del histograma y divide la muestra en dos partes iguales, por lo tanto coincide con la mediana.

El programa estadístico nos informa que el DS es igual a 13,66 kg. ¿Cómo lo calculó?

Lo que hizo el programa fue calcular el área bajo la curva que cubre el 34% a la derecha o a la izquierda de la media y luego se fijó a qué valor de la muestra correspondía este 34%. En este caso el valor correspondiente al 34% es 92 kg. Como la distancia se mide desde la media, lo único que hay que hacer es restar $92 - 78,4 = 13,6$ y obtenemos el valor de 1 DS para esta muestra.

Si ahora calculamos el valor de 2 DS será: $13,6 \times 2 = 27,2$, y si lo sumamos a la media es igual a 105,6. Si hiciéramos el cálculo comprobaríamos que la distancia entre la media y el 2° DS es el 47,5% de área bajo la curva.

Si miramos el histograma, la distancia entre la media y el 1° DS es igual a la distancia entre el 1° y 2° DS, aunque el área bajo la curva que define a cada uno sea diferente.

Esto es así por las propiedades matemáticas del cálculo del área bajo la curva.

Calcular el 3° DS es lo mismo: $13,6 \times 3 = 40,8$. Este valor sumado a la media es igual a 119,2. El valor ya cubre casi toda el área a la derecha de la media.

Veamos ahora qué pasa si agregamos los DS a la izquierda de la media. (Figura 11.3)

Entre ± 1 DS tenemos el 68% de la población, cuyos valores están entre 65 y 92 kg, calculados como: media + 1 DS y media - 1 DS.

Entre ± 2 DS tenemos el 95% de la población, media + 2 DS y media - 2 DS o, dicho correctamente, media + 1,96 DS y media - 1,96 DS. Este, como ya sabemos, es el intervalo de confianza del 95%, que para esta muestra es de 52 - 105 kg.

También podemos calcular el intervalo de confianza del 99% si queremos saber qué valores obtendríamos en el 99% de los casos. Simplemente sumamos y restamos a la media 3 DS, entonces el IC 99% será igual a 37,6 - 120 kg.

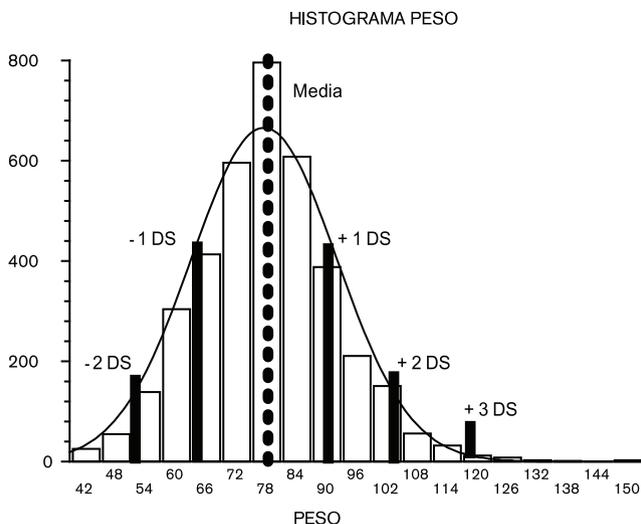


Figura 11.3: Área bajo la curva con $\pm 1, 2$ y 3 DS

Si nos fijamos en el límite inferior de este intervalo, 37,6 kg, en el histograma, veremos que este valor no aparece en la muestra. La primera barra del histograma toma valores que van desde 39 a 45 kg, no obstante el cálculo del DS nos permite predecir cuál sería ese valor.

El cálculo del área bajo la curva, como ya fue mencionado, nos permite derivar probabilidades.

Si tomamos toda el área bajo la curva sabemos que es el 100%; en términos de probabilidades, ¿cuál es la probabilidad de obtener valores entre 39 y 129 kg? (39 y 129 kg son los rangos de esta muestra).

Obviamente, como es el 100% de la muestra, la probabilidad será 1, o muy cercana a la unidad, teniendo en cuenta la variabilidad al azar.

Cuando decimos que la probabilidad de obtener, en el 95% de los casos, x valores, nos estamos refiriendo al 95% del área bajo la curva.

Entonces los porcentajes del área bajo la curva son traducidos a probabilidades.

Si quisiéramos saber cuál es la probabilidad de obtener un peso mayor a 108 kg, lo que tenemos que hacer es calcular el área bajo la curva a la derecha de este valor y si quisiéramos saber la probabilidad de encontrar algún valor por debajo de 108 kg habría que calcular el área bajo la curva a la izquierda de este. (Figura 11.4)

Para el cálculo del área correspondiente para cada caso, lo que se hace es calcular la distancia en las unidades que corresponden, esto es calcular a cuántos desvíos estándar está 108 de la media.

Si la media es 78 kg y el DS 13,6 kg, para medir a cuántos DS está 108 de la media, el cálculo es muy sencillo:

$$N^{\circ} \text{ DS} = \frac{108 - 78}{13,6} = 2,2 \text{ DS}$$

Lo que hicimos primero fue calcular la distancia entre 108 y la media (78) en kg y luego lo dividimos por el valor del DS para obtener el número de desvíos estándares correspondientes.

Ahora, con el número de DS podemos calcular a qué área bajo la curva corresponde la distancia de 2,2 DS, con la ayuda de una tabla que ya tiene calculada el área bajo la curva para cualquier valor de DS.

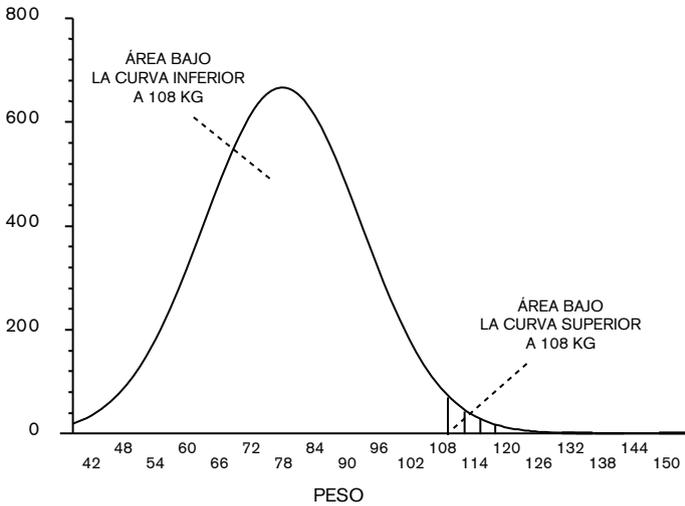


Figura 11.4: Área bajo la curva. Cálculo de probabilidades

Según la tabla, el área bajo la curva a la derecha de 108 es de 1,39%; en términos de probabilidad esto es igual a $p = 0,0139$, casi 1,4 en 100 posibilidades de obtener un valor igual o mayor a 108.

Para calcular la probabilidad de obtener un valor menor nos fijamos en el valor del área bajo la curva a la izquierda, también está en la tabla.

Este valor es de 0,98, lo que es lo mismo al 98% del área. La probabilidad de obtener un valor menor es entonces 98 en 100 o $p = 0,98$.

Cuando calculamos el área bajo la curva sólo a la derecha o izquierda de un determinado valor, decimos que buscamos el área o valor de probabilidad o valor de p a *una cola*.

Luego veremos en qué caso buscamos el valor del área bajo la curva o valor p a dos colas, esto es, calcular el área a la derecha e izquierda simultáneamente.

¿Cuál sería el área bajo la curva por encima de 105 kg y por debajo de 52 kg?

Estos valores son los límites inferior y superior del IC 95%, o lo que es lo mismo, estos valores corresponden a ± 2 DS que ya calculamos, por lo tanto a la derecha de 105 el área será de 0,025 (2,5%) y a la izquierda de 52 también 0,025.

Si queremos saber cuál es la probabilidad de encontrar valores superiores a 105 o menores a 52, el área bajo la curva sumando estos valores es de 5%, la probabilidad o valor p es de 0,05.

Por lo tanto, cuando calculamos el área bajo la curva por encima o por debajo de un determinado valor o rango de valores, lo que estamos calculando es valor p .

Encontrar un valor mayor a 108 kg tiene una probabilidad o área bajo la curva de 0,0139, lo que es igual a decir $p = 0,0139$.

Si el intervalo de confianza del 95% de esta población es tomado como los límites inferior y superior para considerar *normal* el peso de un sujeto, entonces diríamos que el sujeto con peso mayor a 108 kg está por fuera de estos límites y por lo tanto es considerado no normal o, en términos estadísticos, que el valor no es esperable de obtener sólo por azar.

Es lo mismo que decir ¿cuál es la probabilidad de que por azar un sujeto pese más de 108 kg si es normal?: la respuesta es 1,39% o $p = 0,0139$.

Si recordamos que la Estadística asume a priori que todo evento ocurre por azar y que para demostrar lo contrario debemos tener *la evidencia suficiente*, concluimos que este valor no fue encontrado por azar, ya que el valor de p es inferior a 0,05, este hallazgo no es una

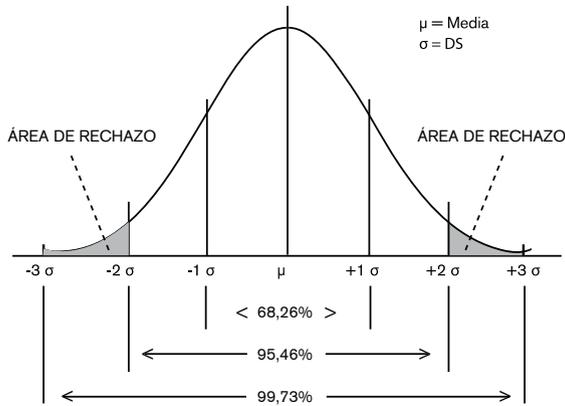


Figura 11.5: Área de rechazo en una distribución normal

variación esperada para el 95% de la muestra, por lo tanto la causa por la que se obtuvo este valor no puede ser atribuida al azar, podríamos decir que la causa es que el sujeto es obeso.

Si consideramos que por convención la evidencia para rechazar la hipótesis de azar o hipótesis nula es obtener un valor de $p < 0,05$, podemos decir que en la curva normal hay un área de rechazo establecida. Cualquier valor que esté dentro de esta porción del área de rechazo será considerado como no producido por el azar. (Figura 11.5)

Siempre que un conjunto de datos muestre una distribución normal, es posible aplicar los conceptos que hemos visto.

Al ser la media el punto a partir del cual se miden los porcentajes de área bajo la curva, esta se transforma en el punto cero, y a partir de este punto se divide la muestra en tramos equidistantes llamados desvío estándar.

El eje x de un histograma podría entonces tomar valores cero como punto central y valores negativos y positivos de DS a los lados de la media.

A esta forma de notación se la denomina curva normal estandarizada. (Figura 11.6)

Esto es lo que posibilita que, con sólo conocer los parámetros de una muestra, el valor de la media y el DS, podamos reconstruir la distribución de los datos.

Si nos dijeran que la media de glucemia es de 90 mg% y el DS de 10 mg%, podríamos ponerle números de valores de glucemia al eje x de la Figura 11.6.

El punto cero es la media cuyo valor sería 90 mg%, el 1° DS a la derecha contendría un valor de glucemia de 100 mg%, 2 DS 110 mg% y 3 DS 120 mg%. Los valores a la izquierda de la media serían: 80 mg% para -1 DS, 70 mg% -2 DS y 60 para -3 DS.

Ahora vemos que el eje x puede tomar cualquier valor o unidad, pero la forma en que se distribuyen los datos, si es normal, será siempre la misma.

¿De qué depende que un grupo de datos tenga una distribución normal?

El primer término depende fundamentalmente del número de sujetos que tenga la muestra y el segundo término, de la naturaleza del dato; hay variables continuas que, por más que la muestra sea de un millón de sujetos, nunca van a mostrar una distribución normal; a la inversa, hay otras variables que con pocos sujetos muestran una forma muy similar a la normal.

En la Figura 11.7 A, se observa la distribución de la superficie corporal de 3 800 sujetos, en B, con 120 sujetos.

Es evidente que el histograma con 120 sujetos ya no se adapta a la forma de la curva normal.

Cuando tenemos muestras con pocos sujetos, uno de los problemas es definir cuántos son pocos sujetos. Para algunos autores es menor a 100, para otros menor a 60, y para otros menor a 30... Obviamente no existe acuerdo.

Cuando la forma de distribución es del tipo de la Figura 11.7 B, se dice que no tiene distribución simétrica o bien que su distribución es no paramétrica. En este caso, como veremos más adelante, usaremos métodos estadísticos que no se basen en la distribución normal.

Si la muestra tiene pocos sujetos pero tiene distribución normal, lo que sucede es que la curva se aplana y los extremos no llegan a tocar el eje x. Esto sucede porque al poseer menor número de sujetos el DS es más grande, recordemos que el DS es inversamente proporcional al número de sujetos, entonces, a menor cantidad de datos, mayor variabilidad de la muestra.

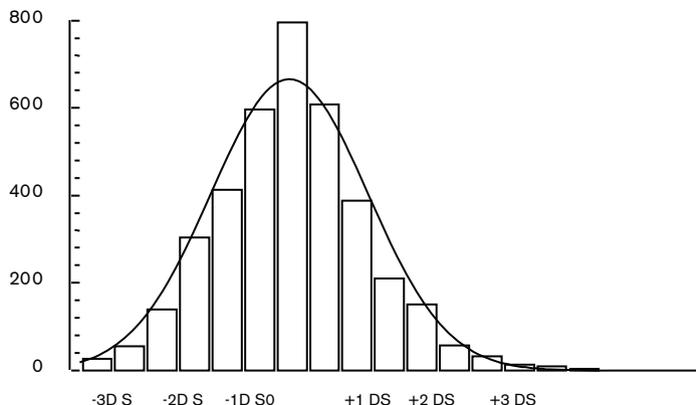
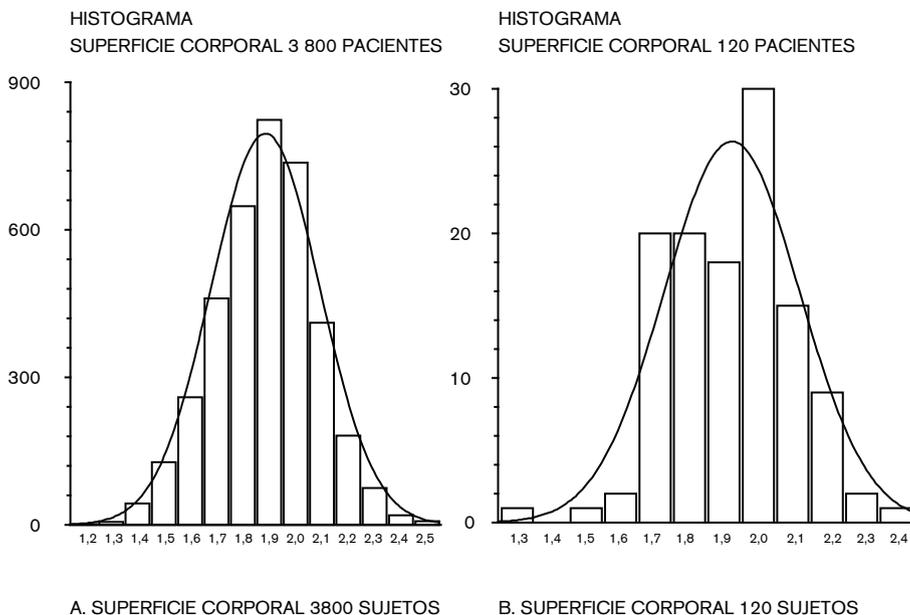


Figura 11.6: Curva normal estandarizada



A. SUPERFICIE CORPORAL 3800 SUJETOS

B. SUPERFICIE CORPORAL 120 SUJETOS

Figura 11.7: Comparación de muestras con diferente número de sujetos

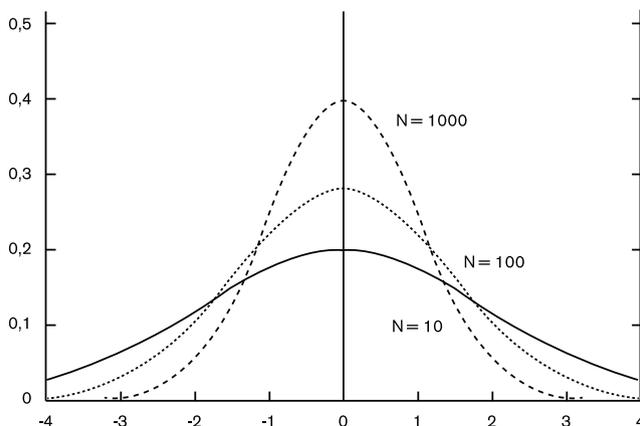


Figura 11.8: Cambio de la forma de la curva normal según el número de sujetos

Como puede observarse en la Figura 11.8, la curva con 1 000 sujetos tiende a ser más estrecha y alta, a medida que el número de sujetos disminuye la curva se aplana y se ensancha sin llegar a cortar el eje x, como se ve en la curva con 10 sujetos.

Este cambio en la forma de la curva hace que el cálculo del área bajo la curva deba ser corregido matemáticamente.

Cuando calculamos a qué distancia está un valor determinado de su media en número de DS, lo hacemos aplicando el razonamiento que hemos visto hasta ahora:

$$\text{n}^\circ \text{ DS} = \frac{\text{valor} - \text{media}}{\text{DS}}$$

En Estadística, a este cálculo se lo conoce como cálculo de Z, entonces:

$$Z = \frac{\text{valor} - \text{media}}{\text{DS}}$$

Siendo Z igual al número de DS.

El número de DS o Z lo utilizamos para buscar en la tabla el área bajo la curva que corresponde a ese valor, como ya vimos anteriormente. A la tabla se la conoce como de distribución normal.

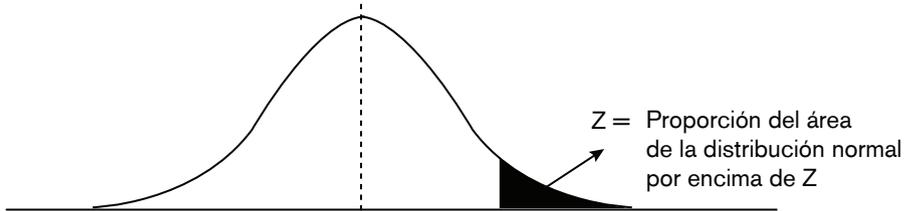
Supongamos que calculamos Z para una muestra cuyo DS es 1,96. Buscando en la tabla correspondiente (Tabla 11.1) entramos primero por la columna Z y buscamos 1,90, luego siguiendo la misma línea nos vamos hasta la 1ª fila donde hallamos el segundo decimal correspondiente, 0,06, para obtener 1,96, el área bajo la curva hallada será de 0,025. Si recordamos que 1,96 corresponde a 2 DS y entre ellos está el 95% de la población, entonces es lógico que por encima de +1,96 DS tengamos 0,025 o 2,5% del área.

Por último, ¿cómo comprobamos si nuestra muestra tiene distribución normal? Hay varios métodos.

Podemos pedirle al programa estadístico que nos calcule las medidas de posición; si estas son bastante cercanas, podríamos pensar que puede tratarse de una distribución normal; luego podemos ver los métodos gráficos como el histograma o *box plot*; si estos son simétricos casi estamos seguros de que se trata de una distribución normal.

Por último existen pruebas estadísticas específicas, como la Wilk Shapiro, que nos confirman o no nuestra primera apreciación.

De cualquier forma, ante la duda, lo mejor es asumir que no hay distribución normal y trabajar con los métodos estadísticos apropiados, la otra opción es transformar los datos, por ejemplo calcular los logaritmos de los datos originales, comprobar si los logaritmos tienen distribución normal y trabajar con estos datos transformados usando pruebas estadísticas que asumen distribución normal.



Segundo decimal de Z

+	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.10	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.20	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.30	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.40	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.50	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.60	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.70	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.80	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.90	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.00	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.10	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.20	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.30	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.40	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.50	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.60	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.70	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.80	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.90	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.00	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.10	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.20	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.30	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.40	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.50	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.60	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.70	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.80	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.90	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.00	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
3.10	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
3.20	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
3.30	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
3.40	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
3.50	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
3.60	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
3.70	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
3.80	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
3.90	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

Tabla 11.1: Distribución normal

ANÁLISIS DE DATOS CUANTITATIVOS. COMPARACIÓN DE MEDIAS



Hasta el momento hemos visto cómo describir de manera resumida los resultados obtenidos, reportando medidas de posición y de dispersión de una muestra o varias muestras, y cómo determinar la probabilidad de que un valor de esa misma muestra pueda o no presentarse por azar.

En los estudios analíticos el objetivo es comparar el resultado de dos o más muestras. En general, lo que se desea saber es si dos medias son iguales o diferentes desde el punto de vista estadístico, siempre y cuando las muestras tengan una distribución normal. La comparación de medias no es válida cuando las muestras son asimétricas. Luego veremos cómo comparar dos medianas, que es la medida de posición que se utiliza para comparar muestras asimétricas.

El razonamiento que se emplea es el mismo que vimos hasta ahora, la pregunta es ¿las medias que queremos comparar están dentro del 95% del área bajo la curva o una de ellas está en el área de rechazo? ¿A qué *distancia estadística* está una de la otra?

En otras palabras, si la media1 es 140 mmHg y la media2 es de 150 mmHg, ¿cuál es la probabilidad de que la media2 sólo represente una variación al azar de la media1 o viceversa?

Cuando estimamos el valor de una media, esta, al igual que los datos, muestra variaciones. Dicho en términos estadísticos, los datos y la media tienen un grado de variabilidad.

Si tomamos 100 muestras de 1 000 sujetos con características similares y calculamos para cada una su media, veremos que las medias obtenidas no son iguales, varían dentro de un rango.

Si graficamos estas 100 medias veremos que su variabilidad es menor a la variabilidad de los datos a los que pertenecen.

Esto significa que los posibles valores que puede adoptar una media tienen un rango más estrecho: si los datos de una muestra tienen un rango de entre 70 y 110, la media tendrá un rango menor de variabilidad, por ejemplo entre 85 y 95.

ERROR ESTÁNDAR

La variabilidad de la media está en relación directa con la variabilidad de los datos, medida por su DS, y es inversamente proporcional a la raíz cuadrada de n de la muestra. Esto significa que mínimas variaciones de n de la muestra impactan mucho en la variabilidad de la media.

Por lo tanto, la variabilidad de la media de una muestra es mucho menor a la variabilidad de sus datos. La medida de variabilidad de la media se denomina *error estándar* (ES).

$$ES = \frac{\sigma}{\sqrt{n}}$$

$\sigma = DS$
 $n = n^\circ$ de la muestra

Ejemplo: para una muestra de 100 sujetos, la media de superficie corporal es de 1,88 m², su varianza es de 0,04. Con estos dos datos podemos calcular el DS y ES.

Como el DS es la raíz cuadrada de la varianza, este es igual a $0,2 (\sqrt{0,04})$ y su error estándar es $0,02 (0,2/\sqrt{100})$

Aquí vemos cómo el valor del DS 0,2 es mucho mayor al ES 0,02, justamente 10 veces mayor, que es la raíz cuadrada de 100.

Veamos qué pasa si aumentamos el n de la muestra.

$$n = 3\ 800$$

$$\text{Media} = 1,89 \text{ m}^2$$

$$\text{Varianza} = 0,036$$

$$\text{SD} = \sqrt{0,036} = 0,19$$

$$\text{ES} = 0,19/\sqrt{3\ 800} = 0,003$$

La media no ha cambiado mucho; tampoco su varianza y por lo tanto su DS. Pero el ES es ahora mucho más chico: es de 0,003 vs. 0,02 comparado con la muestra de 100 sujetos.

La otra particularidad que tiene la distribución de las medias es que siempre adoptan la distribución normal, aunque los datos de donde provengan no sean normales. A esto se lo denomina *Teorema central del límite*. Este teorema es fundamental, porque cuando aplicamos pruebas estadísticas para comparar medias, estas asumen que la dispersión de las medias tiene distribución normal.

Por lo tanto, si graficamos varias medias podemos promediarlas entre sí y obtener la media de las medias. Si tomamos esta *gran media* como punto central veremos que entre ± 1 ES encontraremos el 68% de los valores posibles de obtener por azar, entre ± 2 ES el 95% y entre 3 ES el 99%.

Es lo mismo que para los datos, sólo que aquí la curva es más alta y estrecha debido a la menor variabilidad; pero el área bajo la curva es la misma, representa el 100% y por lo tanto es posible calcular áreas o probabilidades entre diferentes puntos de la misma.

Entonces, cuando queremos saber a qué distancia está una media de otra o queremos conocer cuál es la variabilidad esperada de una media calculando el IC 95% para ella, lo hacemos usando el ES. Es muy importante recordar que la medida de variabilidad de los datos es el DS y la de la media es el ES. Si utilizamos DS en lugar de ES para comparar dos medias o para calcular el IC 95% de la media, cometeremos un error grave.

INTERVALO DE CONFIANZA DE LA MEDIA

Si medimos la tensión arterial sistólica de un grupo de 100 pacientes podríamos obtener una media de 140 mmHg, dentro de una hora podría variar por ejemplo a 145 mmHg y mañana podría ser de 130 mmHg. Como en general lo que hacemos es medir una sola vez la tensión arterial, debemos conocer cuál es el rango esperable de variación de la media, y esto lo logramos calculando el Intervalo de confianza del 95% para la media.

Así como el IC 95% de los datos se calcula tomando el DS, para la media se usa el ES, por lo que el IC 95% es igual a:

$\text{IC 95\% media} = \text{media} \pm 1,96 \text{ ES}$

Usando los datos de los dos ejemplos anteriores calculemos ahora los IC 95% para los datos y para las medias.

$$n = 100$$

$$\text{Media} = 1,88 \text{ m}^2$$

$$\text{Varianza} = 0,04$$

$$DS = 0,2$$

$$ES = 0,02$$

IC 95% Datos:

$$\begin{aligned} \text{Media} \pm 1,96 * DS &= 1,88 - 1,96 * 0,2 = 1,88 - 0,39 = 1,48 \text{ m}^2 \\ &= 1,88 + 1,96 * 0,2 = 1,88 + 0,39 = 2,27 \text{ m}^2 \end{aligned}$$

$$\text{IC 95\% datos} = 1,48 - 2,27 \text{ m}^2$$

IC 95% Media

$$\begin{aligned} \text{Media} \pm 1,96 * ES &= 1,88 - 1,96 * 0,02 = 1,88 - 0,039 = 1,84 \text{ m}^2 \\ &= 1,88 + 1,96 * 0,02 = 1,88 + 0,039 = 1,92 \text{ m}^2 \end{aligned}$$

$$\text{IC 95\% media} = 1,84 - 1,92 \text{ m}^2$$

Como vemos, en el 95% de los casos esperaremos que los valores de los datos estén entre $1,48 - 2,27 \text{ m}^2$, mientras que, en el caso de la media, esperaremos obtener en el 95% de los casos valores de media de entre $1,84$ y $1,92 \text{ m}^2$. Como se observa, los datos tienen mayor variación que su media.

$$n = 3\ 800$$

$$\text{Media} = 1,89 \text{ m}^2$$

$$\text{Varianza} = 0,036$$

$$DS = \sqrt{0,036} = 0,19$$

$$ES = 0,19 / \sqrt{3\ 800} = 0,003$$

IC 95% datos:

$$\begin{aligned} \text{Media} \pm 1,96 * DS &= 1,89 - 1,96 * 0,19 = 1,89 - 0,37 = 1,52 \text{ m}^2 \\ &= 1,89 + 1,96 * 0,19 = 1,89 + 0,37 = 2,26 \text{ m}^2 \end{aligned}$$

$$\text{IC 95\% datos} = 1,52 - 2,26 \text{ m}^2$$

IC 95% media

$$\begin{aligned} \text{Media} \pm 1,96 * ES &= 1,89 - 1,96 * 0,003 = 1,89 - 0,0058 = 1,88 \text{ m}^2 \\ &= 1,89 + 1,96 * 0,003 = 1,89 + 0,0058 = 1,90 \text{ m}^2 \end{aligned}$$

$$\text{IC 95\% media} = 1,88 - 1,90 \text{ m}^2$$

Como se puede observar, los IC 95% para los datos y la media de la muestra con 100 pacientes son más amplios que los IC 95% de la muestra con 3 800 pacientes; este intervalo es más preciso.

El IC 95% de la media de la muestra con 100 pacientes admite un rango posible de valores esperables por azar de 8 m², mientras que el IC 95% de la media con 3 800 pacientes es de sólo 2 m², lo que significa que el *verdadero valor* está entre 1,8 y 1,90 m². Todo valor por debajo o por arriba de este estrecho margen se considerará no esperable por azar.

De la misma manera que se calcula el número de DS o Z para medir la distancia entre un valor y su media, también podemos calcular el número de ES a los que se encuentran 2 medias. Si las muestras tienen un número de pacientes > 100 cada una de ellas, para determinar a qué distancia en ES están dos medias de distribución normal, podemos aplicar el mismo razonamiento que utilizamos para determinar a qué distancia está un dato de su media. Entonces Z será igual a:

$$Z = \frac{\text{media1} - \text{media2}}{\text{ES}}$$

Con el valor de Z obtenido, nos fijamos en la tabla de distribución Z a qué valor de p corresponde.

Como ya mencionamos, cuando el tamaño de la muestra es chico la curva se modifica, haciéndose más plana y sus extremos no llegan a tocar el eje x.

Esto también sucede si graficáramos 100 medias obtenidas de muestras, por ejemplo, de 10 sujetos.

Aunque por Teorema central del límite se asume que las medias tienen distribución normal, esto se aplica para muestras con gran número de sujetos.

W S Gossett demostró que la distribución de las medias para muestras con pocos sujetos ($n < 100$) era similar pero no igual a la distribución normal.

Este matemático fue conocido como *Student*, y a la distribución que él describió se la conoce como *distribución t* o de *Student*.

En este caso, para calcular el número de ES a los que se encuentran 2 medias, aplicamos el mismo cálculo que para Z:

$$t = \frac{\text{media1} - \text{media2}}{\text{ES}}$$

Cuando el cálculo del número de ES se hace en una muestra chica, hay que ajustar el cálculo del área bajo la curva al número de sujetos. Como el área bajo la curva depende del número de sujetos (modifica el DS y por lo tanto al ES) a estos se los denomina grados de libertad y se calcula como $n - 1$.

En la tabla correspondiente, que se denomina *tabla de distribución t de Student* (Tabla 12.1), se busca primero el número de grados de libertad (n° de sujetos - 1) y luego el valor de t. Una vez hallado este valor, la tabla nos informa el valor de p a que corresponde. En este caso, no nos da el valor exacto del área bajo la curva.

Dado que las diferencias con respecto a la curva normal son mínimas en el caso de muestras grandes, siempre que se comparan medias de muestras con gran cantidad de sujetos también se usa la distribución de t.

Supongamos que hemos medido la superficie corporal de un grupo de 24 deportistas, cuya media es 1,95 m², y deseamos compararla con la media de superficie corporal de una población no deportista para saber si la superficie corporal de aquel grupo es estadísticamente mayor a la de la población general.

Tomemos el dato de la media de superficie corporal de los 3 800 pacientes que previamente utilizamos para calcular el IC 95%.

Lo primero que debemos saber es a cuántos ES está una media de la otra, para eso debemos calcular t.

$$t = \frac{\text{media2} - \text{media1}}{\text{ES}} = \frac{1,95 - 1,89}{0,01} = 3$$

$n = 24$

Media2 = 1,95 m²

ES = 0,01

Media1 = 1,89 m²

Error estándar promedio de las dos muestras

Media de referencia

Con el valor de t hallado (3) para una muestra de 24 pacientes, buscamos en la tabla de distribución de t (Tabla 12.1) primero los grados de libertad (24 - 1 = 23), luego el valor de t = 3 y desde ese valor subimos hasta la primera fila donde se encuentran los valores de p.

Como el valor de t = 3 no está, buscamos los más aproximados: 2,80 - 3,48. Para estos valores, p está entre 0,01 y 0,002. Decimos entonces que para t = 3 p es < a 0,002.

La conclusión que obtenemos es que la media de superficie corporal de los deportistas es estadísticamente diferente a la superficie corporal de la población general, debido a que la probabilidad de encontrar un valor de 1,95 m² en una población no deportista es menor a 2 en 1 000.

También podemos deducir que si una media está a 3 ES de otra, obviamente está por fuera del 95% esperado de variación al azar de la media1 y por lo tanto está en el área de rechazo de la hipótesis nula de igualdad de las medias.

Si observamos de nuevo el IC 95% para la media de 3 800 pacientes (IC 95% media = 1,88 - 1,90 m²) también podemos inferir que 1,95 m² es estadísticamente diferente, porque no se encuentra entre el 95% de los valores posibles de obtener por azar para un grupo de pacientes no deportistas.

Debido a que el cálculo del área bajo la curva se ajusta al número de sujetos en muestras pequeñas, el IC 95% también debe ser ajustado: en lugar de usar ± 1,96 ES hay que averiguar a cuántos ES está el 95% de los valores probables de obtener por azar para la media.

Para la muestra de 100 pacientes vimos que el IC 95% de la media tomando ± 1,96 ES era 1,84 m² - 1,92 m².

Corregido por distribución de t será:

Media ± $t_{0,05}$ * ES

$t_{0,05}$ = valor de t para el n° de grados de libertad correspondientes a la muestra y p = 0,05

Para el ejemplo, el IC 95% es: media ± 2,069 * ES

Reemplazamos 1,96 por el valor de t que nos da la tabla. Buscamos n - 1 = 23 y nos fijamos qué valor aparece cuando p es igual a 0,05, en este caso 2,069. (Tabla 12.1)

Tomando este valor, el IC 95% para la muestra será:

Media ± 2,069 * ES

$$1,88 \pm 2,069 * 0,02 = 1,88 - 0,041 = 1,83$$

$$= 1,88 + 0,041 = 1,92$$

IC 95% 1,83 - 1,92

Comparado con el IC 95% calculado asumiendo distribución normal (el que utiliza 1,96 ES), este intervalo no es muy diferente; esto sucede porque la muestra no es tan pequeña. Pero para muestras < 100 se ven diferencias considerables.

En resumen, para comparar medias (y con esto ya estamos asumiendo distribución normal) utilizamos como medida de dispersión o variabilidad al ES.

Para el cálculo del número de ES cuando la muestra es grande (> 100) utilizamos Z y el IC 95% para la media se construye con $\pm 1,96$ ES.

Si la muestra tiene < 100 sujetos utilizamos t para el cálculo de número ES y construimos el IC 95% con el valor de t adecuado al número de sujetos observado en la tabla de distribución t, siempre obtenido de la columna p 0,05.

Si bien la fórmula general para el cálculo de Z o t es igual, el cálculo del ES es diferente, dependiendo de si lo que queremos calcular es Z o t y del tipo de comparación de medias que debamos realizar.

Veremos en el próximo capítulo cómo se calcula el ES para cada caso de comparación de medias.

p		0,20	0,10	0,05	0,02	0,01	0,002	0,001	
GRADOS DE LIBERTAD	1	3,078	6,314	12,71	31,82	63,66	318,3	637	1
	2	1,886	2,920	4,303	6,965	9,925	22,330	31,6	2
	3	1,638	2,353	3,182	4,541	5,841	10,210	12,92	3
	4	1,533	2,132	2,776	3,747	4,604	7,173	8,610	4
	5	1,476	2,015	2,571	3,365	4,032	5,893	6,869	5
	6	1,440	1,943	2,447	3,143	3,707	5,208	5,959	6
	7	1,415	1,895	2,365	2,998	3,499	4,785	5,408	7
	8	1,397	1,860	2,306	2,896	3,355	4,501	5,041	8
	9	1,383	1,833	2,262	2,821	3,250	4,297	4,781	9
	10	1,372	1,812	2,228	2,764	3,169	4,144	4,587	10
	11	1,363	1,796	2,201	2,718	3,106	4,025	4,437	11
	12	1,356	1,782	2,179	2,681	3,055	3,930	4,318	12
	13	1,350	1,771	2,160	2,650	3,012	3,852	4,221	13
	14	1,345	1,761	2,145	2,624	2,977	3,787	4,140	14
	15	1,341	1,753	2,131	2,602	2,947	3,733	4,073	15
	16	1,337	1,746	2,120	2,583	2,921	3,686	4,015	16
	17	1,333	1,740	2,110	2,567	2,898	3,646	3,965	17
	18	1,330	1,734	2,101	2,552	2,878	3,610	3,922	18
	19	1,328	1,729	2,093	2,539	2,861	3,579	3,883	19
	20	1,325	1,725	2,086	2,528	2,845	3,552	3,850	20
	21	1,323	1,721	2,080	2,518	2,831	3,527	3,819	21
	22	1,321	1,717	2,074	2,508	2,819	3,505	3,792	22
	23	1,319	1,714	2,069	2,500	2,807	3,485	3,768	23
	24	1,318	1,711	2,064	2,492	2,797	3,467	3,745	24
	25	1,316	1,708	2,060	2,485	2,787	3,450	3,725	25

Tabla 12.1: Distribución t

ANÁLISIS DE DATOS CUANTITATIVOS.

PRUEBAS DE SIGNIFICACIÓN PARA MEDIAS



En este capítulo veremos los diferentes casos de comparación de medias como los aborda la Estadística.

A modo de introducción y resumen, diremos que antes de analizar una, dos o más muestras de datos continuos debemos:

1. Determinar si la o las muestras tienen una distribución normal.
2. Establecer si los datos están apareados o son independientes.
3. Elegir la prueba estadística correspondiente según la Tabla 13.1.

Nº de muestras	Distribución normal	Apareadas	Test
1	Sí	-	Test de t para una muestra
	No	-	Test de Signo
			<i>Wilcoxon signed rank sum test</i>
2	Sí	No	Test de t para muestras independientes
	No	No	Mann-Whitney
	Sí	Sí	Test de t para muestras apareadas
	No	Sí	<i>Wilcoxon matched pairs signed rank sum test</i>
> 2	Sí	Sí / No	ANOVA

Tabla 13.1: Pruebas estadísticas para comparar medias o medianas

Ya hemos visto cómo la media de una muestra y su ES se usan para construir el IC 95% de la media, que representa los valores probables de obtener por azar de la media para esa población.

También hemos calculado el valor de t para determinar a cuántos ES está una media de la otra y a través de este cálculo establecer la probabilidad de obtener una media dada sólo por azar.

Cuando nos preguntamos si un valor determinado de una media es consistente con cierto valor hipotético de media, lo que hacemos para responder a esta pregunta es utilizar las llamadas *pruebas de significación estadística*, también llamadas *pruebas de hipótesis*.

PRUEBA DE T PARA UNA MUESTRA

En el ejemplo del capítulo anterior nos preguntamos: ¿el valor 1,95 m² es consistente con una media de 1,89 m²? Cuando decimos *consistente* nos referimos a la probabilidad

de obtener 1,95 m² (llamada media muestral), sólo por azar cuando la media hipotética (llamada media poblacional) es de 1,89 m², asumiendo que ambas son iguales desde el punto de vista estadístico (hipótesis nula). En este caso es claro que los datos de las medias provienen de dos muestras de diferentes sujetos, decimos entonces que las medias son independientes.

Aquí tenemos sólo una muestra, que son nuestros 24 deportistas. Al dato de la media de superficie corporal de la población general (media poblacional) contra la cual queremos comparar nuestra media (media muestral) lo hemos obtenido de la literatura o de algún otro estudio que hayamos realizado con anterioridad.

Como ya sabemos, la Estadística asume que los eventos ocurren por azar (hipótesis nula) y para probar lo contrario necesitamos tener *evidencia*. Pues bien, el valor de p que obtenemos de la prueba de significación estadística es la *prueba o evidencia* que necesitamos para demostrar si la hipótesis nula es válida o no.

En otras palabras, el valor de p es la probabilidad de obtener un resultado como el observado cuando la hipótesis nula es cierta.

En términos estadísticos diríamos que la hipótesis nula (notación H₀) para este caso de comparación de medias es que ambas medias son iguales; la hipótesis alternativa (H₁) sería que las medias son diferentes.

En notación estadística:

H₀ media₂ = media₁

H₁ media₂ ≠ media₁

Media₁ = media poblacional

Media₂ = media muestral

La prueba de significación estadística que se utiliza es el llamado *test de Student* o *test de t para una muestra*.

Esta prueba no es otra cosa que el cálculo de t que ya hemos realizado. Como este utiliza media y ES, sólo es aplicable a muestras con distribución normal. En todos los casos que veamos la prueba de t, siempre estaremos hablando de muestras con distribución normal.

En notación estadística:

Prueba de Student para una muestra:

$$t = \frac{x - \mu}{\sigma_X}$$

t = N° de ES

X = media de nuestra muestra (media muestral)

μ = media hipotética (media poblacional)

σ_X = ES de nuestra muestra (ES muestral)

Lo que hacen los programas estadísticos es aplicar la fórmula de la prueba de Student para averiguar a cuántos ES están una media de otra y, una vez obtenido este valor, calculan la probabilidad o valor de p y nos informan su resultado.

Basándonos en el resultado de la prueba decimos que nuestra media es estadísticamente diferente de la media poblacional o no, de acuerdo al valor de p obtenido.

En la Tabla 13.2 se observa el reporte de un programa estadístico comparando la media de 24 deportistas con la media de la población general. El programa informa los grados de libertad (*degree freedom*, $df = 23$), el valor $t = 3$ y dos valores p : uno a 1 cola y otro a 2 colas.

El valor p es la probabilidad de obtener una media al menos tan alejada de la media poblacional como lo observado, asumiendo que ambas medias son iguales (hipótesis nula). La definición estadística dice: p es la probabilidad de obtener un resultado al menos tan extremo como el observado cuando la hipótesis nula es cierta.

En otras palabras, si asumimos que ambas medias son iguales ¿cuál es la probabilidad de obtener una media como la que obtuvimos en nuestra muestra cuando está situada a 3 ES de la media poblacional?

Obviamente, ya sabemos que, estando nuestra media a 3 ES de la media poblacional, se encuentra en el área de rechazo. Recordemos que esta área es de 2,5% a cada lado de la distribución de la media y está por fuera de ± 2 ES.

Como ya vimos, la media puede variar por azar dentro de cierto rango. El IC 95%, para la media 1,89 puede variar entre 1,88 – 1,90 m². Cuando se calcula p a 2 colas, lo que se está haciendo es verificando que el valor observado (el de nuestra muestra de 1,95) no se encuentre dentro del IC 95% de la media poblacional; es decir, busca hacia la derecha y hacia la izquierda de la media poblacional que el valor de nuestra media no esté dentro de 2 ES.

Esto se debe a que la hipótesis alternativa que exploramos supone que la media muestral es diferente, ya sea mayor o menor, al valor de la media poblacional.

En muy pocos casos la hipótesis alternativa supone que un valor es mayor o menor a la media poblacional, lo que es distinto a decir que sólo son diferentes. En ese caso sólo se explora un lado del área bajo la curva y se dice que el valor de p es a una cola.

Por ejemplo, si nuestra hipótesis alternativa fuera que la media de los deportistas es mayor a la de la población general, buscaríamos sólo en el lado derecho del área bajo la curva, para ver cuál es la probabilidad de obtener un valor igual o mayor a 1,95 cuando la hipótesis nula es cierta.

En el caso planteado, sólo nos preguntamos: si las dos medias son iguales ¿la hipótesis nula es cierta? Como el valor de p a 2 colas es menor a 0,05, tenemos evidencia para rechazar la hipótesis nula y decimos que aceptamos la hipótesis alternativa de que las medias son diferentes.

En términos del estudio, la superficie corporal de los deportistas es estadísticamente diferente a la de la población general, en este caso es estadísticamente mayor a la superficie corporal de la población general.

<p>Prueba de t para una muestra Media de la muestra = 1,95 Media de la población = 1,89 Tamaño de la muestra $n = 24$ DS de la muestra = 0,0489 $df = 23$ $t = 3,011018$ Una cola $p < 0,0001$ (p 1 cola) Dos colas $p < 0,0001$ (p 2 colas)</p>
--

Tabla 13.2: Reporte del programa estadístico. Prueba de Student para una muestra

Otra forma de determinar si dos medias son estadísticamente diferentes es a través de la interpretación del IC 95%.

La hipótesis nula sostiene que la media de nuestra muestra es igual a 1,89 (media poblacional). El IC 95% para la media 1,95 es: 1,93 – 1,97 m². ¿Está el valor de la media poblacional dentro de este IC 95%?

¡NO! Por lo tanto, estamos en condiciones de afirmar que estas dos medias son estadísticamente diferentes.

Como regla general para cualquier prueba estadística, si conocemos la hipótesis nula decimos que, si el valor propuesto como referencia NO se encuentra dentro del IC 95% de la media que estamos estimando, estas son estadísticamente diferentes. No sabremos exactamente cuál es el valor de p , pero seguro es menor a 0,05.

Cuando la muestra de sujetos es > 100 , podemos usar la distribución normal, en lugar de la distribución de t .

La fórmula es la misma, lo único que cambia es que el valor de Z se busca en la tabla de distribución normal y el IC 95% de la media se construye como $\text{media} \pm 1,96 \text{ ES}$, ya que en la distribución normal entre $\pm 1,96$ está el 95% de los valores esperados para la media.

PRUEBA DE T PARA MUESTRAS APAREADAS

¿A qué se llaman muestras apareadas?

Hay situaciones en las cuales las medias que se quieren comparar no son independientes, no provienen de dos grupos diferentes de sujetos.

En muchos casos podemos realizar múltiples mediciones en los mismos sujetos.

El ejemplo más común es medir la tensión arterial de un grupo de sujetos antes y después de administrar un tratamiento. Entonces, dado que los datos provienen del mismo grupo, ya no son independientes; los datos de la segunda medición van a depender de la primera. Se dice entonces que estos son datos apareados.

La importancia de esto radica en que el planteo de la hipótesis nula aquí es diferente.

Supongamos que tenemos las mediciones de homocisteína pre y posmedicación de una muestra de 20 sujetos y queremos determinar si esta droga disminuye los valores de homocisteína. (Tabla 13.3)

Cuando los datos están apareados, en lo que estamos interesados es en el promedio de las diferencias entre mediciones de cada sujeto y en su variabilidad.

Como vemos en la Tabla 13.3, por cada par de mediciones sobre el mismo sujeto se calcula la diferencia entre ambas y con todas las diferencias se calcula la *media de las diferencias* y su DS, en este caso la media de las diferencias es de $-4,4$ y su DS 5,5.

Significa que, en promedio, la droga disminuyó 4,4 puntos los valores de homocisteína. Pero, ¿estamos seguros de que si repetimos el experimento esta droga siempre demostrará disminuir la homocisteína? ¿O cabe la posibilidad de que en algún caso veamos que no hay efecto?

Lo que en este caso nos preguntamos es ¿la media de las diferencias es distinta a cero?

Si la medicación no tuviera efecto, la homocisteína no descendería sus valores, y las diferencias individuales serían cero. Por lo tanto, la media de las diferencias también sería cero. Pero si la medicación es eficaz, la homocisteína descendería; en ese caso, la segunda medición sería menor a la primera y las diferencias tenderían a ser negativas. Como consecuencia, la media de las diferencias también sería negativa.

SUJETO	HOM. BASAL	HOM. CONTROL	DIFERENCIA
1	13,3	15,8	2,5
2	14,6	9,8	-4,8
3	12,5	11,6	-0,9
4	17,2	14,3	-2,9
5	15,3	13,3	-2,0
6	10,5	10,2	-0,3
7	17,4	12,9	-4,5
8	19,9	11,1	-8,8
9	22,3	19,0	-3,3
10	14,0	12,3	-1,7
11	13,3	10,8	-2,5
12	15,1	11,6	-3,5
13	15,1	11,0	-4,1
14	30,3	8,0	-22,3
15	24,3	12,4	-11,9
16	12,0	12,2	0,2
17	13,5	10,4	-3,1
18	12,8	10,4	-2,4
19	28,0	16,7	-11,3
20	12,2	11,9	-0,3
MEDIA	16,68	12,28	-4,4
DS	5,5	2,5	5,5

Tabla 13.3: Valor de homocisteína basal y postratamiento

Como ya sabemos que los valores de la media de las diferencias varían por azar, lo que tenemos que determinar es cuál es la probabilidad de encontrar un valor de $-4,4$ para la media de las diferencias, asumiendo que esta es en principio igual a cero.

Es lo mismo que decir que la hipótesis nula es “no hay diferencias entre las mediciones pre y pos” o bien que la media de las diferencias es igual a cero.

$H_0: d = 0$

$d =$ diferencia de las medias

La hipótesis alternativa será en este caso que la media de las diferencias es distinta a cero:

$H_1: d \neq 0$

Para este caso la prueba de t es igual a:

$$t = \frac{d - 0}{es(d)}$$

$d =$ media de las diferencias

$es(d) =$ ES de la media de las diferencias

La fórmula incluye el cero porque la media de la de la hipótesis nula es cero; podríamos también escribir la fórmula como:

$$t = \frac{d}{es(d)}$$

Recordemos que

$$\text{Para este ejemplo } t \text{ es igual a: } ES = \frac{DS}{\sqrt{n}}$$

$$t = \frac{-4,4}{1,23} = 3,57$$

Si buscamos en la tabla de distribución de t, vemos que para 19 grados de libertad ($n - 1$) el valor de $t = 3,57$ y corresponde a un valor de p de 0,002.

Con esto concluimos que la media de las diferencias es estadísticamente diferente a cero. Es lo mismo que decir que rechazamos la hipótesis nula y aceptamos la alternativa. Por lo tanto, para el estudio, se demuestra que la droga disminuyó los valores de homocisteína de forma significativa.

Si construimos el IC 95% de la media de las diferencias tendremos:

$$IC \ 95\% = -4,4 \pm (2,09 * 1,23) = -4,4 \pm 2,57$$

Usamos 2,09 en lugar de 1,96 porque es una muestra pequeña; 2,09 es el valor que corresponde a 19 grados de libertad en la tabla de distribución de t cuando p vale 0,05. En esta muestra vemos que entre $\pm 2,09\%$ del área bajo la curva se encuentra el 95% de los valores posibles de obtener por azar para la media de las diferencias.

$$\text{Límite inferior IC } 95\% = -6,9$$

$$\text{Límite superior IC } 95\% = -1,8$$

Como vemos el IC 95% no contiene al cero (valor de la media propuesta por la hipótesis nula); por lo tanto inferimos que de 100 veces que hagamos este experimento, en 95 oportunidades obtendremos valores de media de las diferencias entre $-6,9$ y $-1,8$.

Como en el test para una sola muestra, también en este caso si la muestra es > 100 usamos Z.

PRUEBA DE T PARA DOS MUESTRAS INDEPENDIENTES

En este caso tenemos dos grupos de 150 sujetos, que no tienen ninguna relación entre ellos.

A un grupo se le administró placebo y al otro droga activa y se midió colesterol LDL al cabo de 8 semanas de tratamiento.

La media de LDL del grupo placebo fue de 130 mg% con DS de 1,3 y la del grupo droga activa 100 mg% con DS 1.

Para este caso la hipótesis nula (H_0) sostiene que la diferencia entre las medias es igual a cero: no habla de la media de las diferencias como el caso anterior.

$$\text{Para este caso la } H_0 \text{ sería: } 130 - 100 = 0$$

Y la hipótesis alternativa:

$$H_1: 130 - 100 = 0$$

La prueba de Student para este caso se plantea como:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{ES(x_1 - x_2)}$$

\bar{X}_1 = media 1

\bar{X}_2 = media 2

$ES(x_1 - x_2)$ = ES promedio de las dos muestras

Como aquí tenemos dos muestras independientes con sus respectivos ES, lo que la prueba utiliza es el error estándar promedio de las dos muestras.

Sólo a modo informativo se muestran las fórmulas para el cálculo del ES promedio de las dos muestras.

$$ES(x_1 - x_2) = s\sqrt{1/n_1 + 1/n_2}$$

S = varianza promedio de las dos muestras

n_1 = número de sujetos en la muestra 1

n_2 = número de sujetos en la muestra 2

La varianza promedio de las dos muestras se calcula como:

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}}$$

s_1^2 = DS de la muestra 1 al cuadrado

s_2^2 = DS de la muestra 2 al cuadrado

n_1 = n° de sujetos muestra 1

n_2 = n° de sujetos muestra 2

Por lo tanto, la prueba de t para dos muestras independientes (n de cada muestra < 100) es:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}} \sqrt{1/n_1 + 1/n_2}} \quad df = n_1 + n_2 - 2$$

df = grados de libertad

Como vimos antes, construyendo el IC 95% podemos saber si hay diferencias estadísticamente significativas entre ambas medias. Si la H_0 establece que $media_1 - media_2 = 0$, debemos construir un IC 95% para esta diferencia. Si la diferencia de medias contiene al cero, sabremos que la diferencia no será significativa.

El IC 95% para la diferencia de las medias con muestras < 100 es:

IC 95% = $media_1 - media_2 \pm (t * ES)$

$$ES(x_1 - x_2) = s\sqrt{1/n_1 + 1/n_2}$$

t = es el valor de t a p 0,05 correspondiente a los grados de libertad, calculado como $df = n_1 + n_2 - 2$

Cuando las muestras tienen > 100 sujetos, como ya sabemos se utiliza z. El cálculo del ES promedio se calcula de forma diferente:

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1^2 / n_1 + s_2^2 / n_2}}$$

s_1^2 = DS de la muestra 1 al cuadrado
 s_2^2 = DS de la muestra 2 al cuadrado

n_1 = n° de sujetos muestra 1
 n_2 = n° de sujetos muestra 2

El IC 95% para muestras > 100 sujetos es igual a:

IC 95% = media1 - media2 ± (1,96 * ES)

Siendo ES = $\sqrt{s_1^2 / n_1 + s_2^2 / n_2}$

En este caso no se usan los grados de libertad.

Volviendo al ejemplo inicial tenemos:

Muestra 1: n = 150 media 130 mg/% DS 1,3

Muestra 2: n = 150 media 100 mg/% DS 1

Como las muestras tienen > 100 sujetos, usaremos z.

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1^2 / n_1 + s_2^2 / n_2}} =$$

$$z = \frac{130 - 100}{\sqrt{1,3^2 / 150 + 1^2 / 150}} =$$

$$z = \frac{30}{\sqrt{1,69 / 150 + 1 / 150}} =$$

$$z = \frac{30}{\sqrt{0,011 + 0,0066}} =$$

$$z = \frac{30}{\sqrt{0,0176}} =$$

$$z = \frac{30}{0,132} = 227$$

Si buscamos en la tabla de distribución normal el valor 227, veremos que no existe. Eso es porque estar a 227 ES significa una distancia tal que no es necesario casi ningún cálculo. Ya sabemos que están muy alejadas y el valor de p será muy chico, en este caso $< 0,00003$.

Para construir el IC 95% de la diferencia de las medias usamos el valor de ES que calculamos para $z = 0,132$

$$\text{IC 95\%} = 30 \pm 1,96 * 0,132 =$$

$$\text{IC 95\% límite superior} = 30 + 0,258 = 30,25$$

$$\text{IC 95\% límite inferior} = 30 - 0,258 = 29,74$$

$$\text{IC 95\% diferencia de las medias} = 29,74 - 30,25$$

Como podemos observar, este intervalo NO contiene al cero.

Por lo tanto, en base al test de z y al IC 95%, decimos que rechazamos la H_0 (diferencia de la medias = cero) y aceptamos la alternativa.

En este estudio diremos que la media de colesterol del grupo tratado fue significativamente menor a la del grupo placebo.

Hasta aquí hemos visto cómo analizar una muestra, dos muestras apareadas y dos muestras independientes, cuando estas tienen distribución normal.

Veremos ahora cómo analizar los mismos casos cuando las muestras no son normales o simétricas.

PRUEBAS NO PARAMÉTRICAS

Cuando la o las muestras a comparar no tienen distribución normal, se recurre a las llamadas pruebas no paramétricas.

Estas pruebas comparan medianas en lugar de medias. El método matemático que estas pruebas emplean está más allá del alcance de este libro; por lo tanto, sólo mencionaremos los tests disponibles de acuerdo con cada caso en particular.

Cuando queremos comparar una muestra se utiliza el llamado test del signo o, más frecuentemente, el test de *Wilcoxon* (*Wilcoxon signed rank sum test*).

Para el caso de muestras apareadas, el test usado es el de *Wilcoxon* para muestras apareadas (*Wilcoxon matched pairs signed rank sum test*).

Por último, para la comparación de dos muestras independientes, el test utilizado es el llamado *Mann-Whitney*.

Veamos un ejemplo.

Se quiere comparar el valor de PCR de dos grupos independientes.

Lo primero que hacemos es verificar si los datos tienen distribución normal.

El gráfico *box plot* nos muestra que ambas muestras son francamente asimétricas. (Figura 13.1)

Veamos qué sucede, si aplicamos en este caso el test de Student para muestras independientes.

TEST DE T PARA MUESTRAS INDEPENDIENTES

Media de PCR Grupo 1 = 8,084553

Media de PCR Grupo 2 = 14,349254

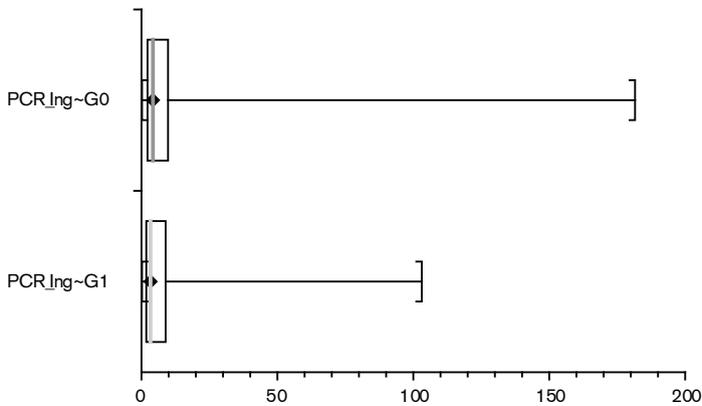


Figura 13.1: Distribución de dos muestras independientes

Error estándar combinado = 2,859258

IC 95% para la diferencia de las medias = -11,895468 a -0,633933

df = 255

t = 2,191023

Valor p = 0,0294

TEST DE T NO VÁLIDO. UTILICE TEST DE MANN-WHITNEY

Como podemos observar, el valor de p es significativo, y por lo tanto concluiríamos que ambas medias son diferentes. Esto no es válido, debido a que las muestras no tienen distribución normal. El mismo programa estadístico detecta este error e informa que el test no es válido y sugiere utilizar el test correspondiente.

No todos los programas estadísticos advierten sobre este error, por lo que es nuestra obligación confirmar el supuesto de normalidad para las dos muestras.

Veamos qué pasa al aplicar el test correspondiente.

TEST DE MANN-WHITNEY

Observaciones PCR G1 = 123 mediana = 3,3

Observaciones PCR G0 = 134 mediana = 4,2

p (2 colas) = 0,1569

IC 95% para la diferencia de las medias = -1,5 a 0,2

Como vemos, cuando aplicamos el test correspondiente, las cosas son bastante diferentes.

Aquí el valor de p ya no es significativo, por lo tanto la conclusión es que ambas medias no son diferentes.

Este ejemplo demuestra lo *peligroso* que puede ser utilizar un test que no está indicado para el tipo de muestras con las que estamos trabajando.

Lo mismo sucede si para datos apareados utilizamos test para muestras independientes.

COMPARACIÓN DE MÁS DE DOS MEDIAS

Otra situación muy frecuente es cuando tenemos más de dos muestras para comparar. En este caso los tests de t y los test no paramétricos que hemos visto hasta ahora no se pueden aplicar.

Un ejemplo de este caso podría ser cuando se prueban 3 o más dosis de una droga en 3 o más grupos de pacientes.

Aquí tendríamos:

Grupo 0: placebo media de tensión arterial 0

Grupo 1: dosis 1 media de tensión arterial 1

Grupo 2: dosis 2 media de tensión arterial 2

Grupo 3: dosis 3 media de tensión arterial 3

En primer término deberíamos saber si existe diferencia entre alguna media en general, y luego entre cada una de las posibles combinaciones, por ejemplo:

Media 0 vs. media 1

Media 1 vs. media 2

Media 2 vs. media 0

Etc.

El test necesario para este tipo de comparación es el llamado *Análisis de la varianza* o *ANOVA*.

Esta prueba estadística es muy complicada desde el punto de vista matemático para explicarla en este texto.

Para cada caso en particular existe un subtipo de análisis de la varianza. Por ejemplo, si los datos son no normales, son apareados o si existe más de un factor que tenga que ser tenido en cuenta para la comparación, por ejemplo 2 drogas con 3 dosis cada una.

Por lo tanto, aquí sólo enfatizamos que para la comparación de más de dos muestras de datos continuos se debe utilizar análisis de la varianza, y para la elección del test adecuado sugerimos la consulta a un profesional en Estadística.

PODER ESTADÍSTICO

Las pruebas de significación estadística proveen, a través del cálculo del valor p , la aceptación o rechazo de la hipótesis nula. En general la hipótesis nula establece que no hay diferencia entre los valores comparados.

En la práctica, lo que se quiere demostrar es si hay o no diferencia estadística entre los grupos a comparar.

Idealmente, la prueba utilizada debe rechazar la hipótesis nula de no diferencia cuando no la hay y rechazarla cuando hay diferencia.

Sin embargo, ninguna prueba es perfecta. Lo que puede suceder se describe en la *tabla de la verdad*. En la Tabla 14.1 las caras felices muestran el comportamiento deseado de una prueba estadística cuando la hipótesis nula es aceptada y realmente no hay diferencia estadística (verdadero negativo) y cuando la hipótesis nula es rechazada y existe diferencia estadística (verdadero positivo).

Las caras tristes muestran la decepción que sentimos cuando una prueba estadística acepta la hipótesis nula de no diferencia cuando realmente la hay (falso negativo) y cuando la prueba nos dice que hay diferencias cuando realmente no las hay (falso positivo).

	H_0 Cierta	H_0 Cierta
Prueba acepta H_0	😊	😞
Prueba rechaza H_0	😞	😊

Tabla 14.1: Tabla de la verdad

H_0 : Hipótesis nula

Una presentación más científica muestra las posibles clasificaciones de un resultado estadístico (aceptación verdadera, rechazo verdadero, falsa aceptación o rechazo de la hipótesis nula). (Tabla 14.2)

Cuando concluimos que no hay una diferencia estadística o que si la hay, esta conclusión puede ser errónea debido a la existencia de dos tipos de errores llamados error tipo I o alfa y error tipo II o beta.

En terminología estadística se denomina error alfa o tipo I al falso rechazo de la hipótesis nula. En los diseños experimentales es frecuente utilizar valores de $\alpha = 0,05$, lo que significa que asumimos la posibilidad de cometer en un 5% un falso rechazo.

En el caso de usar un test de comparación de medias cometemos un error tipo I cuando, siendo la hipótesis nula cierta (no hay diferencias entre medias), concluimos que hay que rechazarla y aceptar la hipótesis alternativa (hay diferencias en las medias).

El error beta o tipo II es la falsa aceptación de no diferencia cuando realmente la hay. El test acepta la hipótesis nula cuando esta es falsa.

	H_0 Cierta	H_0 Cierta
Test acepta H_0	Verdadero negativo ($1 - \alpha$)	Falso negativo Aceptación falsa Error tipo II β
Test rechaza H_0	Falso positivo Falso rechazo Error Tipo I α	Verdadero positivo Rechazo cierto Poder ($1 - \beta$)

Tabla 14.2: Definiciones. Error alfa y beta

El error tipo II es crítico debido a que el interés de un investigador es tener una alta tasa de resultados confiables. Deseamos una alta probabilidad de tener verdaderos rechazos.

A esta probabilidad se la denomina poder estadístico y es la inversa del error beta:

$$\text{Poder} = 1 - \beta$$

Generalmente, el poder estadístico de un estudio varía entre el 80 y el 90%, lo que significa que la probabilidad de cometer un error beta varía entre el 20 y el 10%.

Es frecuente que estos conceptos sean difíciles de incorporar rápidamente, lo relevante es entender las consecuencias de cometer estos errores.

Recordemos que el área de aceptación se encuentra en general por consenso entre ± 2 ES, dejando como área de rechazo un 2,5% a cada lado en la curva de Gauss.

Esta área de rechazo también es la probabilidad de estar errados al decidir si aceptamos o rechazamos una hipótesis nula.

Si el valor de p acordado para rechazar una hipótesis nula es de 0,05, debemos asumir que podemos estar cometiendo un error en un 5% de los casos. Si el valor de p es de 0,01 entonces asumimos que podemos equivocarnos en el 1% de los casos.

Si las distribuciones de dos medias están completamente separadas, no hay posibilidad de cometer ningún error. Los errores tipo I y II, ambos, son igual a cero. (Figura 14.1)

Al no haber superposición el valor de p será $< 0,05$; cuanto más lejos estén las dos distribuciones, menor será el valor de p , tendiendo a cero. Al mismo tiempo, la posibilidad de cometer un error beta es cero, debido a que, al no haber superposición, no hay posibilidad de aceptar erróneamente una hipótesis nula de no diferencia. Si beta es = cero, el poder estadístico es igual a 100%, lo que significa que sabemos que tenemos 100% de posibilidades de detectar una diferencia si esta realmente existe.

En el caso de que las curvas se superpongan (Figura 14.2), la habilidad del test estadístico dependerá de cuánto se superpongan las curvas.

En este ejemplo el error alfa es de 0,05 y la probabilidad de error tipo II es de 0,15 o del 15%, que es igual al área de superposición de ambas curvas. Por lo tanto, el poder estadístico ahora es de 85% ($1 - 0,15$).

Esto indica que hay sólo 85% de probabilidad de detectar una verdadera diferencia, o a la inversa, un 15% de que, habiendo diferencia, esta no pueda detectarse.

Para entender el concepto de poder estadístico hay que mirar cuánto del área de una muestra está más allá de la línea vertical que corresponde al área de rechazo del 0,05 de la otra media.

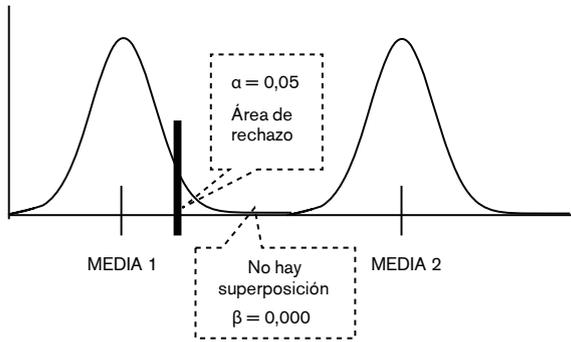


Figura 14.1: Valor β bajo

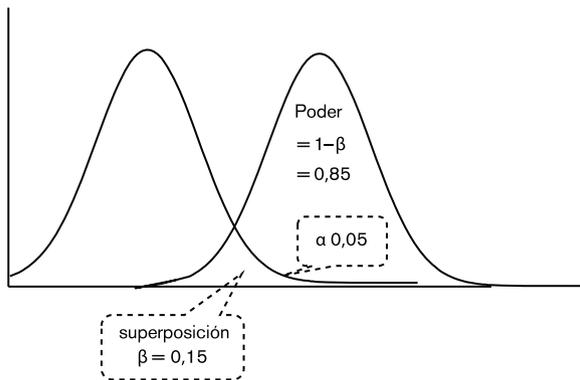


Figura 14.2: Valor β 15%

En la Figura 14.1 toda la curva está a la derecha de la línea vertical correspondiente al área de rechazo. Por lo tanto, aquí el poder estadístico es máximo.

En la Figura 14.2 sólo parte de la curva está a la derecha de la línea vertical, hay un 15% que se superpone, por lo que el poder estadístico está disminuido.

El poder muestra la capacidad de un experimento de detectar una diferencia de determinada magnitud. Cuanto mayor es la diferencia entre las medias, mayor poder para detectar la diferencia. Cuanta mayor superposición, menor poder estadístico. Hay que recordar que la distribución de las medias depende del tamaño de la muestra, a muestras grandes, distribuciones con base más estrecha y más probabilidad de que estén separadas. Por lo tanto, el tamaño de la muestra afecta directamente el error beta.

En conclusión, grandes diferencias serán fáciles de detectar y pequeñas diferencias, más difíciles. Veamos dos ejemplos.

En la Figura 14.3 las muestras tienen un tamaño relativamente grande, debido a la forma de la distribución de las medias. Con un nivel de significación estadística de 0,05, la superposición de las curvas o error beta es de 30, con un poder de 70%. (Figura 14.3)

Si ahora el tamaño de las muestras disminuye, las curvas tendrán una base más ancha y aunque no se modifique el nivel de significancia, el área de superposición aumentará a un error beta de 60, con la consecuente reducción del poder a 40%. (Figura 14.4)

Como vemos, el tamaño de la muestra tiene una influencia muy importante en el error beta y por lo tanto en el poder para detectar una cierta diferencia.

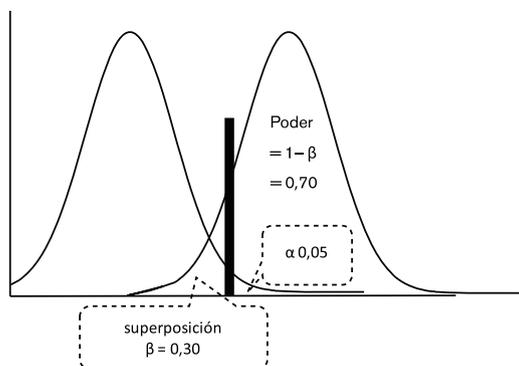


Figura 14.3: Valor β 30%

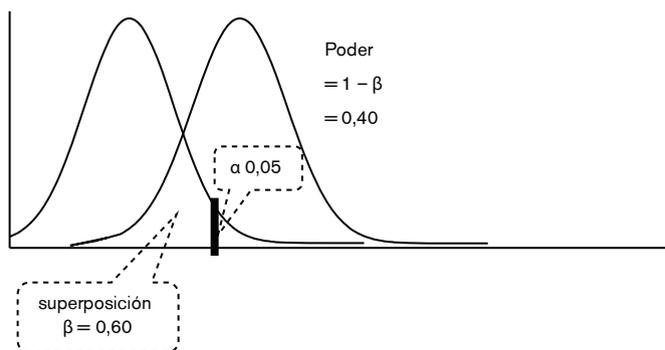


Figura 14.4: Valor β 60%

ELECCIÓN DEL ERROR ALFA O NIVEL DE SIGNIFICACIÓN

Generalmente el nivel de significación estadística o punto de corte para el rechazo de la hipótesis nula se establece en 5%. Existe la posibilidad de disminuirlo a 0,01 o 1%, con lo cual sólo habrá un 1% de posibilidad de rechazar una hipótesis nula verdadera (1% de aceptar diferencia cuando no la hay).

Este cambio de alfa afecta al error beta en la dirección opuesta.

En la Figura 14.5 se fijó el nivel de significación en 0,05, beta es 0,30 y poder estadístico es de 70%.

Si el nivel de significación, con las mismas muestras se cambia a 0,01, ahora el área de superposición será de 0,50 y el poder cae a 50%.

Por lo tanto la probabilidad de detectar una diferencia verdadera ahora es de sólo el 50%. (Figura 14.6)

En conclusión, la elección del nivel alfa implica cambios en los valores de beta y el poder estadístico, que pueden afectar la posibilidad de llegar a una conclusión errónea. De ahí la importancia de entender este tipo de errores y sus consecuencias.

Otra forma de evitar cometer estos errores es calcular correctamente el tamaño de la muestra.

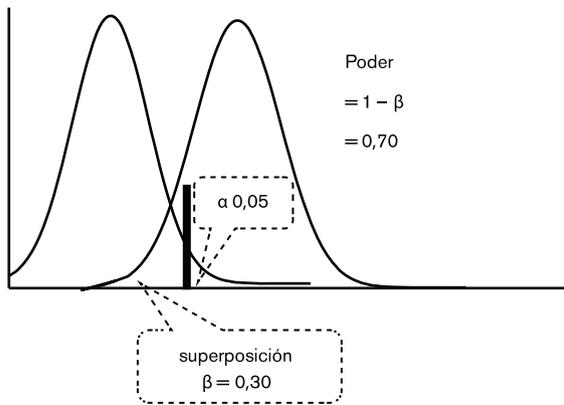


Figura 14.5: $\alpha = 0,05$ - $\beta = 30\%$

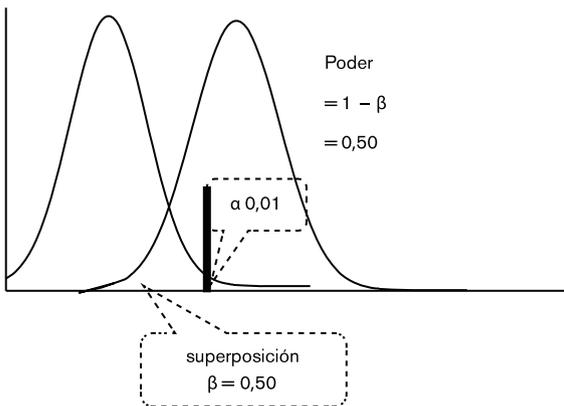


Figura 14.6: $\alpha = 0,01$ - $\beta = 50\%$

CÁLCULO DEL TAMAÑO DE LA MUESTRA



El cálculo correcto del tamaño de la muestra es de vital importancia, dado que de esta podemos reducir al máximo los errores de tipo I y II.

Para calcular el número necesario de sujetos a incluir en un estudio, es condición determinar:

1. ¿Cuál es el diseño del estudio?

El cálculo del número de sujetos que se debe incluir en un estudio para tener la mayor certeza posible de detectar una diferencia cuando esta es cierta, depende del diseño del estudio. Si utilizamos un diseño caso-control donde se ha decidido seleccionar dos controles por cada caso, el tamaño de la muestra debe ser adecuado a esta circunstancia.

2. ¿Qué tipo de variable se desea comparar?

El cálculo del tamaño de la muestra no es igual si uno quiere comparar variables cuantitativas o cualitativas.

3. ¿Cuál es el evento principal?

El evento de interés es determinante, debido a que su incidencia, prevalencia o valor de media en el grupo control influirá en el valor esperado del evento en el otro grupo.

4. ¿Cuál es la mínima diferencia que se quiere detectar?

Cuando realizamos un estudio de tipo analítico, ya sea observacional o experimental, lo que deseamos es determinar si existe diferencia estadística en los eventos de interés entre los grupos que están siendo comparados. Esta diferencia debe ser planteada con anticipación. Para establecer el valor de la diferencia que se desea detectar, ya sea entre medias o porcentajes entre los grupos, se debe estimar cuál sería la posible diferencia que se observaría.

Esto dependerá de varios factores: por ejemplo de la tasa de eventos del grupo control, del efecto biológico de la droga, de las características de la población a estudiar, etc. Por lo tanto, debemos adecuarnos a estos factores y plantear una diferencia posible de detectar bajo estas circunstancias.

En medicina se estima que una diferencia de 20% es biológicamente plausible y tiene un impacto significativo en salud pública.

Existen dos formas de calcular el tamaño de la muestra.

DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA POR PRECISIÓN

Cuando el efecto de la intervención es conocido, se puede desear estimar este efecto con mayor precisión.

Si sabemos que una droga reduce la tasa de complicaciones de una enfermedad entre el 8 y 14% de acuerdo con el IC 95% obtenido en otro estudio, podemos calcular el número de

sujetos necesarios para obtener un ancho del IC 95% menor, con lo cual el valor del efecto de la intervención será estimado con mayor precisión.

Este método es muy poco utilizado en el diseño de estudios de investigación médica.

DETERMINACIÓN DE LA MUESTRA POR PODER

Lo más frecuente es que el efecto de la intervención se desconozca; por lo tanto, el número de sujetos necesarios debe ser tal que nos permita obtener una conclusión estadísticamente válida.

La muestra debe permitir detectar la diferencia deseada. Para ello se determina el poder, que es la probabilidad de detectar una diferencia de determinada magnitud, si esta es verdadera.

El poder estadístico o potencia del estudio representa la probabilidad de observar en la muestra una determinada diferencia o efecto, si existe en la población. Es el complementario del error beta (1-beta).

En un estudio de investigación, el error beta (también llamado error tipo II) aparece cuando el investigador no rechaza la hipótesis nula siendo esta falsa en la población. Es equivalente a la probabilidad de obtener un resultado falso negativo, debido a que el investigador llega a la conclusión de que ha sido incapaz de encontrar una diferencia que en realidad existe. Se acepta en un estudio que el valor del error beta debe estar entre el 5 y 20%.

En sentido estricto, el tipo de error se llama error tipo II; *beta* no es un error, sino la *probabilidad* de cometer ese error (error tipo II).

Si el valor del poder es del 80% (inversa de error beta = $1 - 0,20$), de 100 veces que se realice el estudio, en 80 oportunidades se obtendrá un resultado significativo si la verdadera diferencia fue realmente de la magnitud preespecificada.

El poder de un estudio depende de:

1. El valor de la verdadera diferencia entre los eventos o las medias de los grupos a comparar. Cuando comparamos la incidencia o prevalencia de una enfermedad entre dos grupos, de existir una diferencia, la *verdadera* diferencia nunca puede estimarse; para ello habría que incluir a todos los sujetos de cada grupo con esa enfermedad que existen en el planeta y aun así, cada vez que realicemos una estimación de la enfermedad en cada grupo los valores serán diferentes sólo por azar. Por lo tanto, al realizar una sola vez el estudio, estimamos con cierto grado de certeza cuál sería el valor más probable de esta diferencia. A mayor diferencia, mayor es la probabilidad de detectarla. Si la diferencia de mortalidad entre un grupo y otro es del 50%, por ejemplo, 10% en un grupo y 5% en el otro, la distancia entre ambos valores es grande, por lo que será fácil detectarla. Si se desea detectar una diferencia de 20%, 10% en un grupo vs. 8% en el otro, estos valores están más cerca y se necesitará un gran poder para detectarla. El poder es como una lente: con más aumento es posible ver estructuras más pequeñas y discriminar cuál es cuál. Si las estructuras a observar con una lente (en nuestro caso la diferencia entre grupos) son grandes, se requerirán pocas dioptrías; si las estructuras son muy pequeñas podría requerirse un microscopio electrónico para distinguirlas bien.

2. El tamaño de la muestra

Cuanto mayor sea el número de sujetos mayor será la probabilidad de detectar la diferencia y por lo tanto mayor será el poder. El n de la muestra son las dioptrías de la lente del poder.

3. El nivel de probabilidad al cual la diferencia se considere significativa o error alfa

Como hemos visto, el error alfa influye sobre el poder. Si el área de rechazo o error alfa es pequeña, $p < 0,01$, necesitaremos de un gran poder o una muestra de pacientes grande.

Para calcular el tamaño de la muestra para comparar dos proporciones es necesario especificar:

1. Cuán grande es la verdadera diferencia y cuál es la mínima diferencia que se desea detectar.

No siempre es posible detectar la verdadera diferencia, ya que esta puede ser muy pequeña y se requerirá una enorme cantidad de sujetos. Recordemos que, cuanto más cerca están los valores, la diferencia es más difícil de detectar, por lo que se requiere un gran poder estadístico y un error alfa aceptable, lo cual influye en la cantidad de pacientes requeridos para que estos parámetros sean los correctos y no cometamos errores.

El tamaño de la muestra que seleccionaremos permite detectar diferencias como mínimo del valor especificado. Si la diferencia especificada que se quiere detectar es del 50% (por ejemplo 10% en el grupo control y 5% en el intervención) aunque sospechemos que existe una diferencia verdadera de 30% (10 vs. 7%), el tamaño de la muestra calculado para detectar un 50% de diferencia no nos permitirá detectar diferencias menores a este valor, sólo nos permite detectar diferencias de al menos 50% o más.

La diferencia del 50% contiene al 30%, aunque no podamos corroborar estadísticamente que existe, debido a que se encuentra en el área de superposición. La diferencia estadísticamente significativa que detectaremos siempre estará en el área de NO superposición. (Figura 15.1)

Cuán seguro se desea estar de obtener una diferencia significativa si esta existe. Esto es determinar el valor del error beta o tipo II.

En general los valores más frecuentemente usados están entre el 5 y el 20%, lo que corresponde a valores de Poder de la Muestra del 80 al 95%.

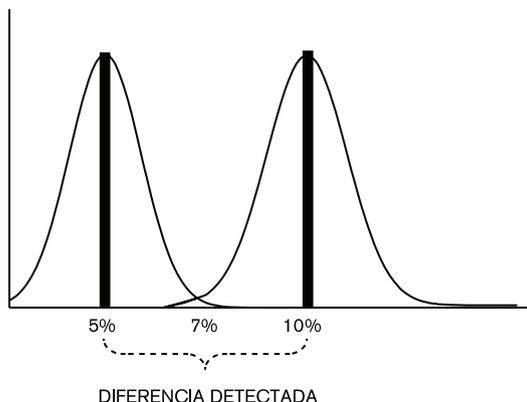


Figura 15.1: Diferencias entre proporciones

2. Por último, hay que determinar cuál es el valor de error alfa que permitiremos en nuestro estudio.

Veamos un ejemplo. En un estudio experimental droga vs. placebo deseamos detectar una diferencia del 30% (10% de mortalidad en el grupo control vs. 7% en el grupo de la droga) con un poder del 95% y error alfa del 5%. Con estos parámetros, el número necesario a reclutar en cada grupo es de 2 243 sujetos.

Como no contamos con presupuesto para realizar este estudio, tenemos varias opciones. Podemos reducir el poder al 80%. Con este cambio necesitaremos 1.356 sujetos en cada grupo.

Como todavía es un número de sujetos muy elevado, decidimos que es útil detectar una diferencia del 40% (10 vs. 6%), con un poder = 95% y alfa = 5%. El número de sujetos ahora es de 1 193 por grupo.

Si reducimos el poder al 80% con alfa 0,05, el $n = 721$ sujetos por grupo. Si bien este es un número de sujetos más factible, no hay que olvidar que estamos sacrificando seguridad en el resultado. Tenemos una probabilidad del 80% de detectar una diferencia del 40%, si esta existe, vs. el 95% de probabilidad que teníamos antes.

Como se puede observar con este ejemplo, la decisión del número de pacientes depende entonces del juicio del investigador: cuál es el monto de diferencia que considera suficientemente útil y cuánto se arriesga a no poder detectarla.

Otro aspecto a tener en cuenta, es que el cálculo de la muestra depende del diseño del estudio y del tipo de variables a comparar.

Si deseamos comparar medias, es necesario conocer el valor del DS de la media del grupo control. En general este dato puede obtenerse de estudios previamente realizados.

AJUSTE POR PÉRDIDA DE SUJETOS

Si deseamos realizar un estudio de cohorte, sabemos que algunos sujetos se perderán durante el seguimiento. Si esto ocurre, al final del estudio no tendremos el número de pacientes necesarios para detectar la diferencia de eventos propuesta: perdimos poder estadístico.

Por lo tanto, el cálculo inicial de la muestra debe contemplar esta pérdida de sujetos e incrementar el tamaño de la muestra para que al final del estudio tengamos suficientes sujetos y por lo tanto, poder estadístico para detectar la diferencia deseada.

Para calcular el tamaño de la muestra, teniendo en cuenta las posibles pérdidas, se utiliza la siguiente fórmula:

$$\text{Muestra ajustada a las pérdidas} = n / (1 - R)$$

n = número de sujetos sin pérdidas

R = proporción esperada de pérdidas

Así, por ejemplo, si en el estudio en el que necesitamos 48 pacientes por grupo prevemos tener un 15% de pérdida de sujetos, el tamaño muestral necesario sería: $48 / (1 - 0,15) = 56$ pacientes en cada grupo.

Como se observa, la muestra se incrementó de forma proporcional a la pérdida esperada de sujetos.

Si bien los programas estadísticos disponibles realizan estos cálculos fácilmente, hay que recordar que una muestra pequeña implica una gran posibilidad de cometer errores. Por eso es mejor consultar con un profesional experto en el tema. Todo el estudio puede verse en peligro si este cálculo se realiza de manera errónea.

CORRELACIÓN Y REGRESIÓN LINEAL

CAPÍTULO 16

Hasta el momento hemos visto la manera de determinar si dos variables cuantitativas son diferentes o no, de acuerdo con el concepto estadístico.

Otro tipo de análisis que muchas veces se requiere con las variables cuantitativas es determinar si están asociadas y si es posible, a partir de una de ellas, predecir el valor de la otra.

Para ello se utilizan dos pruebas estadísticas diferentes.

CORRELACIÓN

Es un método de análisis que se usa para estudiar la relación lineal entre dos variables continuas.

Supongamos que queremos determinar si los valores de colesterol total tienen una relación lineal con los valores de LDL.

Lo primero que hacemos es incluir estos valores en un gráfico *scatter plot*. (Figura 16.1)

Como podemos observar, a medida que los valores de colesterol total aumentan, también lo hacen los de LDL.

Para determinar si esta relación es por azar, se calcula el *coeficiente de correlación de Pearson* (si los datos son de distribución normal).

El valor del coeficiente r para este caso es igual a 0,94 con un IC 95% de 0,90 a 0,96 y $p < 0,0001$, lo que significa que la relación lineal no es debida al azar. Los valores posibles de r están dentro de +1 a -1.

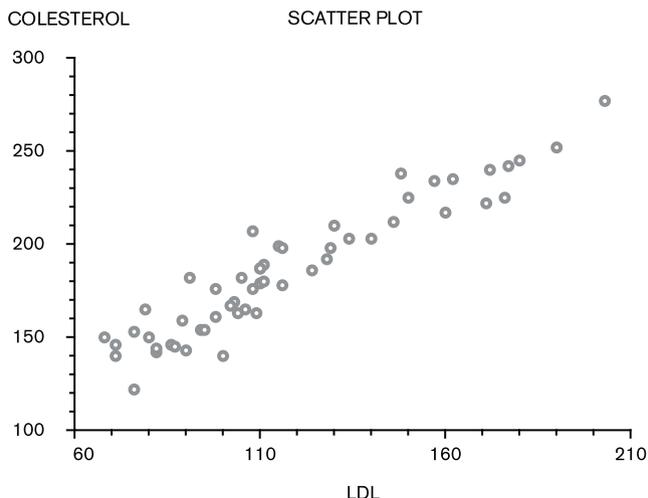


Figura 16.1: Gráfico *scatter plot*. Relación colesterol vs. LDL

Un valor de +1 significa que todos los puntos del *scatter plot* están sobre una recta perfecta. La correlación es positiva si con el aumento de uno de los valores el otro también aumenta, y es negativa si uno de los valores tiende a disminuir con el aumento del otro.

Si el coeficiente de correlación se acerca a cero, significa que no hay asociación lineal.

En el ejemplo, el valor del coeficiente es muy cercano a 1, y visualmente se puede apreciar que los puntos del gráfico están sobre una línea casi recta.

Es importante recalcar que el coeficiente de correlación evalúa si existe una relación lineal. Es posible que dos variables tengan una relación pero no de forma lineal. Pueden describir relaciones de tipo elíptica o curvilínea.

Es por eso que es importante primero observar el gráfico, debido a que un coeficiente bajo ($< 0,70$) no implica una falta total de relación, sino que no hay una relación lineal.

Para que el coeficiente sea válido hay condiciones que deben cumplirse:

- a) Al menos una de las variables debe tener distribución normal.
- b) Los valores deben obtenerse al azar.
- c) Los valores no deben estar apareados.

En el caso de que los valores no tengan una distribución normal, puede usarse el *coeficiente de correlación de Spearman*.

REGRESIÓN LINEAL

Cuando dos variables están asociadas en una relación de tipo lineal es posible, mediante el cálculo de la regresión lineal, obtener la ecuación matemática que describa la recta que las asocia.

En matemática, describir una recta significa determinar dos parámetros. Uno de ellos se denomina intersección (que es el punto donde la recta corta al eje y). El otro parámetro es su pendiente. Con estos dos parámetros se puede dibujar una recta en un eje de ordenadas y abscisas.

Describir la recta permite calcular cuánto aumenta la variable y con el incremento de la variable x , y así calcular el valor de y a partir de cualquier valor de x .

A la variable y se la denomina variable dependiente y a la variable x se la llama variable independiente o explicatoria, porque a partir de sus valores se calculan los valores de y .

Es importante determinar cuál es la variable dependiente y cuál la explicatoria, debido a que, a diferencia de la correlación, el resultado será diferente dependiendo de cuál variable se elija como dependiente.

La ecuación general de la regresión lineal es la siguiente:

$$y = a + bx$$

y es la variable dependiente. Es aquella que deseamos predecir.

a es el parámetro de intersección (llamado *intercept*) o punto donde la recta corta al eje y . Este parámetro nos da el valor de la variable y cuando x vale cero.

b es la pendiente de la recta y nos informa cuántas unidades aumenta la variable y cuando la variable x aumenta una unidad.

x es el valor que toma la variable independiente.

Tomando el ejemplo anterior, supongamos que, a partir de un valor determinado de colesterol total, deseamos calcular el valor de LDL esperado.

Ya hemos comprobado que el colesterol total y la LDL están asociados en una relación lineal por el coeficiente de correlación.

Si calculamos la ecuación de la regresión lineal, obtendremos los parámetros que describen la recta que asocia el colesterol y la LDL.

Como se observa en la Figura 16.2 los puntos que unen colesterol con LDL se hallan sobre una recta que corta al eje y en el valor -52 (si proyectamos la recta hasta cortar el eje y y veremos que lo hace por debajo del valor 60) y que tiene una pendiente cercana a los 45° .

En matemática, una pendiente de 45° equivale a un valor de 1 , debido a que si la pendiente está a 45° y corta el eje de coordenadas en su punto de intersección, la relación entre ambas variables será 1 a 1 . Esto significa que, por cada unidad que aumente la variable x , la variable y también aumentará en una unidad.

Si la pendiente de la recta es igual a cero, significa que no hay pendiente, o que la recta es paralela al eje x , y por lo tanto, por más que aumenten los valores de la variable x no hay aumento en los valores de la variable y . (Figura 16.3)

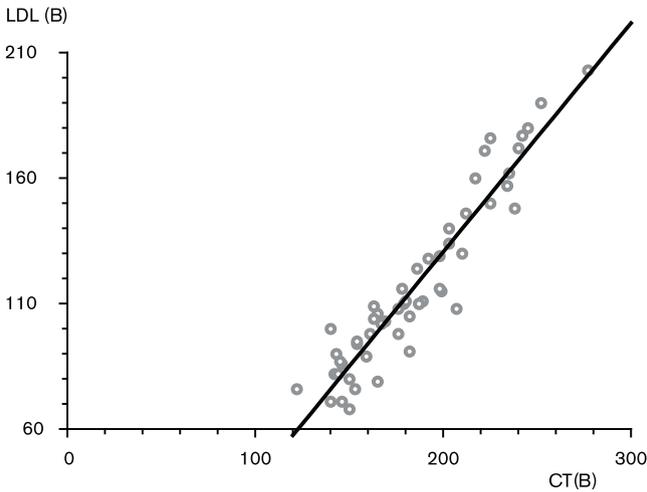


Figura 16.2: Relación LDL vs. colesterol total

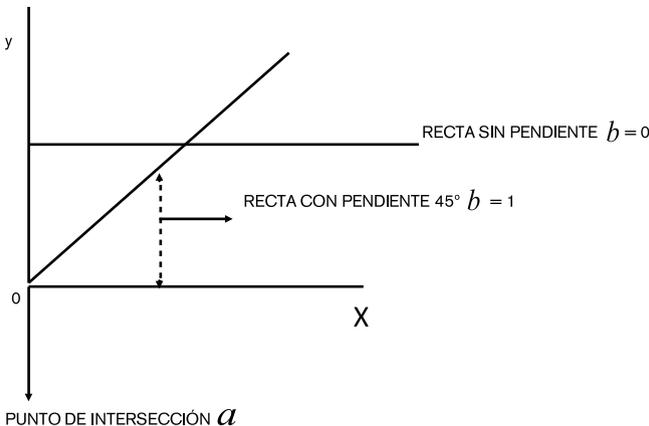


Figura 16.3: Parámetros de una recta

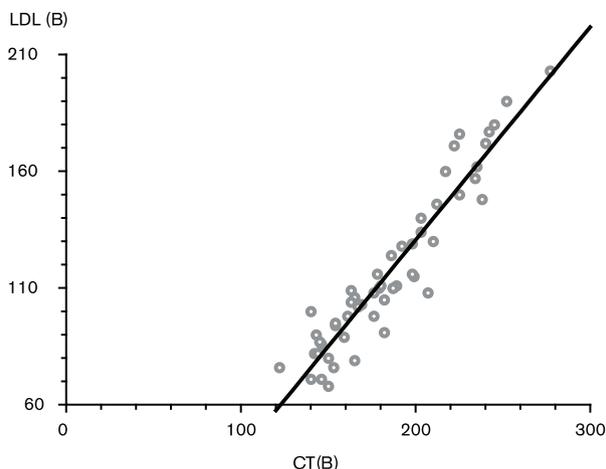


Figura 16.4: Valores de la regresión lineal. Predicción de LDL.

$$\begin{array}{ccccccc}
 \text{Regresión lineal} = & \text{LDL} & = & 0,91 & * & \text{CT} & - 52 & & p < 0.0001 \\
 & \downarrow & & \downarrow & & \downarrow & \downarrow & & \\
 & y & & = b & & * x & - a & &
 \end{array}$$

Si lo que deseamos es predecir el valor de LDL a partir de cualquier valor de colesterol total, utilizamos los parámetros calculados por la regresión lineal.

Entonces, si la pendiente es igual a 0,91 (valor de la pendiente cercano a 1), por cada unidad de aumento del colesterol total, la LDL aumentará 0,91 unidades multiplicado por cualquier valor de colesterol (valor de x) menos 52 (valor de intersección).

Como es lógico, el valor de *intersección* es negativo, debido a que los valores de LDL son menores a los de colesterol total. (Figura 16.4)

La regresión lineal, como prueba estadística, evalúa el valor de la pendiente de la recta.

Asume como hipótesis nula que la recta no tiene pendiente ($b = 0$), es decir que es perpendicular el eje x , con lo cual, por más que el valor de la variable x aumente, no se produce ningún incremento en la variable y .

La regresión lineal tiene que establecer si el valor de la pendiente es estadísticamente diferente a cero y de esta manera obtener la evidencia para rechazar la hipótesis nula.

En otras palabras, tiene que probar que hay pendiente y si la hay, que al incrementar x , aumenta o disminuye y .

El valor de p obtenido en el cálculo de la regresión lineal anterior de LDL es $< 0,001$, con lo que se rechaza la hipótesis nula de no pendiente y se acepta que hay pendiente, en este caso cercana a 45° .

Como toda estimación en la estadística, el valor de la pendiente tiene un IC 95%, en este caso IC 95% de la pendiente = 0,90 a 0,97.

Esto significa que los valores esperables de la pendiente de la recta que asocia colesterol y LDL estarán en el 95% de los casos entre valores que van de 0,90 a 0,97.

En la práctica, si queremos saber qué valor de LDL tendríamos si el paciente tiene un valor de colesterol total de 300 mg/dL, aplicamos la fórmula obtenida:

$$\text{LDL} = 300 * 0,91 - 52 = 221 \text{ mg/dL}$$

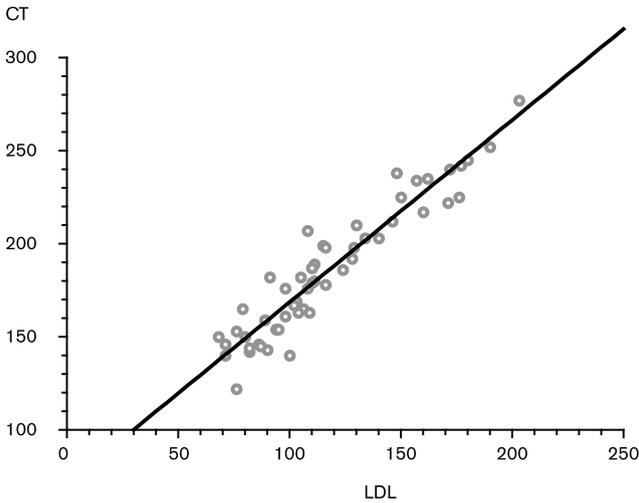


Figura 16.5: Regresión lineal. Predicción de colesterol total
 Colesterol total = $0,98 * LDL + 70,7$ $p < 0,001$

Sólo con fines ilustrativos, veamos qué pasa si lo que queremos predecir esta vez es el valor de colesterol total.

El resultado de la regresión lineal se observa en la Figura 16.5.

En este caso, como es esperable, el valor de a es positivo ($+70,7$), dado que los valores de colesterol total son mayores a los de LDL.

El valor de b $0,98$ significa que, además de ser cercano a 1 , por cada incremento en una unidad de LDL el colesterol aumenta $0,98$ mg/dL.

Aquí también vemos que la pendiente es de casi 45° (valor de b $0,98$) con $p < 0,001$, por lo que confirmamos que la pendiente es diferente a cero (hay pendiente, la recta no es horizontal). (Figura 16.5)

En resumen, la prueba de correlación determina si dos variables tienen relación lineal y la regresión lineal establece el grado de asociación de estas variables.

DATOS CUALITATIVOS. MEDIDAS DE RESUMEN



MEDIDA DE RESUMEN NUMÉRICA

Las variables categóricas son las más fáciles de resumir. Habitualmente reportamos porcentajes de sujetos en cada categoría; por ejemplo, porcentaje de sujetos con infarto previo, porcentaje de sujetos con insuficiencia renal, etc.

Generalmente, si estos datos se reportan como porcentajes, se presentan en forma de tabla. (Tabla 17.1)

Variable	Grupo 1 n = 383 %	Grupo 2 n = 1 708 %
Sexo femenino	22,7	16,7
Cirugía no electiva	40,9	35,5
Cirugía previa	4,3	4,1
Infarto de miocardio previo	29,7	32,1
Diabetes	4,1	2,8
EPOC	10,8	9,5
Hipertensión arterial	63,5	66,5
Edad (media \pm DS)	64,9 \pm 10	62,5 \pm 9,7
Días de internación (mediana)	5	6

Tabla 17.1: Resumen datos categóricos

Si miramos la Tabla 17.1, veremos que también los datos cuantitativos pueden informarse junto con los datos categóricos. Lo que se hace es aclarar que se está informando media \pm DS o mediana y rango intercuartil.

RESUMEN POR MÉTODOS GRÁFICOS

Cuando deseamos presentar los datos categóricos de forma gráfica, recurrimos al diagrama de barras o circular. (Figuras 17.1 y 17.2)

En el caso de los gráficos, es importante que en el de barras el eje **y** comience en cero; de otra manera se puede dar una falsa impresión acerca de la magnitud de las proporciones y la suma de los porcentajes que cada barra representa no necesariamente debe sumar 100%.

En el gráfico circular, recuérdese que este representa el 100% de la muestra, por lo tanto sus partes deben sumar 100%, de lo contrario estamos cometiendo un error. (Figuras 17.1 y 17.2)

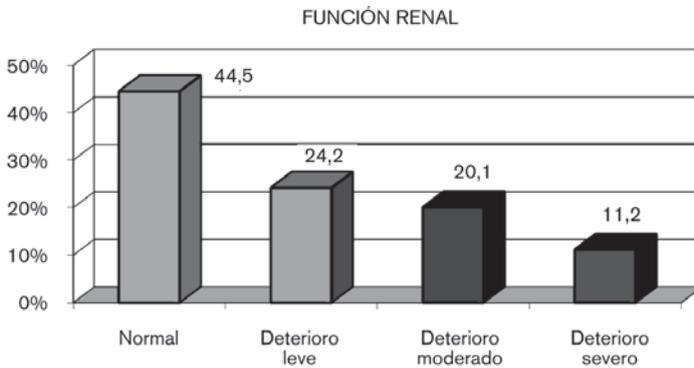


Figura 17.1: Gráfico de barras. Datos categóricos



Figura 17.2: Gráfico circular. Datos categóricos

VARIABILIDAD DE DATOS CATEGÓRICOS

Los porcentajes, como medida de resumen de datos categóricos, también muestran variabilidad.

Si calculamos el porcentaje de diabéticos tratados con angioplastia en varias muestras, veremos que el porcentaje va a variar de muestra en muestra.

Por lo tanto, los porcentajes o proporciones también tienen su DS y ES.

Se puede calcular el DS de una proporción como:

$$SD(p) = \sqrt{p(1-p)}$$

p = proporción

Y su ES como DS dividido la raíz cuadrada de n , como para los datos continuos:

$$ES(p) = \sqrt{\frac{p(1-p)}{n}}$$

Recordar que las proporciones se escriben con decimales.

Por ejemplo, si el porcentaje de diabéticos tratados con angioplastia en una muestra de 100 pacientes es del 15%, la proporción es 0,15.

Su DS será:

$$SD = \sqrt{0,15(1-0,15)} = 0,35$$

$$ES = \sqrt{\frac{0,15(1-0,15)}{100}} = 0,035$$

Si tenemos el valor de ES podemos construir el IC 95% para la proporción.

En este caso IC 95% es: 0,03 a 0,22

Este intervalo tiene el mismo significado que ya hemos visto: de cada 100 veces que realicemos este estudio, en 95 oportunidades encontraremos que el porcentaje de diabéticos está entre el 3 y el 22%.

DATOS CUALITATIVOS. DISTRIBUCIÓN BINOMIAL

CAPÍTULO 18

Los datos categóricos también muestran una distribución de probabilidades como los datos cuantitativos.

Supongamos que tiramos 10 veces una moneda y de 10 tiros en 2 obtenemos cara. En este caso obtuvimos un 20% de caras. Repetimos el experimento y obtenemos 4 caras (40%), y así sucesivamente hasta hacerlo 100 veces.

Cada 10 veces que tiramos la moneda anotamos el resultado. Al final tendremos 100 porcentajes.

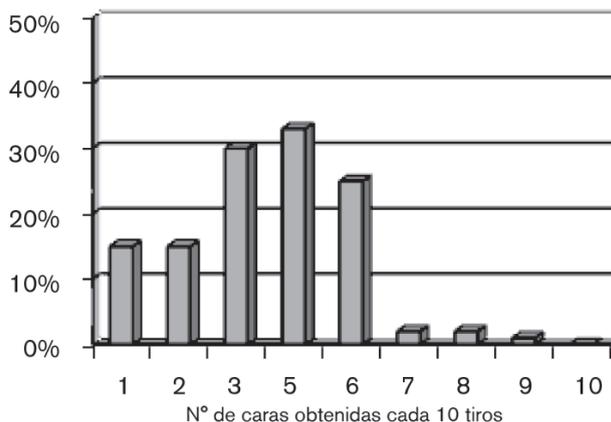


Figura 18.1: Porcentaje de número de caras en 100 experimentos de 10 tiros realizados al azar

Con los resultados obtenidos de estos experimentos realizados completamente al azar podemos graficar el número de veces que obtuvimos cara o el porcentaje de veces que obtuvimos 1, 2, 3 o más caras en las 100 veces que hicimos el experimento. (Figura 18.1)

Como vemos en la Figura 18.1, de las 100 veces que realizamos el experimento por azar, obtuvimos diferentes porcentajes en cada serie de 10 tiros, pero en la mayoría de los casos (30% de las veces) obtuvimos 5 veces cara cuando tiramos 10 veces.

Sabemos que la *verdadera* probabilidad de obtener cara cuando se tira una moneda es de 0,50, o del 50%, ya que sólo existe 1 posibilidad en 2 de obtener este resultado.

A medida que aumentamos el número de experimentos, por ejemplo de 100 a 1 000, nos acercaremos cada vez más a esa *verdadera* proporción o probabilidad.

Si graficamos el porcentaje de veces que obtenemos cara, esta vez en 100 tiros sucesivos realizados 1 000 veces, veremos que la distribución de probabilidades se acerca cada vez más a la verdadera proporción y además toma forma simétrica, muy similar a la distribución normal. (Figura 18.2)

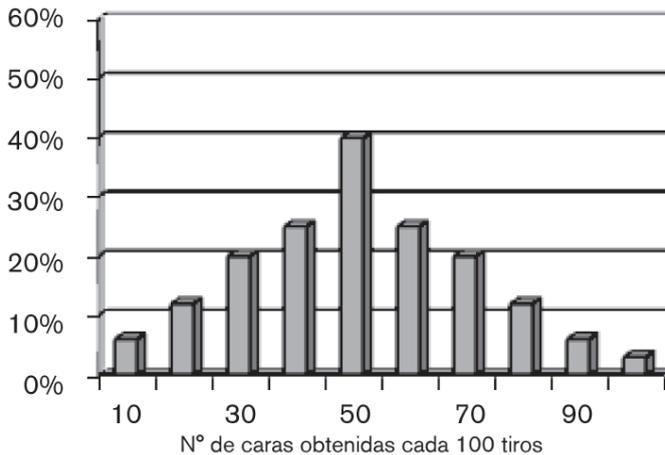


Figura 18.2: Frecuencia de éxitos para 1 000 veces 100 tiros

En la Figura 18.2 se observa que en la mayoría de los casos (serie de 100 tiros sucesivos) se obtuvo la verdadera probabilidad, en este caso 50 veces cara en 100 tiros.

Por lo tanto, para las proporciones también hay una distribución de probabilidades, que se conoce como *binomial*.

Se dice que un experimento aleatorio es binomial cuando:

- 1) Consiste en n repeticiones independientes del mismo ensayo.
- 2) Cada ensayo puede resultar en sólo dos resultados que llamaremos éxito y fracaso.
- 3) La probabilidad de éxito se mantiene constante a través de los n ensayos.

Dicho en otras palabras, el experimento binomial es aplicable a una muestra independiente de n sujetos en los cuales se mide una variable dicotómica cuya probabilidad no cambia en el tiempo.

Esta distribución también tiene parámetros, conocidos como n y p , siendo n el número de sujetos de la muestra y p la verdadera proporción.

La notación matemática para referirse a la distribución binomial es:

Bi (n, p)

Como en la distribución normal, si conocemos n y p podemos construir un gráfico como el de las Figuras 18.1 y 18.2, y aplicar los conceptos de área bajo la curva para calcular probabilidades.

Por ejemplo, si tenemos una caja con 10 bolitas de las cuales 1 es roja y 9 blancas, el porcentaje de bolitas rojas es de 10% o bien $p = 0,10$.

Si sacamos 20 veces una bolita de esta caja (cada vez que sacamos una la volvemos a poner en la caja) tendremos $n = 20$ y $p = 0,10$.

Conociendo estos dos parámetros podemos calcular, en base a la distribución binomial, la probabilidad de ocurrencia de distintos eventos, por ejemplo, la probabilidad de no obtener ninguna bolita roja en 20 extracciones, la probabilidad de obtener 5 bolitas rojas o la probabilidad de obtener todas bolitas rojas.

Estos cálculos son realizados por los programas estadísticos basados en la distribución binomial, que ya está tabulada para cada uno de sus parámetros.

En la vida real, obviamente, no conocemos la verdadera proporción que queremos estimar, no conocemos el número total de bolitas ni cuántas bolitas rojas hay en la caja.

Por lo tanto, lo que hacemos es calcular la probabilidad de obtener un valor determinado asumiendo valores de n y p .

Por ejemplo, si el porcentaje de diabéticos es del 15% en una muestra de 100 sujetos, y el porcentaje reportado en la literatura para este grupo de pacientes es del 10% en 1 000 sujetos, la pregunta es: ¿cuál es la probabilidad de obtener 0,15 cuando $n = 100$ y $p = 0,10$? Asumimos que el verdadero valor de n es el reportado por la literatura (10%).

Para contestar esta pregunta podemos recurrir a las ya conocidas pruebas de significación estadística.

Como se observa en la Figura 18.2, cuando el n de la muestra es suficientemente grande, la distribución binomial se asemeja a la normal.

Cuando $n * p$ Y $n - n * p$ es ≥ 5 , se puede utilizar la distribución normal en lugar de la binomial.

Por ejemplo, si tenemos una muestra de 10 sujetos con un porcentaje de hipertensos de 20% entonces $n * p = 10 * 0,20 = 2$ y $n - n * p = 10 - 2 = 8$

En este caso, como no se cumplen las dos condiciones, debemos utilizar la distribución binomial.

Si aumentamos n , por ejemplo, a 50, será posible entonces utilizar la distribución normal.

$$n * p = 50 * 0,20 = 10 \text{ y } n - n * p = 50 - 10 = 40$$

Ahora, si se cumplen las dos condiciones, utilizamos la distribución normal.

A medida que n aumenta y/o p se acerca a 0,50, la distribución binomial se va asemejando a la normal. (Figura 18-3)

En la Figura 18.3 se observa que, cuando p es igual a 0,50, aunque n sea pequeño, la distribución toma forma simétrica, de la misma forma que, aunque p no sea 0,50, cuando n es grande también se torna simétrica, acercándose a la distribución normal.

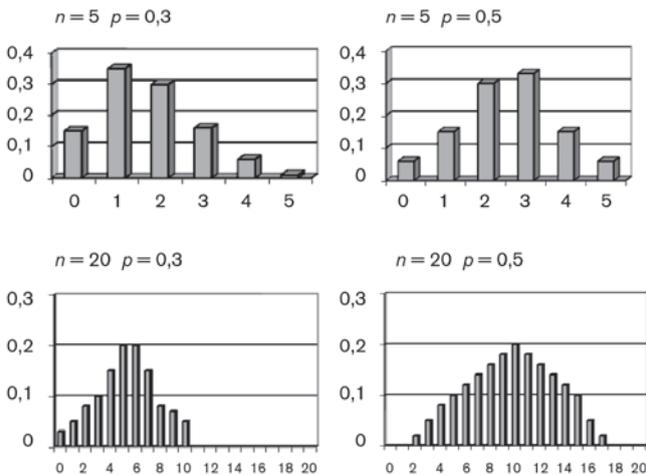


Figura 18.3: Cambio de distribución binomial a normal

DATOS CUALITATIVOS.

PRUEBAS DE SIGNIFICACIÓN ESTADÍSTICA PARA PROPORCIONES



Debido a que la distribución binomial en ciertas condiciones ($n * p$ Y $n - n * p \geq 5$) se asemeja a la normal, es posible entonces aplicar los mismos conceptos que hemos visto para las variables cuantitativas, siempre y cuando estemos trabajando con variables dicotómicas. (Véase *Análisis de datos cuantitativos*)

PRUEBA PARA UNA PROPORCIÓN

Supongamos que hemos observado que el porcentaje de infartos de miocardio al egreso hospitalario en el último año fue del 20% ($n = 1\ 000$ pacientes). Datos del año previo indican un 15% de egresos hospitalarios con diagnóstico de infarto. Nos preguntamos ¿ha aumentado la incidencia de infartos?

Para responder esta pregunta, podemos recurrir a la prueba de hipótesis.

La H_0 propone que $p_1 = p_0$, en este caso 20% = 15%.

La hipótesis nula dice que 20% es una variación encontrada al azar.

La H_1 sería $p_1 \neq p_0 = 20\% \neq 15\%$

Debido a que la muestra es grande, podemos usar la distribución normal para calcular z y de esta forma determinar a cuantos ES está un porcentaje de otro.

Es como asumir que los porcentajes son medias.

Para las proporciones z es igual a:

$$z = \frac{p_1 - p_0}{SE(p_1)}$$

Tenemos entonces:

$$p_0 = 0,15 \text{ (proporción correspondiente a 15\%)}$$

$$p_1 = 0,20 \text{ (proporción correspondiente a 20\%)}$$

$$ES\ p_1 = \sqrt{\frac{P_1(1-P_2)}{n}} = \sqrt{\frac{0,20(1-0,20)}{1\ 000}} = 0,012$$

$$z = \frac{0,20 - 0,15}{0,012} = 3,9\ ES$$

Buscamos ahora el valor de z en la tabla de distribución y encontramos que el valor correspondiente de p es 0,000.

Por lo tanto concluimos que, si la proporción de infartos es de 15% (hipótesis nula), la chance de obtener un 20% es < 1 en 1 000. Lo cual es muy poco probable.

Decimos entonces que existe diferencia estadísticamente significativa y que la incidencia de infartos ha aumentado.

Como esta muestra puede tomarse como normal, podemos construir el intervalo de confianza del 95% para 20%.

Este será en general: $p \pm 1,96 * ES$

Para nuestro caso: $0,20 \pm 1,96 * 0,012 = 0,20 \pm 0,023$

Límite inferior IC 95% = 0,17

Límite superior IC 95% = 0,22

Por lo tanto, el IC 95% para 0,20 es 0,17 a 0,22 o, lo que es lo mismo, el IC 95% de 20% es 17-22%.

Como ya hemos visto antes, si el IC 95% no contiene al valor propuesto por la H_0 ya sabemos que hay diferencia estadísticamente significativa.

En este caso el IC 95% no contiene 15%.

PRUEBA PARA DOS PROPORCIONES

En este caso tenemos dos muestras, un grupo A de 1 000 pacientes al que hemos tratado con una droga y que muestra un porcentaje de infarto del 8% y otro grupo B, de 1 000 pacientes al que le administramos placebo con un porcentaje de infarto de 12%.

La pregunta es: los resultados hallados ¿son variaciones al azar o existe diferencia estadísticamente significativa?

Aquí también planteamos la $H_0: p_1 = p_2$

Para nuestro ejemplo $H_0: 0,08 = 0,12$

Matemáticamente se comprueba cómo $p_1 - p_2 = 0$, es el mismo caso que hemos visto para dos medias: si estas son iguales, la diferencia será cero.

La $H_1: p_1 \neq p_2$ es matemáticamente $p_1 - p_2 \neq 0$

Para el ejemplo, la pregunta en términos matemáticos sería: $0,08 - 0,12 = -0,04$, ¿es diferente a cero?

Para determinar si la diferencia entre estas proporciones es distinta a cero debemos establecer a cuántos errores estándar (ES) está $-0,04$ de cero.

El cálculo de z para dos proporciones es:

$$z = \frac{p_1 - p_2}{ES_{comin}(p_1 - p_2)}$$

Cuando calculamos z para dos proporciones lo que hacemos es transformar la diferencia en ES, $-0,04$ ¿a cuántos ES corresponden?

El ES común para dos proporciones es:

$$ES = \sqrt{p(1-p) \times \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

p es una única proporción que se calcula como:

$$p = \frac{r_1 + r_2}{n_1 + n_2}$$

$$p = \frac{80+120}{1000+1000} = \frac{200}{2000} = 0,1$$

Donde $r_1 + r_2$ es la suma del número de sujetos con la característica a estudiar, en este caso 80 para el grupo A y 120 para el grupo B, y $n_1 + n_2$ es la suma del número de sujetos en cada grupo.

$$\text{Entonces el ES común será} = \sqrt{0,1(1-0,1) \times \left(\frac{1}{1000} + \frac{1}{1000} \right)} = 0,013$$

Y

$$z = \frac{0,08 - 0,12}{0,013} = -2,98$$

Para $-2,98$, el valor de p es igual a $0,0029$

En conclusión, la diferencia entre $0,08$ y $0,12$ ($-0,04$) es estadísticamente diferente de cero.

Para el estudio, la conclusión es que el grupo A tuvo significativamente menos infartos que el grupo B; estas proporciones son estadísticamente diferentes.

Como siempre en Estadística, es posible determinar el IC 95% para la diferencia de proporciones. Si este intervalo contiene al cero, concluiremos que la diferencia de proporciones no es diferente a cero y por lo tanto las proporciones serán iguales, y aceptamos la H_0 .

IC 95% de la diferencia entre proporciones es: $(p_1 - p_2) \pm 1,96 \times \text{SE}(p_1 - p_2)$

Donde el ES que se utiliza es distinto al utilizado para z .

$$\text{ES} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$\text{Para el ejemplo el ES sería: } \sqrt{\frac{0,08(1-0,08)}{1000} + \frac{0,12(1-0,12)}{1000}} = 0,0133$$

El IC 95% de la diferencia = $-0,04 \pm 1,96 * 0,0133$

IC 95% = $-0,013$ a $-0,06$

Este intervalo, como esperábamos, no contiene al cero, por lo que ya sabemos que la diferencia de proporciones, el 95% de las veces, va a mostrar valores entre $-0,013$ y $-0,06$.

El análisis de datos categóricos mediante la prueba de significación, usando proporciones, es útil cuando los datos son dicotómicos y se quiere comparar dos porcentajes.

La manera más usual de analizar frecuencias (número de sujetos con una característica determinada del total de sujetos) o porcentajes es a través de tablas, conocidas como *tablas de contingencia*. En estas, los datos se presentan en columnas y filas que definen las categorías de las variables.

Dependiendo de cuántas categorías existan en cada variable, el número de filas y columnas se adaptarán a ellas.

En términos generales, se habla de tabla $r \times c$ (del inglés *row* y *column*). De acuerdo con esta terminología, tendremos por ejemplo tablas 2×2 (2 filas y 2 columnas), $2 \times c$ (2 filas y más de 2 columnas) y $r \times c$ (más de 2 filas y más de 2 columnas).

Variable 1 Variable 2	Grupo 1	Grupo 2	Grupo 3	Total
Sí	5 (11%)	15 (33%)	25 (55%)	45 (100%)
No	95 (37%)	85 (33%)	75 (29%)	255 (100%)
Total	100 (33%)	100 (33%)	100 (33%)	300 (100%)

Tabla 20.1: Modelo de tabla de contingencia

Cada celda de la tabla corresponde a una combinación particular de características que relacionan dos o más categorías de las variables que se analizan. (Tabla 20.1)

En este ejemplo, la variable 1 podría ser tabaquismo, clasificado como nunca fumó, ex fumador y fumador actual.

La variable 2 podría corresponder a la presencia de EPOC, en este caso como variable dicotómica (Sí / No).

Las celdas muestran la frecuencia (número de sujetos) y/o el porcentaje correspondiente a cada combinación de las categorías de las dos variables. En este caso se calcularon los porcentajes de las filas, porcentaje de EPOC en cada categoría de tabaquismo.

Los porcentajes se calculan de acuerdo con la pregunta que se quiera contestar; en este caso la pregunta es: del total de pacientes con EPOC ¿cuántos corresponden a tabaquistas actuales, cuántos a ex fumadores y cuántos a aquellos que nunca fumaron?

Si la pregunta fuera ¿cuál es el porcentaje de EPOC en cada categoría de tabaquismo?, se tendrían que calcular los porcentajes de las columnas.

Para el grupo 1 es 5%, grupo 2 = 15% y grupo 3 = 25%.

Para contestar estadísticamente estas preguntas, debemos analizar las frecuencias observadas en cada celda y aquellas que *esperaríamos* encontrar bajo la hipótesis nula.

Para el caso de las tablas de contingencia, el supuesto del que partimos o hipótesis nula es que las 2 variables no están relacionadas; dicho de otra manera, son independientes o no están asociadas.

Esto significa que la probabilidad de pertenecer a alguna celda en particular no depende de las categorías de las variables que se están comparando.

Usando el ejemplo, el porcentaje de EPOC en cada celda no depende del estatus de tabaquismo sino simplemente de la ley de probabilidades o depende del azar.

Basados en la ley de probabilidades para proporciones, es posible calcular cuál sería la probabilidad esperada por azar de obtener los valores observados.

Este cálculo se realiza con las proporciones observadas.

Por ejemplo, hay 45 pacientes con EPOC de 300 sujetos, proporción = 0,15 y 100 pacientes que no fumaron nunca de 300 pacientes en total, proporción = 0,33.

Si tabaquismo y EPOC son independientes la proporción en la muestra total de EPOC y no fumadores sería $0,15 \times 0,33 = 0,0495$ o 4,95%.

Si observamos la Tabla 20.1, veremos que el porcentaje de pacientes que nunca fumaron y tienen EPOC es 1,66% (5 pacientes / 300 pacientes total).

Debemos ahora decidir si las proporciones observadas son estadísticamente diferentes a las esperadas por ley de probabilidad. Esto es, decidir con evidencia estadística si las proporciones observadas son sólo una variación esperable de aquellas que surgirían por azar. ¿Cuán posible es obtener 1,66% cuando lo esperado para esta muestra es 4,95%?

En términos de hipótesis nula decimos que las proporciones observadas se deben sólo al azar, mientras que la hipótesis alternativa sostiene que no se deben al azar sino que las proporciones observadas están en relación con las categorías de las variables comparadas.

El porcentaje de EPOC observado está en relación con el estatus de tabaquismo, es el tabaquismo el que determina el porcentaje de EPOC.

De esta manera decimos que estas variables están asociadas, porque la categoría de la variable asumida como posible factor de riesgo determina el porcentaje del evento observado.

La prueba estadística que debemos aplicar para comparar lo *observado* vs. *esperado* es la conocida prueba de chi cuadrado (χ^2 o Chi^2).

El resultado de esta prueba nos ofrece la evidencia necesaria para aceptar o rechazar la hipótesis nula que, como en todos los casos de la estadística, es lo que debemos verificar.

La prueba de chi cuadrado tiene diferentes *formas matemáticas* de acuerdo con el caso en particular que estemos tratando.

Esto depende de:

- El número total de sujetos de la muestra
- El número de categorías de cada variable
- Si estas categorías están ordenadas o no
- Si las variables están apareadas

Entonces, el tipo de prueba de chi cuadrado a utilizar dependerá de la combinación de estos puntos.

A continuación veremos los diferentes casos para la aplicación correcta de la prueba de chi cuadrado.

PRUEBA DE CHI CUADRADO



PRUEBA DE CHI CUADRADO PARA DOS PROPORCIONES

En el caso más simple de comparación de dos proporciones se relacionan dos variables, cada una con dos categorías (dicotómicas) usando la tabla de contingencia 2 x 2.

Supongamos que tenemos una muestra de 2 000 sujetos; un grupo de 1 000 sujetos recibe una droga y el otro placebo. El evento de interés es infarto de miocardio. (Tabla 21.1)

Las preguntas clínicas aquí son: ¿la droga disminuye la incidencia de infarto? ¿La eficacia de la droga es mayor al placebo?

En términos estadísticos la pregunta es: ¿las proporciones observadas de infarto están relacionadas con el tipo de tratamiento asignado (droga o placebo) o sólo es producto del azar? ¿8% de mortalidad en el grupo tratamiento es igual a 12% en el grupo placebo?

La prueba de chi cuadrado se usa para verificar si hay *asociación* entre ambas variables (tratamiento e infarto) y por lo tanto si los porcentajes obtenidos son estadísticamente diferentes.

El *análisis por prueba de significación de dos proporciones* estudia si los porcentajes son estadísticamente diferentes; no establece asociación entre variables.

Establecer asociación no necesariamente implica establecer la causa, para ello debemos contar con otro tipo de evidencia que veremos más adelante.

Infarto	Grupo tratamiento	Grupo placebo	Total
Sí	80 (8%)	120 (12%)	200
No	920 (92%)	880 (88%)	1 800
Total	1 000	1 000	2 000

Tabla 21.1: Proporción de infarto y grupo de tratamiento

Para decidir si hay suficiente evidencia para asegurar que la droga fue efectiva o si los porcentajes encontrados fueron sólo al azar usamos la *prueba de chi cuadrado*.

Bajo la hipótesis nula, de no asociación entre droga e infarto, podríamos pensar que tanto la droga como el placebo tienen la misma eficacia. Por lo tanto, del total de infartos, la mitad correspondería al grupo tratamiento y la otra mitad al grupo placebo.

Si las proporciones o el número de sujetos observados fueran por azar, podemos calcular cuáles podrían ser.

Para el cálculo del número de sujetos *esperados* en cada celda se multiplica el total de la fila por el total de la columna correspondiente a cada celda y se divide por el total de pacientes en el estudio. (Tabla 21.2)

Para el grupo tratamiento, el número esperado de individuos con infarto es:
 $(200 \times 1\,000) / 2\,000 = 100$ pacientes

Infarto	Grupo tratamiento	Grupo placebo	Total
Si	100 (50%)	100 (50%)	200
No	900 (50%)	900 (50%)	1 800
Total	1 000	1 000	2 000

Tabla 21.2: Número de sujetos *esperados*

Para el grupo tratamiento, el número esperado de individuos sin infarto es:
 $(1\ 800 \times 1\ 000) / 2\ 000 = 900$ pacientes

El mismo cálculo se realiza para cada celda del grupo placebo.

También podríamos calcular cuál es el porcentaje de infartos esperados en el grupo tratamiento y en el grupo placebo. Como vemos en la Tabla 21.3, los porcentajes esperados son iguales. Esto es debido a que el test asume como hipótesis nula, no asociación y por lo tanto no diferencia entre los porcentajes de infarto entre los grupos tratamiento y placebo.

En términos generales, las celdas de la tabla 2 x 2 se designan con letras.

Infarto	Grupo tratamiento	Grupo placebo	Total
Si	100 (10%)	100 (10%)	200
No	900 (90%)	900 (90%)	1 800
Total	1 000	1 000	2 000

Tabla 21.3: Número de porcentajes *esperados* de infarto según tratamiento o placebo

Variable	Grupo 1	Grupo 2	Total
Si	a	b	a + b
No	c	d	c + d
Total	a + c	b + d	n

Tabla 21.4: Tabla de contingencia. Notación general

Entonces, el valor esperado de número de sujetos para cada celda es:

Para la celda a = $(a + b) \times (a + c) / n$

Para la celda b = $(a + b) \times (b + d) / n$

Para la celda c = $(c + d) \times (a + c) / n$

Para la celda d = $(c + d) \times (b + d) / n$

Para que la prueba de chi cuadrado sea válida debe cumplir con una regla: el 80% de las celdas deben tener un valor esperado igual o mayor que 5. En nuestro ejemplo, ninguna de las celdas tiene valores esperados menores que 5.

Las preguntas que contesta la prueba de chi cuadrado son: ¿las proporciones observadas son estadísticamente diferentes a las esperadas? ¿Hay una asociación entre la droga y la tasa de eventos?

Matemáticamente la prueba de chi cuadrado es igual a:

$$X^2 = \sum \frac{(O-E)^2}{E}$$

$O = n^\circ$ de sujetos observados
 $E = n^\circ$ de sujetos esperados

La prueba utiliza los valores observados y esperados de cada celda, por eso se emplea el símbolo \sum (sumatoria). La diferencia entre O y E está elevada al cuadrado para que los valores finales sean siempre positivos.

Para la prueba de chi cuadrado, existe una distribución de probabilidades que le pertenece. Al calcular el valor de X^2 estamos buscando en su tabla de distribución a qué valor de p corresponde y entonces podemos determinar si las variables analizadas están asociadas o bien si son variaciones al azar.

Como las frecuencias esperadas en cada celda son el producto del total de la columna y la fila de interés dividido por el número total de la muestra, la prueba de chi cuadrado está condicionada por estos totales. Por lo tanto, de acuerdo con el número de columnas y de filas que tenga, la tabla de la distribución de chi cuadrado se modifica.

Es por eso que la prueba de chi cuadrado utiliza también grados de libertad para determinar su valor; en este caso los grados de libertad corresponden al número de filas y columnas de la tabla.

El cálculo de los grados de libertad para las tablas de contingencia es igual a:

$$(r - 1) \times (C - 1)$$

Para este caso de tabla 2 x 2, el grado de libertad es igual a 1.

$$(r - 1) \times (C - 1) = (2 - 1) \times (2 - 1) = 1$$

Una vez hallado el valor de X^2 se busca en la tabla de distribución de chi cuadrado, de acuerdo con los grados de libertad, el valor de p . (Ver tabla distribución X^2 al final del capítulo)

Para nuestro ejemplo, el valor de chi cuadrado es:

$$X^2 = \sum \frac{(80-100)^2}{100} + \frac{(120-100)^2}{100} + \frac{(920-900)^2}{900} + \frac{(880-900)^2}{900} = 8,88$$

Para $X^2 = 8,88$ con un grado de libertad, el valor de p no se encuentra exactamente para este valor de X^2 , el valor más cercano de X^2 es 10,8 que corresponde a un valor $p = 0,001$.

En la práctica, los programas estadísticos calculan el valor de X^2 y valor de p exacto.

Para el ejemplo, el resultado indica que el número de sujetos con infarto (o el porcentaje) observado es diferente al esperado; la distribución de sujetos no fue por azar o por ley de probabilidades, sino que se debió al grupo de tratamiento al que pertenecían. El valor de p nos da la evidencia necesaria para rechazar la hipótesis nula de no asociación.

Concluimos entonces que el tratamiento es más efectivo que el placebo y por lo tanto existe *asociación* entre droga y tasa de infarto.

A este tipo de prueba de X^2 se lo denomina *chi cuadrado simple*.

Este caso, como hemos visto, podría analizarse usando la prueba de significación para 2 proporciones calculando Z, debido a que la muestra es > 100 . Si hacemos este cálculo vemos que el valor de chi cuadrado hallado corresponde a Z al cuadrado.

$$Z = \frac{0,08 - 0,12}{0,013} = -2,98$$

Verificamos que $2,98^2$ es igual a 8,88 que fue el valor de X^2 hallado.

PRUEBA DE CHI CUADRADO PARA MUESTRAS PEQUEÑAS

Cuando la muestra tiene pocos sujetos o algunas celdas quedan con pocos casos, la distribución de chi cuadrado no se calcula bien. En general tiende a arrojar valores de X^2 grandes que se corresponden con valores de p pequeños.

Para corregir este problema se introduce un cambio en la formula de X^2 .

Al término $O - E$ se le resta $\frac{1}{2}$ y se reemplaza por $|O - E| - \frac{1}{2}$.

A esta modificación se la conoce como *corrección por continuidad de Yates*.

Es muy frecuente que los programas estadísticos informen el valor de p no corregido y el corregido por Yates.

En los casos de muestras grandes en las cuales algunas celdas tienen pocos pacientes, suele verse que algunos de estos valores de p son menores a 0,05 y otros no, entre ellos el corregido por Yates.

Infarto	Grupo tratamiento	Grupo placebo	Total
Sí	137	10	147
No	2 838	104	2 942
Total	2 975	114	3 089

Tabla 21.4: Tabla 2 x 2 para muestra con pocos sujetos en algunas celdas

Este es el valor que debe elegirse para informar el valor de p , debido a que podemos estar asumiendo una asociación cuando en realidad no la hay. (Tabla 21.4)

Chi² no corregido = 4,206039 $p = 0,0403$

Chi² Yates = 3,336914 $p = 0,0677$

Como puede verse en el ejemplo, aun con muestras grandes puede pasar que alguna celda quede con pocos sujetos (grupo placebo con infarto 10 sujetos), y el valor de Chi² corregido y no corregido serán diferentes: cuando se corrija el valor de p resultará mayor a 0,05 mientras que sin la corrección se asumirá una p significativa y una asociación entre las variables.

La fórmula completa para una tabla 2 x 2 se muestra a continuación sólo con fines informativos.

$$X^2 = \frac{\left(|ad - bc| - \frac{1}{2} N \right)^2 \times N}{(a + b)(c + d)(a + c)(b + d)} \text{ grados de libertad} = 1$$

N = número total de la muestra

Como ya mencionamos, para que la prueba de chi cuadrado sea válida, debe cumplir con una regla: el 80% de las celdas deben tener un valor esperado igual o mayor a 5.

En una tabla 2 x 2 se requeriría que esto se cumpla en todas las celdas. Cuando esto no sea posible se utilizará la llamada *prueba exacta de Fisher*.

Esta prueba también se basa en las frecuencias observadas y esperadas, pero su cálculo matemático es mucho más complicado.

El valor de p calculado con la prueba de Fisher es más parecido al valor de p calculado con corrección de Yates. (Tabla 21.5)

Infarto	Grupo tratamiento Observado	Grupo tratamiento Esperado	Grupo placebo Observado	Grupo placebo Esperado	Total Observado
Sí	1 a	3,3	5 b	2,6	6
No	8 c	5,6	2 d	4,3	10
Total	9		7		16

Tabla 21.5: Tabla 2 x 2 para pocos sujetos valores Observados y Esperados

Como se observa en la Tabla 21.5, los valores esperados para las celdas a, b y d son menores a 5 y por lo tanto debemos usar la prueba de Fisher.

Si calculamos el valor de p por prueba de chi cuadrado sin corrección es muy probable que estemos cometiendo un gran error.

En este caso, el valor de p por chi cuadrado sin corrección es de 0,01, mientras que el corregido por Yates es de 0,051 y el de Fisher 0,049.

Aquí vemos cómo el valor de p calculado por corrección de Yates y el calculado por prueba de Fisher son muy diferentes al de chi cuadrado sin corrección. Este último es muy significativo, mientras que los otros están muy cercanos al punto de corte de 0,05.

Por lo tanto, cuando analizamos tablas 2 x 2 con prueba de chi cuadrado y los valores de p informados por el programa son diferentes, el valor correcto es el de Yates. Si la muestra es muy pequeña, entonces usamos la prueba de Fisher.

PRUEBA DE CHI CUADRADO PARA DATOS APAREADOS

Hasta ahora hemos visto cómo analizar dos variables categóricas dicotómicas independientes.

En el caso en que los datos estén apareados, el método estadístico a utilizar es diferente.

El caso más común es el de los estudios caso-control, donde los controles son elegidos de forma apareada al caso (*matched* en inglés), en general de acuerdo con el sexo y la edad.

Como los casos y los controles tienen igual edad y sexo se dice que están *macheados* o *apareados*.

Supongamos que deseamos determinar si el tabaquismo se asocia con infarto de miocardio.

Elegimos 100 casos de pacientes con infarto de miocardio y buscamos su control, paciente sin infarto, de igual sexo y edad.

Obtendremos al final 100 pares de pacientes, 1 caso y 1 control, y les preguntamos si fuman o no.

De estos 100 pares tendremos las siguientes posibilidades:

Caso fuma - control fuma

Caso fuma - control no fuma

Caso no fuma - control fuma

Caso no fuma - control no fuma

		Casos (con infarto)		Total de pares
		Fuma	No fuma	
Controles (sin infarto)	Fuma	15 a	15 b	30
	No fuma	55 c	15 d	70
	Total de pares	70	30	100

Tabla 21.6: Tabla 2 x 2 datos apareados.

Estos datos se representan en tablas de 2 x 2 tal como muestra la Tabla 21.6.

De 100 pares de pacientes en total, 30 pares muestran concordancia, 15 casos fuman y su control también y en otros 15 pares, ni el caso ni el control fuman.

Los pares discordantes muestran que en 55 pares el caso fuma y el control no y en 15 pares el caso no fuma y el control sí. (Tabla 21.6)

Sobre estos últimos pares es que se calcula la posible asociación. Dado que los pares concordantes no son informativos, no podemos saber si hay más tabaquismo en un grupo o en otro. Pero en los pares discordantes es donde veremos si el tabaquismo es más frecuente en los casos, en los controles o es igualmente frecuente en ambos.

La hipótesis nula que se postula en este caso es que las celdas discordantes tienen el mismo número de pares discordantes, asumiendo que, del total de pares discordantes, el 50% estará en la celda b y el otro 50% en la celda c.

La prueba que se aplica a este caso es la llamada *prueba de McNemar* o *prueba de chi cuadrado para muestras apareadas*.

$$X^2_{\text{apareado}} = \frac{(|c - b| - 1)^2}{c + b} \quad 1 \text{ grado de libertad}$$

Donde *b* (15) y *c* (55) son el número de pares discordantes.

Aplicado a nuestro ejemplo tenemos:

$$X^2_{\text{apareado}} = \frac{(|15 - 55| - 1)^2}{15 + 55}$$

Chi² no corregido = 22,857143 (1 DF) p < 0,0001

Chi² corregido por Yates = 21,728571 (1 DF) p < 0,0001

Como vemos, aquí también se calculan 2 valores de *p*, el no corregido y el corregido por Yates. Como la muestra es grande, los valores de Chi² son semejantes, por lo que los valores de *p* son iguales.

En este caso la conclusión, de acuerdo con el valor de *p*, es que no hay la misma cantidad de pares discordantes en las celdas b y c, hay más pares donde el caso fuma y el control no.

Decimos entonces que, según esta prueba, el tabaquismo está asociado con el infarto.

PRUEBA DE CHI CUADRADO PARA ANÁLISIS DE TABLAS R X C

Antes de decidir el tipo de tabla a utilizar y el análisis respectivo, hay dos consideraciones que debemos hacer.

1. Si una o ambas variables tienen sólo dos categorías o más.
2. Si las categorías de una o ambas variables están ordenadas (variables categóricas ordinales).

De acuerdo con estos dos puntos podríamos tener las posibilidades que muestra la Tabla 21.7.

Número de categorías		Tipo de tabla	Test de Chi ²
Variable 1	Variable 2		
2	2	2 x 2	Chi ²
2	3 o más no ordinales	2 x c	Chi ² múltiple
2	3 o más ordinales	2 x c	Chi ² trend
3 o más no ordinales	3 o más no ordinales	r x c	Chi ² múltiple
3 o más ordinales	3 o más no ordinales	r x c	Chi ² trend
3 o más ordinales	3 o más ordinales	r x c	Chi ² trend

Tabla 21.7: Diferentes tipos de tablas de contingencia

Como se desprende de la tabla, en los casos en que al menos una variable tenga 3 o más categorías, hay dos tests de chi cuadrado que deben emplearse de acuerdo con que estas categorías sean ordinales o no.

PRUEBA DE CHI CUADRADO MÚLTIPLE

En general, es raro encontrar en medicina tablas 2 x c o r x c en las que al menos una de las variables no tenga datos categóricos ordinales, pero es posible encontrar el caso.

Supongamos que se ha estudiado la composición de las placas de colesterol de 200 pacientes, las cuales fueron clasificadas como: fibrosas, lipídicas, calcificadas y mixtas.

Se desea determinar si existe alguna asociación entre los diferentes tipos de placas y el antecedente de hipertensión arterial. Dicho de otra manera, si la hipertensión arterial es un factor de riesgo para el desarrollo de determinado tipo de placa aterosclerótica.

Presentamos los datos en forma de tabla 2 x 4 (en general se la designa como tabla 2 x k). (Tabla 21.8)

Vemos en la tabla que el 86% de los pacientes que tienen placas lipídicas es hipertenso y el 64% con placas fibrosas es hipertenso.

Si aplicamos la prueba de chi cuadrado múltiple, el resultado es:

$$\text{Chi}^2 \ 32,9 \quad p < 0,001$$

¿Qué significa p estadísticamente significativa en este contexto?

Lo único que nos responde la prueba de chi cuadrado múltiple es que *al menos una* de las proporciones es significativamente diferente a las proporciones esperadas.

En otras palabras, la proporción de hipertensos para cada categoría de placa de acuerdo con el tamaño de esta muestra en algún caso no es la esperada por azar.

Tipo de placa	Número de sujetos	Número de HTA	% HTA
Calcificada	50	20	40
Fibrosa	50	32	64
Lipídica	50	43	86
Mixta	50	18	36
Total	200	113	56,5

Tabla 21.8: Prevalencia de hipertensión arterial de acuerdo con el tipo de placa aterosclerótica

HTA: hipertensión arterial

La hipótesis nula de esta prueba es que la proporción de hipertensos en cada categoría de placa es la misma y por lo tanto las variaciones observadas son por azar.

Recordemos que la prueba compara las proporciones observadas vs. las esperadas. En este caso se calcularon todas las proporciones esperadas vs. las observadas y la prueba determinó que al menos una no está dentro del rango de lo esperado. Pero, ¿cuál es esa proporción?

Si miramos la tabla, podríamos decir que la proporción de hipertensos en los grupos de placas fibrosas y lipídicas es mayor.

Para confirmar si esta asociación es estadísticamente significativa recurrimos a la llamada *corrección de Bonferroni*.

Para aplicar esta corrección se transforma la tabla 2 x 4 en varias tablas 2 x 2 en las que se comparan las categorías entre sí y se calcula el valor de Chi² para cada comparación.

El resto de las comparaciones se muestra en la Tabla 21.10.

Una vez que hemos realizado las comparaciones de a dos, debemos multiplicar cada valor de p obtenido en cada una por el número de comparaciones totales que realizamos, en este caso 6.

Es así como el valor de p de 0,01 de la comparación entre placa fibrosa vs. placa lipídica se transforma en 0,06, siendo no significativa.

HTA	Calcificada	Fibrosa
Sí	40	64
No	60	36

Tabla 21.9: Porcentaje de hipertensos con placa calcificada vs. con placa fibrosa

p Yates 0,02

Comparación % de hipertensos	Valor p	Valor p con corrección
Calcificada vs. lipídica	< 0,0001	< 0,0006
Calcificada vs. mixta	0,8	4,8
Fibrosa vs. lipídica	0,01	0,06
Fibrosa vs. mixta	0,009	0,054
Lipídica vs. mixta	< 0,0001	< 0,0006

Tabla 21.10: Comparación múltiple. Aplicación de la corrección de Bonferroni

Para la comparación entre placa calcificada vs. placa fibrosa el valor de p corregido es de 0,12. Ahora, con los valores de p corregidos, podemos decir que las placas lipídicas están significativamente asociadas a hipertensión arterial.

La forma más correcta de realizar comparaciones múltiples es tratar de reagrupar las categorías para disminuir el número de comparaciones.

En este caso podríamos agrupar las placas calcificadas con las fibrosas en un grupo y de esta manera obtendríamos menos grupos para comparar.

El criterio para agrupar categorías siempre depende del investigador, del objetivo del estudio y del criterio fisiopatológico empleado.

En resumen, la prueba de Chi^2 múltiple nos ofrece la evidencia para rechazar o aceptar la hipótesis nula de que todas las proporciones son iguales a través de cada categoría.

Para determinar cuál es el grupo que muestra una asociación estadística debemos emplear la corrección de Bonferroni.

PRUEBA DE CHI CUADRADO TREND

Cuando en una tabla $r \times c$ las variables categóricas están ordenadas, lo que se desea determinar es si existe una tendencia hacia un incremento o hacia una disminución de la variable de interés a través de la variable categórica ordinal.

El ejemplo más común que podemos citar es cuando queremos determinar si al aumentar la edad aumenta la tasa de mortalidad.

Si bien la edad es una variable cuantitativa, es muy común agrupar los sujetos en categorías que abarcan 5 a 10 años. (Tabla 21.11)

Edad	N° de pacientes (%)	Sobrevida a 5 años (%)
≤ 49	3 084 (4,40)	10,5
50-59	9 474 (13,51)	7,84
60-69	20 615 (29,40)	6,16
≥ 70	36 947 (52,69)	2,83

Tabla 21.11: Sobrevida al cáncer de esófago de acuerdo con grupo etario

Chi-cuadrado trend = 41,33034

Grados de libertad = 1

$p < 0,0001$

Como se observa en la tabla, a medida que la edad aumenta, la sobrevida disminuye. La pregunta es: ¿existe una tendencia a menor sobrevida con el aumento de la edad? O bien ¿hay una asociación inversa entre edad y sobrevida o este *aparente* aumento de la mortalidad es sólo una variación esperable por azar?

La prueba de chi cuadrado trend tiene como hipótesis nula que no existe tal tendencia. Esta prueba comprueba si existe relación *lineal* entre las variables de interés; dicho en otras palabras, si el aumento o la disminución de una variable se acompaña de un cambio lineal (en cualquier sentido) de la otra variable.

En el ejemplo de la Tabla 21.11, el valor de p obtenido mediante la prueba de chi cuadrado trend nos dice que existe una relación lineal inversa entre aumento de la edad y sobrevida, no esperada por azar.

VALOR P					
DF	0,10	0,05	0,025	0,01	0,001
1	2.706	3.841	5.024	6.635	10.828
2	4.605	5.991	7.378	9.210	13.816
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.125
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588
11	17.275	19.675	21.920	24.725	31.264
12	18.549	21.026	23.337	26.217	32.910
13	19.812	22.362	24.736	27.688	34.528
14	21.064	23.685	26.119	29.141	36.123
15	22.307	24.996	27.488	30.578	37.697
16	23.542	26.296	28.845	32.000	39.252
17	24.769	27.587	30.191	33.409	40.790
18	25.989	28.869	31.526	34.805	42.312
19	27.204	30.144	32.852	36.191	43.820
20	28.412	31.410	34.170	37.566	45.315
21	29.615	32.671	35.479	38.932	46.797
22	30.813	33.924	36.781	40.289	48.268
23	32.007	35.172	38.076	41.638	49.728
24	33.196	36.415	39.364	42.980	51.179
25	34.382	37.652	40.646	44.314	52.620
26	35.563	38.885	41.923	45.642	54.052
27	36.741	40.113	43.195	46.963	55.476
28	37.916	41.337	44.461	48.278	56.892
29	39.087	42.557	45.722	49.588	58.301
30	40.256	43.773	46.979	50.892	59.703
31	41.422	44.985	48.232	52.191	61.098
32	42.585	46.194	49.480	53.486	62.487
33	43.745	47.400	50.725	54.776	63.870
34	44.903	48.602	51.966	56.061	65.247
35	46.059	49.802	53.203	57.342	66.619
36	47.212	50.998	54.437	58.619	67.985
37	48.363	52.192	55.668	59.893	69.347
38	49.513	53.384	56.896	61.162	70.703
39	50.660	54.572	58.120	62.428	72.055
40	51.805	55.758	59.342	63.691	73.402
41	52.949	56.942	60.561	64.950	74.745
42	54.090	58.124	61.777	66.206	76.084
43	55.230	59.304	62.990	67.459	77.419
44	56.369	60.481	64.201	68.710	78.750
45	57.505	61.656	65.410	69.957	80.077
46	58.641	62.830	66.617	71.201	81.400

Continúa en página siguiente.

Capítulo 21 Prueba de chi cuadrado

VALOR P					
47	59.774	64.001	67.821	72.443	82.720
48	60.907	65.171	69.023	73.683	84.037
49	62.038	66.339	70.222	74.919	85.351
50	63.167	67.505	71.420	76.154	86.661
51	64.295	68.669	72.616	77.386	87.968
52	65.422	69.832	73.810	78.616	89.272
53	66.548	70.993	75.002	79.843	90.573
54	67.673	72.153	76.192	81.069	91.872
55	68.796	73.311	77.380	82.292	93.168
56	69.919	74.468	78.567	83.513	94.461
57	71.040	75.624	79.752	84.733	95.751
58	72.160	76.778	80.936	85.950	97.039
59	73.279	77.931	82.117	87.166	98.324
60	74.397	79.082	83.298	88.379	99.607
61	75.514	80.232	84.476	89.591	100.888
62	76.630	81.381	85.654	90.802	102.166
63	77.745	82.529	86.830	92.010	103.442
64	78.860	83.675	88.004	93.217	104.716
65	79.973	84.821	89.177	94.422	105.988
66	81.085	85.965	90.349	95.626	107.258
67	82.197	87.108	91.519	96.828	108.526
68	83.308	88.250	92.689	98.028	109.791
69	84.418	89.391	93.856	99.228	111.055
70	85.527	90.531	95.023	100.425	112.317
71	86.635	91.670	96.189	101.621	113.577
72	87.743	92.808	97.353	102.816	114.835
73	88.850	93.945	98.516	104.010	116.092
74	89.956	95.081	99.678	105.202	117.346
75	91.061	96.217	100.839	106.393	118.599
76	92.166	97.351	101.999	107.583	119.850
77	93.270	98.484	103.158	108.771	121.100
78	94.374	99.617	104.316	109.958	122.348
79	95.476	100.749	105.473	111.144	123.594
80	96.578	101.879	106.629	112.329	124.839
81	97.680	103.010	107.783	113.512	126.083
82	98.780	104.139	108.937	114.695	127.324
83	99.880	105.267	110.090	115.876	128.565
84	100.980	106.395	111.242	117.057	129.804
85	102.079	107.522	112.393	118.236	131.041
86	103.177	108.648	113.544	119.414	132.277
87	104.275	109.773	114.693	120.591	133.512
88	105.372	110.898	115.841	121.767	134.746
89	106.469	112.022	116.989	122.942	135.978
90	107.565	113.145	118.136	124.116	137.208
91	108.661	114.268	119.282	125.289	138.438
92	109.756	115.390	120.427	126.462	139.666
93	110.850	116.511	121.571	127.633	140.893

Continúa en página siguiente.

VALOR P					
94	111.944	117.632	122.715	128.803	142.119
95	113.038	118.752	123.858	129.973	143.344
96	114.131	119.871	125.000	131.141	144.567
97	115.223	120.990	126.141	132.309	145.789
98	116.315	122.108	127.282	133.476	147.010
99	117.407	123.225	128.422	134.642	148.230
100	118.498	124.342	129.561	135.807	149.449

Tabla 21.12: Tabla de distribución X^2

DF = grados de libertad

En Epidemiología, la cuantificación de la frecuencia de un evento o de un factor de exposición es la base para la comparación de estas estimaciones entre diferentes poblaciones. Si determinamos la incidencia de cáncer de pulmón en una población expuesta a un determinado factor de exposición y la comparamos con la incidencia de cáncer de pulmón en otra población no expuesta, podemos determinar si la exposición a este factor constituye un factor de exposición asociado a cáncer de pulmón.

Para hacerlo más eficiente e informativo, las dos frecuencias a comparar pueden resumirse en un solo valor que estima el riesgo de desarrollar la enfermedad.

El objetivo es determinar si existe un aumento del riesgo o tal vez una reducción del riesgo entre las poblaciones a comparar, con y sin el factor de exposición a estudiar.

RIESGO ABSOLUTO

A la incidencia de una enfermedad en una población se la denomina *riesgo absoluto*. Este puede indicar la magnitud del riesgo de sufrir una enfermedad en un grupo de personas con exposición a un factor de exposición.

El riesgo absoluto de morir un año después de un infarto de miocardio es la mortalidad calculada en un grupo de pacientes que han sufrido un infarto de miocardio, por ejemplo 8%.

Como este valor no se compara con el riesgo de la población no expuesta, no indica si existe asociación entre el factor de exposición y el evento.

No obstante, el riesgo absoluto es importante a nivel clínico y de salud pública. Un paciente podría preguntarnos cuál es su expectativa de vida al año de haber sufrido un infarto de miocardio.

RIESGO RELATIVO

Para determinar si existe asociación entre un factor de exposición y un evento, o bien si hay un aumento o disminución del riesgo en el grupo expuesto, debemos comparar el riesgo absoluto del grupo expuesto con el riesgo absoluto del grupo no expuesto.

Esta comparación se realiza calculando el cociente entre el riesgo absoluto de los expuestos y el riesgo absoluto de los no expuestos.

$$\text{Riesgo relativo (RR)} = \frac{\text{Riesgo absoluto expuestos}}{\text{Riesgo absoluto no expuestos}}$$

En un estudio de cohortes se han seguido por 3 años, sujetos expuestos a una dieta con alto contenido en grasas y sujetos expuestos a una dieta baja en grasas para determinar la incidencia de enfermedad coronaria en ambos grupos. La Tabla 22.1 muestra los datos y la notación general que se utiliza en una tabla 2 x 2.

2° Evento 1° Exposición	Enfermedad coronaria +	Enfermedad coronaria -	Total	Incidencia
Dieta grasa	40 <i>a</i>	160 <i>b</i>	200 (<i>a</i> + <i>b</i>)	20% <i>a</i> / (<i>a</i> + <i>b</i>)
Dieta no grasa	20 <i>c</i>	180 <i>d</i>	200 (<i>c</i> + <i>d</i>)	10% <i>c</i> / (<i>c</i> + <i>d</i>)
Total	60	340		400

Tabla 22.1: Estudio de cohortes. Cálculo de incidencia

El riesgo absoluto o incidencia de enfermedad coronaria en los expuestos al cabo de 3 años es del 20% mientras que en los no expuestos es de 10%, el riesgo relativo es:

$$\text{Riesgo relativo} = \frac{20}{10} = 2$$

Esto significa que el grupo expuesto tiene el doble de riesgo de sufrir enfermedad coronaria que el grupo no expuesto. O, lo que es lo mismo, el grupo expuesto tuvo un 100% más de enfermedad coronaria sobre el grupo no expuesto.

Como el riesgo absoluto o la incidencia sólo se pueden calcular en estudios longitudinales, ya sea de cohortes o experimentales, el cálculo del riesgo relativo para determinar asociación se utiliza sólo en este tipo de diseño de estudios.

Cuando valoramos el riesgo absoluto de una enfermedad en una población, lo que hacemos es estimar un valor que, como ya hemos explicado, puede tomar un rango dependiendo del azar. Es por esto que también para los valores de riesgo absoluto o incidencia existe un intervalo de confianza del 95%.

El riesgo de enfermedad coronaria en los expuestos a dieta grasa podría estar entre el 18 y el 22%. Estos son los valores probables de obtener 95 veces si midiéramos 100 veces la incidencia de enfermedad coronaria en la misma población. El riesgo absoluto en los no expuestos podría estar entre valores del 8 al 12%.

Por lo tanto, el riesgo relativo también tiene un intervalo de confianza del 95%, en este caso varía entre 1,5 a 2,5.

Esto significa que el aumento del riesgo en el grupo expuesto puede ser 1,5 veces mayor (50%) hasta 2,5 veces (250%) comparado con el grupo no expuesto.

Supongamos que el riesgo absoluto de enfermedad coronaria en ambos grupos es igual (20%); en este caso el riesgo relativo sería 1. No habría un aumento del riesgo en el grupo expuesto y por lo tanto no habría asociación entre el factor de riesgo y la enfermedad.

Si, en cambio, el riesgo relativo es mayor a 1, esto significa que el grupo expuesto tiene un riesgo absoluto mayor, y por lo tanto un aumento del riesgo.

Para determinar con certeza si existe asociación entre el factor de exposición y el evento, debemos estar seguros de que el valor del riesgo relativo siempre está por encima de 1.

Para esto, el cálculo del intervalo de confianza del 95% del riesgo relativo nos es de gran ayuda. Si este intervalo no contiene el valor 1 (no diferencia entre riesgos absolutos) decimos que el grupo expuesto tiene más riesgo y que existe asociación entre el factor de riesgo y la enfermedad, y si realizamos una prueba estadística el valor de *p* hallado será inferior a 0,05, aunque no sepamos el valor de *p* exacto.

Para el ejemplo anterior, el riesgo relativo de 2 tiene un intervalo de confianza que comprende 1,5 a 2,5. Como este intervalo no contiene el 1, estamos 95% seguros de que el grupo expuesto tiene más riesgo. Si, en cambio, el intervalo 95% fuera de 0,8 a 2,5, al contener el valor 1 cabría la posibilidad de que en alguna medición ambos grupos tuvieran

la misma incidencia de enfermedad coronaria, por lo que la asociación con el factor de exposición ya no se sostiene.

Si el valor del riesgo relativo es menor que 1, significa que el grupo expuesto tuvo un riesgo absoluto menor al del grupo no expuesto, implicando que el factor de exposición es en realidad un factor protector.

Por ejemplo, un grupo expuesto a una droga podría tener menor mortalidad (10%) comparado con el grupo no tratado con la droga (20%).

$$\text{Riesgo relativo} = \frac{10}{20} = 0,5 \text{ IC } 95\% \text{ } 0,3-0,7$$

Aquí vemos que el grupo expuesto a la droga tuvo la mitad de muertes que el grupo no expuesto (50% menos de muertes).

Para determinar si esta diferencia de muertes se asocia con la exposición a la droga, basta con calcular el intervalo de confianza del riesgo relativo (IC 95% 0,3-0,7); si no contiene el valor 1, entonces decimos que existe asociación entre exposición a la droga y disminución significativa de la mortalidad, por lo que el valor de p hallado con la prueba estadística será $< 0,05$.

El valor del riesgo relativo informa acerca de la fuerza de la asociación entre un factor de exposición y un evento; cuanto más grande es este valor, más seguros podríamos estar de esta asociación.

En el caso en que el factor de exposición fuera protector, el valor debería estar muy alejado del valor 1 hacia el cero.

También el ancho del intervalo de confianza contribuye a determinar el grado de certeza de la asociación; cuanto más chico es el intervalo, mayor certeza.

No es lo mismo un intervalo de confianza del RR entre 1,5 a 2,5 que otro que varía entre 1,8 a 2,2.

En el primero, el riesgo puede variar entre 1,5 veces a 2,5 veces, mientras que en el segundo los valores se acercan mucho más al valor 2, obtenido como riesgo relativo.

En resumen: Para calcular el RR (estudio cohorte o experimental) primero se debe calcular la incidencia o riesgo absoluto de la enfermedad en el grupo expuesto y en el no expuesto.

Luego se deberá calcular el cociente = incidencia en expuestos/incidencia no expuestos. Si:

RR = 1 riesgo en expuestos es igual al de los no expuestos. No hay asociación.

RR < 1 riesgo en expuestos menor a los no expuestos. Asociación si IC 95% no contiene el valor 1.

RR > 1 riesgo en expuestos mayor a los no expuestos. Asociación si IC 95% no contiene el valor 1.

REDUCCIÓN DEL RIESGO RELATIVO

La Reducción del Riesgo Relativo (RRR) es un índice que se reporta muy habitualmente en los estudios experimentales, generalmente estudios clínicos aleatorizados.

En la Tabla 22.2 se muestran los resultados de un estudio en el que se evaluó una droga vs. placebo y cuyo punto final fue mortalidad.

- El riesgo absoluto de muerte o incidencia de muerte en el grupo placebo fue:

$$404/1325 = 30\% \text{ (IC } 95\% \text{ } 28-33)$$

	Muertos	Vivos	Total
Droga	350	974	1.324
Placebo	404	921	1.325
Total	754	1 895	2.649

Tabla 22.2: Mortalidad droga vs. placebo

En este grupo la mortalidad fue del 30%, pudiendo variar entre el 28 y el 33%.
El riesgo absoluto de muerte o incidencia de muerte en el grupo droga fue:

$$350/1\ 324 = 26\% \text{ (IC 95\% 24-28)}$$

Aquí la mortalidad fue inferior, 26%, pudiendo encontrarse entre el 24 y el 28%.

- El riesgo relativo (expuestos a la droga vs. no expuestos) fue:

$$26/30 = 0,86 \text{ (IC 95\% 0,82-0,98)}$$

El riesgo relativo droga vs. placebo fue de 0,86, el grupo expuesto tuvo menor mortalidad, por lo tanto menor riesgo.

Si partimos del supuesto de no diferencias en la mortalidad es decir $RR = 1$, vemos que con la droga el riesgo se redujo a 0,14.

- Reducción de Riesgo Relativo (RRR) = $1 - 0,86 = 0,14$

Si consideramos $RR = 1$ como 100% decimos entonces que la droga redujo el RR un 14%.

- Reducción de Riesgo Relativo (RRR) = $100\% - 86\% = 14\%$

Como hemos visto, el RR tiene un IC 95%, en este caso entre 0,82 a 0,98, ¿cuál sería la reducción del riesgo relativo si este hubiera sido 0,98?

- $RRR = 100\% - 98\% = 2\%$

Como conclusión podríamos decir que la droga reduce el RR entre un 2 y un 18%, lo cual en términos de mortalidad parece bastante amplio.

Siempre hay que tener en cuenta el IC 95% del RR y de la RRR para determinar cuál es el rango posible de beneficio de una droga o intervención.

Supongamos ahora que en otro estudio la mortalidad con droga fue del 4,3 vs. el 5% en el grupo placebo, el RR sería también 0,86 y la RRR otra vez 14%.

$$\text{Estudio 1: } 26\% \text{ vs. } 30\% = RR \ 0,86$$

$$\text{Estudio 2: } 4,3\% \text{ vs. } 5\% = RR \ 0,86$$

$$RRR \ 14\%$$

Si bien en ambos casos la RRR es de 14%, no es lo mismo reducir 14% la mortalidad de un grupo que tiene 30% de mortalidad (muertos en el primer estudio) que reducir el 14% de un grupo que tiene un 5% mortalidad (segundo estudio).

Por lo tanto, cuando se informa la RRR, se está omitiendo informar el riesgo del grupo control y la verdadera diferencia de riesgos entre los dos tratamientos.

En el primer estudio la diferencia de riesgos absoluta fue del 4% que es el 14% de 30%, mientras que en el segundo 0,7%. (14% de 5%)

Calcular la diferencia absoluta de riesgo es muy importante, ya que de ello se desprende el beneficio que se obtiene con una intervención y permite el cálculo de otras variables muy importantes a la hora de valorar la utilidad de una intervención.

NÚMERO DE PACIENTES NECESARIOS A TRATAR. NNT

Uno de los índices que permite determinar el beneficio de un tratamiento o intervención en una población determinada, además del riesgo relativo, es el cálculo de cuántos pacientes hay que tratar para evitar un evento. A este índice se lo conoce como NNT, número necesario a tratar (*Number Needed to Treat*).

El término evento está referido al punto final primario determinado en cada estudio.

A veces el evento es muerte, pero generalmente el punto final de los estudios aleatorizados es la combinación de varios eventos como muerte y/o reinternación y/o desarrollo de insuficiencia cardíaca, etc.

En el ejemplo utilizado anteriormente, la droga tuvo una mortalidad de 26% y el placebo de 30%, la diferencia absoluta de riesgo es de 4%.

El NNT es la inversa de la diferencia absoluta de riesgos multiplicado por 100:

$$\text{NNT} = 1 / \text{diferencia absoluta de riesgos} \times 100$$

$$\text{NNT} = 1 / 4 = 0,25 \times 100 = 25 \text{ pacientes}$$

En este caso decimos que, para evitar una muerte con la droga, es necesario tratar a 25 pacientes por el mismo período de tiempo en que fueron tratados durante el estudio y a la misma dosis.

En el segundo ejemplo, la mortalidad con la droga fue del 4,3% y con el placebo del 5%. La diferencia absoluta de riesgo es del 0,7%.

$$\text{El NNT es: } 1 / 0,7 \times 100 = 143 \text{ pacientes}$$

En este caso hay que tratar 143 pacientes para evitar una muerte, por el mismo período de tiempo en que se trataron los pacientes del estudio y a la misma dosis que se administró la droga.

En el caso de que el estudio tenga como punto final sólo un evento, el cálculo del NNT es fácil de interpretar; pero cuando el punto final es combinado, sólo podemos decir que necesitamos tratar, por ejemplo, 143 pacientes para evitar una muerte y/o una hospitalización y/o desarrollo de insuficiencia cardíaca, etc.

En este caso la interpretación del beneficio es mucho más dificultosa.

Como vemos, dependiendo de la incidencia del evento en cada grupo y de su diferencia, el NNT es muy diferente.

Si consideramos esta diferencia, el tiempo en que hay que tratar a los pacientes, costos de la droga, posibles efectos adversos y análisis de costo/efectividad, podremos determinar el real beneficio de implementar una intervención.

NÚMERO DE PACIENTES NECESARIOS A DAÑAR. NNH

Así como es posible determinar el NNT para evitar un evento (o punto final combinado) también es posible determinar el número de pacientes necesarios a tratar para producir un evento adverso. A este índice se lo conoce como NNH, número necesario a dañar (*Number Needed to Harm*).

Esto se difundió mucho en los estudios con trombolíticos, donde el riesgo de accidente cerebrovascular hemorrágico es una posibilidad con consecuencias graves.

El cálculo es la inversa de la diferencia absoluta de eventos adversos serios de cada rama de tratamiento.

Supongamos que la droga tuvo un 18% de eventos adversos serios y el placebo un 10%. La diferencia es del 8% y el NNH:

$$\text{NNH} = 1/8 \times 100 = 12,5 \text{ pacientes}$$

Si el evento adverso serio es sangrado mayor, entonces, tratando 12,5 pacientes por el mismo período de tiempo y a la misma dosis de la droga que se administró en el estudio esperaríamos producir un sangrado mayor.

Si analizamos el NNT vs. el NNH podremos determinar la chance de beneficiar vs. dañar con un determinado tratamiento.

En el primer ejemplo el NNT fue de 25 pacientes y el NNH de 12,5.

La probabilidad de beneficiar es de $1/25 = 0,04 = 4\%$ vs. la probabilidad de dañar $1/12,5 = 0,08 = 8\%$.

Por lo tanto, hay más probabilidad de dañar que de beneficiar.

Esto es lógico, si vemos que para evitar una muerte hacen falta 25 pacientes y para producir un sangrado mayor sólo 12,5 pacientes.

Cuando una droga incrementa la incidencia de eventos, entonces lo que calculamos es el NNH de la droga, ya no podemos calcular la RRR ni el NNT porque la droga no fue beneficiosa.

- Como conclusión, a la hora de interpretar los resultados de una intervención, para determinar el posible beneficio a nivel individual de cada paciente, hay que analizar los riesgos absolutos, su diferencia y el RR. Recordemos que la RRR es sólo una medida relativa del beneficio producido.
- Para determinar el beneficio a nivel de asignación de recursos, el cálculo del NNT y NNH es de suma utilidad.

ODDS RATIO

Ya hemos mencionado que para calcular el RR debemos tener los valores de la incidencia en expuestos y no expuestos. Esto sólo es posible de obtener en estudios longitudinales.

En los estudios caso-control y transversales, no podemos conocer la incidencia, debido a que partimos de la población enferma (casos) y la sana (controles) y luego calculamos el porcentaje de expuestos en cada grupo y en los estudios transversales obtenemos datos de prevalencia de factor de exposición y/o evento de interés.

2° Evento 1° Exposición	Evento +	Evento -	Total	Incidencia
Expuestos	<i>a</i>	<i>b</i>	<i>a + b</i>	$a / (a + b)$
No expuestos	<i>c</i>	<i>d</i>	<i>c + d</i>	$c / (c + d)$

Tabla 22.3: Estudio de cohortes. Cálculo de incidencia

1° seleccionar 2° Exposición	Enfermos casos	No enfermos controles
Expuestos	<i>a</i>	<i>b</i>
No expuestos	<i>c</i>	<i>d</i>
Total	<i>a + c</i>	<i>b + d</i>
% Expuestos	$a / (a + c)$	$b / (b + d)$

Tabla 22.4: Estudio caso-control. Proporción de expuestos

Por lo tanto, para calcular si existe un exceso de riesgo y asociación debemos recurrir a otra medida conocida como *odds ratio*.

El *odds ratio* puede obtenerse en cualquier tipo de estudio, incluidos los longitudinales, aunque no es la medida de efecto correcta para estos estudios.

En los estudios de cohortes partimos identificando expuestos y no expuestos y luego determinamos la cantidad de eventos en cada uno de estos grupos. (Tabla 22.3)

En los estudios de caso-control partimos identificando enfermos (casos) y no enfermos (controles) y calculamos la proporción en cada uno de los expuestos. (Tabla 22.4)

Una vez obtenida la proporción de expuestos en los casos y en los controles, ¿cómo sabemos si existe un aumento del riesgo en los expuestos? y ¿este exceso está asociado al factor de exposición?

Se dice que la posibilidad de un evento (*odds*) puede definirse como el cociente del número de veces que el evento puede ocurrir vs. número de veces que el evento no puede ocurrir.

Supongamos que un caballo tiene una probabilidad (*p*) de ganar una carrera de 60%: por lo tanto, la probabilidad de no ganar es del 40% ($1 - p$ o $100 - 60$), por lo tanto el *odds* de que el caballo gane es igual a:

$$Odds = \frac{\text{Posibilidad de que el caballo gane}}{\text{Posibilidad de que el caballo no gane}}$$

$$Odds = \frac{60\%}{40\%} = 1,5$$

Es importante diferenciar entre probabilidad y posibilidad (*odds*).

Probabilidad de ganar = 60%

Posibilidad de ganar = 1,5

Veamos ahora cómo aplicar este concepto a estudios de cohorte y caso-control.

ODDS RATIO PARA ESTUDIOS CASO-CONTROL

La pregunta que nos hacemos en un estudio caso-control es ¿cuál es la proporción de casos expuestos al factor de exposición?

En la Tabla 22.4 vemos que esta proporción es igual a: $a / (a + c)$

En segundo término, nos preguntamos ¿cuál es la posibilidad de que un caso fuera expuesto? A esto lo respondemos calculando el *odds* o posibilidad de que un caso fuera expuesto.

La posibilidad (*odds*) de que un caso haya sido expuesto es = a (posibilidad de ganar la carrera).

La posibilidad (*odds*) de que un caso no haya sido expuesto es = c (posibilidad de no ganar la carrera).

Por lo tanto el *odds* es = $\frac{a}{c}$ (posibilidad de ganar vs. no ganar la carrera)

Para el grupo control, la proporción de expuestos es = $b/b + d$

El *odds* de que un control haya sido expuesto es = b

El *odds* de que un control no haya sido expuesto es = d

Por lo tanto, el *odds* es = $\frac{b}{d}$

Si ahora calculamos el cociente entre el *odds* de los casos vs. *odds* de los controles, obtenemos el *odds ratio*.

$$\text{Odds ratio} = \frac{\frac{a}{c}}{\frac{b}{d}}$$

O lo que es lo mismo:

$$\text{Odds ratio} = \frac{ad}{bc}$$

Al cálculo del odds ratio también se lo denomina *cocientes de los productos cruzados*, debido a que multiplicamos la celda a por d y lo dividimos por b por c .

a	b
c	d

Supongamos que en un estudio caso-control los resultados son los que se muestran en la Tabla 22.5.

El porcentaje de casos expuestos es del 45% y el de controles expuestos del 20%. ¿Cómo podemos saber si el factor de exposición está relacionado con la enfermedad?

Calculamos entonces el OR.

$$\text{OR} = (45 \times 80) / (55 \times 20) = 3,2 \text{ (IC 95\% 1,7 a 6,5)}$$

El OR también tiene su IC 95%, debido a que toda estimación en estadística tiene su grado de incertidumbre. En este caso la posibilidad de haber estado expuesto entre casos y controles es de 3,2 veces más posible para los casos.

En otras palabras, los casos estuvieron 3,2 veces más expuestos que los controles.

	Casos	Controles
Expuestos	45	20
No expuestos	55	80
Total	100	100
% expuestos	45	20

Tabla 22.5: Resultados estudio caso-control

¿Cómo sabemos, basados en este valor del OR, si la exposición está asociada a la enfermedad?

Si el *odds* de casos fuera igual que el *odds* de controles, el OR sería = 1, en esta situación concluiríamos que el factor de exposición, al ser igual en casos y controles, no contribuyó a producir la enfermedad.

Si el OR es > 1 y su IC 95% no contiene al valor 1, el *odds* de los casos es mayor al de los controles, por lo tanto concluimos que el factor de exposición contribuyó a producir la enfermedad con un 95% de certeza.

Si el IC 95% contiene el valor 1 significa que hay alguna posibilidad de que ambos *odds* sean iguales, por lo tanto no podemos concluir con certeza que el factor de exposición contribuya a producir la enfermedad.

Si el OR es < 1 y su IC 95% no contiene el valor 1, el *odds* de los casos es menor al de los controles; por lo tanto, concluimos que el factor de exposición está relacionado negativamente con la enfermedad con un 95% de certeza, no hay posibilidad de que por azar los *odds* fueran iguales.

Un ejemplo de esto podría ser que los casos fueran sujetos con bocio y los controles sin bocio, el factor de exposición podría ser suplemento de yodo en la sal. Los casos estarían menos expuestos (tendrían menos suplemento de yodo) y los controles más expuestos (más suplemento de yodo).

ODDS RATIO PARA ESTUDIOS CASO-CONTROL APAREADOS

En muchos casos, los estudios caso-control se realizan apareando los sujetos enfermos (casos) con los controles (sanos), esto significa que por cada caso se elige un control que tenga ciertas características del caso, generalmente sexo y edad.

En esta situación, lo que se analizan son pares caso-control en lugar de sujetos.

La forma general de presentar estos datos se observa en la Tabla 22.6.

La celda *a* representa al par caso-control donde ambos fueron expuestos.

La celda *b* representa al par caso-control donde el caso fue expuesto y el control no.

La celda *c* representa al par caso-control donde el caso no fue expuesto y el control sí.

La celda *d* representa al par caso-control donde ambos no fueron expuestos.

Casos	Controles	
	Expuestos	No expuestos
Expuestos	<i>a</i>	<i>b</i>
No expuestos	<i>c</i>	<i>d</i>

Tabla 22.6: Estudio caso-control apareado

Por lo tanto, las celdas a y d representan los pares concordantes, donde ambos, casos y controles, fueron expuestos o no expuestos, como hay concordancia esto no aporta información acerca de si los casos fueron más o menos expuestos.

Las celdas b y c al ser las discordantes son las que aportan información acerca de si los casos estuvieron más expuestos que los controles o viceversa.

Por lo tanto, para calcular el OR en los estudios caso-control apareados sólo necesitamos los valores de las celdas b y c .

$$OR \text{ para datos apareados} = \frac{b}{c}$$

Supongamos que tenemos, en un estudio caso-control apareado, cero pares expuestos concordantes y 62 pares no expuestos concordantes, sólo los pares discordantes son informativos. (Tabla 22.7)

Vemos que los casos expuestos con controles no expuestos son 9 pares y 2 pares donde los casos no fueron expuestos y sus controles sí.

Se observa que en estos grupos los casos fueron más expuestos.

Aplicando la fórmula para casos apareados, el $OR = \frac{9}{2} = 4,5$

Concluimos entonces que los casos tienen más posibilidad de haber estado expuestos que los controles.

Para verificar si existe asociación, debemos observar el IC 95% = 0,93 a 42,7.

Como el IC 95% contiene al valor 1, la conclusión final es que en este estudio el factor de exposición no está estadísticamente asociado con la enfermedad de forma significativa.

ODDS RATIO PARA ESTUDIOS DE COHORTES

En un estudio de cohortes lo primero que nos preguntamos es ¿cuál es la probabilidad de que los sujetos expuestos desarrollen la enfermedad?

La respuesta está en calcular la incidencia de la enfermedad en los sujetos expuestos. (Tabla 22.8)

Casos	Controles	
	Expuestos	No expuestos
Expuestos	0	9
No expuestos	2	62

Tabla 22.7: Estudio caso-control apareado

2° evaluar eventos 1° Exposición	Enfermos +	Enfermos -	Total	Incidencia
Expuestos	a	b	$a + b$	$a / (a + b)$
No expuestos	c	d	$c + d$	$c / (c + d)$

Tabla 22.8: Cálculo de incidencia y odds ratio. Estudio de cohortes

La incidencia en sujetos expuestos es igual = $\frac{a}{a+b}$

Luego la pregunta es ¿cuál es la posibilidad (*odds*) de que la enfermedad se desarrolle en el grupo de expuestos?

Si observamos la Tabla 22.8 vemos que hay $a + b$ sujetos expuestos, por lo tanto el *Odds* de desarrollo de la enfermedad es = $\frac{a}{b}$

Vemos que a es la probabilidad de tener la enfermedad en el grupo expuesto (equivalente a la probabilidad de ganar la carrera) y que b es la probabilidad de no tener la enfermedad en el grupo expuesto (probabilidad de no ganar la carrera).

Por lo tanto, la posibilidad (*odds*) de desarrollar la enfermedad es $\frac{a}{b}$

Lo mismo ocurre para el grupo no expuesto.

La probabilidad de desarrollar la enfermedad (incidencia) es $\frac{c}{c+d}$

Y la posibilidad de desarrollar la enfermedad (*odds*) = $\frac{c}{d}$

Así como el cociente de la incidencia entre expuestos vs. no expuestos (riesgo relativo) puede ser utilizado como medida de asociación entre factor de exposición y enfermedad, también el cociente entre el *odds* de los expuestos vs. el *odds* de los no expuestos puede ser usado con el mismo fin.

A este cociente se lo denomina *Odds Ratio* (OR), y se lo define como el cociente del *odds* de desarrollar la enfermedad en los sujetos expuestos vs. *odds* de desarrollar la enfermedad en los sujetos no expuestos.

$$\text{Odds ratio} = \frac{\frac{a}{b}}{\frac{c}{d}} \text{ o lo que es lo mismo } \text{odds ratio} = \frac{ad}{bc}$$

Igual que el obtenido para los estudios caso-control no apareados.

Veamos cómo calcular el *odds ratio* utilizando los datos de la Tabla 22.9.

Ya hemos visto que el RR de expuestos vs. no expuestos es de $26/30\% = 0,86$ (IC 95% 0,82-0,98)

El OR es = $(350 \times 921) / (404 \times 974) = 0,82$ (IC 95% 0,69 a 0,97)

Como vemos el valor del OR es muy similar al RR y también se calcula su IC 95%.

	Muertos	Vivos	Total
Droga	350	974	1 324
Placebo	404	921	1 325
Total	754	1 895	2 649

Tabla 22.9: Mortalidad droga vs. placebo

¿Cuándo el OR puede utilizarse en lugar del RR?

En los estudios caso-control no es posible calcular el RR, por lo tanto nos basamos en el cálculo del OR.

El OR obtenido puede ser una buena aproximación al RR si se cumplen 3 condiciones:

1. Cuando los casos son representativos con respecto a la historia del factor de exposición de todos los casos de la población de donde fueron tomados.
2. Cuando los controles son representativos con respecto a la historia del factor de exposición de todos los controles de la población de donde fueron tomados.
3. Cuando la enfermedad en estudio es infrecuente.

Ejemplo: en un estudio de cohortes cuya incidencia de una enfermedad es baja (Tabla 22.10) el RR y OR serán igual a:

$$RR = \frac{0,02}{0,01} = 2 \text{ y el } OR = \frac{200 \times 9900}{100 \times 9800} = 2,02$$

La incidencia de muerte en la población estudiada es de 2%.

En un estudio con alta incidencia de la enfermedad podríamos obtener los siguientes resultados. (Tabla 22.11)

La incidencia de muerte en la población total estudiada es de 50%.

$$RR = \frac{50}{25} = 2 \text{ y el } OR = \frac{50 \times 75}{25 \times 50} = 3$$

En conclusión:

1. El OR es útil como medida de asociación, tanto en estudio de cohortes como en caso-control. Esto es de utilidad al momento de analizar la relación causal entre un factor de exposición y un evento.
2. En los estudios de cohortes el RR puede calcularse directamente.
3. En estudios caso-control, el RR no puede calcularse; por lo tanto se usa el OR como medida de asociación, y es un buen estimador del RR cuando la enfermedad es infrecuente.
4. Tanto el RR como el OR informan acerca de la posibilidad o probabilidad de que un sujeto expuesto desarrolle una enfermedad o de que un sujeto considerado como caso haya estado expuesto al factor de exposición. El RR y el OR informan riesgo a nivel individual.

	Muertos	Vivos	Total
Droga	200	9.800	10.000
Placebo	100	9.900	10.000

Tabla 22.10: RR y OR estudio de cohortes. Enfermedad infrecuente

	Muertos	Vivos	Total
Droga	50	50	100
Placebo	25	75	100

Tabla 22.11: RR y OR Estudio de cohortes. Enfermedad frecuente

En este capítulo nos referiremos a una pregunta diferente: ¿cuánta de la cantidad de eventos, que ocurre en una población, puede atribuirse a cierto factor de exposición?

Por ejemplo ¿cuántos de los casos con cáncer de pulmón pueden atribuirse al hábito de fumar?

A este índice se lo denomina *Riesgo Atribuible (RA)*, y se define como la cantidad o proporción de una enfermedad en una población que puede atribuirse a un específico factor de exposición.

El riesgo atribuible es importante en la práctica clínica y en la salud pública, debido a que resuelve la cuestión de cuánto del riesgo de una enfermedad puede prevenirse si se elimina el factor de riesgo en una determinada población.

Para que el cálculo del RA sea de utilidad, primero debemos estar seguros de que existe una relación causal entre el factor de exposición y el evento, por eso podemos hablar de factor de riesgo.

Para ello debemos tener evidencia suficiente, a partir de diferentes estudios epidemiológicos, de que el factor de exposición produce la enfermedad. En próximos capítulos veremos cuáles son los criterios para establecer causalidad.

El RA puede calcularse en los sujetos expuestos; en este caso calculamos cuántos sujetos enfermos se podrían evitar si eliminamos el factor de riesgo.

También podemos calcular el RA para la población general, llamado *riesgo atribuible poblacional*. En este caso podemos calcular cuántos casos de la enfermedad en la población total (enfermos + sanos) podríamos prevenir si eliminamos de la población el factor de exposición.

RIESGO ATRIBUIBLE PARA SUJETOS EXPUESTOS

Si consideramos un grupo expuesto y otro no expuesto (Figura 23.1), el riesgo total de enfermedad en el grupo expuesto es mayor que en el grupo no expuesto.

Podemos preguntarnos, en el grupo expuesto ¿cuánto del total del riesgo de enfermedad es en realidad a causa del factor de exposición? En un grupo de pacientes tabaquistas (expuestos) ¿cuántos casos de cáncer de pulmón se deben al tabaquismo?

Para contestar esta pregunta debemos observar que en el grupo no expuesto hay un cierto riesgo de padecer la enfermedad (Figura 23.1) aunque menor al del grupo expuesto.

En realidad, el riesgo de sufrir una enfermedad nunca es cero, ni siquiera en un grupo no expuesto.

Aun en pacientes no fumadores hay riesgo de cáncer de pulmón, posiblemente debido a otros factores.

A este riesgo se lo denomina *riesgo basal*. Por lo tanto, cada persona comparte este riesgo basal, independientemente de estar expuesto o no al factor de exposición en estudio.

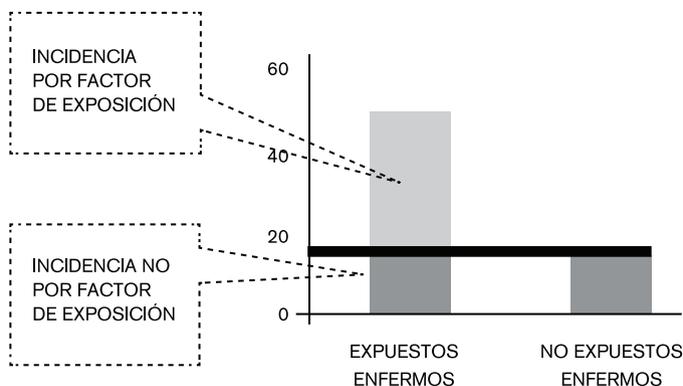


Figura 23.1: Riesgo atribuido a un factor de exposición en una población de enfermos

El riesgo total de enfermedad en el grupo expuesto es la suma del riesgo basal + el riesgo conferido por el factor de exposición.

Si el riesgo en expuestos es del 50% y en los no expuestos del 15%, el riesgo debido al factor de exposición de interés es 35%.

Si queremos conocer cuánto del riesgo total en el grupo expuesto se debe al factor de exposición, debemos restar el riesgo basal al riesgo total de los expuestos.

Por lo tanto el RA en expuestos:

$$RA = (Incidenca\ en\ expuestos) - (Incidenca\ en\ no\ expuestos)$$

Supongamos que estamos estudiando el impacto de amamantar en la incidencia de diarrea en niños menores a un año.

La incidencia o riesgo de diarrea en el grupo que amamanta es = 150 por cada 1.000 niños.

La incidencia de diarrea en el grupo que no amamanta es = 280 por cada 1 000 niños.
El RA será = 280 - 150 = 130 por 1 000

Si pudiéramos evitar el factor de riesgo, en este caso no amamantar, podríamos evitar 130 casos de diarrea por cada 1 000 niños con diarrea menores de un año.

Esto también puede expresarse como porcentaje: ¿cuál es la proporción de casos de diarrea que se podrían evitar amamantando?

A esto se lo denomina *Fracción Atribuible (FA)*

$$FA = \frac{(Incidenca\ en\ expuestos) - (Incidenca\ en\ no\ expuestos)}{(Incidenca\ en\ expuestos)}$$

Aplicado al ejemplo: $\frac{280 - 150}{280} = 46\%$

Decimos entonces que amamantando evitaríamos el 46% de los casos de diarrea en niños menores de un año que padecen diarrea.

Como vemos, el RA y la FA son indicadores de prevención potencial que se lograría en la población de enfermos con la eliminación del factor de riesgo; datos muy importantes a la hora de elegir e implementar políticas de salud para la población.

Es importante resaltar la diferencia entre *RR* y *RA*.

Si un paciente nos pregunta cuál es el riesgo de padecer cáncer si sigue fumando, lo que responderemos es el *RR*; es decir, si sabemos que fumar aumenta 5 veces la probabilidad de padecer cáncer de pulmón comparado con otra persona que no fuma, le diremos que tiene un riesgo 5 veces mayor que un sujeto que no fuma.

Pero si la pregunta es qué factor de riesgo debemos eliminar para reducir el número de casos de cáncer de pulmón en una población de pacientes con esta enfermedad, recurrimos al *RA*.

Otro punto para resaltar es que tanto el *RA* como la *FA* se refieren al número o proporción de casos evitables en la población enferma y no en la población total.

A continuación veremos cómo calcular este índice.

RIESGO ATRIBUIBLE POBLACIONAL

Asumimos que sabemos cómo eliminar el tabaquismo y para eso necesitamos fondos para implementar la campaña.

La pregunta que nos harán será ¿cuál será el impacto del programa de cesación de tabaquismo sobre la incidencia de cáncer de pulmón en la población general?

¿Qué proporción de la incidencia de una enfermedad en una población total puede atribuirse a un factor de exposición específico?

¿Cuál sería el impacto total con un programa de prevención?

Esta pregunta es diferente a la anterior. Antes preguntamos cuál es el impacto en los enfermos de cáncer de pulmón; ahora preguntamos sobre el impacto en la población total (expuesto y no expuesto).

En una población general hay sujetos expuestos y no expuestos. Por lo tanto, el impacto que produce evitar el factor de exposición va a depender de la cantidad de expuestos que haya en esa población.

Para calcular el *RA* de la población total (*RAP*), el cálculo es similar al *RA* de los expuestos; pero en este caso usamos la incidencia de la enfermedad en la población total (expuestos y no expuestos) y le restamos el riesgo basal (no expuestos).

$$RA \text{ Poblacional} = (\text{Incidencia en la población total}) - (\text{Incidencia en no expuestos})$$

También es posible calcular la *Fracción Atribuible Poblacional (FAP)*

$$FAP = \frac{(\text{Incidencia en la población total}) - (\text{Incidencia en no expuestos})}{\text{Incidencia en los no expuestos}}$$

Para calcular el *RA* o *FA* poblacional debemos conocer la incidencia de la enfermedad en la población total y la incidencia de la enfermedad en no expuestos o bien los tres índices siguientes:

1. Incidencia en expuestos
2. Incidencia en no expuestos
3. Proporción de la población total que están expuestos (Prevalencia del factor de riesgo)

Supongamos que la incidencia de enfermedad coronaria en fumadores es de 28 por 1 000 sujetos y en los no fumadores de 17 por 1 000. Como no conocemos la incidencia

de enfermedad coronaria en la población total, buscamos información acerca de la prevalencia de tabaquistas en esta población.

Por ejemplo, encontramos que el 44% de la población fuma, por lo tanto, el 56% no es fumador.

Con estos datos ahora es posible calcular el *RAP*.

La incidencia en la población total puede calcularse de la siguiente manera:

$$\text{Incidencia en la población total} = \left(\text{Incidencia en expuestos de la enfermedad} \times \text{prevalencia del factor de exposición en expuestos} \right) + \left(\text{Incidencia de la enfermedad en no expuestos} \times \text{prevalencia del factor de exposición en no expuestos} \right)$$

$$\text{Incidencia en la población total} = \left(\frac{28}{1000} \times 0,44 \right) + \left(\frac{17}{1000} \times 0,56 \right) = \frac{22}{1000}$$

Ahora podemos calcular *RAP* y *FAP*

$$\text{RAP} = \frac{22}{1000} - \frac{17}{1000} = \frac{5}{1000}$$

$$\text{FAP} = \frac{22 - 17}{22} = 21,3\%$$

¿Qué concluimos con estos dos datos?

1. Si logramos eliminar el tabaquismo evitaremos en la población total 5 casos de enfermedad coronaria por cada 1 000 habitantes.
2. Reduciremos la incidencia de enfermedad coronaria en esta población en un 21,3% si evitamos el tabaquismo.

¿Qué sucedería si la prevalencia de tabaquismo fuera menor, por ejemplo, del 10%?

$$\text{En ese caso la incidencia total sería} = \left(\frac{28}{1000} \times 0,10 \right) + \left(\frac{17}{1000} \times 0,90 \right) = \frac{18}{1000}$$

$$\text{RAP} = \frac{18}{1000} - \frac{17}{1000} = \frac{1}{1000}$$

$$\text{FAP} = \frac{18 - 17}{18} = 5,5\%$$

Como vemos, ahora el impacto de la intervención es mucho menor, por lo tanto la prevalencia del factor de exposición en la población es un determinante fundamental del *RAP* y *FAP*.

COMPARACIÓN ENTRE RIESGO RELATIVO Y RIESGO ATRIBUIBLE

En la Tabla 23.1 se observan los datos que relaciona la mortalidad por cáncer de pulmón y enfermedad coronaria en fumadores y no fumadores.

Examinemos la mortalidad por cáncer de pulmón. La mortalidad en fumadores es de 140 por 100 000 sujetos y en no fumadores es de 10 por 100 000, el riesgo relativo es: $140/10 = 14$.

	Mortalidad por 100.000		Riesgo relativo	Riesgo atribuible	Fracción atribuible%
	Fumadores	No fumadores			
Cáncer de pulmón	140	10	14	130	92
Enfermedad coronaria	669	413	1,6	256	38

Tabla 23.1: Mortalidad por cáncer de pulmón y enfermedad coronaria

En el caso de la enfermedad coronaria, el riesgo relativo de muerte entre fumadores y no fumadores es $= 669/413 = 1,6$.

Veamos ahora el *RA* y *FA* en fumadores.

¿Cuántas muertes del total de muertos en fumadores puede atribuirse al hábito de fumar?

RA para cáncer de pulmón $= 140 - 10 = 130$ y la *FA* $= 92\%$

RA para enfermedad coronaria $= 669 - 413 = 256$ y la *FA* $= 38\%$

Concluimos que los riesgos relativos para cáncer de pulmón y enfermedad coronaria son muy diferentes.

El tabaquismo está mucho más asociado con cáncer de pulmón que con enfermedad coronaria; por lo tanto, en los fumadores es mayor el riesgo de cáncer de pulmón que de enfermedad coronaria.

Sin embargo, el riesgo atribuible para enfermedad coronaria es casi el doble que para cáncer de pulmón.

Si elimináramos el tabaquismo, podríamos evitar 256 muertes por enfermedad coronaria de un total de 669 muertes por 100 000 habitantes y en el caso de cáncer de pulmón, 130 muertes de 140 por 100 000.

Si lo vemos en forma de porcentaje diríamos que el 92% de las muertes por cáncer de pulmón podría evitarse y sólo el 38% de las muertes por enfermedad coronaria, si elimináramos el tabaquismo.

A primera vista podríamos decir que el impacto de un programa efectivo antitabaco beneficiaría más al cáncer de pulmón (*RR* $= 14$ y *FA* 92%); pero si miramos el beneficio neto en muertes evitadas vemos que se previenen dos veces más muertes por enfermedad coronaria (256 vs. 130 por 100 000 habitantes).

Esto se debe a que la mortalidad basal por enfermedad coronaria es más alta que la mortalidad por cáncer de pulmón (669 vs. 140) y que el riesgo atribuible es mayor en la enfermedad coronaria (256 vs. 130).

En síntesis, el riesgo atribuible nos informa la cantidad de casos que se podrían evitar en una población expuesta al factor de exposición o de riesgo, mientras que el riesgo atribuible y la fracción atribuible poblacional nos informa sobre el impacto que tendría en la población total el retiro del factor de exposición o de riesgo, lo que es de suma utilidad para planear estrategias públicas de prevención de enfermedades o eventos perjudiciales relacionados con la salud.

Cuando en los estudios de cohortes la Incidencia de una enfermedad o de eventos es baja o no varía mucho con el correr del tiempo, los métodos para evaluar riesgo o una asociación, como riesgo relativo, pueden aplicarse sin cometer errores.

Sin embargo, cuando el evento varía rápido con el tiempo, por ejemplo, muerte o remisión luego del diagnóstico de cáncer de mama, el interés está en cuantificar el tiempo transcurrido desde el momento del diagnóstico hasta la muerte o hasta la remisión de la enfermedad o tiempo al evento.

A este tipo de análisis se lo denomina *análisis de supervivencia*.

Este tipo de análisis se usa para describir la historia natural de una enfermedad, evaluar nuevas intervenciones y (a nivel demográfico) calcular la expectativa de vida de una población.

En general, a estos estudios se los denomina estudios pronósticos.

Algunos ejemplos son:

1. Tiempo hasta la muerte por alguna causa determinada desde el diagnóstico (muerte por insuficiencia cardíaca o reinfarto después de un infarto de miocardio)
2. Tiempo a la recuperación (tiempo hasta la curación en la hepatitis A)
3. Tiempo de supervivencia en población mayor de 60 años
4. Prolongación del tiempo hasta la muerte con alguna intervención (quimioterapia luego de diagnóstico de cáncer) comparado con tiempo de muerte en otra población sin intervención

TIEMPO DE SOBREVIDA

En la mayoría de los estudios de seguimiento hay individuos que se pierden del seguimiento, por ejemplo, porque se mudan o sufren otro evento que no es el de interés para el estudio (muerte por otra causa diferente a la definida como evento). Como el evento de interés no pudo observarse, no sabemos si el sujeto hubiera sufrido o no el evento una vez terminado el seguimiento.

En este caso se dice que el tiempo al evento está censurado.

Para un sujeto que se perdió en el seguimiento o sufrió otro evento, el verdadero tiempo al evento o supervivencia se desconoce.

En un estudio de seguimiento, lo único que podemos decir es que la probabilidad de supervivencia para los sujetos incluidos es, *al menos*, el tiempo desde el ingreso al estudio hasta:

1. El final del seguimiento, si continuó en el estudio sin eventos.
2. El momento en que se perdió o sufrió otro evento.
3. El momento en que sufrió el evento de interés.

El principal objetivo del análisis de sobrevida es obtener una estimación de la sobrevida experimentada por una población a partir de una muestra representativa. Se desea estimar el pronóstico de una determinada población en un lapso determinado de tiempo. Por ejemplo, el pronóstico de muerte a 5 años a partir del diagnóstico de cáncer de mama en una población tratada con quimioterapia.

Para lograr esto se calcula la probabilidad de sobrevida teniendo en cuenta el número de sujetos que inicialmente ingresaron al estudio (sujetos en riesgo de sufrir el evento) y la cantidad de sujetos que sufrieron el evento de interés ajustados al número de sujetos censurados (perdidos o que sufrieron otro evento).

La probabilidad de muerte o sobrevida se calcula ajustando el número de sujetos en riesgo según la pérdida de sujetos que se observó en el estudio.

Si al inicio del estudio hay 200 sujetos y al final del estudio, por ejemplo 2 años, se perdieron 20 y se produjeron 40 muertes, el cálculo de sobrevida a 2 años se hará sobre 140 sujetos.

Esto nos da una idea de resumen cruda de la probabilidad de sobrevida al final del estudio. Dependiendo de si se recabó la fecha en que ocurrió el evento o no, se utilizan dos métodos diferentes: *tablas de sobrevida* y *curvas Kaplan-Meier*.

TABLAS DE SOBREVIDA

Si no se recabó la fecha en que ocurrió el evento, se utilizan las tablas de sobrevida.

Si el tiempo de seguimiento es de un año, podemos dividirlo en meses. Para calcular la probabilidad de eventos por cada mes necesitamos saber:

1. El número de sujetos en riesgo al inicio de cada mes
2. El número de muertes en ese período
3. El número de sujetos censurados en ese período
4. El número de sobrevivientes

Supongamos que el estudio comienza con 109 sujetos. En el primer mes murieron 4 (evento de interés), ninguno fue censurado; por lo tanto quedan al final del primer mes 105 sujetos.

La probabilidad de muerte es:

$$\text{Probabilidad de muerte} = \frac{n^\circ \text{ de muertes}}{n^\circ \text{ sujetos en riesgo} - (0,5 * n^\circ \text{ de sujetos censurados})}$$

Para el ejemplo la probabilidad de muertes es:

$$\text{Probabilidad de muerte} = \frac{4}{109 - (0,5 * 0)} = 0,037$$

Si la probabilidad de muerte al final del primer mes es de 3,7%, *la probabilidad de sobrevida es* $= 1 - \text{probabilidad de muerte} = 1 - 0,037 = 0,96$ o 96%.

En el segundo mes el número de sujetos en riesgo ahora es de 105, y durante este mes ocurrieron 5 muertes sin sujetos perdidos o censurados; por lo tanto, al final del segundo mes quedan 100 sujetos, aplicando la fórmula la probabilidad de muerte para este período es de 0,048 y la sobrevida de 0,95.

En el tercer mes que se inicia con 100 sujetos, se mueren 4 y se pierden de seguimiento 2 y sufren otro evento 3 sujetos, por lo que hay 5 sujetos censurados. Al final del tercer mes quedan 91 sujetos. La probabilidad de muerte es de 0,041 y la probabilidad de sobrevida de 95,9%.

Entonces la probabilidad de sobrevida en el primer mes es del 96%, en el segundo mes del 95% y en el tercero de casi el 96%; la probabilidad acumulativa de sobrevida desde el inicio del estudio hasta el tercer mes es el producto de las probabilidades por separado = $0,96 * 0,95 * 0,96 = 0,88$.

La probabilidad de sobrevivir a 3 meses es de 88%.

En otras palabras, el pronóstico de sobrevida de esta población es del 88% a 3 meses.

En la segunda columna de la Tabla 24.1 se observa el número de sujetos vivos al inicio de cada intervalo de tiempo en que se dividió el seguimiento total. En la tercera columna vemos el número de sujetos que sufrió el evento de interés en cada intervalo de tiempo y en la cuarta columna el número de sujetos perdidos o censurados.

El cálculo del número de personas en riesgo al inicio de cada intervalo de tiempo es la resta de la cuarta columna menos la segunda, dividido 2, lo cual es equivalente a multiplicar por 0,5 como en la fórmula de cálculo de probabilidad de muerte. Se divide por 2 porque se considera que la pérdida de sujetos ocurrió en la mitad del periodo de tiempo.

El riesgo de muerte se calcula dividiendo el número de muertos por el número de sujetos vivos de ese período y la probabilidad de sobrevida como su inversa matemática.

La probabilidad de muerte acumulativa se calcula multiplicando sucesivamente las probabilidades individuales de cada período.

A este método de derivar la sobrevida acumulativa a datos agrupados se lo denomina tablas de sobrevida y es una forma de presentar estos datos tabulados. Otra forma de presentarlos es a través de gráficos.

Intervalo de tiempo (meses) (1)	N° sujetos vivos al inicio del período (2)	Muertos durante el período (3)	N° de perdidos en el período (4)	N° de personas en riesgo (5) = (2)-(4) / 2	Riesgo de muerte en el período (6) = (3)/(5)	Chance de sobrevivir (7) = 1 -(6)	Chance acumulativa de sobrevivir desde el inicio (8) = (8, intervalo previo) x (7)
1	240	12	0	240	0,050	0,95	0,95
2	228	9	0	228	0,039	0,96	0,91
3	219	17	1	218	0,077	0,92	0,84
4	201	36	4	199	0,180	0,81	0,68
5	161	6	2	160	0,037	0,96	0,66
6	153	18	7	149,5	0,120	0,87	0,58
7	128	13	5	125,5	0,103	0,89	0,52
8	110	11	3	108,5	0,101	0,89	0,47
9	96	14	3	94,5	0,148	0,85	0,40
10	79	13	0	79	0,164	0,83	0,33
11	66	15	4	64	0,234	0,76	0,25
12	47	6	1	46,5	0,129	0,87	0,22

Tabla 24.1: Tabla de sobrevida para 240 pacientes con diagnóstico de cáncer de mama tratados con radioterapia

CURVAS DE SOBREVIDA O KAPLAN-MEIER

En la mayoría de los estudios de cohortes es muy probable poder obtener la información precisa de la fecha en que ocurrió el evento y del día en que se censuró un sujeto.

En esta situación se pueden utilizar las curvas de Kaplan-Meier para estimar la supervivencia acumulativa inmediatamente antes de producido cada evento individual.

Esta estimación se usa para construir una curva que tiene como característica ser escalonada, a diferencia de la curva que se podría construir con los datos de las tablas de supervivencia que es una curva suave sin saltos. (Figuras 24.1 y 24.2)

En este caso la probabilidad de supervivencia se calcula en cada día en que ocurrió un evento, justo antes de que ocurra. En los días en que no hay eventos la supervivencia se considera 100%.

A diferencia de las tablas de supervivencia, los intervalos de tiempo no son prefijados: están definidos por la ocurrencia de los eventos.

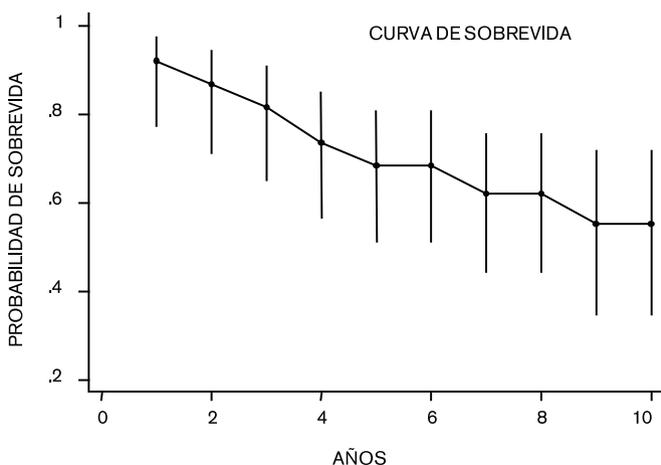


Figura 24.1: Curvas de tablas de supervivencia y Kaplan-Meier

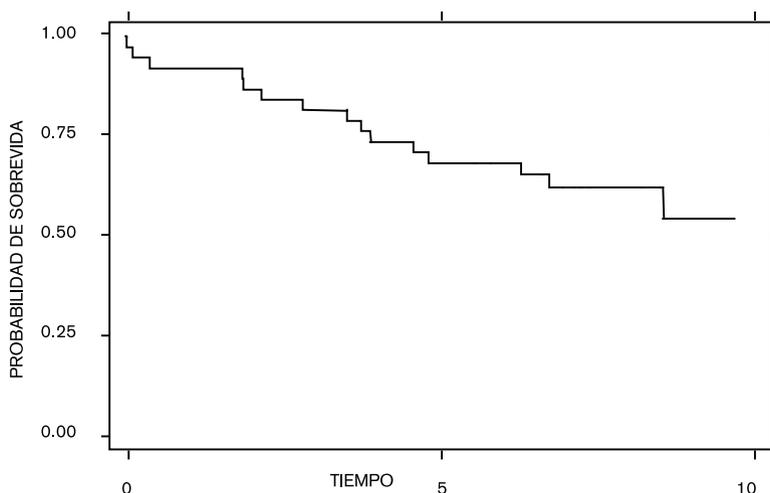


Figura 24.2: Curva de Kaplan-Meier

Intervalo de tiempo (días)	Nº de sujetos vivos	Nº de eventos	Nº de censurados
0	109	0	0
18	109	1	0
207	108	1	0
284	107	1	0
285	106	0	1
300	105	0	0

Tabla 24.2: Datos para curvas de Kaplan-Meier

En la Tabla 24.2 se observa el número de sujetos vivos en cada intervalo de tiempo en que ocurrió un evento, los días en que ocurrió un evento, el número de eventos, el número de sujetos censurados.

Al inicio del estudio hay 109 sujetos, tiempo cero, sin eventos ni sujetos censurados, la sobrevida es 100%.

El día 18 ocurre un evento (intervalo de tiempo desde el día 0 al 18). Justo antes del evento sigue habiendo 109 sujetos; si aplicamos la fórmula:

$$\text{Probabilidad de muerte} = \frac{n^{\circ} \text{ de muertes}}{n^{\circ} \text{ sujetos en riesgo} - (0,5 * n^{\circ} \text{ de sujetos censurados)}}$$

Tendremos una probabilidad de muerte igual a:

$$\text{Probabilidad de muerte} = \frac{1}{109 - (0,5 * 0)} = 0,00917$$

$$\text{y una sobrevida} = 1 - 0,00917 = 0,99$$

Si calculamos la probabilidad de muerte para el período desde el día 18 al 207, en que ocurre otro evento, tenemos 108 sujetos y no contamos el evento nuevo. Calculamos la probabilidad de muerte como:

$$\text{Probabilidad de muerte} = \frac{1}{108 - (0,5 * 0)} = 0,00926$$

La sobrevida para este período (día 18 a 207) es de 0,99.

Para el período entre el día 207 y el 284, la probabilidad de muerte es de 0,0093 y la de sobrevida es del 0,99.

La sobrevida acumulativa al día 284 es el múltiplo de cada cálculo de sobrevida de cada intervalo de tiempo en que ocurrió un evento = $0,99 \times 0,99 \times 0,99 = 0,97$

Si calculamos la probabilidad de muerte en el período donde no ocurrió ningún evento, por ejemplo el día 300, observaremos que esta es igual a cero y la sobrevida igual a 1.

$$\text{Probabilidad de muerte} = \frac{0}{105 - (0,5 * 0)} = 0,00$$

Las observaciones censuradas no causan reducción de la sobrevida, pero sí se tienen en cuenta para el cálculo del número de sujetos en riesgo en el preciso momento en que ocurre una censura.

Las curvas de Kaplan-Meier siempre comienzan en 1 o 100% de sobrevida, y llegan a cero sólo si todos los sujetos han muerto o han sufrido el evento de interés.

Como toda estimación estadística, el cálculo de la probabilidad de muerte o sobrevida también tiene una variabilidad.

Es por esto que, cuando informamos que un grupo tiene una probabilidad de sufrir un evento del 97%, debemos informar el intervalo de confianza del 95%; es decir que la probabilidad puede estar en el 95% de los casos entre dos valores, por ejemplo de 95 a 100%.

El valor de probabilidad de muerte también tiene un error estándar con el cual calculamos el IC 95% y una mediana, que en este caso se define como el momento en que la mitad de los sujetos han muerto o sufrido el evento de interés. (Tabla 24.3)

Intervalo de tiempo	Vivos	Muertos	Perdidos	Sobrevida	Error St	IC 95%
0	38	3	0	0,9211	0,0437	0,7749 0,9738
1	35	2	0	0,8684	0,0548	0,7123 0,9430
2	33	2	9	0,8158	0,0629	0,6521 0,9076
3	31	3	0	0,7368	0,0714	0,5661 0,8488
4	28	2	0	0,6842	0,0754	0,5115 0,8067
5	26	0	3	0,6842	0,0754	0,5115 0,8067
6	23	2	3	0,6206	0,0807	0,4430 0,7561
7	18	0	6	0,6206	0,0807	0,4430 0,7561
8	12	1	6	0,5516	0,0968	0,3462 0,7163
9	5	0	5	0,5516	0,0968	0,3462 0,7163

Tabla 24.3: Tabla de sobrevida

COMPARACIÓN DE CURVAS DE KAPLAN-MEIER. PRUEBA LOG RANK TEST

Supongamos que tenemos dos grupos de sujetos, un grupo con enfermedad coronaria y otro sin ella.

Como vemos en la Figura 24.3, los sujetos con enfermedad coronaria tienen menos sobrevida a 10 años (aproximadamente 55 vs. 80%).

Si queremos saber si estas diferencias son estadísticamente significativas, la prueba que debemos aplicar se llama *log rank test*.

Es muy parecido, desde el punto de vista matemático, a la prueba de chi cuadrado. Compara las muertes esperadas (muertes ocurridas sólo por azar) vs. las observadas en cada grupo y determina si existen diferencias significativas a través del valor de p . (Tabla 24.4)

De acuerdo con el valor de p obtenido, podemos decidir si las curvas son diferentes o no. Lo que se desea comprobar es si la diferencia de sobrevida se debió al azar o a la presencia de enfermedad coronaria.

Si bien con el *long rank test* podemos definir si las curvas son diferentes o no, lo que no podemos es cuantificar la diferencia.

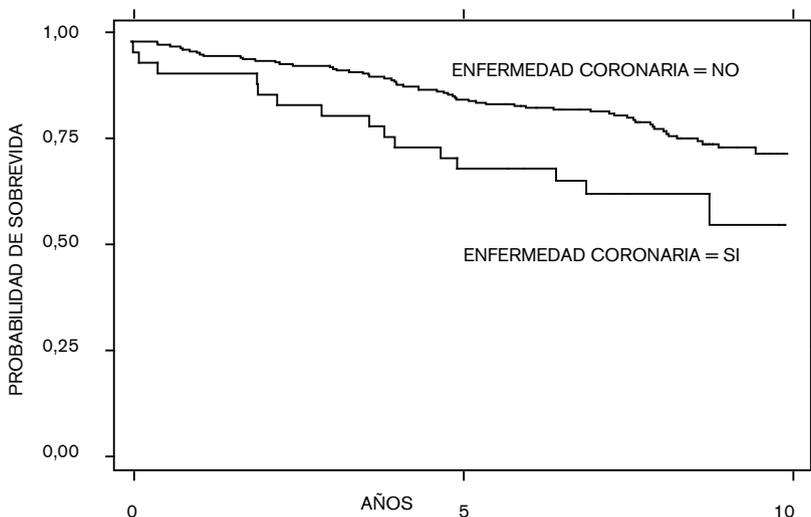


Figura 24.3: Curvas de Kaplan-Meier según enfermedad coronaria

Enfermedad coronaria	Muertes observadas	Esperadas
No	59	66
Sí	15	8
Total	74	74

Tabla 24.4: Ejemplo de *long rank test*

Chi cuadrado = 3,66 $p = 0,0099$

HAZARD RATIO

Para determinar el riesgo de muerte que tienen los enfermos coronarios vs. los no coronarios recurrimos a la estimación de un parámetro, muy parecido al riesgo relativo en su interpretación, llamado *Hazard Ratio* (HR).

Esta medida de estimación de riesgo se obtiene mediante un cálculo complejo en el cual se utiliza la regresión de Cox.

Para el ejemplo anterior, el riesgo de muerte medido por el HR de los pacientes con enfermedad coronaria es de 2,07 con IC 95% de 1,17 a 3,66.

Concluimos entonces que los pacientes con enfermedad coronaria tienen entre un 17 y un 360% más de probabilidad de morir comparado con los sujetos sin enfermedad coronaria.

DEFINICIÓN

Se define como tamizaje o *screening* al proceso de identificar a través de un estudio o prueba de cualquier categoría (examen físico, electrocardiograma, análisis de sangre, etc.) a individuos asintomáticos portadores de estadios tempranos de una enfermedad o algún precursor de ella.

Por ejemplo, la detección de cáncer de cuello de útero por PAP (detecta lesiones precancerosas), el tacto rectal para la detección de cáncer de próstata, la búsqueda de factores de riesgo cardiovascular, etc.

La importancia de realizar un *screening* es que en algunas enfermedades reduce la morbimortalidad.

Muchas veces se utiliza la palabra tamizaje o *screening* como sinónimo de *detección precoz* de una enfermedad. La diferencia es que en este último caso los sujetos ya tienen algún síntoma pero no han consultado y, por lo tanto, se desconoce su existencia. En este caso debemos aplicar una prueba diagnóstica en lugar de una prueba de *screening*.

Por ejemplo, podríamos hacer diagnóstico de casos de diabetes, si el individuo ya tiene síntomas de esta enfermedad. Hacer un *screening* o tamizaje de diabetes sería determinar la glucemia en ayunas en aquellos pacientes que no presentan síntomas o signos de esta enfermedad.

Las pruebas de tamizaje se ofrecen a poblaciones que están aparentemente sanas. Una *prueba diagnóstica* se realiza en individuos que presentan síntomas para confirmar la existencia de una enfermedad.

El tamizaje se refiere a la aplicación de una prueba en una población asintomática con el propósito de clasificarla en relación a la *probabilidad* que tienen los sujetos de padecer o no una determinada enfermedad.

La prueba de tamizaje *no diagnóstica por sí misma*; aquellos individuos con pruebas positivas deberán realizarse estudios diagnósticos para determinar si están realmente enfermos o no.

Una asunción implícita es que la detección precoz, en el período asintomático de la enfermedad, determina un mejor pronóstico. Dada esta presunción, la detección precoz de las enfermedades se considera hoy en día una práctica de rutina y un importante aspecto de buena calidad médica, que también implica prevención de otras enfermedades.

Sin embargo, esto no significa que el tratamiento precoz siempre determine un mejor pronóstico. Está demostrado que la supervivencia del cáncer de pulmón no cambia con su detección precoz.

Antes de aplicar una prueba de tamizaje masivamente es necesario tener en cuenta, no sólo las características de la enfermedad, los riesgos y beneficios de la prueba y sus costos sino futuras consecuencias de los estudios diagnósticos que se deben aplicar a la población que resultó positiva para el tamizaje.

CRITERIOS PARA INTRODUCIR UN PROGRAMA DE TAMIZAJE

Las enfermedades que aplican para implementar una prueba de tamizaje son aquellas que cumplen con alguna de las siguientes condiciones:

1. Las consecuencias de la enfermedad son graves, no sólo en términos de morbimortalidad sino en costo/efectividad.

La asignación de recursos para un programa de *screening* o tamizaje debe ser justificada en términos de eliminar o disminuir consecuencias adversas en la salud. No sólo es importante que la enfermedad sea grave en términos de una alta mortalidad, sino también en términos de las consecuencias que pueda causar su no tratamiento. Por ejemplo, el hipotiroidismo congénito, si bien es infrecuente y no tiene alta mortalidad, trae consecuencias devastadoras; por otro lado la litiasis biliar es muy común pero no es grave y puede ocurrir que nunca se transforme en sintomática, por lo tanto una prueba de tamizaje no estaría justificada.

Con relación a los aspectos éticos, la consecuencia de no diagnosticar y tratar de forma temprana una enfermedad debe balancearse con los riesgos y molestias que la prueba de tamizaje le pueda provocar al paciente y el impacto en salud pública de la enfermedad.

2. El tratamiento iniciado en el período asintomático debe tener beneficios claros en términos de reducción de morbimortalidad, comparado con un tratamiento iniciado en la fase sintomática de la enfermedad.

El tratamiento temprano debe demostrar que modifica el pronóstico de la enfermedad. Para evaluar este punto es necesario considerar la historia natural de la enfermedad. Esta debe tener un período preclínico detectable por algún método. En el caso del cáncer de pulmón el tratamiento administrado en la fase preclínica no modifica el pronóstico comparado con el tratamiento administrado en la fase sintomática. A diferencia del cáncer de pulmón, el cáncer de piel es completamente curable si se detecta tempranamente.

3. Su prevalencia en la población donde se aplica la prueba de tamizaje debe ser alta.

Esto se refiere a la prevalencia de las lesiones preclínicas y guarda relación con los costos de un programa de *screening* (número de casos detectados vs. número de sujetos testeados). Es por eso que los programas de *screening* se realizan en poblaciones consideradas de alto riesgo para la enfermedad que se quiere prevenir. Por ejemplo, cáncer de vejiga en trabajadores expuestos a sustancias carcinógenas, enfermedad coronaria en sujetos con múltiples factores de riesgo coronario.

Un ejemplo de enfermedad que cumple con los tres criterios es la hipertensión arterial. Es grave, debido a que puede producir accidente cerebrovascular con secuelas serias y muerte. El tratamiento en la etapa temprana de la enfermedad cambia el pronóstico de la misma: disminuye la tasa de accidentes cerebrovasculares, hipertrofia ventricular, aparición de arritmias, etc. Por último, la prevalencia de la hipertensión arterial es muy alta.

No todos los criterios mencionados son necesarios para justificar un programa de tamizaje.

La fenilcetonuria es una enfermedad de muy baja prevalencia, pero sus consecuencias son muy graves e irreversibles.

La combinación de gravedad con la disponibilidad de una prueba de tamizaje simple y barata justifica el *screening* a nivel de salud pública.

Otro aspecto fundamental en los programas de *screening* son ciertas características que debe poseer la prueba utilizada.

En la mayoría de los casos es probable que el estadio preclínico de una enfermedad tenga una prevalencia baja, y esto tiene implicancias muy importantes para la selección de una prueba apropiada y la interpretación de sus resultados.

La prueba de tamizaje ideal sería aquella que detectara la mayoría de los casos con la enfermedad como prueba positiva y la mayoría de los casos sin enfermedad como prueba negativa.

Para que un programa de *screening* sea exitoso no basta con aplicar la prueba a una población con determinadas características, además debe utilizarse una prueba adecuada, que detecte de manera confiable la población tanto enferma como sana.

Una prueba de *screening* debe ser barata, fácil de implementar y de administrar al paciente, conllevar un mínimo riesgo y su resultado debe ser *válido, confiable y repetible*.

Al igual que una prueba de *screening*, las pruebas diagnósticas también deben poseer características de validez, confiabilidad y repetibilidad, pero a diferencia de los estudios de *screening*, pueden no ser baratos y conferir ciertos riesgos. En este caso se admiten estas últimas características porque estamos frente a una situación de diagnóstico que decide el futuro del paciente.

SENSIBILIDAD Y ESPECIFICIDAD

La validez de una prueba de *screening* o de diagnóstico se evalúa a través de su *sensibilidad* y su *especificidad*.

Dicho en otras palabras, se espera que la prueba haga lo que tiene que hacer: discriminar personas con lesiones preclínicas como prueba positiva y sin lesiones como prueba negativa.

ESTUDIOS CON RESULTADOS DICOTÓMICOS (POSITIVO O NEGATIVO)

En términos matemáticos, la sensibilidad es la proporción de sujetos con prueba positiva obtenidos de una muestra donde todos son sujetos enfermos.

$$\text{Sensibilidad} = \frac{\text{n}^\circ \text{ de sujetos con prueba positiva}}{\text{n}^\circ \text{ total de sujetos enfermos}} \times 100$$

Si aplicamos una prueba de diagnóstico a 100 pacientes enfermos de hepatitis B y sólo 50 dan positivo, la sensibilidad de esta prueba es de 50% y la proporción de falsos negativos es de 50%. Esta prueba no tiene una gran capacidad para discriminar enfermos; la mitad de estos no se detecta.

La especificidad de una prueba es la proporción de sujetos con prueba negativa obtenidos de una muestra donde todos son sujetos sanos.

$$\text{Especificidad} = \frac{\text{n}^\circ \text{ de sujetos con prueba negativa}}{\text{n}^\circ \text{ total de sujetos sanos}} \times 100$$

En el mismo ejemplo anterior, si ahora tenemos 100 sujetos sanos y sólo 50 dan negativo, la prueba tiene 50% de especificidad. En este caso, 50% de los sujetos serán positivos para la prueba, por lo que la proporción de falsos positivos es muy alta. En este ejemplo

Resultado de la biopsia			
Resultado de la nueva prueba	Enfermos	Sanos	Total
Positiva	80	100	180
Negativa	20	800	820
Total	100	900	1.000
Sensibilidad	80%		
Especificidad		89%	

Tabla 25.1: Cálculo de sensibilidad y especificidad

estamos sobreclasificando sujetos enfermos (falsos positivos), que luego serán derivados a una prueba diagnóstica, que puede no sólo ser costosa sino implicar riesgos para el paciente. Por lo tanto, las consecuencias de aplicar esta prueba son inaceptables.

Es importante entender que el criterio de enfermedad va variando a medida que nuevas pruebas se van desarrollando. La población denominada como enferma o sana resulta de aplicar una prueba que es considerada como *gold standar* o *método de referencia* para la detección de esa patología.

En la práctica, una nueva prueba de *screening* o diagnóstica se compara siempre con la prueba que hasta el momento se considera con mejor sensibilidad y especificidad para la clasificación de sujetos como sanos o enfermos.

En una época pudo haber sido la auscultación cardíaca el método de referencia para detectar sujetos con estenosis mitral y ahora lo es el ecocardiograma.

Supongamos que estamos probando una nueva prueba para implementar un *screening* de cáncer de próstata. El método de referencia que se toma es la punción biopsia y la nueva prueba es la característica de la próstata a la palpación. Tomamos entonces 1 000 sujetos, de los cuales la biopsia dio positiva en 100 y en 900 dio negativa. (Tabla 25.1)

La sensibilidad de la nueva prueba (tacto rectal) es de: $80/100 = 80\%$. Dado que la sensibilidad es el número de sujetos enfermos sobre total de enfermos, en este ejemplo tenemos 80 sujetos clasificados por la nueva prueba como positivos de un total de 100 detectados por biopsia.

La especificidad es del 89% (800/900).

La nueva prueba o método detectó como sanos a 800 sujetos de un total de 900 clasificados como sanos por la biopsia.

Test de referencia (<i>Gold standar</i>)		
Resultado del nuevo test	Enfermos	Sanos
Positivo	Tiene la enfermedad y el nuevo test + Verdaderos positivos (VP)	No tienen la enfermedad y el nuevo test + Falsos positivos (FP)
Negativo	Tienen la enfermedad y el nuevo test – Falsos negativos (FN)	No tienen la enfermedad y el nuevo test – Verdaderos negativos (VN)

Tabla 25.2: Comparación de resultados de un test con variable dicotómica vs. verdadero estatus de la enfermedad

Cuando se conduce un programa de *screening*, este se aplica a una gran cantidad de sujetos, algunos enfermos y otros sanos realmente (verdaderos positivos y verdaderos negativos).

La nueva prueba, al no ser perfecta, clasificará algunos sujetos sanos como enfermos (falso positivo) y algunos enfermos como sanos (falso negativo). (Tabla 25.2)

$$\boxed{\text{Sensibilidad} = \frac{VP}{VP + FN} \quad \text{Especificidad} = \frac{VN}{VN + FP}}$$

¿Cuántos falsos negativos produjo la prueba en el ejemplo anterior? Sobre un total de 100 enfermos, la nueva prueba clasificó a 20 sujetos como negativos, siendo estos verdaderos enfermos. La tasa de falsos negativos fue de 20%.

Matemáticamente se verifica que:

$$\boxed{1 - \text{Sensibilidad} = \text{Falsos negativos}}$$

En el ejemplo = $1 - 0,80 = 0,2 \times 100 = 20\%$

El problema de tener falsos negativos radica en que, si la enfermedad es grave y existe un tratamiento efectivo, estos sujetos se ven privados de él. Si el cáncer de próstata fuera totalmente curable en el período preclínico, estos sujetos podrían morir por falta de detección precoz y tratamiento adecuado.

La importancia de un resultado falso negativo depende de la naturaleza y gravedad de la enfermedad y de la disponibilidad de un tratamiento precoz efectivo.

La proporción de falsos positivos fue del 11%, ya que detectó como positivos a 100 sujetos de 900 con biopsia negativa.

En este caso se puede obtener la tasa de falsos positivos como:

$$\boxed{1 - \text{Especificidad} = \text{Falsos positivos}}$$

Aplicado a los resultados anteriores:

Falsos positivos = $1 - 0,89 = 0,11 \times 100 = 11\%$

El problema con la cantidad de falsos positivos que detecte una prueba es que toda esta población de falsos positivos será sometida a estudios diagnósticos más sofisticados, caros y algunas veces con más riesgo y molestias para el paciente. Esto redundará en posibles complicaciones graves para los pacientes y un gasto en salud innecesario.

Además, estas personas sufren períodos de ansiedad y angustia prolongados hasta que la prueba diagnóstica descarta el resultado de la prueba de *screening*.

Muchas veces el rótulo conferido por la prueba de *screening* es muy difícil de borrar.

Por ejemplo, si a un chico se le detecta un soplo cardíaco y luego el ecocardiograma dice que es funcional, quizás toda su vida haga mención de este primer diagnóstico y en el colegio siempre le soliciten otras evaluaciones aunque los padres digan que se ha descartado alguna enfermedad cardíaca.

La decisión de aceptar una nueva prueba como válida depende de un balance entre su sensibilidad y su especificidad.

Hay que pesar el riesgo de sobrediagnosticar vs. no detectar algunos casos enfermos.

Si la prueba es diagnóstica, se le pedirá mucha más especificidad que sensibilidad. Lo que queremos es descartar, con la mayor seguridad posible, la enfermedad en la mayor cantidad de pacientes posibles. Por supuesto que siempre tendremos falsos negativos y falsos positivos, pero en una proporción muy diferente a la que pueden presentar las pruebas de *screening*.

ESTUDIOS CON RESULTADOS CONTINUOS

Hasta ahora hemos analizado pruebas de *screening* o de diagnóstico que sólo tienen dos resultados posibles: positivos o negativos.

Muy frecuentemente se realiza *screening* de enfermedades que se valoran a través de variables continuas, como la tensión arterial para identificar hipertensos o la glucemia para detectar diabéticos tempranos.

En estos casos se debe determinar un punto de corte, desde el cual diremos que un sujeto es hipertenso o es diabético.

En una población de 20 diabéticos y 20 no diabéticos a la cual se le realiza una determinación de glucemia, se puede observar que los diabéticos tienen valores más elevados de glucemia que los no diabéticos, sin un punto de corte claro entre ambos. (Figura 25.1 A)

No obstante, debemos seleccionar un punto de corte para clasificar a la población como positiva o negativa.

Supongamos que el punto de corte que se establece es alto, digamos 180 mg% de glucemia para decir que la prueba es positiva y el paciente es diabético. (Figura 25.1 B)

En este caso muchos diabéticos no serán identificados como tales y muchos no diabéticos serán correctamente clasificados como negativos.

Aplicando los resultados a una tabla 2 x 2 se obtiene una sensibilidad de 25% (5/20) y una especificidad de 90% (18/20).

Con estos parámetros se concluye que pocos diabéticos serán diagnosticados como tales, mientras que muchos no diabéticos serán correctamente clasificados.

Si en cambio determinamos un punto de corte bajo, por ejemplo 90 mg%, muy pocos diabéticos serán mal clasificados, pero muchos no diabéticos serán tomados como positivos.

En este caso la sensibilidad es de 85% (17 enfermos de 20 fueron correctamente clasificados) y la especificidad de sólo 30%: sólo 6 sanos de 20 fueron clasificados correctamente o lo que es lo mismo, 14 sujetos fueron clasificados como enfermos (70% de falsos positivos). (Figura 25.1 C)

Como vemos, el punto de corte seleccionado influencia claramente la sensibilidad y la especificidad de una prueba *screening* o diagnóstica. También se puede ver cómo al aumentar un parámetro el otro disminuye. Podemos elegir un punto de corte con alta sensibilidad; en ese caso casi todos los enfermos se clasificarán como tales, pero también se clasificará erróneamente a muchos sanos (baja especificidad), por lo que la tasa de falsos positivos será alta.

Como toda estimación en estadística, el valor hallado no es *absoluto*. La sensibilidad y especificidad tienen su IC 95%, en el caso mostrado, la sensibilidad del punto de corte de 90 mg/dL para glucemia podría variar entre 80 a 90% y la especificidad entre 25 a 35%.

El valor del IC 95% nos ofrece una idea de entre qué valores podría estar la verdadera sensibilidad y especificidad de una prueba y el grado de precisión con la que hemos estimado estos parámetros de validez.

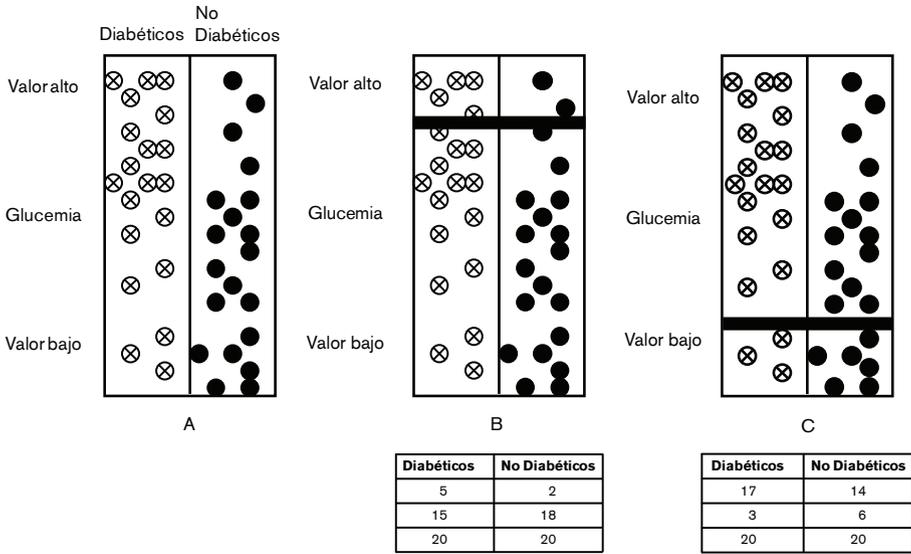


Figura 25.1: Efectos de la elección de distintos puntos de corte para una prueba de *screening* o diagnóstica

CURVAS ROC

En la vida real no existe una clara separación de valores entre enfermos y sanos. Por lo tanto el punto de corte debe seleccionarse teniendo en cuenta el valor que mejor sensibilidad y especificidad ofrezca.

Esto se realiza mediante la construcción de una curva conocida como *curva ROC (Receiver Operating Characteristic)*.

Lo que hace esta curva es determinar qué punto de corte abarca la mayoría de los sujetos bien clasificados, obteniéndose de esta manera la mejor sensibilidad y especificidad.

Supongamos que estamos haciendo un *screening* para detectar tempranamente insuficiencia renal, o queremos una prueba diagnóstica más fácil de aplicar para el diagnóstico de insuficiencia renal. La nueva prueba que usaremos será los valores de creatinina en sangre y el método patrón o *gold standar* será la medición del filtrado glomerular.

¿Cuál es el mejor punto de corte a elegir?

Lo que hacemos es dividir a los sujetos en sanos y enfermos de acuerdo con sus valores de filtrado glomerular. La curva busca el punto de corte de la nueva prueba que incluya la mayoría de los sujetos, tanto sanos como enfermos, correctamente clasificados. (Figura 25.2)

Si la nueva prueba fuera perfecta cubriría el 100% de la superficie de la curva, clasificaría perfectamente enfermos y sanos. Como esto no es posible, en biología se busca el punto de corte que contenga a la mayoría de los sujetos bien clasificados.

En otras palabras, se compara la sensibilidad vs. 1-especificidad, que es una forma de expresar los falsos positivos.

El punto de corte de creatinina tendrá solo un 4% de falsos positivos y un 14% de falsos negativos.

Dadas las circunstancias, aunque la prueba no sea perfecta, es el mejor balance que se puede conseguir.

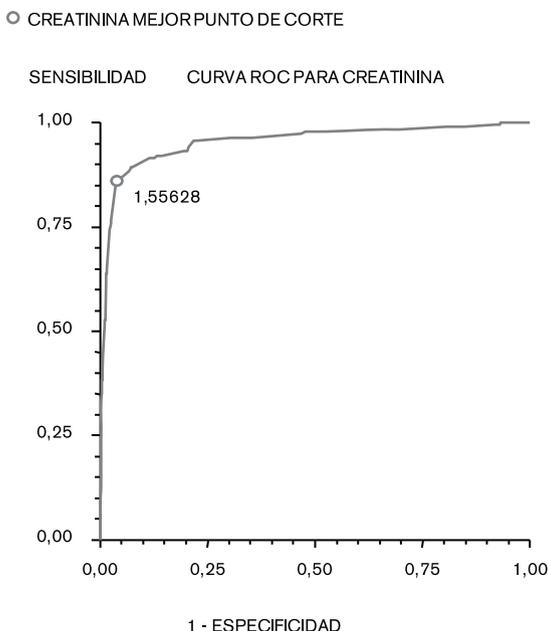


Figura 25.2: Curva ROC. Determinación del mejor punto de corte para creatinina plasmática

Por lo tanto, es de vital importancia para la validez de una prueba (evaluada a través de su sensibilidad y especificidad) la elección del punto de corte.

Los resultados de la curva de la Figura 25.2 muestran que el valor 1,5 tiene la mejor sensibilidad (86%) y la mejor especificidad (96%) y un área bajo la curva de 82% IC 95% 78-85%, lo que significa que el 82% de los sujetos serán bien clasificados (sanos como sanos y enfermos como enfermos) con este punto de corte de creatinina, pudiendo variar esta correcta clasificación entre 78 a 85% solo por azar.

Para que una curva ROC sea considerada como buena, en términos de clasificar bien a los sujetos, debe tener un área bajo la curva $\geq 70\%$. Por debajo de estos valores la curva no se considera útil para discriminar correctamente a los sujetos sanos de los enfermos.

En la práctica habitual se trata de elegir pruebas con alta sensibilidad para un primer *screening* y luego una prueba con alta especificidad para realizar el diagnóstico definitivo, que sólo se realiza en aquellos que dieron positivo en el *screening*.

El mejor ejemplo es la detección de HIV: primero se utiliza el método de ELISA, que es muy sensible, y sólo a aquellos sujetos que fueron positivo se les realiza la prueba de *western blot* que es altamente específica. La primera prueba utilizada es simple, causa mínimas molestias y es relativamente barata, cumpliendo con los requisitos de una buena prueba de *screening*. La segunda prueba ya es más cara, pero esto está justificado, dada la importancia de hacer un correcto diagnóstico.

CRITERIOS DE CONFIABILIDAD

VALOR PREDICTIVO POSITIVO Y NEGATIVO

Además de tener en cuenta la validez de una prueba de *screening* o de diagnóstico, hay que evaluar la credibilidad de su resultado.

Además de preguntarnos cuán buena es la prueba para discriminar entre sanos y enfermos, debemos preguntar, si el resultado de la prueba es positiva ¿cuán confiados podemos estar de que este resultado es un verdadero positivo?

La credibilidad del resultado de una prueba se valora a través de los que se denomina *valor predictivo positivo* y *valor predictivo negativo* del resultado de la prueba.

Aquí lo que es objeto de duda es el valor del resultado de la prueba y no la prueba en sí misma, aunque ambas cosas estén íntimamente relacionadas.

Si tomamos el ejemplo del cáncer de próstata veremos que la nueva prueba arrojó en total 180 resultados positivos y 820 negativos.

¿Cuántos de los resultados positivos corresponden a sujetos realmente enfermos?

Porcentaje de resultados positivos en población enferma = $(80/180) \times 100 = 44\%$

Con esta prueba que tiene 80% de sensibilidad y 89% de especificidad, sólo el 44% de los resultados positivos corresponde a sujetos verdaderamente enfermos. La nueva prueba dio un total de 180 resultados positivos, de los cuales 80 fueron verdaderos positivos.

¿Cuál es entonces el porcentaje de resultados negativos que corresponden a la población de sanos?

Porcentaje de resultados negativos en población sana = $(800/820) \times 100 = 97,6\%$

Con los mismos parámetros de validez, la proporción de resultados negativos que corresponde a sujetos verdaderamente sanos es del 97,6%.

El total de resultados negativos que informó la nueva prueba fue de 820, y 800 correspondieron a verdaderos negativos.

Por lo tanto, de 100 resultados positivos del test sabemos que sólo 44 serán enfermos y que 97 negativos corresponderán a sujetos sanos.

A estos parámetros de credibilidad del resultado se los denomina:

1. Valor predictivo positivo: definido como el porcentaje de resultados positivos que corresponden realmente a sujetos enfermos. El valor predictivo positivo se calcula como:

$$\text{VPP} = \frac{a}{a + b}$$

2. Valor predictivo negativo: definido como el porcentaje de resultados negativos que corresponden realmente a sujetos sanos:

$$\text{VPN} = \frac{d}{c + d}$$

Una pregunta que surge es ¿por qué si la prueba tiene sensibilidad (80%) y especificidad (89%) relativamente buenas su valor predictivo positivo es tan bajo (44%)?

Si observamos la población, veremos que la prevalencia de enfermedad es del 10% (100 sujetos de 1 000 fueron diagnosticados como enfermos por el método *gold standar*).

Por lo tanto es esperable que la mayoría de los resultados de esta población sean negativos.

Resultado de la biopsia			
Resultado del nuevo test	Enfermos	Sanos	Total
Positivo	80 a	100 b	180 a + b
Negativo	20 c	800 d	820 c + d
Total	100	900	1 000

Tabla 25.3: Valor predictivo positivo y negativo del resultado de una prueba de *screening* o diagnóstica

Como la prueba no es perfecta, clasifica algunos sujetos sanos como positivos y, por la baja prevalencia de la enfermedad, el resultado positivo no es muy confiable.

Supongamos que realizamos en una población de niñas menores de 13 años un *screening* de cáncer de cuello uterino con Papanicolau.

¿Cuántos resultados positivos para cáncer de cuello de útero esperaría encontrar? Seguramente ningún resultado positivo. No obstante, seguro obtendremos resultados positivos que por propio criterio clínico, nos parecerán sospechosos.

Esta es la explicación por la cual la prevalencia de la enfermedad afecta tanto el valor predictivo del resultado.

Veamos algunos ejemplos.

Supongamos que tenemos una nueva prueba para detectar alfa-fetoproteína como predictor de un defecto del cierre del tubo neural que posee una sensibilidad del 90% y una especificidad del 95%.

Elegimos primero una población de 10 000 mujeres de alto riesgo y 100 000 de bajo riesgo. (Tabla 25.4)

En el grupo de alto riesgo la prevalencia de defectos del tubo neural fue del 1%. Por lo tanto, a alta prevalencia mayor VPP (82,9%).

En el segundo grupo, de bajo riesgo, la prevalencia fue de 0,14%. Por lo tanto, el VPP es bajo (41%).

Si tomamos de nuevo el ejemplo del cáncer de próstata (Tabla 25.5), la prueba tuvo una sensibilidad del 80% y una especificidad del 89%, un VPP del 44% y un VPN de 97% en una población con 10% de prevalencia.

Eventos al nacimiento				
Alfa-fetoproteína	Defecto del tubo neural	Sano	Total	VPP (%) VPN (%)
Mujeres de alto riesgo				
Anormal	87	18	105	82,9%
Normal	13	9 882	9 895	99,9%
Total	100	9 900	10 000	
Mujeres de bajo riesgo				
Anormal	128	179	307	41,7%
Normal	19	99 674	99 693	99,98%
Total	147	99 853	100 000	

Tabla 25.4: Resultados nuevo test para alfa-fetoproteína

Resultado de la biopsia				
Resultado del nuevo test	Enfermos	Sanos	Total	VPP VPN
Positivo	80	100	180	44%
Negativo	20	800	820	97,6%
Total	100	900	1 000	
Sensibilidad	80%			
Especificidad		89%		

Tabla 25.5: Cálculo de VPP y VPN en población de baja prevalencia de la enfermedad

Resultado de la biopsia				
Resultado del nuevo test	Enfermos	Sanos	Total	VPP VPN
Positivo	240	77	317	75%
Negativo	60	623	683	91%
Total	300	700	1 000	
Sensibilidad	80%			
Especificidad		89%		

Tabla 25.6: Cálculo de VPP y VPN en población de alta prevalencia de la enfermedad

Si ahora cambiamos la prevalencia a 30% porque elegimos hombres mayores de 70 años, tendremos los resultados de la Tabla 25.6.

Ahora, con un 30% de prevalencia, el VPP se elevó al 75% sin casi modificarse el VPN (91 vs. 97%).

Por lo tanto, el valor de la prevalencia de la enfermedad es vital para la confiabilidad de los resultados de la prueba.

Otro factor que afecta el valor predictivo de una prueba es la especificidad del método.

Ya hemos visto en las Tablas 25.5 y 25.6 cómo el cambio de prevalencia afecta el VPP de la prueba. Siguiendo con el mismo ejemplo, ahora utilizamos una prueba con una menor especificidad. (Tabla 25.7)

Resultado de la biopsia				
Resultado del nuevo test	Enfermos	Sanos	Total	VPP VPN
Positivo	240	350	590	40%
Negativo	60	350	410	85%
Total	300	700	1 000	
Sensibilidad	80%			
Especificidad		50%		

Tabla 25.7: Influencia de la especificidad en el VPP del test

Ahora tenemos una prueba con la misma sensibilidad (80%) pero con mucha menor especificidad ($350/700 \times 100 = 50\%$). Esta prueba sólo discriminó correctamente como sanos al 50% de los sujetos sanos.

Esto trae como consecuencia que el número de resultados falsos positivos haya aumentado mucho, y por lo tanto el valor del VPP cayó a 40%.

En resumen, para que una prueba de *screening* o diagnóstica sea válida, su sensibilidad y su especificidad deben ser máximas, dependiendo del punto de corte seleccionado como criterio de positividad.

En los estudios diagnósticos, para que el resultado sea confiable hay que buscar en primer término poblaciones con alta prevalencia de la enfermedad y usar una prueba con muy buena especificidad.

CRITERIOS DE REPETIBILIDAD

Otra característica muy importante en una prueba de tamizaje o diagnóstica es su repetibilidad.

¿Podemos obtener el mismo resultado si repetimos la prueba?

Independientemente de la sensibilidad, la especificidad y los valores predictivos de una prueba, si sus resultados no son reproducibles, su valor y utilidad son mínimos.

Las fuentes que producen variaciones en los resultados de una prueba son dos.

VARIABILIDAD INTRASUJETO

Los resultados de una prueba en un mismo individuo varían durante el día. No es lo mismo medir la glucemia en ayunas que después de las comidas, ni tomar la tensión arterial en diferentes horas del día.

También las condiciones bajo las que se realiza la prueba (postejercicio, posprandial, en el consultorio, en casa) producen variabilidad de los resultados.

Es por eso que, a la hora de evaluar un resultado, hay que tener en cuenta las condiciones en las que se realizó la prueba, y es por esto mismo que muchos estudios se deben realizar bajo ciertas condiciones para que sus resultados sean comparables.

VARIABILIDAD INTEROBSERVADOR

No todos los examinadores de ciertos estudios llegan al mismo resultado. Esto es particularmente frecuente en los diagnósticos por imágenes. Otros ejemplos son el examen físico de un paciente, o el equipo con el que se hace una determinación de laboratorio; de ahí la importancia de tener patrones de calibración estándar.

PRUEBA DE KAPPA

La variabilidad interobservador se evalúa mediante una prueba denominada *kappa* o *prueba de concordancia*.

Con esta prueba se comparan los resultados reportados por dos o más observadores y se calcula la probabilidad de que estos observadores hayan concordado sólo por azar.

Lo importante para nosotros es demostrar que los observadores no han concordado por azar, sino porque el método es bueno y repetible, independientemente del observador.

		Patólogo A		Total B
		Grado II	Grado III	
Patólogo B	Grado II	41	3	44 (58,6%)
	Grado III	4	27	31 (41,4%)
Total A		45 (60%)	30 (40%)	

Tabla 25.8: Concordancia en el diagnóstico entre dos observadores

Constituye un error grave contar el número de concordancias entre los observadores y dividirlo por el número total de observaciones, porque que en muchos casos esto puede deberse a las características de la población estudiada.

Veamos un ejemplo. En la Tabla 25.8 se muestra la clasificación en grado de 75 tumores determinadas por dos patólogos.

Estos dos patólogos han concordado en 41 tumores clasificándolos con grado II y en 27 clasificándolos como grado III.

Si cometemos el error de sumar las concordancias y dividir las por el número total de tumores obtendremos:

$$\text{Concordancia: } (41 + 27)/75 \times 100 = 90,7\%$$

Este 90% es falso, debido a que no toma en consideración la posibilidad de concordancia por azar.

Si ahora aplicamos la prueba de kappa obtendremos un 81%, que es mucho menor al obtenido anteriormente.

La prueba de kappa se calcula como:

$$\text{Kappa} = \frac{(\text{porcentaje de concordancia observada}) - (\text{porcentaje de concordancia esperada solo por azar})}{100\% - (\text{porcentaje de concordancia esperada sólo por azar})}$$

El cálculo de los porcentajes esperados por azar es realizado por el programa estadístico.

Lo importante es saber que para determinar el grado de concordancia entre observadores, se debe utilizar la prueba de kappa.

Valores de kappa por encima del 75% se consideran buenos, entre 75 y 40% intermedios y menores al 40% malos.

Si el resultado de la nueva prueba es una variable cuantitativa, como por ejemplo creatinina o filtrado glomerular, las dos pruebas se comparan mediante la prueba de Bland & Altman.

Esta prueba se realiza mediante el empleo de un programa estadístico, pero en síntesis, lo que busca determinar es si los valores encontrados en la nueva prueba no superan los límites del IC 95% de la prueba de referencia.

Cuando se introduce una nueva prueba de *screening* o de diagnóstico, los estudios realizados para determinar sus características de validez, confiabilidad y repetibilidad deben estar bien diseñados.

Todos los pacientes deben realizarse los dos estudios y la comparación del resultado de la nueva prueba vs. la prueba estándar debe hacerse de forma ciega, para no introducir sesgos de información. Si se conoce el resultado de la prueba estándar, esto influenciará el resultado de la nueva prueba.

Hay circunstancias en las cuales la prueba de referencia es invasiva o tiene riesgos, y desde el punto de vista ético no sería correcto realizarla en pacientes sin indicación. En este caso los pacientes con resultado negativo en la nueva prueba son seguidos en el tiempo para evaluar si desarrollan la enfermedad y de esta manera se les evita el riesgo de la prueba estándar.

La nueva prueba debe evaluarse en diferentes poblaciones con todos los estadios de la enfermedad para determinar en qué grupo de pacientes se consigue el mejor costo/beneficio.

Una vez determinadas sus características y costo/beneficio, el estudio debe repetirse en un segundo grupo independiente de pacientes, para determinar con mayor confiabilidad si los resultados pueden ser duplicados.

En resumen, lo que se desea con una buena prueba es que, a través de su validez, mida o detecte con la mayor exactitud el evento de interés y con la menor variabilidad posible.

Todas las características explicadas en este capítulo deben ser tenidas en cuenta a la hora de valorar la utilidad de una nueva prueba de *screening* o diagnóstica, y a veces la de una prueba que ya está en uso pero que no ha sido ampliamente estudiada desde este punto de vista.

EVALUACIÓN PROGRAMAS DE SCREENING



Una vez evaluada una prueba de tamizaje, hay que demostrar, a través de estudios epidemiológicos, que la introducción de un programa de *screening* es de utilidad para la población.

Si el objetivo de detectar fases tempranas de una enfermedad o lesiones precursoras es disminuir la morbilidad, debe demostrarse desde el punto de vista epidemiológico.

También es necesario evaluar su factibilidad, la que involucra aspectos como aceptabilidad de la prueba por parte de los sujetos y un análisis de costo/efectividad.

ACEPTABILIDAD

La aceptabilidad de una prueba de tamizaje se mide por parámetros tales como sencillez y facilidad en su implementación y también en relación con las molestias que pueda producir al sujeto. Todo esto hace que una prueba sea aceptada o no por la población.

Para valorar aceptabilidad podemos calcular:

$$\text{Aceptabilidad} = \frac{\text{n}^\circ \text{ de personas que realizaron la prueba}}{\text{n}^\circ \text{ total de la población a la cual está destinada}}$$

La relación costo/efectividad se mide como recursos destinados al programa vs. número de casos detectados. Este cálculo es mucho más complejo y excede el propósito de este texto.

Para evaluar eficacia de un programa de *screening*, los posibles diseños a utilizar son los habituales de la Epidemiología.

Eficacia en un programa de *screening* significa demostrar una reducción de la morbilidad mediante la detección de un estadio preclínico de una enfermedad.

Siempre comparamos un grupo al que se le realiza el *screening* vs. otro que no y medimos, por ejemplo, mortalidad o remisión a 5 años. Podemos realizar estudios caso-control, estudios de cohortes analíticos, diseños experimentales y diseños ecológicos. En cada uno de ellos se deben tener en cuenta los posibles sesgos que se puedan introducir al estudio.

SESGOS EN ESTUDIOS DE EVALUACIÓN DE PROGRAMAS DE TAMIZAJE

En general se describen tres sesgos fundamentales a tener en cuenta a la hora de realizar un estudio de eficacia de un programa de *screening*.

SESGO DE SELECCIÓN

Puede suceder que los sujetos que acepten participar tengan más conciencia de salud que los sujetos que no participan. Estos sujetos seguramente concurren de forma regular a controles médicos, por lo que puede ser una población de bajo riesgo.

El efecto de este sesgo depende de la enfermedad. Se sabe que el cáncer de cuello de útero está asociado a un nivel socioeconómico bajo. Las mujeres que realizan controles periódicos para su prevención en general pertenecen a otra clase social, por lo tanto son de menos riesgo. Como ya sabemos, la prevalencia es importante para el rendimiento de la prueba y por lo tanto del programa, por lo que los resultados del estudio pueden ser totalmente erróneos.

Por ejemplo, la mortalidad a 5 años por cáncer de cuello de útero será mucho menor en el grupo de pacientes de mejor nivel socioeconómico, más allá del hecho de haberlo detectado precozmente.

También puede suceder que los sujetos con más alto riesgo sean los que más participan, en este caso el resultado del estudio será que los sujetos sometidos a *screening* tienen mayor mortalidad.

EXISTEN TRES SESGOS PARTICULARES DE LOS ESTUDIOS DE SCREENING

SESGO POR ADELANTO O LEAD TIME BIAS

Si el objetivo es detectar fases tempranas de una enfermedad, todos aquellos sujetos diagnosticados tempranamente tendrán mayor sobrevida a 5 años desde el diagnóstico, comparado con los sujetos diagnosticados en la fase sintomática.

Es obvio que el proceso desde la aparición de una lesión temprana hasta la aparición de la fase sintomática toma tiempo. El período que transcurre desde la aparición de la lesión preclínica hasta el diagnóstico espontáneo por síntomas se denomina *lead time* o *tiempo de adelanto*, y a la aparente mejoría de la sobrevida se la llama *lead time bias* o *sesgo por adelanto*.

Si no se calcula este período, cuando se compara la mortalidad entre los dos grupos, se concluirá erróneamente que el grupo *screening* tiene mejor sobrevida, sólo por el hecho de haber sido diagnosticado antes. (Figura 26.1)

Hay dos formas de solucionar este problema.

La primera es comparar el tiempo de sobrevida desde el diagnóstico a la muerte pero calculado por grupos etarios.

Por ejemplo, en los sujetos de ambos grupos que murieron a los 45 años ¿cuál fue la sobrevida del cáncer de mama desde el diagnóstico en el grupo *screening* y grupo no *screening* a los 45 años?

La segunda es: si sabemos cuál es el período de latencia o de adelanto, se lo sumamos al tiempo de sobrevida del grupo *screening*. Si el cáncer de mama tiene un *lead time* de un

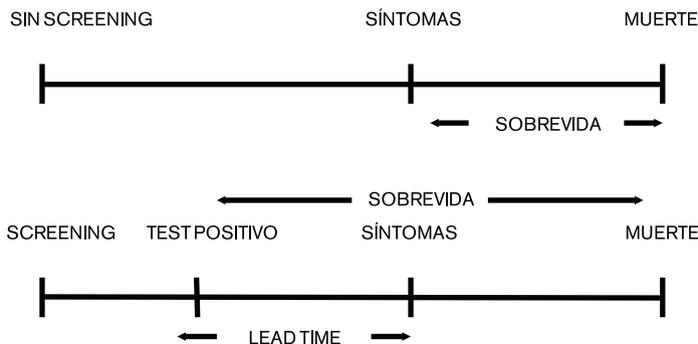


Figura 26.1: *Lead time bias*

año, lo que calculamos es la tasa de mortalidad en el grupo *screening* a 6 años y en el grupo no *screening* a 5 años de seguimiento y luego las comparamos.

Esta última opción no siempre es posible de realizar, ya que el cálculo del *lead time* es difícil.

SESGO DE DURACIÓN O LENGHT TIME BIAS

El segundo sesgo característico de los programas de *screening* es el *lenght time bias* o *sesgo de duración*.

Este sesgo se refiere al hecho de que en muchos casos el *screening* tiene más probabilidad de detectar lesiones con fases preclínicas más largas, por lo que la progresión a la enfermedad sintomática es mucho más larga. Obviamente estos individuos van a tener mejor sobrevida y el error es asumir que el *screening* mejora la sobrevida, cuando en realidad se debe a la detección de lesiones que progresan mucho más lentamente.

Este tiempo de progresión es muy difícil de cuantificar, la forma de controlar este sesgo es realizar la prueba de *screening* de forma seriada e ir comparando los eventos entre ambos grupos.

SESGO POR SOBREDIAGNÓSTICO

En la aplicación de programas de *screening* se corre el riesgo de clasificar sujetos como en riesgo de padecer una enfermedad en el futuro, cuando en realidad por diferentes factores no padecerán la enfermedad. Esto se debe a que muchas lesiones preclínicas nunca progresarán a fases sintomáticas o no son fatales, lo que resulta en mejor sobrevida para el grupo sometido a *screening*. A este sesgo se lo denomina *sesgo por sobrediagnóstico*.

Al diseñar un estudio epidemiológico con el que se quiera evaluar la eficacia de un programa de *screening* hay que evaluar todos estos sesgos y los propios de cada diseño.

Por ejemplo, en los estudios caso-control el sesgo de recuerdo es importante a tener en cuenta, al igual que el sesgo por pérdida de seguimiento en los estudios de cohortes.

Los estudios ecológicos, donde se comparan poblaciones sometidas a *screening* en lugar de sujetos, son de utilidad debido a que eliminan la posibilidad de contaminación.

La contaminación entre sujetos ocurre en los estudios de diseño experimental, cuando una persona asignada al grupo no *screening* se somete a la prueba, por ejemplo por referencias de otra persona asignada a la intervención; en este caso se pierde la asignación por aleatorización y los grupos dejan de ser comparables.

ASPECTOS A TENER EN CUENTA EN UN PROGRAMA DE SCREENING

Si bien el objetivo de un programa de *screening* es reducir la mortalidad, este no está exento de perjuicios, como incremento de la morbilidad, sobretratamiento en aquellos casos que no progresen a fases sintomáticas, riesgos propios de la prueba, ansiedad por resultados positivos que luego se confirman como falsos positivos, riesgo de no detectar lesiones por falsos negativos; todo esto con la consecuencia del incremento en los costos en salud por sobrediagnóstico y sobretratamiento.

Cuando un programa de *screening* se introduce por primera vez en una población, la prevalencia de lesiones preclínicas es alta, debido a que no han sido detectadas con anterioridad. Al transcurrir el tiempo habrá cada vez menos sujetos con lesiones preclínicas, por lo que la prevalencia disminuye. Esto trae como consecuencia que las características de validez y confiabilidad de la prueba se alteren.

Al mismo tiempo, la difusión en la población de la existencia de un programa de detección precoz hará que más sujetos concurren espontáneamente a realizarlo. Por lo tanto, la evaluación de un programa de *screening* debe realizarse periódicamente para evaluar su utilidad con los cambios de las características poblacionales. Es por esto que muchas pruebas de *screening* que antes se realizaban masivamente ahora sólo tengan indicación en poblaciones de alto riesgo.

En la actualidad, con el incremento exponencial de la investigación y de las publicaciones médicas, resulta prácticamente imposible emplear los métodos tradicionales para integrar los resultados obtenidos de estudios sobre un tema específico.

Es muy frecuente hallar en la literatura estudios clínicos que tratan de responder una misma pregunta científica, con resultados contradictorios.

En algunas áreas de la medicina no es fácil realizar estudios con gran cantidad de pacientes debido a que la patología es escasa o bien implica procedimientos muy riesgosos o invasivos, con lo cual la incorporación de pacientes es muy difícil. Por esto la muestra de sujetos es pequeña y, como ya hemos visto, el estudio no tiene poder para detectar un determinado efecto.

Algunos ejemplos son los estudios de trombolíticos en el tromboembolismo de pulmón.

En otros casos, los estudios epidemiológicos tienen diseños diferentes y sesgos que les son propios, además de incluir poblaciones con características diferentes, por lo que sus resultados pueden no ser coincidentes.

Más allá de las razones por las cuales distintos estudios llegan a diferentes conclusiones ante la misma pregunta, lo que necesitamos es arribar a una conclusión concordante en los casos de resultados contradictorios, o bien a una conclusión válida cuando los estudios no han podido hallar una respuesta concluyente.

Cuando existen pocos estudios sobre un tema, es posible hacer una revisión llamada *revisión cualitativa*, en la que el método de integración no incorpora el uso de procedimientos estadísticos formales. Por lo general, el revisor toma cada estudio por su valor aparente e intenta encontrar una teoría que reconcilie los resultados contradictorios, o bien intenta llegar a una conclusión que él cree válida, sin certeza de que así sea.

Cuando los estudios son numerosos, algo usual en nuestros días, el revisor tiene ante sí varios problemas. Debe enfrentarse a una lista enorme de conclusiones que niegan o confirman un mismo resultado, diferentes objetivos de los estudios para valorar la eficacia de una intervención, diferentes diseños de estudios, etc.

Por esto, ante tamaña dificultad, dicho revisor concluye el trabajo con la recomendación “se necesita más investigación...”.

En síntesis, la limitante fundamental que tiene la revisión cualitativa es su incapacidad para tratar una gran cantidad de resultados (estudios) a la vez.

Existe un método de integración de resultados de distintas investigaciones, que utiliza herramientas estadísticas en sus análisis y al que se llama **metanálisis**. El metanálisis consiste en el análisis estadístico de una gran cantidad de resultados extraídos de trabajos individuales con el propósito de integrar los hallazgos obtenidos y llegar a una conclusión válida.

OBJETIVOS DEL METANÁLISIS

a) Resolver incertidumbres

Cuando los resultados de diferentes estudios son contradictorios, o bien existen varios estudios pero sin conclusiones definitivas, el metanálisis hace posible llegar a una conclusión válida.

Al incorporar pacientes de muchos estudios aumenta la potencia estadística y con esto hace posible detectar un efecto determinado (por ejemplo, beneficio o perjuicio de una intervención) con mayor certeza estadística.

b) Mejorar la estimación del tamaño del efecto de una intervención

Cuando existen varios estudios de pocos pacientes, el resultado obtenido en cada uno de ellos puede ser muy amplio, por ejemplo, el IC 95% en la reducción de la mortalidad de una droga podría estar entre 1 y 20%, con lo cual la estimación de este efecto es muy amplia. Al incorporar los pacientes de distintos estudios se reduce el IC 95% y la estimación es más precisa. Al mismo tiempo se pueden reconocer efectos más moderados, que sólo es posible detectar con una gran cantidad de pacientes.

c) Contestar interrogantes no considerados como objetivos principales o bien no tenidos en cuenta antes de diseñar el estudio

Esto sucede cuando se encuentra un efecto especial en un subgrupo de pacientes: si se toman todos los pacientes pertenecientes a este subgrupo, provenientes de diferentes estudios, es posible determinar cuál es el efecto de la intervención sobre este subgrupo con mayor certeza.

También cuando en los estudios aleatorizados se examinan puntos finales secundarios y no hay suficiente cantidad de pacientes para responder las preguntas de estos objetivos secundarios, es posible analizar estos puntos finales secundarios con pacientes provenientes de distintos ensayos clínicos.

Otra circunstancia, muy frecuente, consiste en que una vez terminado un estudio aleatorizado, comiencen a publicarse estudios que tratan de responder una pregunta relativamente nueva para el ámbito médico.

Si el estudio aleatorizado tiene las variables necesarias, se puede realizar un análisis que permita contestar la *nueva* pregunta. Lo que se realiza es un análisis *post hoc* o retrospectivo (como una cohorte retrospectiva) para responder la pregunta, con todas las falencias que tiene un análisis retrospectivo. En este caso también puede ocurrir que cada estudio tenga pocos pacientes, y que sea necesario analizar muchos pacientes provenientes de diferentes estudios.

El metanálisis permite realizar una revisión objetiva, obtener un único resultado a partir de varios resultados de distintos estudios, estimar el efecto promedio de una intervención e investigar si los estudios son consistentes.

METODOLOGÍA

a) Definir el problema a resolver. ¿Cuál es la pregunta que se desea responder?

Los metanálisis responden preguntas generales (por ejemplo, la cirugía bariátrica ¿es efectiva para reducir el peso?), más allá de la población incluida en los estudios que integran el metanálisis.

b) Identificar los estudios que hayan estudiado el problema a resolver en la literatura médica

Se deben identificar todos los estudios que hayan abordado la pregunta de interés, independientemente de su resultado, y del idioma en que hayan sido publicados. También se necesita identificar aquellos estudios que no fueron publicados. Como veremos más adelante, la omisión de estudios por idioma o por su resultado resulta en un sesgo que invalida el resultado del metanálisis.

c) Establecer criterios de inclusión y exclusión de los estudios que se incorporarán al metanálisis, en relación con el tipo de estudios que se desea incorporar (estudios aleatorizados, observacionales, prospectivos, etc.), a los criterios de inclusión y exclusión de cada estudio, definición del factor de exposición y evento de interés y metodología estadística empleada en cada estudio.

MÉTODOS ESTADÍSTICOS

Existen dos métodos estadísticos para realizar el metanálisis y así obtener un resultado general único.

MÉTODO DE EFECTO FIJO

Asume que cada estudio evaluó el mismo efecto *verdadero*. Por ejemplo, todos los estudios asumieron que una droga disminuyó la mortalidad de una determinada patología y que los resultados dispares sólo se debieron al azar. O bien, todos los estudios mostraron reducción de eventos, pero los resultados discordantes se debieron al azar.

Para estar seguros de que todos los estudios detectaron el mismo efecto hay que determinar lo que se denomina *heterogeneidad*.

Puede suceder que no todos los estudios muestren el mismo efecto, en algunos puede haber un efecto a favor y en otros en contra, debido a diferentes características de los pacientes incorporados en cada estudio, diferente definición de factor de exposición y/o eventos, diferentes dosis de una droga, etc. Lo que puede denominarse como *variabilidad clínica*.

También podemos observar *variación metodológica*, debido a las diferencias en los diseños de los estudios y sus sesgos.

La variabilidad del efecto de la intervención evaluada en diferentes estudios es la conocida *heterogeneidad estadística* o heterogeneidad, que es consecuencia de la variabilidad clínica y metodológica.

La heterogeneidad se refiere a la diferencia en el efecto observado de una intervención o de la magnitud del efecto de una intervención entre los estudios seleccionados para el metanálisis.

Para detectar heterogeneidad formalmente se aplica una prueba estadística, muy similar a la prueba de chi cuadrado, conocida como prueba de heterogeneidad (en inglés *heterogeneity*, en notación matemática I^2).

La hipótesis nula de esta prueba es que las diferencias en el efecto de una intervención observadas en los diferentes estudios se debieron al azar.

I^2 puede ser reportado como el porcentaje de la variabilidad en la estimación del efecto que es causada por heterogeneidad.

Un porcentaje entre cero y 40% de variabilidad puede no ser importante, entre 30 a 60% representa heterogeneidad moderada, entre 50 a 90% sustancial heterogeneidad y entre 75 a 100% considerable heterogeneidad.

Esta prueba tiene poco poder cuando los estudios seleccionados cuentan con un tamaño de la muestra chico o cuando se seleccionan pocos estudios para ser incluidos en el metanálisis.

Esto significa que, mientras el resultado de I^2 es estadísticamente significativo, puede indicar un problema de heterogenicidad; un resultado no significativo no debe ser tenido en cuenta como evidencia de no heterogenicidad.

Por este motivo es que para esta prueba se considera un valor de p de 0,1 en lugar de 0,05, con el propósito de determinar significancia estadística.

Otro problema con esta prueba es que cuando se incluyen muchos estudios en el metanálisis, la prueba tiene alto poder para detectar heterogenicidad de poca magnitud, la que puede no ser clínicamente importante.

Ante la presencia de heterogenicidad moderada o más no se puede utilizar el método de efecto fijo, hay que usar otro método estadístico para realizar el metanálisis.

MÉTODO DE EFECTO AL AZAR

Este método tiene en cuenta la posible heterogenicidad de los estudios incluidos en el metanálisis.

Debido a que no es real asumir que el verdadero efecto es el mismo en cada estudio, porque los diseños de protocolos son diferentes, el método de efecto al azar (*random effect*) es el más correcto para aplicar en un metanálisis.

Este método asume que los diferentes estudios estimaron diferentes efectos de la intervención.

En este método se tiene en cuenta el tamaño de la muestra de cada estudio y se le otorga un *peso* de contribución al metanálisis. El *peso* que se otorga a cada estudio es inverso al tamaño de la muestra de cada uno.

Como vemos, la elección del método a utilizar para realizar el metanálisis y, a través de él, lograr encontrar un resultado concluyente o una mejor estimación del efecto de una intervención, es necesario determinar si existe heterogenicidad y luego seleccionar el método estadístico adecuado.

El metanálisis se presenta generalmente con un gráfico, llamado *forrest plot*, que muestra el efecto encontrado en cada estudio con cuadrados, de tamaño correspondiente al tamaño de la muestra de cada estudio (a cuadrado más grande, mayor n de la muestra de cada estudio), y el efecto global encontrado (en forma de diamante) al analizar todos los estudios con el método de metanálisis.

Las líneas a los lados de los cuadrados y la figura de diamante representan el IC 95% de cada estimación en cada estudio y del metanálisis. De esta manera es más fácil observar en cuáles estudios se observó un beneficio (cuadrado y líneas a la izquierda de la línea vertical central que representa el valor de OR, RR o HR = 1) y en cuántos casos el hallazgo tuvo significación estadística. Si cruzan la línea vertical, que corresponde a OR, RR o HR = 1, el resultado es no significativo. (Figura 27.1)

En el gráfico se observa el resultado de 7 estudios, representados por sus cuadrados como medida de efecto encontrada y tamaño de la muestra y las líneas que representan sus IC 95%. Como puede observarse solo un estudio mostró un resultado significativo.

El test de heterogenicidad I^2 no muestra evidencia de heterogenicidad importante (34,2%), no obstante se utilizó el método al azar para realizar el metanálisis que, como mencionamos, asume una situación más real.

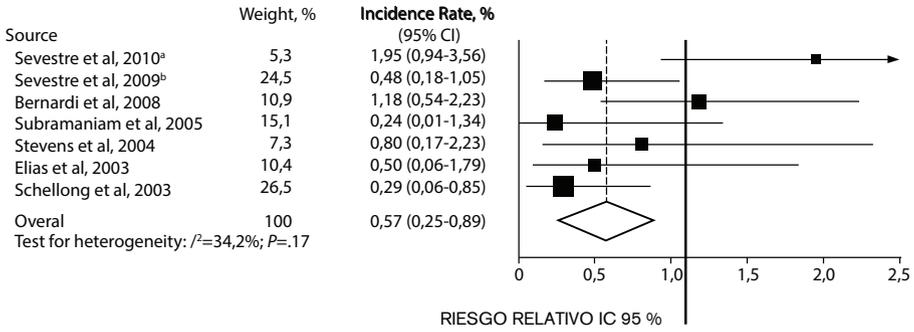


Figura 27.1: Metanálisis. Utilidad de la ecografía de miembros inferiores con compresión negativo para predicción de riesgo de trombosis venosa profunda

El resultado del metanálisis, representado por la figura en diamante, muestra un resultado conclusivo, debido a que los extremos del diamante no alcanzan el valor de $RR = 1$.

También se pueden observar los resultados numéricos de cada estudio y del metanálisis, así como el peso otorgado a cada estudio de forma directa, de acuerdo al tamaño de la muestra de cada estudio. Cabe aclarar que en la fórmula matemática del metanálisis se utiliza la inversa del peso de cada estudio.

Traducido a términos clínicos, la realización de un estudio de ecografía de miembros inferiores con compresión, con resultado negativo, reduce significativamente el riesgo de trombosis venosa profunda.

SESGOS EN METANÁLISIS

Como hemos mencionado, la metodología para realizar un metanálisis correcto está muy relacionada con los posibles sesgos que puedan ser introducidos al seleccionar los estudios a incluir en el metanálisis.

SESGOS DE PUBLICACIÓN

La detección de estudios de investigación a incluir en un metanálisis está influenciada por varios factores.

1. Sus resultados

Los estudios con resultados favorables a la hipótesis del estudio son más frecuentemente publicados que los estudios con resultados negativos.

Por ejemplo, para el tratamiento del cáncer con quimioterapia existe un registro de estudios clínicos. Cuando se compara la cantidad de estudios publicados con los registrados, se observa que en una gran proporción sólo se publicaron los estudios con resultados positivos. Si un metanálisis se basa sólo en los estudios publicados con resultados positivos, su resultado será que la droga es beneficiosa, lo cual no es real, debido a que no se han incluido los estudios con resultados negativos.

2. Idioma

Los estudios originarios de países de habla inglesa son los más publicados. En general también son los que corresponden a resultados positivos.

En Alemania, el 63% de los estudios randomizados publicados en inglés tenían resultados positivos vs. 35% con resultados negativos que fueron publicados sólo en alemán.

En un análisis multivariado se observó un OR de 3,8 para publicaciones en inglés si estos tenían resultados positivos. Esta asociación cambió muy poco cuando se ajustó por número de pacientes incluidos, calidad del estudio y diseño. Lo que significa que el idioma inglés es un predictor de publicación y de resultado positivo.

3. Bases de datos

La mayoría de las revistas que se publican en idiomas diferentes al inglés NO están indexadas en *Medline*, por lo tanto hay mucha información de la cual no se dispone.

Las revistas médicas de países no europeos corresponden sólo al 2% de las citas encontradas en este tipo de bases de datos.

4. Citas bibliográficas

Cuando se están buscando los estudios a incluir en un metanálisis, se recurre frecuentemente a citas de otros estudios hallados.

Cuando se utilizan estas listas de referencias bibliográficas se puede introducir un sesgo.

Los estudios que reportan beneficios son citados con más frecuencia que los que reportan datos negativos, independientemente de la calidad del estudio y del tamaño de la muestra.

SESGO POR PUBLICACIÓN MÚLTIPLE

Es muy frecuente que los estudios multicéntricos sean publicados en diferentes revistas con autores diferentes, y con alguna modificación en relación al estudio principal. Es por esto que, a la hora de buscar citas bibliográficas, encontramos muchas más citas que estudios. Un mismo estudio es citado varias veces con diferentes autores.

Si no se lee bien el estudio, puede tomarse como otro diferente siendo el mismo. De esta manera podemos alterar los resultados del metanálisis por inclusión múltiple de un mismo estudio que seguramente, por lo visto anteriormente, será favorable al punto final del metanálisis.

SESGOS DE SELECCIÓN

Aun habiendo localizado la mayoría de los estudios realizados en el área de interés, se puede introducir un sesgo al determinar los criterios de inclusión de los estudios.

Si estos criterios son desarrollados por un investigador familiarizado con el tema, su selección puede estar influenciada por el conocimiento de los resultados de un grupo de estudios.

La manipulación de estos criterios puede llevar a la inclusión selectiva de estudios favorables y a la exclusión de aquellos con resultados negativos.

La forma más apropiada de seleccionar estudios es incluir todos los disponibles que tengan al menos los mínimos criterios de inclusión requeridos y luego hacer un análisis más sensible de los criterios de inclusión para el metanálisis.

Cualquier conclusión de un metanálisis con criterios de selección muy estrictos debe tomarse con cuidado.

EVALUACIÓN DE SESGOS

GRÁFICO CONO INVERTIDO O FUNNEL PLOT

Para evaluar la posible presencia de sesgos en la inclusión de estudios en un metanálisis se dispone de un test gráfico llamado *funnel plot*. Este gráfico muestra el efecto de una intervención en relación con el tamaño de la muestra del estudio.

Los estudios más pequeños se observan en la base del gráfico y muestran mayor dispersión, a medida que se asciende en el gráfico la dispersión es menor porque los estudios tienen más pacientes.

En un metanálisis sin sesgos el gráfico es simétrico con forma de cono invertido. (Figura 27.2)

Si el gráfico es simétrico significa que no se incluyó algún tipo de sesgo, ya sea de publicación o de localización de estudios.

En cambio, si el gráfico es asimétrico, como muestra la Figura 27.3, se considera que se introdujo algún tipo de sesgo.

Nótese que en este gráfico los puntos se concentran a la izquierda del cono invertido.

Otro factor que afecta la simetría del gráfico es el tipo de metodología estadística empleada en cada estudio. Los estudios con menos pacientes, en general, se analizan con metodología menos rigurosa que los estudios con más pacientes. Los estudios con pobre calidad de análisis tienden a mostrar efecto de mayor magnitud.

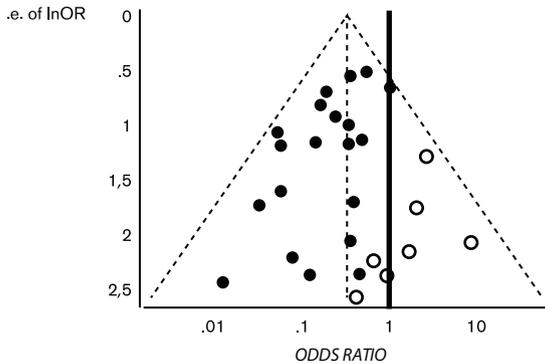


Figura 27.2: Evaluación de sesgos. *Funnel plot*. No evidencia de sesgos

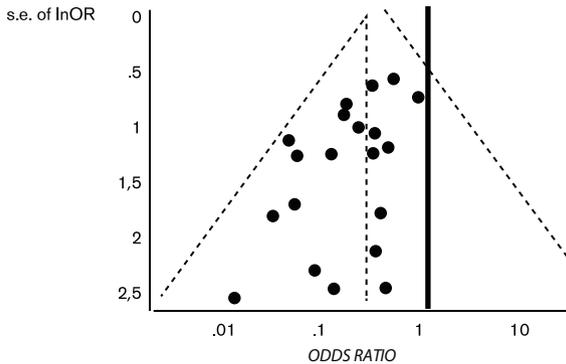


Figura 27.3: Evaluación de sesgos. *Funnel plot*. Evidencia de sesgos

En conclusión, el metanálisis es un método muy útil a la hora de obtener una conclusión válida y objetiva en un terreno donde existen incertidumbres pero, como todo diseño epidemiológico, puede tener errores. Por lo tanto, es de suma importancia evaluar la metodología empleada y determinar si existen sesgos importantes.

INTRODUCCIÓN AL ANÁLISIS MULTIVARIABLE



PRUEBA DE MANTEL-HAENZSEL. INTERACCIÓN

Una vez encontrada una asociación entre un factor de exposición y un evento, debemos estar seguros de que esta asociación es *verdadera*, tomando en cuenta la posibilidad de que otras variables puedan *confundir* la asociación estudiada (variables de confusión) o bien que esta asociación cambie a través de las categorías de otra variable (interacción).

Cuando deseamos controlar una asociación por una sola variable confundidora de tipo dicotómica es posible utilizar una prueba estadística conocida como prueba de Mantel-Haenszel (M-H), que calcula la medida de efecto utilizada para describir la asociación (RR, OR o cociente de tasas) *ajustada* por la presencia de otra variable posiblemente confundidora.

En un estudio se investigó la asociación entre diabetes tipo 2 y mortalidad; entre las variables recabadas contamos con información sobre tabaquismo, que fue clasificado como nunca fumó, ex tabaquista, tabaquista entre 1-14 cigarrillos/día, 14-24 cigarrillos día y ≥ 25 cigarrillos/día.

En primer término deseamos saber si la variable tabaquismo es un posible confundidor de la asociación diabetes y muerte.

Para que una variable pueda ser considerada como potencial confundidor, debe estar asociada al factor de exposición en estudio –en este caso a diabetes– asociado al evento –en este caso muerte– y no debe estar en la vía causal entre el factor de exposición y eventos (diabetes y muerte).

Para determinar si el tabaquismo está asociado a diabetes y a muerte, lo que hacemos es tomar la variable tabaquismo como dicotómica (tabaquista Sí/No) y realizar una prueba de chi cuadrado 2×2 para determinar si hay asociación entre tabaquismo y diabetes y entre tabaquismo y muerte.

Si concluimos que el tabaquismo es un potencial confundidor, procedemos de la siguiente manera.

Primero calculamos la medida de efecto de asociación entre diabetes y muerte, por ejemplo, calculamos el RR, que en este caso es de 5,4. A este RR lo denominamos RR **crudo**.

Luego calculamos el RR de muerte en diabéticos estratificados por tabaquistas y no tabaquistas.

El RR de muerte para diabéticos en fumadores es de 3,33 y el RR de muerte para diabéticos en no fumadores es de 2,41.

Como vemos, los RR estratificados son muy diferentes al RR crudo (5,4 vs. 3,33 y 2,41).

La prueba de Mantel-Haenszel calcula el RR promedio entre el RR en fumadores y no fumadores, que para este caso es de 3,2. A este RR lo denominamos RR ajustado, porque tuvo en cuenta la variable tabaquismo en la relación entre diabetes y muerte.

Si el RR crudo es diferente al RR ajustado, como en este caso, decimos que el tabaquismo es un factor de confusión en la asociación diabetes y muerte. No existe prueba estadística para determinar si el RR crudo es estadísticamente diferente al RR **ajustado**.

Tabaquismo	RR	Prueba de interacción
Nunca	1,82	< 0,05
Ex tabaquismo	1,90	
1-14 cigarrillos/día	1,87	
15-24 cigarrillos/día	2,44	
≥ 25 cigarrillos/día	3,11	

Tabla 28.1: Riesgo relativo de muerte para diabéticos en cada categoría de tabaquismo

También deberíamos averiguar si la variable tabaquismo, independientemente de que sea un factor de confusión, es una variable que, de acuerdo a sus categorías, incremente la mortalidad de diabéticos. Lo que queremos es determinar si existe interacción entre muerte y las distintas categorías de tabaquismo, para lo cual debemos realizar una prueba de interacción o de homogeneidad.

Como se observa en la Tabla 28.1 los RR de muerte se incrementan en cada categoría de tabaquismo y por lo tanto existe interacción (la prueba de interacción arrojó un valor $p < 0,05$).

Entonces decimos que existe una asociación de tipo dosis-respuesta.

Hasta acá, el tabaquismo es una variable asociada a muerte, un confundidor de la asociación entre diabetes y muerte y modifica RR de muerte en diabéticos a través de sus categorías (hay interacción).

Ahora bien, el tabaquismo es una variable asociada a mortalidad en sí misma. Dejando de lado la presencia o ausencia de diabetes, la mortalidad por tabaquismo podría incrementarse a través de las categorías de esta variable.

Tomando como referencia a los sujetos que nunca fumaron, podemos calcular el cociente de riesgo de muerte en las otras categorías.

Lo que hemos hecho acá es tomar la tasa de muerte de los sujetos que nunca fumaron y de cada categoría de tabaquismo y calcular, usando como referencia la tasa de muerte de los sujetos que nunca fumaron, el cociente de tasa (similar al RR) de muerte para cada categoría de tabaquismo. Como vemos, el cociente de tasas se incrementa a medida que aumenta el número de cigarrillos fumados por los sujetos. Podríamos pensar el resultado como lo hemos hecho cuando estudiamos la prueba de chi cuadrado de tendencia.

En resumen, hemos trabajado con el factor de exposición (diabetes), el evento (muerte) y una variable (tabaquismo) y, mediante diferentes cálculos, determinamos que el tabaquismo es un confundidor de la asociación entre diabetes y muerte, que tiene interacción con el

Tabaquismo	Tasa de muerte /1000 pers/año	Cociente de tasas
Nunca	25,3	1 (referencia)
Ex tabaquismo	43,8	1,7 (43,8/25,3)
1-14 cigarrillos/día	76,2	3,0 (76,2/25,3)
15-24 cigarrillos/día	212,1	8,4 (212,1/25,3)
≥ 25 cigarrillos/día	355,6	14,0 (355,6/25,3)

Tabla 28.2: Evaluación del efecto de fumar sobre la tasa de mortalidad

riesgo de muerte, y además, analizada en sí misma, es una variable asociada a muerte que muestra una tendencia a mayor riesgo de muerte a medida que la variable incrementa su nivel.

También deberíamos determinar si la asociación entre tabaquismo y muerte está confundida por la presencia de diabetes, como hicimos antes cuando consideramos al tabaquismo como confundidor de la asociación entre diabetes y muerte.

	Grupo 1 n = 87	Grupo 2 n = 167	Valor p
Edad	59,2 ± 11,3	31,5 ± 12,2	0,1
Masculino, n (%)	59 (67,8)	90 (53,8)	0,01
HTA, n (%)	33 (37,9)	82 (49,1)	0,09
DBT, n (%)	4 (4,6)	8 (4,7)	0,9
TBQ, n (%)	28 (32,1)	55 (32,9)	0,9
IAM previo, n (%)	5 (5,7)	15 (8,9)	0,3
BCPIAO, n (%)	0	2 (1,2)	0,1
Coronariopatía, n (%)	13 (14,9)	64 (38,3)	0,01
Cirugía previa, n (%)	3 (3,4)	10 (5,9)	0,3
FSVI < 50%, n (%)	12 (13,7)	24 (14,3)	0,9
FSVI 30-49%, n (%)	8 (9,1)	14 (8,3)	0,9
FSVI < 30%, n (%)	4 (4,5)	10 (5,9)	0,4
DDVI, mm	59,7 ± 8,1	56,7 ± 7,8	0,005
DSVI, mm	35,7 ± 7,3	34,6 ± 7,5	0,3
AI, mm	50,3 ± 9,4	50,4 ± 12,1,4	0,9
Septum, mm	11 ± 2,3	11,1 ± 2,2	0,8
Creatinina, mg/dl	1,11 ± 0,1	1,23 ± 0,5	0,06
ACV previo, n (%)	3 (3,4)	10 (5,9)	0,3
FA, n (%)	15 (17,2)	44 (26,3)	0,1
Efectivo, n (%)	86 (98,8)	123 (73,6)	0,001
BGC, n (%)	7 (8)	32 (19,1)	0,02
FA, n (%)	13 (14,9)	43 (25,7)	0,04
ACV, n (%)	0 (0)	2 (1,2)	0,3
Mediastinitis, n (%)	0 (0)	1 (0,6)	0,3
IRA, n (%)	3 (3,4)	10 (5,9)	0,3
Dialisis, n (%)	1 (1,1)	8 (4,7)	0,1
ARM, n (%)	1 (1,1)	14 (8,3)	0,02
Reoperación, n (%)	5 (5,7)	11 (6,5)	0,5
Obito, n (%)	2 (2,3)	16 (9,5)	0,03
Complicación mayor, n (%)	12 (13,7)	46 (27,5)	0,01

Tabla 28.3: Comparación de las características basales de las poblaciones

HTA: Hipertensión arterial. **DBT:** Diabetes. **TBQ:** Tabaquismo. **FSVI:** Función sistólica del ventrículo izquierdo. **BCPIAO:** Balón de contrapulsación intraaórtico. **FA:** Fibrilación auricular. **DDVI:** Diámetro diastólico del ventrículo izquierdo. **DSVI:** Diámetro sistólico del ventrículo izquierdo. **AI:** Aurícula izquierda. **ACV:** Accidente cerebrovascular. **BGC:** Bajo gasto cardíaco. **FA:** Fibrilación auricular. **ACV:** Accidente cerebrovascular. **IRA:** Insuficiencia renal aguda. **ARM:** Asistencia respiratoria mecánica. **Reoperación.** Sangrado quirúrgico. **Complicación mayor:** Véase el apartado Definición de las complicaciones.

Ahora imaginen la cantidad de cálculos que deberíamos hacer para evaluar cada variable recabada en un estudio que, como han visto, siempre son muchas con relación a su asociación con el evento, a su papel como confundidor, si actúa como factor de interacción o si el riesgo de evento se incrementa a través de todas las variables categóricas ordinales del estudio.

Todos estos cálculos son posibles de realizar al mismo tiempo utilizando lo que se denomina un modelo de análisis multivariable.

Veamos a continuación qué nos permite hacer los modelos de análisis multivariable.

Utilidad del análisis multivariable

- Controla la asociación entre factor de riesgo y evento por diferentes confundidores.
- Determina la relativa contribución de diferentes factores de riesgo (otras variables) sobre la ocurrencia de un evento (determinación de predictores independientes).
- Permite el desarrollo de puntajes pronósticos.
- Iguala características basales.

CONTROL DE CONFUNDIDORES

En un estudio de cohorte retrospectivo, se evaluó la incidencia de complicaciones y muerte intrahospitalaria en pacientes operados por reemplazo valvular mitral de acuerdo con su clase funcional (CF I/II = grupo 1, CF III/IV = grupo 2).

En la Tabla 28.3 se muestran algunas de las variables relevadas (características basales) de cada grupo de clase funcional.

Como se puede observar en la Tabla 28.4, existe una asociación entre clase funcional III/IV y óbito.

Si calculamos el OR crudo entre clase funcional y óbito obtenemos un valor = 4,6

Ahora analicemos las características del grupo 2. Este grupo tiene mayor prevalencia de hipertensión arterial, infarto previo, enfermedad coronaria, cirugía cardíaca previa, accidente cerebrovascular previo, fibrilación auricular, menos cirugía electiva y mayor incidencia de complicaciones posoperatorias intrahospitalarias.

La pregunta que nos hacemos es, ¿la clase funcional más severa está realmente asociada a óbito? o ¿el hecho de que el grupo 2 esté más enfermo y haya tenido más complicaciones fue lo que determinó la mayor mortalidad?

En otras palabras, la asociación entre clase funcional III/IV y óbito, ¿es *real*? o ¿está confundida por la presencia de las variables mencionadas?

Para evaluar la asociación entre clase funcional III/IV y óbito tomando en cuenta la presencia de estas otras variables posiblemente confundidoras, realizamos un análisis multivariable.

El resultado de este análisis se observa en la Tabla 28.4.

Hospitalaria	OR	Valor p	IC 95%
Disnea	2,39	0,02	1,10-5,16
Edad	1,03	0,01	1,00-1,02
Tiempo CEC	1,01	0,01	1,00-1,06
Plástica	0,51	0,06	0,26-1,02

Tabla 28.4: Análisis multivariable de regresión logística

En la Tabla 28.4 observamos que el valor del OR de disnea CF III/IV vs. clase funcional I/II es de 2,39, muy diferente al OR crudo = 4,6, este OR está ajustado por todas las variables que suponíamos eran posibles confundidoras. Si bien el riesgo de muerte se redujo de 4,6 a 2,39, la asociación continuó siendo significativa, debido a que el valor de p reportado es = 0,02 y el IC 95% no contiene al 1.

DETERMINACIÓN DE PREDICTORES INDEPENDIENTES

En la Tabla 28.4 que muestra el análisis multivariable, observamos otras variables como edad y tiempo de circulación extracorpórea (CEC), que también muestran asociación con muerte hospitalaria, ajustadas por las mismas variables posiblemente confundidoras que utilizamos para evaluar la asociación entre disnea III/IV y muerte, debido a que todo se realizó en el mismo análisis.

Estas variables que muestran asociación con muerte una vez ajustadas por otras variables confundidoras, se denominan predictores independientes, porque están asociadas a muerte independientemente de la presencia de otras variables.

El tipo de cirugía (plástica) pudo no haber sido un confundidor en el análisis univariado o no ajustado, si bien en el análisis multivariable no muestra asociación estadísticamente significativa, el autor la consigna debido a la gran reducción de muerte que se observa (OR = 0,5) y a la posibilidad de que no se haya podido detectar asociación estadísticamente significativa debido a la presencia de error beta.

Muchas variables pueden estar asociadas al evento, sin ser confundidores, y por lo tanto es útil incluirlas en el análisis multivariable para determinar su efecto sobre el evento ajustado por la presencia de otras variables confundidoras.

Para realizar un análisis multivariable también es necesario contar con un mínimo de sujetos que nos asegure que podemos detectar predictores independientes.

A mayor número de sujetos incluidos en el análisis, mejor será la predicción, debido a que el IC 95% será más estrecho.

DESARROLLO DE PUNTAJES PRONÓSTICOS

Los puntajes o *scores* son muy utilizados en medicina para determinar la posibilidad que tiene una persona con determinadas características de sufrir un evento.

Entre los más conocidos podemos citar el puntaje de Framingham para predecir riesgo de sufrir eventos cardiovasculares, APGAR de los recién nacidos después del parto para determinar la necesidad inmediata de cualquier tratamiento adicional o emergencia médica, CHADS2 para predecir el riesgo de accidente cerebrovascular en pacientes con fibrilación auricular, la escala de Glasgow, etc.

Estos puntajes surgen a partir de determinar en primer lugar cuáles son las variables que se asocian de forma independiente con el evento (predictores independientes) y luego tomar el valor del OR, o la medida de efecto que se haya utilizado, para otorgarle un puntaje a cada variable, debido a que la medida de efecto nos informa del *peso* que tiene la variable sobre el evento.

De acuerdo con la suma total del puntaje que presente el sujeto será el riesgo de sufrir el evento.

Por ejemplo, si quisiéramos construir un puntaje de riesgo de muerte intrahospitalaria poscirugía de reemplazo valvular mitral, podríamos otorgarle a la presencia de disnea III/IV un valor de 2,5 puntos (redondeamos el valor del OR), a la edad un valor de 0,3 puntos por

cada año de edad del paciente (si el paciente tiene 50 años el puntaje será de 15, si tiene 75 años será de 22,5). Lo mismo hacemos con el tiempo de circulación extracorpórea, por cada minuto de CEC multiplicamos por 0,1, si la duración fue de 90 minutos el puntaje será de 9.

En cada paciente buscamos la presencia o no de disnea III/IV, si no tiene disnea III/IV, asignamos cero a esta variable, sumamos el puntaje según edad y tiempo de circulación extracorpórea y calculamos el puntaje total.

Muchos puntajes utilizan el valor numérico total y determinan el riesgo del evento de interés mostrado como porcentaje de sufrir el evento.

En otras ocasiones, resulta más fácil categorizar el puntaje. Por ejemplo, los puntajes que determinen un riesgo bajo (posibilidad de sufrir el evento $< 5\%$), supongamos entre 0 y 20 puntos, se agrupan en la categoría riesgo bajo; los puntajes intermedios, ejemplo entre 21 a 40 puntos, en riesgo moderado (posibilidad de sufrir el evento entre 5 a 10%) y los puntajes altos > 40 puntos, en riesgo elevado ($> 10\%$ de sufrir el evento).

Aclaremos que este es un ejemplo al azar. La construcción de puntajes pronósticos depende del tipo de evento, de su incidencia total y de los valores de OR u otra medida de efecto que se haya utilizado en el análisis multivariable.

Una vez desarrollados, estos puntajes deben ser validados en una población diferente a aquella de la que se obtuvo el puntaje, para verificar su poder de predicción, en otras palabras, si es aplicable a otras poblaciones.

Muchos de los puntajes que se utilizan en la práctica diaria médica fueron obtenidos de poblaciones de otros países, que tiene características diferentes a las nuestras, por eso es importante validar estos puntajes en cada población en la que se desea utilizar.

IGUALACIÓN DE CARACTERÍSTICAS BASALES

Los estudios aleatorizados nos permiten igualar las características de las poblaciones a estudiar e igualar las variables confundidoras; por este motivo, los estudios experimentales no requieren de un análisis multivariable para ajustar por confundidores. No obstante, la utilización de un modelo multivariable podría ser muy útil para desarrollar un puntaje de riesgo.

En muchos casos no es posible realizar un estudio experimental para determinar el beneficio o perjuicio de una intervención y, por lo tanto, se realiza un estudio observacional, estos estudios siempre mostrarán diferencias en alguna de las características de las poblaciones a analizar y de los posibles confundidores.

Para igualar las características basales de las poblaciones se debe realizar un procedimiento llamado *propensity score*, que se basa en la técnica de análisis multivariable. Con esta técnica estadística se trata de *simular* la asignación aleatoria de una intervención a una población, como si fuera un estudio experimental.

Si se desea saber si el uso de profiláctico reduce el contagio de HIV, por un problema ético no podemos realizar un estudio experimental aleatorizado, en el cual se asigne el uso de profiláctico a un grupo de sujetos y al otro grupo no.

A través de la realización de un análisis multivariable de tipo de regresión logística se identifican las variables asociadas de forma independiente a la intervención, en este caso utilización de profiláctico.

Muchas variables podrían estar asociadas a la utilización de profiláctico: edad, situación socioeconómica, no uso de drogas, etc.

Una vez identificados estos predictores, se agrupa la población con similares predictores independientes, utilicen o no profiláctico, para igualar las características basales de la población. Una vez agrupada la población por predictores independientes, entonces sí se

la divide en uso de profiláctico y no uso y se procede a determinar si el uso de profiláctico se asocia o no a menor contagio de HIV.

Esta técnica es mucho más compleja, lo que excede el alcance de este texto.

TIPOS DE MÉTODOS ESTADÍSTICOS PARA ANÁLISIS MULTIVARIABLE

De acuerdo con el tipo de estudio clínico y del tipo de variable que sea el evento (variables cuantitativas, categóricas dicotómicas u ordinales), es el tipo de análisis multivariable que debemos utilizar.

En cualquiera de estos modelos se denomina al evento variable dependiente, y a los posibles confundidores u otras variables asociadas, variables independientes o covariables.

Lo que deseamos al construir un modelo de análisis multivariable es predecir lo mejor posible el riesgo de sufrir un evento a través de un conjunto de diferentes variables, por lo que el riesgo de evento dependerá de estas variables.

Los resultados de estos análisis no son perfectos, debido a que muchos posibles confundidores pudieron no ser recabados en el estudio. Es por esto que para algunos estudios los predictores independientes para un evento son diferentes a los de otros estudios.

En los próximos capítulos veremos los dos métodos multivariados más utilizados.

Evento (tipo de variable)	Tipo de estudio	Medida de efecto	Análisis multivariable
Dicotómica	Caso-control / transversal	OR	Regresión logística
Dicotómica	Caso-control / transversal	OR	Regresión logística condicional
Dicotómica	Cohorte	HR	Regresión de Cox
Cuantitativa	Cualquiera		Regresión lineal

Tabla 28.5: Métodos de análisis multivariable

En la regresión logística múltiple estamos interesados en evaluar la asociación entre un factor de exposición (que puede ser cualquier tipo de variable) y un evento, que en este caso es una variable categórica dicotómica, ajustada por posibles confundidores, saber si esta relación muestra interacción con diferentes categorías de alguna otra variable y determinar si otras variables están asociadas de forma independiente con el evento.

Ya hemos estudiado en el capítulo 16 cómo se evalúa la asociación entre dos variables continuas, a través de la ecuación de la recta.

Ahora bien, ¿qué sucede en un modelo de regresión logística simple?

En la regresión logística la variable evento, en la que estamos interesados, es una variable dicotómica (variable dependiente), que tomará un valor = 0 cuando no se produzca el evento y un valor 1 cuando este ocurra.

Si cuantificamos la ocurrencia del evento (accidente cerebrovascular) entre hipertensos y no hipertensos, algunos sujetos presentarán el evento y otros no. Por lo que la probabilidad del evento será 1 o 0, sin valores intermedios. No tendremos una relación lineal, sino dicotómica, como vemos en la Figura 29.1.

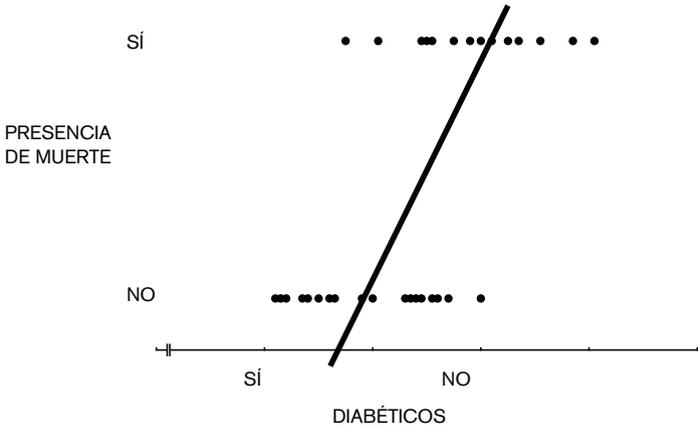


Figura 29.1: Relación entre variables dicotómicas

Si la frecuencia del evento no muestra relación lineal con el factor de exposición, tampoco lo harán las medidas de efecto de esta asociación (OR, RR o HR).

Para lograr una relación no dicotómica necesitamos *transformar* esta relación mediante la utilización de logaritmos. En este caso especial con el logaritmo natural (ln).

Entonces observaremos una relación de tipo exponencial, donde cambios en los valores de la variable dependiente (x) producirán cambios en la probabilidad del evento (y). (Figura 29.2)

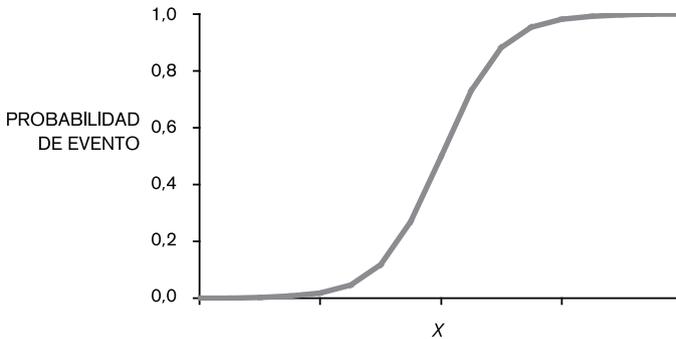


Figura 29.2: Función logística

En un modelo de regresión logística simple evaluamos sólo la asociación entre el factor de exposición y el evento (variable dicotómica) sin tener en cuenta la presencia de otras variables, a través de su asociación medida como OR.

Para determinar el valor del OR de la asociación estudiada, partimos de la fórmula de OR y luego veremos cómo se construye una ecuación que represente la función logística.

$$\text{Si el OR} = \frac{\text{odds en expuestos}}{\text{odds en no expuestos}}$$

Entonces el *odds* en expuestos (posibilidad de sufrir un evento vs. no sufrirlo) será igual a

$$\text{Odds expuestos} = \text{OR} \times \text{odds no expuestos}$$

(pasamos el término *odds* en expuestos al otro lado del signo =)

Si trabajamos con el ln del odds en expuestos, la fórmula anterior se transforma en:

$$\text{Log odds en expuestos} = \text{log OR} + \text{log odds en no expuesto}$$

(OR x *odds* en no expuestos pasa a ser la suma de estos dos términos, debido a que en los logaritmos la multiplicación cambia por suma, la división por resta y para obtener valores de *odds* debemos calcular la exponencial de los logaritmos)

En la regresión logística lo que predécimos es el *log odds* en expuestos (logaritmo de la posibilidad de sufrir el evento vs. no sufrirlo en sujetos expuestos), lo que sería semejante a predecir la variable Y en la regresión lineal, a partir de una constante, alfa, que en este caso sería el *odds* en no expuestos, debido a que es el riesgo basal que siempre estará presente, más el *log* de OR de la asociación, lo que sería el valor de la pendiente de la regresión lineal (coeficiente beta).

A través de la predicción del *log* del *odds* en expuestos, podemos calcular el OR de la asociación.

Función logística:

$$\text{Log odds} = \alpha + \beta_1 x_1$$

α = *log odds* del evento en no expuestos

β = *log odds ratio* asociado con el factor de exposición

x = factor de exposición (ausente = 0, presente = 1)

Como vemos, hemos creado una fórmula semejante a la ecuación de la recta.

Entonces, si el evento es accidente cerebrovascular y el factor de exposición hipertensión, el *log odds* en hipertensos (posibilidad de que los hipertensos sufran un accidente cerebrovascular vs. que no lo sufran) será:

$$\text{Log odds hipertensos} = \text{log odds no hipertensos} + \text{log OR} * x$$

Supongamos que el OR = 2,4 para la asociación hipertensión y accidente cerebrovascular y el *odds* en no hipertensos es de 1,05 (riesgo de accidente cerebrovascular en no hipertensos).

El *log* del OR es el ln del OR = 0,8810 y el *log* del *odds* en no hipertensos = 0,0511.

Por lo tanto, el *log* del *odds* en hipertensos es = 0,0511 (constante alfa) + 0,8810 (coeficiente beta)* hipertensos (presente = 1, ausente = 0).

Si la variable hipertensión está presente (valor = 1), entonces el *log* del *odds* en expuestos es = 0,0511 + 0,8810*1 = 0,93

Si la variable hipertensión está ausente (valor = 0), entonces el *log* del *odds* en expuestos es = 0,0511, sólo la constante, porque 0,8810*0 = 0.

Para interpretar estos valores debemos calcular su exponencial (que es la inversa del ln).

Odds accidente cerebrovascular en hipertensos = *odds* en no hipertensos (exponencial de 0,511) x (exponencial de 0,8810).

Odds accidente cerebrovascular en hipertensos 1,05 * 2,4 = 2,52

Entonces el *odds* o posibilidad de sufrir un accidente cerebrovascular en hipertensos es 2,52, si la variable hipertenso está presente.

Si la variable hipertenso está ausente, no se puede calcular el OR y por lo tanto sólo queda el valor de la constante alfa (1,05), que es el *odds* en no hipertensos.

Veamos ahora un reporte de la regresión simple entre accidente cerebrovascular e hipertensión:

Accidente cerebrovascular	Coficiente	Std. Err	z	P> z	IC 95%
Hipertensión	0,881	,0010174	0,03	0,01	0,4064 - 0,92345
Constante	0,0511	,001222	12,36	0,000	0,0378 - 0,07812

En los programas estadísticos podemos calcular directamente el valor del OR:

Accidente cerebrovascular	OR	Std. Err	z	P> z	IC 95%
Hipertensión	2,52412	,1945221	0,75	0,03	2,2564-2,7124

Hasta aquí, vemos que el accidente cerebrovascular se asocia a muerte. El valor p es 0,03 y el IC 95% no contiene al 1.

Como todo cálculo en Estadística, el coeficiente tiene un error estándar (Std. Err); el cociente entre el coeficiente y su error estándar da un valor que vemos como Z , que mediante una prueba llamada prueba de Wald (*Wald test*) calcula el valor p de la asociación entre accidente cerebrovascular e hipertensión. También se informa el IC 95% de coeficiente.

Estos datos transformados (elevados a la exponencial) nos dan los valores de OR, error estándar e IC 95% para que puedan ser interpretados. Lo que no se reporta es el valor de la constante, debido a que esta se calcula a través de los logaritmos.

Ahora vamos a agregar una variable posiblemente confundidora, el sexo:

Accidente cerebrovascular	Coficiente	Std. Err	z	P> z	IC 95%
Hipertensión	,0359266	,1748587	0,21	0,037	-,30679 ,3786433
Sexo	-,53663	,230376	-2,33	0,020	-,9881587 -,0851013
Constante	-1,777392	,1249245	-14,23	0,000	-2,02224 -1,532545

Aquí el *log* del odds es:

$$\log \text{ odds en hipertensos} = 1,777392 - 0,53663 \cdot 1 + 0,0359266 \cdot 1 = -2,2780954$$

Accidente cerebrovascular	OR	Std. Err	z	P> z	IC 95%
Hipertensión	1,03658	,1812549	0,21	0,037	7358051 1,460302
Sexo	,5847154	,1347044	-2,33	0,020	,3722615 ,9184192

El OR para hipertensión y accidente cerebrovascular, ajustado por sexo, será:

$$\text{Exp coeficiente de hipertensión} = \exp 0,0359266 = 1,03658$$

El OR para sexo (hombres vs. mujeres) ajustado por hipertensión es la exponencial del coeficiente de sexo = $\exp -0,53663 = 0,5847154$

Observemos cómo han cambiado los coeficientes y por lo tanto los OR; sin ajustar el OR de la asociación entre hipertensión y accidente cerebrovascular era de 2,52 y luego del ajuste por sexo, el OR ajustado es = 1,03, por lo que el sexo es un confundidor de la asociación entre accidente cerebrovascular e hipertensión.

La variable sexo (en este caso mujeres vs. hombres) muestra que reduce significativamente el OR de accidente cerebrovascular (OR = 0,58), ajustado por hipertensión, con un valor $p = 0,02$ y un IC 95% que no contiene al 1.

Por lo tanto, la hipertensión se asocia de forma independiente a accidente cerebrovascular, aunque luego de ajustar por sexo el valor del OR haya cambiado, pero no se perdió la asociación, como lo muestra el valor de p y el IC 95 % del OR.

El sexo también está asociado a accidente cerebrovascular, pero de forma beneficiosa: ser mujer reduce el riesgo de sufrir una accidente cerebrovascular y por lo tanto es un predictor independiente, luego de ser ajustado por hipertensión, y además es un factor de confusión entre la asociación hipertensión y accidente cerebrovascular.

Ahora bien ¿cómo sabemos si la introducción de la variable sexo mejoró el cálculo del OR para la asociación en estudio?

Lo que se hace es comparar el modelo inicial vs. el modelo con la variable agregada con diferentes pruebas estadísticas, como el estadístico *C*, *likelihood ratio*, o el cálculo de ajuste de Hosmer y Lemeshow.

Si estas pruebas arrojan un valor $p < 0,05$, significa que el modelo con la variable agregada es mejor que el modelo sin la variable.

A medida que vamos agregando variables, de a una, debemos comparar el modelo con y sin la variable agregada y decidir con cuál modelo se continúa, por ejemplo, el agregado de sexo al modelo podría no mejorar el cálculo del log del odds, y entonces volvemos al modelo inicial y a partir de ahí agregamos otra variable que puede ser cuantitativa o cualitativa de cualquier tipo.

De acuerdo con el valor p obtenido con la prueba de *likelihood-ratio*, que fue $< 0,05$, lo que interpretamos es que el agregado de sexo mejora la predicción del odds en expuestos.

Agreguemos ahora una variable categórica ordinal, como edad estratificada en 4 categorías.

Como hemos visto, cuando se desea calcular una medida de efecto de una variable categórica ordinal, recordamos que la categoría menor (en este caso los de menor edad) se toma como referencia, lo que equivaldría a que la variable de referencia se codifica como cero y por lo tanto desaparece el término de esa categoría.

Accidente cerebrovascular	Coefficiente	Std. Err	z	P> z	IC 95%
Hipertensión	-,0417511	,1790853	-0,23	0,036	-,3927517 ,3092495
Sexo	-,5559814	,2315306	-2,40	0,016	-1,009773 -,1021897
Edad cat 2	,1457143	,4084611	0,36	0,721	-,6548548 ,9462834
Edad cat 3	,7417695	,3903714	1,90	0,057	-,0233443 1,506883
Edad cat 4	-,1310124	,8319345	-0,16	0,875	-1,761574 1,499549
Constante	-2,216948	,3757655	-5,90	0,000	-2,953434 -1,480461

Accidente cerebrovascular	OR	Std. Err	z	P> z	IC 95%
Hipertensión	,9591085	,1717622	-0,23	0,036	,6751963 1,362402
Sexo	,5735091	,1327849	-2,40	0,016	,3643016 ,9028582
Edad cat 2	1,156866	,4725346	0,36	0,721	,5195175 2,576117
Edad cat 3	2,099647	3,8196422	1,90	0,057	,976926 4,512644
Edad cat 4	,8772069	,7297787	-0,16	0,875	,1717743 4,47967

Ahora vemos que la hipertensión sigue asociada a accidente cerebrovascular, con un OR aún menor que en el modelo anterior, por lo tanto la adición de la edad muestra que es un confundidor de la relación entre hipertensión y accidente cerebrovascular.

El sexo sigue estando asociado a accidente cerebrovascular, con un valor de OR casi igual, por lo que el agregado de la edad no es un confundidor del OR de sexo.

Las categorías de edad muestran un incremento del OR en cada categoría, comparadas con la categoría basal o de referencia, pero no están asociadas a accidente cerebrovascular.

Cada categoría de edad se compara con el grupo asignado como referencia y están ajustadas por hipertensión y sexo.

El agregado de la variable edad contribuyó positivamente al modelo, de acuerdo con el valor p obtenido con la prueba de *likelihood-ratio*.

Por lo tanto, aunque algunas variables no estén asociadas a accidente cerebrovascular, es importante mantenerlas en el modelo debido a que mejoran la predicción del *log* del *odds*.

La fórmula general del modelo de regresión logística múltiple sería:

$$\text{Log Odds} = \alpha + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5$$

α = es la constante, valor del log del odds en no hipertensos

β_1 = es el coeficiente de la variable X1 hipertensión (que toma valores 1 o 0)

β_2 = es el coeficiente de la variable X2 sexo (que toma valores 1 o 0)

β_3 = es el coeficiente de la variable X3 edad por categoría 2 (que toma valores 1 o 0)

β_4 = es el coeficiente de la variable X3 edad por categoría 3 (que toma valores 1 o 0)

β_5 = es el coeficiente de la variable X3 edad por categoría 4 (que toma valores 1 o 0)

Si aplicamos esta fórmula a un sujeto con hipertensión, mujer y cuya edad está en la categoría 3, obtendremos la siguiente fórmula:

$$\text{Log odds} = \alpha + \beta_1x_1 + \beta_2x_1 + \beta_3x_0 + \beta_4x_1 + \beta_5x_0$$

Resumiendo quedaría:

$$\text{Log odds} = \alpha + \beta_1x_1 + \beta_2x_1 + \beta_4x_1$$

Como conclusión de este modelo podemos decir que las variables hipertensión y sexo son predictores independientes de accidente cerebrovascular.

El sexo es un confundidor de la asociación hipertensión-accidente cerebrovascular y la introducción de la variable edad, como variable categórica, no es confundidor de ninguna asociación, ni se asocia a accidente cerebrovascular.

El agregado de las otras variables permitió ajustar la relación entre accidente cerebrovascular e hipertensión y la de otras variables con accidente cerebrovascular. Todas las variables incluidas en el modelo se ajustaron unas con otras.

El modelo de regresión logística múltiple puede ser construido agregando una variable (*forward*) o bien empezando con todas las variables e ir quitando de a una (*backward*). A este método se lo denomina *stepwise* (modelo a pasos).

Los procedimientos que hay que realizar antes de llegar a desarrollar un modelo de regresión logística múltiple son, en primer lugar, realizar un análisis de todas las variables para determinar su asociación con el evento.

Las variables que se seleccionan para ingresar al modelo de regresión logística son aquellas que han mostrado asociación con el evento y aquellas que clínicamente tienen impacto sobre el evento, aunque no se haya demostrado asociación estadística con él. Algunos autores deciden qué variables ingresar al modelo, tomando como punto de corte un valor $p \leq 0,1$, de esa forma se incluyen más variables al modelo, pero también el modelo es más complicado de desarrollar.

El mejor modelo es aquel que tiene la menor cantidad de variables posibles, de acuerdo con la comparación de los diferentes modelos, debido a que aumenta su poder estadístico. Por ejemplo, podemos reducir la cantidad de variables incluyendo las variables cuantitativas como tales en lugar de introducirlas al modelo como categóricas.

En conclusión, el modelo de regresión logística múltiple permite calcular el valor del OR de la asociación estudiada, controlado por confundidores, evaluando posibles interacciones con otras variables y determinar variables asociadas de forma independiente con el evento.

ANÁLISIS MULTIVARIABLE. REGRESIÓN DE COX



Para los estudios de cohorte y experimentales, en los cuales se evalúa la incidencia o la tasa de incidencia de uno o varios eventos durante un período determinado de seguimiento, las medidas de efecto que se utilizan son el RR o el cociente de tasas (*rate ratio*), las que se miden en un punto intermedio predeterminado, o bien al final del seguimiento.

También, como vimos en el análisis de sobrevida, es posible determinar la probabilidad de sufrir un evento en un punto determinado de tiempo y comparar esta probabilidad de sufrir un evento entre sujetos expuestos y no expuestos a un factor de exposición, mediante la construcción de curvas de Kaplan-Meier y su comparación con la prueba *log rank*.

Ahora bien, el riesgo de sufrir un evento durante el período de seguimiento cambia a medida que transcurre el tiempo de seguimiento, por lo tanto necesitamos una medida de efecto que tenga en cuenta el tiempo como covariable. Dicho en otras palabras, el tiempo es una variable que influencia la ocurrencia de eventos.

También, como ocurre en cualquier otro diseño observacional, es necesario ajustar el riesgo de sufrir un evento por posibles confundidores, valorar la interacción de la asociación evaluada con variables ordinales, determinar si otras variables se asociaron al evento de forma independiente y desarrollar puntajes de riesgo, ya sea en estudios observacionales de cohorte, como el conocido puntaje de riesgo de eventos cardiovasculares a 10 años obtenido del estudio de cohorte de Framingham, o como puntajes de riesgo derivados de estudios experimentales.

Para realizar un análisis multivariable en este tipo de estudios, que nos permita lograr las mismas utilidades que la regresión logística, necesitamos un método que tenga en cuenta el tiempo de seguimiento y por lo tanto la incidencia de eventos.

Este método se denomina regresión de Cox.

DESCRIPCIÓN DEL MÉTODO

Cuando el tiempo de seguimiento es largo y cuando la tasa de un evento cambia muy rápido en el tiempo, el modelo de regresión adecuado para analizar estos datos es la regresión de Cox.

En general, la regresión de Cox se utiliza a partir de un mes de seguimiento si la tasa de eventos se produce rápidamente, por ejemplo, si estudiamos posibles factores de exposición asociados a infección respiratoria por virus sincicial respiratorio en niños, veremos que en un mes el número de niños infectados (evento) se incrementa rápidamente; al contrario, si deseamos determinar la tasa de incidencia de diálisis en pacientes recientemente diagnosticados con diabetes tipo 2, lo más probable es que en un mes no tengamos ningún evento.

Por lo tanto, la aplicación de la regresión de Cox dependerá del tipo de evento y del tiempo de seguimiento.

Como en todo análisis multivariable, lo que se desea predecir es el riesgo de sufrir un evento, lo que denominamos variable dependiente, a partir de uno o varios factores de exposición denominados variables independientes o covariables.

La variable dependiente que se desea predecir en la regresión de Cox es la tasa de un evento y esto se debe a que los sujetos no contribuyen con la misma cantidad de tiempo al seguimiento.

MODELO MATEMÁTICO DE COX

La fórmula matemática de la regresión de Cox es muy complicada, por ese motivo omitiremos referirnos a ella, pero vamos a describir su razonamiento.

El modelo de Cox puede describirse como:

$$\text{Tasa del evento} = \text{tasa basal} \times \text{factor de exposición}$$

La tasa basal del evento es la de los sujetos no expuestos, que cambia durante el tiempo de seguimiento.

El factor de exposición es el efecto de este sobre la tasa de eventos. Por ejemplo, el cociente de tasas (*rate ratio*) del grupo expuesto vs. el grupo no expuesto.

Ambos parámetros son utilizados para describir el modelo, sin embargo lo único que se estima es el efecto del factor de exposición sobre la tasa de eventos, a diferencia de la regresión logística en la cual se calcula el log del odds en expuestos o riesgo de sufrir el evento en sujetos expuestos.

A diferencia del método de regresión logística, no se calcula una constante, que representa el riesgo en el grupo no expuesto.

Por ejemplo, impacto que tiene la pobreza sobre la tasa de virus sincicial respiratorio vs. una población no expuesta a la pobreza, podría incrementar un 30% la tasa de virus sincicial respiratorio durante el período de seguimiento del estudio.

HAZARD RATIO

En la regresión de Cox la tasa es referida como *hazard* (riesgo o peligro en inglés) y el *rate ratio* como *Hazard Ratio* (HR), que puede interpretarse de la misma forma que el RR o el OR.

Si el IC 95% de HR contiene al 1, sabremos que no hay asociación entre el factor de exposición y el evento, debido a que el factor de exposición puede reducir el evento, tener la misma tasa de eventos o aumentar la tasa de eventos, con respecto al grupo no expuesto.

Como hemos visto en la construcción de las curvas de Kaplan-Meier, en estos estudios tenemos el dato de la fecha de ocurrencia del evento y de la cantidad de sujetos censurados.

La regresión de Cox toma estos datos y divide el tiempo de seguimiento en períodos de tiempo muy pequeños, tal que sólo admite la ocurrencia de un evento como máximo, y a este período de tiempo pequeño se lo denomina *time click*, asumiendo que dentro de este *time click* la tasa de evento es constante, ya que sólo ocurrió un evento. Con un evento que se produce rápidamente, la tasa del evento no puede considerarse constante para períodos más prolongados de tiempo.

La regresión de Cox toma en cuenta los datos de los sujetos con eventos y censurados de acuerdo con estos períodos de tiempo pequeños o *time clicks*.

Para el cálculo del *rate ratio* en cada *time click* el modelo utiliza como numerado el valor 1, debido a que ocurrió un evento y, como denominador, no el cálculo de personas/tiempo en riesgo, sino el grupo de personas en riesgo de sufrir el evento en ese momento, lo que se denomina grupo en riesgo (*risk group*) porque la variable tiempo está siendo tenida en cuenta a través de los *time clicks*. En general se observa debajo de las curvas de Kaplan-Meier como número de sujetos en riesgo (*n° at risk*).

Al dividir el tiempo total de seguimiento en pequeños *time clicks* finalmente se asume que entre cada *time click* la tasa de evento va cambiando con el tiempo.

Dentro de cada *time click* la tasa de eventos será = $\frac{1}{n^\circ \text{ personas cuando ocurre el evento}}$

Cuando no ocurre un evento, la tasa de eventos es = 0 y no hay *time click*, hasta que se produzca un nuevo evento.

CRITERIO DE PROPORCIONALIDAD

La ventaja de utilizar la regresión de Cox es que podemos modelar la relación entre un factor de exposición y un evento o su tasa y el tiempo, de forma de tener en cuenta el cambio de tasa de evento durante el tiempo de seguimiento.

La asunción de la regresión de Cox es que el efecto del factor de exposición es proporcional durante el tiempo de seguimiento. Lo que significa que durante todo el estudio, por ejemplo, la pobreza incrementará la tasa de virus sincicial respiratorio un 30% durante todo el seguimiento, aunque la tasa basal en no expuestos cambie con el tiempo.

En otras palabras, el *rate ratio* o HR entre expuestos y no expuestos es constante durante todo el seguimiento, aunque la tasa en no expuestos cambie con el tiempo. (Figura 30.1)

Por este motivo es que se conoce a la regresión de Cox como método de riesgos proporcionales.

Determinar la proporcionalidad de los riesgos es fundamental para aplicar este método; si esta asunción no se cumple, no se puede utilizar regresión de Cox para analizar los datos.

Existen dos formas de verificar este supuesto.

1) Observando las curvas de Kaplan-Meier, que deben mostrar una divergencia constante de las curvas. (Figura 30.2)

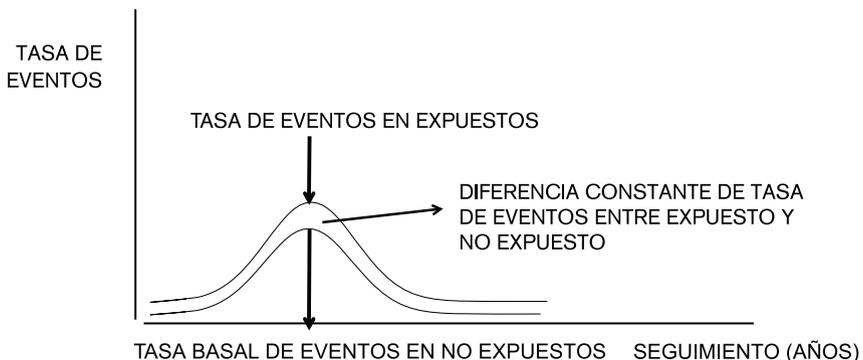


Figura 30.1: Riesgo proporcional

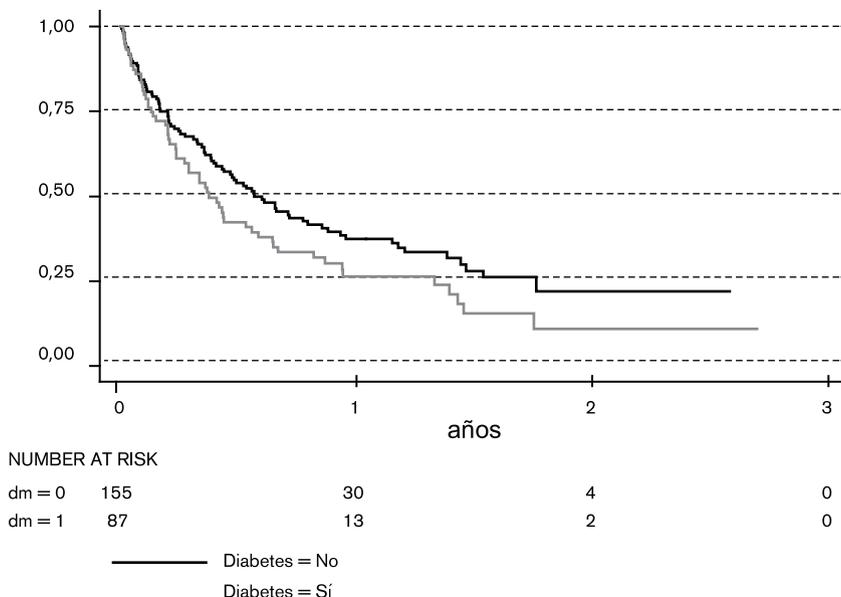


Figura 30.2: Curvas de Kaplan-Meier. Sobrevida de acuerdo a presencia de diabetes

Como se ve en la Figura 30.2, los pacientes con diabetes durante todo el seguimiento tienen menor sobrevida que los pacientes sin diabetes.

En cambio, si las curvas se cruzan, significa que no se cumple con el supuesto de proporcionalidad de los riesgos. (Figura 30.3)

Como vemos en la Figura, durante el seguimiento el sexo masculino tiene al principio mejor sobrevida que las mujeres y luego cruza la curva de sobrevida de estas, para mostrar peor sobrevida.

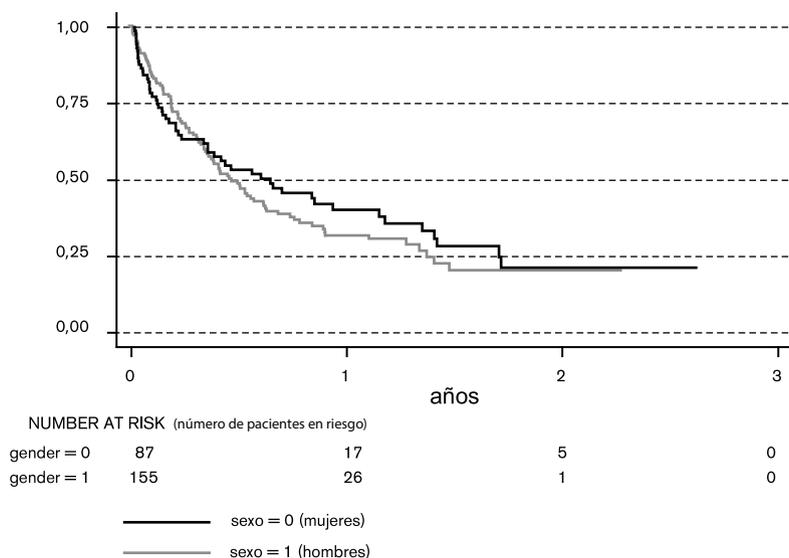


Figura 30.3: Curvas de Kaplan Meier. Sobrevida de acuerdo a sexo

En este caso no se cumple con la asunción de proporcionalidad de los riesgos.

Este método es subjetivo, por lo que existe una prueba formal para determinar la proporcionalidad de los riesgos.

2) La forma objetiva de determinar proporcionalidad es examinar el log de la tasa acumulativa de eventos entre expuestos y no expuestos.

Las curvas deben mostrar que son paralelas durante todo el período de tiempo de seguimiento. A esta comparación logarítmica del *hazard* acumulativo se la denomina curvas de Nelson Aalen.

Si este supuesto no se cumple, se debe utilizar otro método estadístico multivariable.

REGRESIÓN DE COX SIMPLE

En un estudio de cohorte se investiga la asociación entre la función ventricular izquierda, medida por la fracción de eyección (*fev*), y mortalidad a largo plazo. El tiempo de seguimiento fue de 11 años.

Realizamos entonces una regresión de Cox simple, cuyo resultado se observa en la siguiente tabla.

Variable	Hazard ratio	ES	Z	p	IC 95%
Fev	,9599	,0087	-4,46	,000	,9429348 ,9773541

El resultado muestra que el HR es de 0,9599, con un valor $p = 0,000$, lo que significa que por cada porcentaje de incremento de la *fev* (la variable fue introducida como variable continua) el riesgo de muerte a 11 años se reduce un 4%. Tomando el valor de $HR = 1$, restamos el valor del HR obtenido (0,96) resulta en 0,04, multiplicado por 100, la reducción del riesgo es de 4%.

Además vemos que el valor p es $< 0,05$, lo que significa que existe asociación entre el factor de exposición y el evento. Esto lo confirma el valor del IC 95% del HR, el que no contiene el valor = 1.

Como vemos, este método nos informa el impacto de la presencia de la variable (*fev* > 35%) sobre el evento (muerte), sin informar el riesgo de los pacientes no expuestos.

Este valor de HR no está ajustado por ninguna variable posiblemente confundidora, por lo que a este HR lo denominamos HR crudo.

También podríamos introducir la variable como dicotómica, agrupando los pacientes con valores con *fev* $\leq 35\%$ vs. $> 35\%$.

En este caso obtendríamos el siguiente resultado:

Variable	Hazard ratio	ES	Z	p	IC 95%
Fev > 35%	,46157	,090	-3,95	,000	,3144842 ,6774627

En este caso comparamos el riesgo de muerte a 11 años de seguimiento entre pacientes con *fev* > 35% vs. $\leq 35\%$. Los pacientes con *fev* > 35% muestran una reducción de la mortalidad a 11 años de 54%. La asociación sigue siendo significativa, como lo demuestra el valor p y el IC 95% del HR.

La forma de informar el impacto de la variable (continua o dicotómica) dependerá del criterio del investigador.

REGRESIÓN DE COX MÚLTIPLE

Veamos ahora si la asociación cambia con el agregado de una variable posiblemente confundidora.

Variable	Hazard ratio	ES	Z	p	IC 95%
<i>Fey</i> > 35%	,4677455	,0915	-3,88	,000	,318687 ,6865226
Edad	1,037687	,0189	2,02	,043	1,001174 1,075532

Como vemos, la introducción de la edad como variable continua no modifica el HR de la *fey*, o sea que el HR de *fey* ajustado por edad, no está confundido por la edad.

Con relación a la edad, se observa que por cada año de incremento de edad el riesgo de muerte a 11 años se incrementa un 3%. Además, la edad se asocia significativamente a mortalidad y de forma independiente, ya que el HR de edad está ajustado por *fey*, y el valor de p es $< 0,05$ y del IC 95% del HR no contiene el valor = 1. Podemos decir que la edad es un predictor independiente de muerte a largo plazo y no es confundidor de la relación entre *fey* y muerte.

Podemos seguir agregando posibles variables confundidoras.

Variable	Hazard ratio	ES	Z	p	IC 95%
<i>Fey</i> > 35%	,4772978	,0935	3,77	,000	3250452 ,7008663
Edad	1,035481	,0190	1,90	,057	,9989015 1,073399
Sexo femenino	,5942261	,1533	-2,02	0,044	,358311 ,9854697
Enf. coronaria	,7619037	,1437	-1,44	0,150	,526339 1,102896

En este modelo de regresión de Cox, vemos que el HR de la *fey* se ha modificado muy poco, por lo que la introducción de sexo femenino y de la presencia de enfermedad coronaria son confundidores *leves* de la asociación entre *fey* y muerte.

La introducción de estas variables definitivamente es confundidora de la relación entre edad y muerte; si bien el valor del HR de edad no se modificó, sí se perdió la asociación entre edad y muerte debido a que el valor p es $> 0,05$ y el IC 95% del HR contiene el valor = 1.

Entonces el HR de edad ajustado por *fey*, sexo femenino y enfermedad coronaria ya no tiene impacto sobre la mortalidad.

El sexo femenino vs. sexo masculino muestra que reduce la mortalidad a largo plazo un 40% de forma significativa (ver valor de p e IC 95% HR). Por lo que el sexo femenino es un predictor independiente de supervivencia, ya que su HR está ajustado por la presencia de las otras variables.

La presencia de enfermedad coronaria no está asociada a mortalidad (ver valor p e IC 95% HR) y es confundidor de la asociación entre edad y muerte.

Cuando se evaluó la introducción de cada una de las variables posiblemente confundidoras, la prueba estadística arrojó un resultado significativo, motivo por el cual las variables, aunque no se asocian a muerte, mejoran el valor de predicción de muerte del modelo. Esto es importante si se desea desarrollar un sistema de puntaje que prediga el evento a largo plazo.

Como hemos visto, la regresión de Cox posibilita determinar si la asociación de interés, una vez ajustada por confundidores, se mantiene y nos permite evaluar el impacto del factor de exposición de interés sobre el evento a largo plazo.

Permite determinar la existencia de otras variables asociadas, independientemente del evento (predictores independientes).

La regresión de Cox se debe utilizar en estudios que tengan como mínimo un mes de seguimiento y la tasa del evento varíe rápidamente en el tiempo.

El método informa el impacto del factor de exposición sobre el evento, a través del cálculo del HR, que se interpreta de la misma forma que el RR u OR.

Este método cumple con las mismas funciones que la regresión logística, la diferencia está en que debe aplicarse a un determinado tipo de estudio y su método matemático es diferente; no obstante, su interpretación es la misma que en la regresión logística.

Uno de los objetivos más importantes de la Epidemiología es investigar la etiología o las causas de las enfermedades.

Hemos visto cómo, a través del correcto diseño y análisis de un estudio epidemiológico, podemos establecer la asociación entre un factor de exposición y una enfermedad.

Cuando obtenemos el resultado de un estudio de investigación debemos preguntarnos si esta asociación es real o ficticia.

Las causas por las cuales una asociación entre un factor de exposición o factor de riesgo y enfermedad puede no ser cierta pueden deberse a un mal diseño del estudio por introducción de sesgos, un mal análisis que no haya tenido en cuenta factores de confusión o a que el análisis estadístico fue mal realizado y la asociación es sólo por azar. Así, al realizar otro estudio esta asociación no puede comprobarse de nuevo.

Si asumimos que el resultado de un estudio es cierto (existe una verdadera asociación entre factor de exposición y enfermedad) debemos preguntarnos si este factor de riesgo es la causa de la enfermedad. Que un factor de riesgo esté asociado con una enfermedad no implica que sea la causa de la misma.

TIPOS DE RELACIÓN CAUSAL

La relación causal entre un factor de riesgo y una enfermedad puede ser directa o indirecta. En una relación causal directa el factor A es causa de la enfermedad sin intermediarios. En una relación indirecta el factor causal A produce la enfermedad sólo a través de pasos intermedios.

En biología humana este tipo de relación causal es la más frecuente. Siempre se desencadenan pasos intermedios entre la introducción de un factor de riesgo y una enfermedad, sin los cuales la enfermedad no se produce.

Existen cuatro tipos de relaciones causales propuestas como modelos etiológicos:

1. Causa necesaria y suficiente para producir la enfermedad
2. Causa necesaria pero no suficiente para producir la enfermedad
3. Causa suficiente no necesaria para producir la enfermedad
4. Causa no necesaria ni suficiente para producir la enfermedad

La causa necesaria y suficiente es aquella en la que, en ausencia del factor de riesgo, la enfermedad nunca se produce. El factor de riesgo siempre es necesario y en presencia del mismo la enfermedad siempre se produce, es decir que es suficiente para desarrollar la enfermedad. Podríamos decir que hablamos de una relación directa entre agente causal y enfermedad. (Figura 31.1A)

Esta situación es muy rara. Por ejemplo, en las enfermedades infecciosas se puede observar que en una misma familia donde hay integrantes infectados por tuberculosis, otros

miembros de la misma familia que conviven no desarrollan la enfermedad. Si se asume que la exposición de la familia al bacilo fue la misma, entonces debe pensarse que existen diferentes estados de inmunidad, susceptibilidad genética u otros factores que determinan que algunos miembros de la familia desarrollen la enfermedad y otros no.

Otro modelo causal consiste en asumir que cada factor es *necesario* pero no suficiente en sí mismo para producir la enfermedad. Por lo tanto se requieren múltiples factores con una secuencia temporal determinada.

Por ejemplo, la carcinogénesis se considera un proceso que involucra diferentes etapas. Un agente iniciador actúa primero y luego un agente promotor estimula los cambios necesarios para que resulte una lesión neoplásica. La acción de iniciador o promotor por sí solos no producen cáncer.

En el ejemplo de la tuberculosis, el bacilo es claramente necesario, no obstante su presencia puede no ser suficiente para producir la enfermedad. Se necesitan los otros factores, por ejemplo genéticos. (Figura 31.1B)

El modelo que propone una *causa suficiente pero no necesaria* se basa en que el factor puede producir la enfermedad, pero otros factores también pueden hacerlo. La radiación o el benceno producen leucemia por sí solos, sin necesidad de que ambos estén presentes.

Aun en esta situación, no todos los sujetos expuestos a la radiación o al benceno desarrollan leucemia. Por lo tanto, ambos factores no son necesarios: otros cofactores probablemente sí sean necesarios. Dicho en otras palabras, en presencia de cofactores es suficiente con exponerse al benceno para desarrollar leucemia, sin la presencia de estos cofactores la presencia de benceno no es necesaria para desarrollar leucemia. (Figura 31-1C)

Cuando la causa no es ni suficiente ni necesaria, significa que diferentes combinaciones de distintos factores pueden producir la misma enfermedad. Este es el modelo más complejo, pero seguramente el más aplicable a la etiología de las enfermedades no transmisibles humanas. (Figura 31.1D)

Como hemos visto en la mayoría de los casos, no existe una sola causa para una enfermedad. Hoy sabemos que existe la *multicausalidad*.

Como modelo más sencillo podríamos decir que el factor genético es la causa suficiente (confiere susceptibilidad) y que los factores ambientales (agentes infecciosos, drogas, tóxicos, etc.) son la causa necesaria.

Bajo el modelo de multicausalidad decimos entonces que las enfermedades necesitan un *complemento causal* para producirse.

Cuando tenemos las causas suficiente y necesaria presentes, el mecanismo causal está completo.

Por ejemplo, por un lado tenemos el virus HIV y la susceptibilidad propia de un individuo a ser infectado, por otro tenemos las vías de transmisión (sexual, sanguínea y materna); cuando ambos componentes están presentes la enfermedad se producirá. Pero si las vías de transmisión no están, no habrá forma de infectarse.

Entonces, la prevalencia de una enfermedad depende de la prevalencia de su complemento causal en la población; de ahí que diferentes poblaciones tengan diferentes prevalencias de la misma enfermedad.

La prevalencia del HIV depende entonces del porcentaje de sexo seguro que tenga una determinada población, o de la prevalencia de drogadicción o detección durante el embarazo.

El conocer el complemento causal de una enfermedad es de suma importancia a la hora de hacer prevención de la misma.

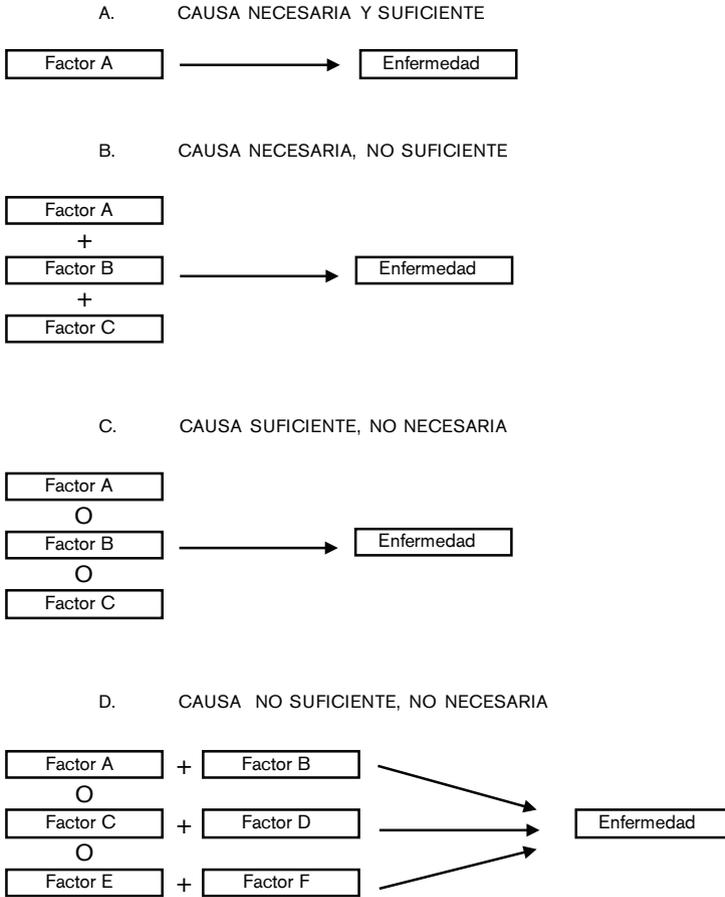


Figura 31.1: Tipos de relación causal

Asumiendo multicausalidad decimos que:

1. La causa de una enfermedad específica es cualquier factor que juegue un rol esencial en la producción de la ocurrencia de la enfermedad.
2. Una sola causa no es suficiente para producir una enfermedad.
3. Diferentes factores causales actúan juntos para causar una enfermedad específica.

Si bien puede decirse que un factor de riesgo está asociado con una enfermedad dada la complejidad de las relaciones causales, este solo hecho no es suficiente para decir que es causa: se necesita de varios otros criterios para juzgar si esta asociación es causal.

CRITERIOS DE BRADFORD HILL

Para decidir si una asociación es causa de enfermedad, existen varios criterios o características de la asociación que deben verificarse. No todas estas características deben cumplirse para sostener que una asociación es causal, pero cuantos más criterios cumpla una asociación, más fuerte será la evidencia a favor de causalidad.

- Relación temporal entre el factor de exposición y el desarrollo de enfermedad

Es claro que, si creemos que un factor es causa de una enfermedad, la exposición al mismo debe ocurrir antes de que se desarrolle la enfermedad.

Los estudios de cohortes son particularmente importantes a la hora de comprobar este criterio.

Si seguimos en el tiempo a dos grupos de sujetos, uno de expuestos al asbesto y otro no y comprobamos que el grupo expuesto desarrolla más frecuentemente cáncer de pulmón, habremos establecido el criterio de temporalidad.

Es importante también determinar el período de latencia entre la exposición y la aparición de la enfermedad. Si observamos que el cáncer de pulmón se desarrolla 15 o 20 años luego de la exposición al asbesto, los casos que se desarrollen antes, por ejemplo a los 3 años, seguramente se deberán a otra causa.

- Fuerza de asociación

Hemos visto que la asociación entre un factor de riesgo y enfermedad puede valorarse a través de diferentes medidas de asociación, como *odds ratio*, riesgo relativo, *hazard ratio*, etc.

Cuanto más fuerte sea esta asociación, más evidencia existe para aceptar su relación causal.

- Relación dosis-respuesta

A medida que la dosis del factor de exposición aumenta, el riesgo de enfermedad es cada vez mayor.

Por ejemplo, está demostrado que el número de cigarrillos por día tienen relación directa con el incremento de la mortalidad.

Ese es uno de los criterios más fuertes a favor de la causalidad. Por supuesto que su ausencia no descarta causalidad, ya que existen casos donde hay un umbral por encima del cual no se incrementa el riesgo de enfermedad.

- Replicabilidad

Si la relación es causal, es esperable que se observe en estudios con diferentes diseños y en poblaciones diferentes. Hay que tener en cuenta que es muy poco probable que diferentes estudios lleguen a la misma conclusión sobre asociación sólo por azar, debido a que cada uno de ellos tiene diferentes sesgos, confundidores y métodos de análisis.

También es esperable que la asociación causal se verifique en subpoblaciones.

- Plausibilidad biológica

Este criterio se refiere a la coherencia que tiene la asociación hallada con el conocimiento de la biología actual. Existen muchos ejemplos de asociaciones halladas antes de poder hallar una explicación biológica plausible, las cuales han impulsado el conocimiento de nuevos mecanismos etiológicos. Por lo tanto, este criterio depende del conocimiento que se tenga del mecanismo de la enfermedad en ese momento.

- Consideración de una explicación alternativa

Cuando se interpreta una asociación observada siempre existe la posibilidad de una explicación alternativa, que tiene que ser tenida en cuenta y suficientemente descartada con evidencia para asegurar que una determinada asociación es causal y no otra que sea confundidora.

- Cesación del factor de exposición o evidencia experimental

Si un factor es la causa de una enfermedad, es de esperar que el riesgo de la misma disminuya con la reducción o eliminación del factor de exposición. Por ejemplo, en las

ciudades donde se ha logrado el control de la contaminación ambiental, la tasa de trastornos respiratorios infantiles ha mostrado una disminución muy importante.

- **Especificidad de la asociación**

Esto indica que un factor de exposición está asociado con una sola enfermedad. Este criterio es uno de los más débiles, debido a que, como sabemos, la etiología de la mayoría de las enfermedades es multicausal, aun las infecciosas, donde factores genéticos y de inmunidad predisponen o no a contraer la enfermedad. Otro ejemplo es el cigarrillo, responsable de varios tipos de cáncer, enfermedad cardiovascular y respiratoria.

- **Analogía**

Si la asociación es causal, esperaríamos que los hallazgos sean consistentes con datos de otros procesos patológicos similares.

Si sabemos que un factor produce una determinada enfermedad, otro factor de la misma clase producirá una enfermedad similar.

Si bien todos estos criterios no permiten una estimación cuantitativa a favor o en contra de una relación causal, ellos son de gran ayuda a la hora de decidir si hay evidencia suficiente para asegurar que una asociación es causal.



ÍNDICE ANALÍTICO

A

- Ajuste (*véase confundidores*)
- Aleatorización 43, 46
 - estratificada 47
 - por bloques 47
 - simple 46
- Análisis de la varianza 137
 - de sobrevida 199
 - multivariable 237
- ANOVA (*véase análisis de varianza*) 137
- Aparear 40
- Área de rechazo 116
- Asociación 19

B

- Box-Plot* 102

C

- Cálculo de probabilidad 90
 - de Z 118,164
- Casos 37,80
- Causalidad reversa 30
- Centilos 102
- Coefficiente de correlación 149
- Controles 64
- Corrección de Bonferroni 176
 - por continuidad de Yates 172
- Cuartiles 102
- Curva normal estandarizada 117

D

- Datos apareados 130
- Diagrama de barras 155
- Diseño Cross-Over 48
 - factorial 48
 - paralelo 48
- Distancia intercuartiles 102
- Distribución asimétrica 100
 - simétrica 100
 - gaussiana 112
- Doble ciego 43

E

- Efectividad 45
- Eficacia 45
- Error estándar 121
 - Tipo I / Alfa 139
 - Tipo II / Beta 139
- Estimación 89
- Estratificación 71
- Estudio de cohorte retrospectivo 34
- Estudios controlados aleatorizados 43
 - de prevalencia 25

F

- Factor de exposición 18
- Factores de confusión 31
- Falacia ecológica 57
- Falso negativo 139
 - positivo 139
- Fracción atribuible 194
 - atribuible poblacional 196
- Fuerza de asociación 260
- Funnel plot* 231

G

- Grados de libertad 124

H

- Heterogeneidad 227
- Hipótesis alternativa 91
 - nula 91
- Histograma 97

I

- Incidence Rate* 75
- Risk* 75
- Intención de tratar 50

K

- Kappa (*véase prueba de concordancia*) 218

L

Lead Time Bias 222
Length Time Bias 223
Log Rank test 204

M

Mann-Whitney 135
Muestras apareadas 130
 independientes 132
Muestreo al azar 30

O

Objetivo primario 44
Objetivos secundarios 44

P

Pearson (*véase coeficiente de correlación*) 149
Poder de la muestra 147
 estadístico 148
Potencia del estudio 146
Promedio 95
Proporción 74
Prueba de Hipótesis 163
 de McNemar 174
 de significación estadística 128
 de Student 128,133
 de Wilcoxon 135
 del signo 135
 diagnósticas 207
 de Fisher 173
Punto final primario 44
 primario combinado 45
Puntos finales surrogantes 45

Q

Quintilos 104

R

Rango Intercuartilo 102
Riesgo absoluto 181
Reducción del riesgo relativo 183

Registro (*véase diseño prospectivo*) 22

Regresión de Cox 249
 lineal 150
 logística 241

Relación causal 257

Riesgo atribuible 193
 atribuible poblacional 195

S

Scatter Plot 149
Sensibilidad 209
Sesgo de asignación 43
 de duración 223
 de recuerdo 65
 de seguimiento 36
 por adelanto 222
Sesgos de publicación 229
 de respuesta 60
 de vigilancia 61
Sobrevida acumulativa 201
Spearman 150 (*véase correlación*)
Screening 221 (*véase tamizaje*)

T

Tabla de distribución normal 119
 t de Student 124
Tiempo al evento 199

V

Valor de p 91
Variabilidad de la media 121
Variable dependiente 150
 independiente 150
Variables categóricas 86
 dicotómicas 85
 discretas 85
 nominales 86

Varianza 105

W

Wilk Shapiro 118