



DISEÑO Y ANÁLISIS EN INVESTIGACIÓN



Carlos Ochoa Sangrador





DISEÑO Y ANÁLISIS EN INVESTIGACIÓN



Carlos Ochoa Sangrador

Servicio de Pediatría.

Complejo Asistencial de Zamora.

CERSO (Clinical Epidemiology

Research Support Office)

Objetivo general

- Aprender a diseñar y analizar un proyecto de investigación.

Objetivos específicos

- Aprender a elaborar la memoria de un proyecto de investigación, apartados y criterios de calidad.
- Aprender a elegir el diseño y los tipos de variables.
- Practicar el diseño de bases de datos clínicas.
- Revisar las técnicas de muestreo y aprender a calcular el tamaño muestral.
- Revisar las medidas de frecuencia, riesgo e impacto para cada tipo de estudio y aprender a estimarlas con calculadoras epidemiológicas.
- Revisar los fundamentos de la estadística descriptiva e inferencial.
- Aprender a realizar estadística básica. Aprender a elegir el test estadístico.
- Revisar los errores epidemiológicos más comunes y los métodos de ajuste.

© Carlos Ochoa Sangrador.

Diseño, realización y coordinación editorial:



Alberto Alcocer 13, 1.º D
28036 Madrid
Tel.: 91 353 33 70 • Fax: 91 353 33 73
www.imc-sa.es • imc@imc-sa.es

ISBN: 978-84-7867-685-9
Depósito Legal: M-15006-2019

Ni el propietario del copyright, ni el coordinador editorial, ni los patrocinadores, pueden ser considerados legalmente responsables de la aparición de información inexacta, errónea o difamatoria, siendo los autores los responsables de la misma.

El autor declara no tener conflicto de intereses.

Reservados todos los derechos. Ninguna parte de esta publicación puede ser reproducida, transmitida en ninguna forma o medio alguno, electrónico o mecánico, incluyendo las fotocopias, grabaciones o cualquier sistema de recuperación de almacenaje de información, sin permiso escrito del titular del copyright.



ÍNDICE



5

CAPÍTULO 1

- La memoria de investigación.
- Definir la pregunta de investigación (PICO).
- La medida principal de efecto.

27

CAPÍTULO 2

- Búsquedas bibliográficas y gestores de citas bibliográficas.

41

CAPÍTULO 3

- Diseños epidemiológicos.
- Muestreo y aleatorización.

59

CAPÍTULO 4

- Tipos de variables.
- Diseño de bases de datos.

77

CAPÍTULO 5

- Estadística descriptiva.
- Medidas de frecuencia, riesgo e impacto en epidemiología.

97

CAPÍTULO 6

- Estadística inferencial.
- Tamaño muestral e intervalos de confianza.

113

CAPÍTULO 7

- Contraste de hipótesis.
- Elección del test estadístico.

133

CAPÍTULO 8

- Errores metodológicos

146

ANEXO

- Formulario para la Memoria de un Proyecto de Investigación.



1

La memoria de investigación

Definir la pregunta de investigación (PICO)

La medida principal de efecto

Objetivos docentes:

- Aprender a elaborar la memoria de un proyecto de investigación
- Conocer los fundamentos de la memoria de investigación
- Aprender a definir la pregunta de investigación estructurada (PICO).
- Saber elegir la medida principal de efecto
- Conocer los apartados de una memoria de investigación.



La memoria de investigación

La memoria o protocolo de investigación es un documento en el que se describe cómo se va a pasar de la concepción de un problema de investigación (idea abstracta planteada por una o más personas) a la puesta en marcha de la investigación en sí misma (reclutamiento de colaboradores, contratación de personal, reparto de tareas, adquisición de material, recogida y análisis de datos, etc.).

Toda investigación debe llevarse a cabo siguiendo métodos rigurosos y precisos que garanticen la validez de los resultados, tal y como dicta el método científico. El protocolo de investigación será el marco de referencia donde se plantea la pregunta de investigación y donde se detallan todas las definiciones operativas y procedimientos de trabajo a ejecutar en el proyecto.

Pero los investigadores no solo elaboran la memoria de investigación como manual de trabajo interno, la principal razón por la que la hacen es poder recabar permisos por parte de comisiones o comité éticos de investigación o para poder competir en convocatorias de agencias de financiación.

La preparación de un protocolo de investigación comprende tres fases: la conceptualización del problema a estudiar, la elección de una estrategia o diseño de investigación y la planificación operativa de la investigación.

En la **conceptualización**, el investigador especifica el problema a estudiar, revisa el estado de los conocimientos, define el modelo teórico que adopta y formula sus hipótesis.

A continuación, el investigador elige la **estrategia de investigación**, es decir, el tipo de diseño epidemiológico con el que pretende contrastar su hipótesis. Finalmente, en función del diseño elegido, el investigador realiza la **planificación operativa**, en la que se describen los métodos y procedimientos a seguir, se define la población a estudiar, las variables a recoger, los análisis a llevar a cabo y las consideraciones de factibilidad, éticas y económicas del proyecto.

La calidad de la memoria de investigación será el principal argumento con el que contará el investigador para “vender” su proyecto. Aunque la experiencia y currículum profesional es un importante “aval” del proyecto, una buena memoria de investigación será fundamental a la hora de obtener permisos o financiación. A lo largo del presente libro repasaremos los elementos clave a tener en cuenta en la planificación y ejecución de la investigación, lo que nos permitirá estar preparados para elaborar memorias de calidad que sean evaluadas favorablemente. Asimismo, una adecuada planificación será la principal garantía del desarrollo y culminación de nuestro proyecto.

Los **mensajes clave** que abordaremos, y que constituyen los criterios de calidad de un buen proyecto de investigación, son:

1. Elaborar una memoria de investigación antes de iniciar la recogida de datos.
2. Definir claramente la pregunta de investigación siguiendo la estructura: Paciente, Intervención (exposición), Comparación y Resultado (*outcome*) (PICO o PICoR).
3. Elegir la medida principal de efecto (jerarquizar su importancia clínica).
4. Hacer una búsqueda bibliográfica sistemática (usar programas de gestión de citas) para justificar el proyecto.
5. Seleccionar el diseño epidemiológico más apropiado.
6. Elegir métodos adecuados de muestreo y aleatorización.
7. Definir con precisión las variables y los procedimientos de recogida de datos.
8. Emplear bases de datos con formularios y variables codificadas.
9. Utilizar medidas descriptivas adecuadas en función del tipo de distribución de las variables.
10. Estimar medidas de frecuencia, riesgo e impacto.
11. Estimar intervalos de confianza de las principales medidas.
12. Evaluar la importancia clínica y diferenciar de la significación estadística.
13. Estimar correctamente el tamaño muestral necesario (considerar la diferencia clínicamente importante y los errores tipo I y II).
14. Elegir el test estadístico más apropiado según las escalas de medida de las variables y su distribución.
15. Prevenir los principales errores metodológicos.
16. Planificar el control de factores de confusión e interacción.

Comenzaremos revisando los fundamentos de la **pregunta de investigación**, ya que es el eje en el que se sustenta el diseño y análisis de todo proyecto. Resulta fundamental seleccionar la población de estudio, la intervención o exposición evaluada, la intervención o exposición con que se va a comparar y la variable o medida de resultado. En esta misma sección destacaremos la importancia de seleccionar cuidadosamente la **medida principal de efecto**. Sobre ella se va a plantear el contraste de hipótesis, el cálculo del tamaño muestral y los análisis estadísticos principales de nuestro proyecto. Los principales errores que podemos cometer y que comprometerán la aprobación de nuestro proyecto tienen que ver con estas cuestiones. A continuación, repasaremos los principales **apartados de una memoria de investigación**.

Intercalaremos preguntas entre las explicaciones, que nos servirán como evaluación de los conocimientos previos. Tras cada pregunta se darán las **respuestas comentadas, aunque solo se explicarán en profundidad en los capítulos sucesivos de este libro**. Nuestro objetivo es que preguntas similares puedan ser contestadas por el lector al final de la lectura de este libro.

La pregunta de investigación estructurada (PICO)

Todo proyecto de investigación debe tratar de buscar respuesta a una pregunta. De la pregunta surgen directamente la hipótesis de estudio y los objetivos. Asimismo, una pregunta bien planteada facilita la selección del diseño epidemiológico, las variables de estudio, el análisis de resultados y la generación de conclusiones.

Hay distintas preguntas para distintas dudas. En ocasiones tenemos dudas sobre la frecuencia de un determinado factor de riesgo o enfermedad, y nuestra pregunta tiene un objetivo descriptivo. En otras ocasiones queremos saber si un factor de exposición o intervención influye en el riesgo de una enfermedad y nos planteamos preguntas analíticas. En el primer caso solo necesitamos un grupo de sujetos de estudio, en el segundo necesitamos, al menos, dos grupos de sujetos que vamos a comparar.

Podemos plantearnos preguntas como meros observadores de la realidad o preguntas sobre nuestra capacidad para modificar la realidad mediante intervenciones controladas. En el primer caso necesitaremos diseños de estudios observacionales, mientras que en el segundo supuesto los diseños tendrán que ser experimentales.

Una buena pregunta de investigación debe ser pertinente y concreta. La pertinencia se valora en función del acrónimo **FINER**: factible, interesante, novedosa, ética y relevante:

- **Factible.** ¿Somos capaces de contestar a la pregunta? ¿Contamos con los recursos materiales y humanos para desarrollar el proyecto de investigación que la contestación a dicha pregunta requiere?
- **Interesante.** ¿Está la pregunta de investigación relacionada con nuestros intereses y campo de trabajo? ¿Estamos lo suficientemente interesados en el proyecto para dedicar tiempo y esfuerzo en su ejecución?

- **Novedosa.** ¿Está nuestra pregunta de investigación suficientemente contestada en la literatura científica? ¿Qué aportamos al conocimiento disponible?
- **Ética.** ¿Hemos evaluado los conflictos éticos implicados en la realización del estudio? ¿Va a ser la metodología de estudio lo suficientemente válida para justificar su realización? (lo que no es metodológicamente correcto no es éticamente justificable).
- **Relevante.** ¿Es previsible la utilidad o aplicabilidad práctica de los resultados del estudio?

Asimismo, la pregunta de investigación debe ser concreta. Para ello recomendamos la estructura ligada al acrónimo PICO (o PICO_R): Paciente, Intervención (exposición), Comparación y Resultado (*outcome*):

- **Paciente.** Los pacientes o sujetos de estudio a los que queremos extrapolar nuestros resultados. Resulta fundamental describir en los criterios de inclusión y exclusión sus características de edad, sexo, ámbito, diagnóstico, etc.
- **Intervención (exposición).** El factor de estudio, exposición o intervención que será evaluado.
- **Comparación.** Cuando el estudio sea analítico, tendremos que definir el factor de comparación, tanto si son factores de exposición como si son intervenciones. La pregunta de investigación cambiará en función de la comparación elegida (por ejemplo: placebo vs. intervención estándar).
- **Resultado (*outcome*).** El último elemento del acrónimo PICO corresponde a la palabra en inglés "*outcome*", aunque puede sustituirse por la española "resultado", adoptando el acrónimo PICO_R. De igual manera a como se comentó con la comparación, la medida de resultado elegida es un elemento clave de la pregunta de investigación.

Veamos un ejemplo.



Supuesto 1.1

Queremos saber si la asistencia a guardería se asocia a un mayor o menor riesgo de desarrollar sibilantes recurrentes en la primera infancia.

La pregunta podría redactarse de forma textual, como: ¿En lactantes sanos, se asocia la asistencia a guardería a un mayor riesgo de sibilantes recurrentes?

La pregunta de forma estructurada sería:

- Paciente: lactantes sanos.
- Intervención (exposición): asistencia a guardería.
- Comparación: cuidados en domicilio familiar.
- Resultado (*outcome*): episodios de sibilantes recurrentes (al menos tres) en los 2 primeros años de vida.

En esta pregunta se podrían haber precisado más algunos componentes. Así, indicar en paciente: "lactantes sanos seguidos regularmente en centros de salud desde el nacimiento". En intervención podríamos haber concretado la exposición en función del momento de inicio (por ejemplo: inicio de guardería antes de los 6 meses de edad) o el nivel de exposición (por ejemplo: número de horas por semana, número de niños por aula, etc.). Hemos elegido como comparación los cuidados en domicilio familiar (no asistencia a guardería), aunque podríamos haber seleccionado otros criterios de comparación (por ejemplo: inicio tardío de guardería, cuidados en unidades de guardería con menos de seis niños, etc.). Finalmente, hemos optado por establecer como medida de resultado la aparición de, al menos, tres episodios de sibilantes recurrentes, pero podríamos haber seleccionado otras medidas de efecto cualitativas (por ejemplo: episodios de sibilantes que hayan precisado consulta a Urgencias o ingreso) o cuantitativas (por ejemplo: número de episodios de sibilantes recurrentes).

Practicemos con otros supuestos la elaboración de la pregunta de investigación según la estructura PICO. Estos mismos supuestos se emplearán en sucesivos ejercicios de este capítulo. Al final de cada capítulo se presentan las respuestas correctas a cada pregunta y un comentario explicativo.



Supuesto 1.2

No sabemos si el suero salino hipertónico nebulizado resulta eficaz en la bronquiolitis aguda. Vamos a plantear una pregunta de investigación para esta duda, siguiendo la estructura PICO.

Pregunta 1.1

Queremos plantear una pregunta de investigación para resolver la duda de si el suero salino hipertónico nebulizado resulta eficaz en la bronquiolitis aguda atendida en Urgencias. Elija la pregunta de investigación estructurada más correcta:

- Paciente: lactantes con bronquiolitis tratados con nebulización de suero salino hipertónico; Intervención/Exposición: ingreso hospitalario; Comparación: alta a domicilio; Resultado: cambios en la escala de gravedad.
- Paciente: lactantes con bronquiolitis aguda; Intervención/Exposición: nebulización de suero salino hipertónico 3 %; Comparación: no tratamiento; Resultado: riesgo de ingreso hospitalario.
- Paciente: lactantes que acuden a Urgencias con bronquiolitis aguda (criterios de McConochie) con una afectación moderada (escala Sant Joan de Deu 6-10); Intervención/Exposición: nebulización de suero salino hipertónico 3 %; Comparación: nebulización de suero salino fisiológico; Resultado: ingreso hospitalario.
- Paciente: lactantes que acuden a Urgencias con bronquiolitis aguda (criterios de McConochie) con una afectación moderada (escala Sant Joan de Deu 6-10); Intervención/Exposición: nebulización de suero salino hipertónico 3 %; Comparación: nebulización de broncodilatador de acción corta; Resultado: saturación de oxígeno al alta.



Supuesto 1.3

No conocemos la frecuencia de hipertensión arterial en la población infantil en nuestro medio. Nos planteamos determinar la presión arterial con dispositivos de medición oscilométricos y confirmación con métodos auscultatorios y monitorización ambulatoria de la presión arterial (MAPA).

Pregunta 1.2

Queremos plantear una pregunta de investigación para resolver la duda de cuál es la frecuencia de hipertensión arterial en la población infantil. Elija la pregunta de investigación estructurada más correcta:

- Paciente: niños con presión arterial sistólica o diastólica superior al percentil 95; Intervención/Exposición: obesidad o sobrepeso; Comparación: normopeso; Resultado: frecuencia de hipertensión.
- Paciente: niños entre 4 y 14 años; Intervención/Exposición: medición de la presión arterial; Comparación: no medición de la presión arterial; Resultado: frecuencia de hipertensión arterial.
- Paciente: niños sin sobrepeso ni obesidad; Intervención/Exposición: ninguna; Comparación: ninguna; Resultado: frecuencia de hipertensión (presión arterial sistólica o diastólica superior al percentil 95 confirmada).
- Paciente: niños entre 4 y 14 años; Intervención/Exposición: ninguna; Comparación: ninguna; Resultado: frecuencia de hipertensión arterial (presión arterial sistólica o diastólica superior al percentil 95 confirmada).



La medida principal del efecto

Tal y como hemos visto en los supuestos previos, es frecuente que una pregunta de investigación tenga varias medidas de resultado posibles. La elección de la medida principal de efecto resulta fundamental y debe establecerse *a priori*. En función de la medida elegida, plantearemos el tamaño muestral necesario y los principales análisis. Entre los requisitos habituales de las normas reguladoras de los ensayos clínicos se establece que debe haber, al menos, una medida principal de efecto. Ese mismo requisito debe aplicarse a cualquier otro diseño epidemiológico. La existencia de medidas de efecto secundarias permite completar el análisis y ayuda a evaluar la consistencia interna de los resultados.

A la hora de elegir la medida principal deberíamos tener en cuenta su importancia clínica y la factibilidad de su medición. En ocasiones queremos medir una variable con gran importancia clínica (por ejemplo: mortalidad), pero debemos renunciar a ello porque no vamos a contar con suficiente número de eventos para su medición. Cuando lleguemos al capítulo de cálculo del tamaño muestral veremos las implicaciones de esta limitación. Si cometemos el error de elegir una variable cuya medición no es factible, estaremos comprometiendo la credibilidad de nuestro proyecto.

Una buena aproximación a la valoración de la importancia clínica de una medida de efecto es la propuesta por el grupo de trabajo GRADE (*Grading of Recommendations Assessment, Development and Evaluation*; www.gradeworkinggroup.org) para evaluar los resultados de estudios publicados. En la fase inicial de la evaluación de la evidencia, GRADE propone consensuar la importancia clínica de medidas alternativas en una escala de 1 a 9, asignando tres niveles de importancia:

- 7-9: resultado **clave (o crítico)** para la toma de decisiones.
- 4-6: resultado **importante, pero no clave**, para la toma de decisiones.
- 1-3: resultado **no importante**. Estos desenlaces no juegan un papel relevante en la formulación de las recomendaciones.

Podemos entender mejor esta jerarquización utilizando el ejemplo del supuesto 1.2 (suero salino hipertónico en la bronquiolitis aguda). Resultados claves serían: la mortalidad o la necesidad de ventilación mecánica; resultados importantes, pero no claves, ingreso hospitalario o necesidad de suplementación de oxígeno, y resultados no importantes, cambios en una escala de signos-síntomas o cambios en saturación de oxígeno. Si en la muestra de pacientes a incluir en el estudio no esperamos un riesgo suficiente de medidas de resultado clave, tendremos que optar por medidas de menor importancia.

Lamentablemente, criterios de factibilidad hacen que nos sea más fácil realizar estudios que aborden medidas de efecto importantes, pero no claves, que medidas de efecto claves o críticas. Esto mismo se refleja en la literatura científica publicada.

Una vez elegida la medida principal de efecto, debemos establecer la magnitud del efecto esperada y su importancia clínica. Asumamos que en nuestro supuesto 1.2 (suero salino hipertónico en la bronquiolitis aguda) elegimos como medida de efecto el ingreso hospitalario. Ahora es el momento de que elijamos la diferencia clínicamente importante. Si el riesgo de ingreso en nuestros pacientes con bronquiolitis moderada fuera del 30 %, parece razonable aceptar que una reducción del 10 % de dicho riesgo sería clínicamente relevante. Esta magnitud puede ser, como tal, formulada en la hipótesis operativa del estudio y será necesaria a la hora de realizar el cálculo del tamaño muestral.

Apartados de una memoria de investigación

La memoria de investigación consta de una serie de apartados, que en esta presentación se han inspirado en las propuestas por el Fondo de Investigación Sanitario (actualmente Instituto de Salud Carlos III) para evaluación de proyectos de investigación:

- a) Título, tipo de proyecto y equipo investigador.
- b) Resumen.
- c) Antecedentes, justificación y bibliografía.

- d) Hipótesis.
- e) Objetivos.
- f) Métodos.
- g) Utilidad esperada del estudio.
- h) Limitaciones del estudio.
- i) Aspectos éticos y legales.
- j) Cronograma y reparto de tareas.
- k) Experiencia previa del equipo investigador.
- l) Medios disponibles para la realización del trabajo.
- m) Presupuesto del estudio.

En un documento anexo se presenta un formulario preconfigurado para cumplimentar la memoria de proyectos de investigación. La versión presentada contiene instrucciones de cumplimentación en cada apartado. Suprimiendo dichas recomendaciones, se puede usar para elaborar una nueva memoria.

Título, tipo de proyecto y equipo investigador

La memoria del proyecto comienza con un apartado en el que consta el título y tipo de proyecto (individual, coordinado, multicéntrico) y su duración.

Se recomienda redactar el título del proyecto una vez elaborados los principales apartados de la memoria, ya que algunos aspectos, como el tipo de diseño o las variables elegidas, van a condicionar su contenido. Se aconseja que el título sea informativo, pudiendo contener en el mismo los elementos de la pregunta de investigación y el diseño epidemiológico. En ocasiones se incluye en el título algún acrónimo del proyecto o del grupo investigador. No conviene emplear títulos largos de redacción y lectura complejas.

Definimos un proyecto como **individual** cuando se realiza en un solo centro por un solo grupo investigador. Cuando en el estudio participan diferentes centros o grupos y cada uno de ellos realiza una tarea o fase del proyecto, se considera **coordinado**. Si el mismo proyecto se realiza en más de un centro o por más de un grupo investigador, se considera **multicéntrico**. La presentación de proyectos coordinados o multicéntricos es favorecida en la evaluación de algunas convocatorias de investigación.

La **duración** del proyecto se refiere no solo a la recogida de datos, sino a todo el proceso de diseño, organización, recogida de datos, análisis y presentación de resultados. En algunas convocatorias se establece una duración determinada (1 o 3 años), aunque se pueden justificar periodos diferentes en función de las características del proyecto.

Todo proyecto debe contar con un investigador principal que ha de permanecer vinculado al mismo durante toda su ejecución. La vinculación a un proyecto como investigador

principal implica restricciones a la participación en otros proyectos, coincidente en el tiempo, en las convocatorias de financiación. El investigador principal es el responsable de presentar la memoria, coordinar las tareas y presentar los resultados.

Los investigadores colaboradores deben cumplir los requisitos de autoría establecidos para proyectos y publicaciones científicos. Cada colaborador ha de conocer el proyecto, haber contribuido a su elaboración y estar de acuerdo con sus métodos. En el momento del reclutamiento cada colaborador debe conocer su misión en el proyecto, que se detallará más adelante en el reparto de tareas.

Resumen

La memoria debe contar con un resumen estructurado, que suele incluir los siguientes apartados:

- **Objetivos:** en él se presenta el objetivo principal del estudio, que debería contener los componentes de la pregunta de investigación estructurada (PICO).
- **Diseño:** en este apartado se debe detallar el tipo de diseño epidemiológico (estudio transversal, estudio ecológico, estudio de casos y controles, estudio de cohortes, ensayo clínico, etc.). En diseños mixtos se debe describir si el proyecto es descriptivo o analítico, concurrente o histórico, observacional o experimental.
- **Ámbito de estudio:** debe indicarse si el ámbito de estudio es hospitalario, comunitario o mixto y si el proyecto es unicéntrico o multicéntrico.
- **Población de estudio:** se puede indicar la población diana a la que se quieren extrapolar los resultados y la población de estudio de la que se obtendrá la muestra de estudio.
- **Muestra de estudio:** debe indicarse el tamaño muestral estimado y, cuando sea pertinente, la estrategia de muestreo.
- **Instrumentación:** en este apartado se detallan las principales variables de exposición y/o efecto y los procedimientos de recogida.
- **Determinación (análisis):** en este apartado se concretarán las principales medidas de frecuencia, riesgo e impacto a estimar y, cuando sea de interés, las técnicas estadísticas a aplicar.

Debe hacerse un esfuerzo de síntesis en la elaboración del resumen del proyecto. Una extensión aceptable sería de 250 a 300 palabras.

Antecedentes, justificación y bibliografía

En este apartado debe incluirse información que permita al evaluador conocer el estado de conocimiento actual sobre el problema en estudio. Se recomienda no hacer

revisiones extensas, existiendo limitaciones de espacio para este apartado en algunas convocatorias. La revisión debe sustentarse en referencias bibliográficas actualizadas, y resulta útil mencionar los detalles y extensión de la búsqueda bibliográfica realizada para su documentación. Si el grupo investigador tiene aportaciones en el tema, pueden incluirse en los antecedentes, aunque más adelante existe un apartado específico para destacar la experiencia del grupo.

Este apartado debe concluir con una justificación de la pregunta de investigación, de su relevancia, interés, originalidad y factibilidad (ética y material). Esta información se podrá completar más adelante en un apartado específico de la memoria.

Las referencias bibliográficas deben presentarse siguiendo el estilo Vancouver, recomendándose en algunas convocatorias que el investigador haga una breve reseña del objeto o importancia de los principales estudios citados. Se recomienda la utilización de un programa de gestión de citas bibliográficas (por ejemplo: Zotero, Mendeley, EndNote, etc.).

Hipótesis

En este apartado debe hacerse la formulación de la hipótesis de estudio siguiendo los componentes de la pregunta de investigación (PICO o PICO_R), tal y como hemos visto en un apartado previo de este capítulo.

Para el supuesto 1.2 (suero salino hipertónico en bronquiolitis) visto previamente sería:

- En pacientes con bronquiolitis aguda moderada (P), el tratamiento nebulizado con suero salino hipertónico (I), en comparación con suero salino fisiológico (C), reduce el riesgo de ingreso hospitalario (O o R).

En ocasiones no se puede plantear una hipótesis previa. Así ocurre en estudios descriptivos como el del supuesto 1.3 (frecuencia de hipertensión arterial en la infancia). Cuando ocurre esto, puede hacerse un planteamiento de contrastar nuestros resultados (prevalencia de hipertensión) con la información disponible en la literatura (estimaciones de otras áreas o épocas).

Objetivos

En este apartado debe desarrollarse la hipótesis de estudio en forma de objetivos, que variarán según el tipo de pregunta y diseño. A continuación recogemos algunas opciones disponibles en función del tipo de objetivo de la pregunta de investigación:

- **Frecuencia:** estimación de incidencia acumulada, densidad de incidencia o prevalencia.
- **Riesgo:** estimación de riesgo relativo, *odds ratio*, cocientes de riesgos instantáneos (*hazard ratio*), etc.
- **Impacto:** estimación del riesgo o proporción atribuible o prevenible.

- **Eficacia/efectividad de procedimientos terapéuticos:** estimación de la reducción absoluta del riesgo, número necesario a tratar, etc.
- **Coste:** estimación de coste-efectividad, coste-utilidad, coste-beneficio, etc.
- **Validez:** estimación de sensibilidad, especificidad, cocientes de probabilidades, áreas bajo la curva ROC, etc.
- **Precisión:** estimación de concordancia (índice kappa, coeficiente de correlación intraclase, etc.).

Podemos diferenciar, en función de la medida de efecto principal, objetivos principales relacionados con ella y objetivos secundarios.

Los objetivos deben ser operativos, concretando cada componente de la pregunta de investigación. Para el supuesto 1.2 (suero salino hipertónico en bronquiolitis) podríamos concretar el objetivo de esta manera: “estimar si el riesgo de ingreso por bronquiolitis en pacientes tratados con suero salino hipertónico es un 10 % menor que en los tratados con suero salino fisiológico”.

Métodos

En próximos capítulos repasaremos cuestiones metodológicas aquí recogidas de forma breve.

Diseño

El apartado de metodología debe empezar definiendo el diseño epidemiológico elegido, mediante descripción de sus características o asignando un tipo de diseño modelo:

- **Características del estudio:** descriptivo/analítico (existe grupo de comparación), concurrente (prospectivo)/histórico (retrospectivo), observacional/experimental.
- **Tipo de diseño:** estudio transversal/estudio ecológico/estudio de casos y controles/estudio de cohortes/ensayo clínico.
- **Otros estudios:** evaluación de pruebas diagnósticas/revisión sistemática.

Sujetos de estudio

En este apartado debemos hacer una descripción de la población accesible y susceptible de entrar en el estudio, que puede ser diferente de la población diana a la que queremos extrapolar nuestros resultados.

Muestra de estudio

- **Tamaño muestral:** debemos realizar en este momento la justificación del tamaño muestral en función de la hipótesis y objetivos del estudio. Para el cálculo del tamaño

muestral se debe detallar (a partir de estudios previos) la medida principal de efecto, su precisión o la diferencia esperada entre grupos.

- Ejemplo del supuesto 1.2 (suero salino hipertónico en bronquiolitis): diferencia esperada en el porcentaje de pacientes con bronquiolitis que requieren ingreso en función de que sean tratados con nebulización de suero salino hipertónico frente a suero fisiológico del 10 % (riesgo basal en tratados con suero fisiológico, 30 %; riesgo esperado con suero hipertónico, 20 %), para un riesgo alfa (error tipo I) del 5 % y riesgo beta (error tipo II) del 20 %. Con estos datos podemos estimar, con el programa *online* gratuito Granmo (<https://www.imim.cat/ofertadeserveis/software-public/granmo/>), un tamaño muestral de 292 sujetos por grupo.

Podemos ver en la figura 1.1 el cálculo realizado con el programa.

Figura 1.1

Cálculo del tamaño muestral para el supuesto 1.2 (suero salino hipertónico en bronquiolitis) con el programa *online* Granmo

Disponible en <https://www.imim.cat/ofertadeserveis/software-public/granmo/>

- **Criterios de inclusión:** los criterios de inclusión deben describirse de tal manera que permitiera a una tercera persona reproducir el estudio o juzgar si un determinado paciente estaría incluido en él.
- **Criterios de exclusión:** los criterios de exclusión deben definirse *a priori*. Debemos tener en cuenta que estos criterios pueden condicionar la aplicabilidad práctica de los resultados del estudio, ya que pueden cuestionar la representatividad de la muestra de estudio.
- **Sistemática de muestreo:** en los estudios en los que se haga una selección de sujetos entre la población de estudio debe detallarse el tipo de muestreo. Más adelante, en este libro, aprenderemos a diferenciar entre muestreos aleatorios (todos los sujetos de la población de estudio tienen la misma probabilidad de ser incluidos en la muestra de estudio) y no aleatorios, entre los muestreos estratificados y por conglo-

merados, y entre los muestreos sistemáticos y consecutivos. Resulta especialmente importante mencionar los procedimientos operativos de muestreo, principalmente si se emplea algún tipo de enmascaramiento en la asignación aleatoria (ocultación de la secuencia de aleatorización).

Vamos a explorar sus conocimientos previos sobre tipos de diseño y tipos de muestreo con varias preguntas (al final de cada capítulo se presentan las respuestas más correctas y un comentario explicativo). A continuación comentaremos brevemente las respuestas, refiriéndoles a capítulos sucesivos del libro para más información.

Pregunta 1.3

Qué tipo de diseño epidemiológico es el más idóneo para la pregunta de investigación del supuesto 1.1: ¿La asistencia a guardería aumenta el riesgo de sibilantes recurrentes en los 2 primeros años de vida?:

- a. Estudio de cohortes.
- b. Estudio transversal.
- c. Ensayo clínico.
- d. Estudio de casos y controles.

Pregunta 1.4

Qué tipo de diseño epidemiológico es el más idóneo para la pregunta de investigación del supuesto 1.2 (suero salino hipertónico en bronquiolitis): ¿Es eficaz el suero salino hipertónico nebulizado en la bronquiolitis aguda?:

- a. Estudio de cohortes.
- b. Estudio transversal.
- c. Ensayo clínico.
- d. Estudio de casos y controles.

Pregunta 1.5

Qué tipo de diseño epidemiológico es el más idóneo para la pregunta de investigación del supuesto 1.3: ¿Cuál es la frecuencia de hipertensión arterial en la población infantil?:

- a. Estudio de cohortes.
- b. Estudio transversal.
- c. Ensayo clínico.
- d. Estudio de casos y controles.

Pregunta 1.6

Para estimar la frecuencia de hipertensión arterial en la infancia (supuesto 1.3), seleccionamos, siguiendo un procedimiento aleatorio, aulas de los colegios de nuestra área de salud, midiendo la presión arterial a todos los alumnos de cada una de dichas aulas. ¿Qué técnica de muestreo hemos empleado?:

- a. Muestreo aleatorio simple.
- b. Muestreo aleatorio estratificado.
- c. Muestreo aleatorio por conglomerados.
- d. Muestreo sistemático.

Variables de estudio

En este apartado debemos recoger todas las definiciones operativas de las variables de estudio. Las variables principales deben definirse en este apartado, pudiendo completarse esta información en anexos de la memoria, en forma de listado o cuaderno de recogida de datos.

Para cada variable debemos mencionar su nombre, tipo de variable (cualitativa-discreta/cuantitativa-continua), su escala de medida (nominal/ordinal/continua de intervalos o razones) y sus valores posibles (categorías o unidades de medida). La escala de medida de cada variable será el elemento fundamental a la hora de planificar el análisis. Si una variable es intrínsecamente cuantitativa (por ejemplo: edad), con un número de valores posibles infinitos, la escala de medida será habitualmente de intervalos o razones (en este libro las denominaremos "continuas"). Si una variable es discreta (valores posibles finitos), puede tener una escala de medida ordinal, si los posibles valores están ordenados entre sí (por ejemplo: test de Apgar), o nominal, si los posibles valores no están ordenados (por ejemplo: sexo).

Deben diferenciarse las variables, cuando corresponda, en:

- Variables de Exposición/Intervención (variables independientes). Podrían ser más de una.
- Variables de Efecto/Resultado (variable dependiente):
 - Principal. Debe haber una variable de efecto o resultado principal.
 - Secundarias.

Veamos cómo quedaría este apartado para el supuesto 1.2 (suero salino hipertónico en bronquiolitis):

- Variable de Exposición/Intervención: tipo de suero en el aerosol; variable cualitativa, discreta, de escala de medida nominal dicotómica (1: suero hipertónico; 2: suero fisiológico).
- Variable de Efecto/Resultado principal: ingreso; variable cualitativa, discreta, de escala de medida nominal dicotómica (1: Sí; 2: No).
- Variable de Efecto/Resultado secundaria: escala de gravedad de trabajo respiratorio (escala RDAI); variable cuantitativa, discreta, de escala de medida ordinal (valores enteros entre 0 y 17).

En los estudios experimentales debe haber un apartado en el que se detallen los procedimientos de asignación aleatoria de tratamientos (ocultación de la secuencia de aleatorización) y de ocultación o enmascaramiento (ciego) de su aplicación.

Recogida de datos

Debe indicarse en este apartado la sistemática de recogida de datos: quién recoge los datos (investigador, colaborador, autocumplimentado por el paciente, etc.), por qué medio (revisión de historias, registros informatizados, llamada telefónica, etc.).

Asimismo, debe especificarse si se emplea algún método de enmascaramiento en la recogida de datos. Esto es especialmente importante en los estudios analíticos y, sobre todo, en los ensayos clínicos. En ocasiones, la intervención no puede enmascarse, pero sí la evaluación del efecto (evaluadores ciegos al tratamiento asignado).

Análisis estadístico

El análisis estadístico se planificará empezando por la estadística descriptiva y siguiendo por la analítica con estimación de medidas de riesgo, impacto, validez o precisión y contrastes de hipótesis.

- **Descriptivo:** contempla las medidas de frecuencia de variables discretas y medidas de centralización y dispersión de variables continuas (media/mediana; desviación típica/rango intercuartílico). Asimismo, se incluyen aquí la estimación de medidas de frecuencias de enfermedad (prevalencia, incidencia acumulada, densidad de incidencia, etc.). Todas las estimaciones deberán acompañarse de sus intervalos de confianza del 95 %.
- **Analítico:** el análisis implica comparación entre variables o grupos. La elección de la medida de riesgo, impacto, validez o precisión más adecuada para cada tipo de estudio, así como del test estadístico más adecuado para cada análisis, se abordará en próximos capítulos del libro. El principal criterio a considerar en la elección es el tipo de escala de medida que siguen las variables analizadas. A continuación resumimos las principales opciones:
 - Medidas de riesgo/impacto/validez/precisión:
 - Medidas de riesgo: riesgo relativo/*odds ratio*/cociente de riesgos instantáneos (*hazard ratio*), etc.
 - Medidas de impacto: riesgo o proporción atribuible o prevenible.
 - Eficacia/efectividad de procedimientos terapéuticos: reducción absoluta del riesgo, número necesario a tratar.
 - Validez: sensibilidad, especificidad, cocientes de probabilidades, áreas bajo la curva ROC.
 - Precisión: concordancia (índice kappa, coeficiente de correlación intraclase, etc.).
 - Contraste de hipótesis; opciones más frecuentes:
 - Dos variables discretas: test de ji cuadrado o pruebas exactas (Fisher) (para grupos relacionados, test de McNemar).
 - Contraste de medias: test de la t de Student para muestras no relacionadas o relacionadas, análisis de la varianza o pruebas no paramétricas alternativas.

- Supervivencia: función de supervivencia (Kaplan Meier) y *Log-rank test*.
- Correlación entre variables cuantitativas: coeficiente de correlación de Pearson o Spearman.
- Métodos de ajuste multivariante: regresión lineal múltiple (variable dependiente continua), regresión logística (variable dependiente discreta) o regresión de Cox (supervivencia).

Vamos a explorar sus conocimientos previos sobre la selección de test estadísticos según la pregunta de investigación y las variables implicadas. A continuación comentaremos brevemente las respuestas, refiriéndoles a capítulos sucesivos del libro para más información.

Pregunta 1.7

¿Qué medida de **impacto** es la más apropiada para estimar el efecto del tratamiento (supuesto 1.2) con suero salino hipertónico nebulizado (con respecto al suero salino fisiológico) sobre el riesgo de ingreso por bronquiolitis?:

- a. Número necesario a tratar.
- b. Riesgo relativo.
- c. *Odds ratio*.
- d. Diferencia de medias.

Pregunta 1.8

¿Qué **test estadístico** es el más apropiado para ver si el tipo de tratamiento (supuesto 1.2; suero salino hipertónico vs. fisiológico) influye en el riesgo de ingreso por bronquiolitis?:

- a. Test de la t de Student para muestras independientes.
- b. ANOVA (análisis de la varianza).
- c. Test de McNemar.
- d. Test de la ji cuadrado (o exacto de Fisher).

Pregunta 1.9

¿Qué **test estadístico** es el más apropiado para ver si hay diferencias en la edad de los pacientes con bronquiolitis incluidos en los grupos tratados con suero salino hipertónico frente a los tratados con suero fisiológico (supuesto 1.2; suero salino hipertónico en bronquiolitis)?:

- a. Test de la t de Student para muestras independientes.
- b. ANOVA (análisis de la varianza).
- c. Test de McNemar.
- d. Test de la ji cuadrado (o test exacto de Fisher).

Utilidad esperada del estudio

Si no se ha mencionado anteriormente en el apartado de antecedentes-justificación, en este apartado debe hacerse una reflexión sobre lo que puede aportar el estudio en cuanto al conocimiento previo, la práctica clínica, el paciente, el sistema sanitario o la sociedad.

Las agencias de financiación de la investigación valoran especialmente la potencial utilidad económica de los resultados de investigación: patentes, reducción de costes, etc.

Limitaciones del estudio

Deben comentarse en este apartado, si corresponde, las limitaciones del proyecto en relación con el tamaño muestral (falta de potencia estadística), la sistemática de muestreo (sesgos de selección), las variables medidas (sesgos de clasificación) o el análisis (sesgos de confusión o interacción). Conviene anticipar posibles críticas al proyecto, justificando que dichas limitaciones no van a interferir en la validez de los resultados. Asimismo, se recomienda planificar *a priori* estrategias de control de sesgos incluidas ya en el análisis.

Aspectos éticos y legales

En este apartado se debe considerar la adecuación ética de los procedimientos a realizar en el proyecto. Se recomienda que todo proyecto cuente con un documento de consentimiento informado y la aprobación por el Comité Ético de Investigación Clínica. Asimismo, debe hacerse una declaración de confidencialidad de los datos personales a utilizar y detallar los procedimientos que han de garantizar su seguridad.

Cronograma y reparto de tareas

Se debe detallar el cronograma de cada fase del estudio, incluyendo la fase de análisis y explotación de resultados.

Otro requisito habitual en las memorias es el reparto de tareas de los colaboradores en el estudio.

Experiencia previa del equipo investigador

La descripción de la experiencia previa de los investigadores en el tema de estudio (experiencia profesional, estudios realizados y su producción científica) es un elemento muy valorado en la evaluación de proyectos de investigación.

Medios disponibles para la realización del trabajo

Otro aspecto de gran importancia es la descripción de medios personales y materiales disponibles y su adecuación a los objetivos propuestos. Asimismo, es importante mencionar toda fuente de financiación complementaria del proyecto.

Presupuesto del estudio

En los trabajos en los que se solicita ayuda económica debe hacerse una justificación detallada de cada partida económica solicitada. Los principales capítulos de gasto son:

- Gastos de personal: becarios y contratación de servicios. Antes de solicitar dinero para becarios, asegurarse de que la institución tiene recursos para su contratación (el personal del centro no puede recibir dinero en concepto de beca).
- Viajes y dietas: gastos de desplazamiento para reuniones o trabajo de campo. Se incluye habitualmente alguna cantidad para la asistencia a reuniones científicas, aunque algunas convocatorias limitan estas partidas.
- Material inventariable (equipos): las ayudas para proyectos no deben usarse como vía de financiación de infraestructura, debe existir un equilibrio en la cantidad solicitada en este capítulo y otros gastos.
- Material fungible: debe ser acorde con el tamaño muestral y los procedimientos previstos.
- Otros gastos.

En los trabajos en los que no se solicita ayuda económica, debe haber una justificación de los costes estimados en carga de trabajo y material, si estos consumen recursos de la institución en la que se lleva a cabo.

Si el proyecto se ha diseñado para más de 1 año, debe detallarse por separado la financiación de cada año.

Anexos

Son anexos habituales en las memorias de investigación:

- Cuestionario de recogida de datos.
- Escalas auxiliares de recogida de datos.
- Consentimiento informado.
- Hoja de firmas o Cartas de adhesión de los colaboradores.
- Permisos de responsables de instituciones.
- Aprobación por el Comité Ético de Investigación Clínica (cuando corresponda).
- Currículum de los investigadores.

Respuestas y comentarios a las preguntas

Comentario a la pregunta 1.1

La opción más correcta es la "c".

Paciente: lactantes que acuden a Urgencias con bronquiolitis aguda (criterios de McConochie) con una afectación moderada (escala Sant Joan de Deu 6-10); Intervención/Exposición: nebulización de suero salino hipertónico 3 %; Comparación: nebulización de suero salino fisiológico; Resultado: ingreso hospitalario.

En ella se concretan los pacientes a los que queremos extrapolar los resultados del estudio y además se detalla un criterio de inclusión (gravedad moderada atendida en Urgencias). En la opción "a" el paciente incluye la intervención y en la opción "b" no se concreta qué gravedad deben tener los pacientes para ser incluidos.

La intervención se detalla correctamente en todas las opciones menos en la opción "a", en la que se menciona una posible medida de resultado (ingreso). La comparación con más sentido clínico es el empleo de un suero salino fisiológico (vehículo común de la administración de medicamentos en nebulización). Podría haberse optado por una comparación con ausencia de nebulización, tal y como consta en la opción "b", aunque ello impediría realizar enmascaramiento de la intervención. De hecho, esta es la comparación habitual en estudios clínicos publicados. La alternativa dada en la opción "d", broncodilatador, puede interferir en la valoración de la respuesta.

En cuanto al resultado, hemos optado por el ingreso hospitalario, tal y como consta en las opciones "b" y "c", por su importancia clínica, pero podíamos haber optado por medidas de resultado alternativo, tal y como consta en las opciones "a" y "b" (cambios en la escala de gravedad o saturación de oxígeno al alta). Otras opciones podrían haber sido: duración de la estancia en Urgencias, necesidad de oxígeno suplementario, etc.

Comentario a la pregunta 1.2

La opción más correcta es la "d".

Paciente: niños entre 4 y 14 años; Intervención/Exposición: ninguna; Comparación: ninguna; Resultado: frecuencia de hipertensión arterial (presión arterial sistólica o diastólica superior al percentil 95 confirmada).

La pregunta del supuesto 1.3 no contiene más que dos elementos del acrónimo PICO. El paciente y el resultado. Podríamos haber planteado alguna pregunta más compleja en la que incluyéramos factores de riesgo de hipertensión, como la obesidad. En ese caso tendríamos los otros dos elementos de la pregunta: Intervención/Exposición: obesidad/sobrepeso; Comparación: normopeso. Su inclusión o no en la pregunta de investigación va a condicionar el cálculo del tamaño muestral, ya que tendremos que prever suficien-

tes sujetos para estimar la frecuencia de hipertensión con precisión en cada grupo. En todo caso, la elección de una u otra pregunta depende del criterio del investigador.

En cuanto al resultado, en las opciones "c" y "d" se concreta el criterio de hipertensión arterial. Podríamos haber optado por una medida cuantitativa, como la media de presiones arteriales sistólica, diastólica o media.

Comentario a la pregunta 1.3

La respuesta más correcta es la "a", Estudio de cohortes.

El estudio de cohortes es un tipo de estudio observacional (sin intervención controlada), analítico (se comparan dos o más grupos, en este caso con o sin asistencia a guardería) y anterógrado (se evalúa primero el factor de exposición, asistencia a guardería, y luego el efecto, sibilantes recurrentes). Se diferencia del estudio de casos y controles en que los grupos se seleccionan en función de la exposición; en un estudio de casos y controles hubiéramos seleccionado niños con sibilantes recurrentes y otro grupo sin sibilantes recurrentes.

Comentario a la pregunta 1.4

La respuesta más correcta es la "c", Ensayo clínico.

El ensayo clínico es un tipo de estudio experimental (se modifica la historia natural con una intervención controlada), analítico (se comparan, al menos, dos grupos con intervenciones distintas), anterógrado (se evalúa primero el factor de exposición, la intervención, y luego el efecto, por ejemplo, ingreso hospitalario) y concurrente (el investigador asiste de forma prospectiva a la exposición y al efecto).

Comentario a la pregunta 1.5

La respuesta más correcta es la "b", Estudio transversal.

El estudio transversal es un tipo de estudio observacional (sin intervención controlada), descriptivo (no se seleccionan grupos diferentes a comparar), sin direccionalidad (no hay evaluación seriada de exposición y efecto) e histórico (el efecto evaluado ya existe en el momento que se realiza el estudio).

Comentario a la pregunta 1.6

La respuesta más correcta es la "c", Muestreo aleatorio por conglomerados.

La técnica de muestreo es aleatoria, porque todos los sujetos de la población de estudio tienen la misma probabilidad de ser incluidos en el estudio. Para facilitar la selección cuando no se dispone de una lista completa de los sujetos de estudios, pero sí de

grupos de los mismos (por ejemplo: aulas de colegios), la selección se aplica a dichos grupos, que llamamos conglomerados. Esta estrategia facilita también el trabajo de campo, por concentrar la recogida de datos. El muestreo sistemático es aquel en el que se aplica una sistemática fija de selección entre sujetos disponibles (uno de cada dos o tres pacientes).

Comentario a la pregunta 1.7

La respuesta más correcta es la "a", Número necesario a tratar.

La forma más intuitiva de describir el impacto del tratamiento es estimar a cuántos sujetos he tenido que tratar con la intervención experimental (en este caso, suero salino hipertónico nebulizado) para evitar un evento (en este caso, ingreso), lo que se conoce como número necesario a tratar (NNT). Esta medida se calcula fácilmente a partir de la reducción absoluta del riesgo. Imaginemos que un 20 % (0,20) de los tratados con suero hipertónico y un 30 % (0,30) de los tratados con suero fisiológico han precisado ingreso; eso supone una reducción absoluta del riesgo de 0,10 (10 %). El NNT se estima a partir del inverso de la reducción absoluta: $1/0,10$ igual a 10. He tenido que administrar la intervención experimental a 10 sujetos para beneficiar a uno.

Comentario a la pregunta 1.8

La respuesta más correcta es la "d", Test de la ji cuadrado (o test exacto de Fisher).

Las variables implicadas en este análisis son ambas discretas dicotómicas (dos valores posibles: suero salino hipertónico o fisiológico, ingreso sí o no), con una escala de medida nominal. Veremos más adelante cómo uno de los test que nos permite estimar la probabilidad de que la diferencia encontrada se deba al azar es el test de la ji (chi) cuadrado (con muestras pequeñas puede ser necesario el test exacto de Fisher).

Comentario a la pregunta 1.9

La respuesta más correcta es la "a", Test de la t de Student para muestras independientes.

Las variables implicadas en este análisis son una discreta dicotómica (dos valores posibles: suero salino hipertónico o fisiológico) y otra continua (edad). Cuando la variable continua sigue una distribución normal, este análisis equivale a una comparación de medias (de edad) por grupos (suero hipertónico vs. fisiológico). El test más apropiado es el de la t de Student para muestras independientes. Si la variable continua no siguiera una distribución normal, deberíamos emplear pruebas no paramétricas alternativas.

BIBLIOGRAFÍA RECOMENDADA

- Abramson JH. Survey Methods in Community Medicine. 2nd ed. Edinburgh. New York: Churchill Livingstone; 1990.
- Argimón Pallás JM, Jiménez Villa J. Métodos de investigación aplicados a la Atención primaria de salud. Barcelona: Harcourt; 2000.
- Contandriopoulos AP, Champagne F, Potvin L, Denis JL, Boyle P. Preparar un proyecto de investigación. Barcelona: Editores SG; 1991.
- Cummings. Diseño de la investigación clínica. Un enfoque epidemiológico. Barcelona: Ed. Doyma; 1993.
- Hulley SB, Cummings SR, eds. Designing clinical research: an epidemiologic approach. Baltimore: Williams & Wilkins; 1987. [Traducción al castellano publicada por la editorial Doyma. Barcelona. 1982].
- Kleinbaum DG, Kupper LL, Morgensten H. Epidemiologic Research. Principles and quantitative Methods. New York: Van Nostrand Reinhold Company; 1982.
- Ministerio de Sanidad y Consumo. Fondo de Investigación Sanitaria. Diseño de un protocolo de investigación. Convocatoria de acciones 1992. Madrid: Instituto Nacional de la Salud; 1991.
- Riegelman RK, Hirsh RP. Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica. 2.ª ed. Washington, D.C.: Organización Panamericana de Salud; 1992. (Publicación Científica n.º 531).

2

Búsquedas bibliográficas

Gestores de citas bibliográficas

Objetivos docentes:

- Conocer los pasos de una búsqueda bibliográfica para documentar una pregunta de investigación.
- Conocer los tipos de fuentes de información.
- Ser capaz de elaborar y depurar una búsqueda bibliográfica básica.
- Conocer las funciones de un gestor de citas bibliográficas.
- Ser capaz de crear una base de datos de referencias bibliográficas a partir de fuentes de información.
- Ser capaz de insertar referencias bibliográficas en documentos de texto.



Búsquedas bibliográficas en investigación

La investigación científica no puede plantearse sin tener en cuenta el conocimiento previo existente sobre el tema de estudio. Este conocimiento se basa fundamentalmente en la literatura científica publicada, aunque pueda existir también información importante no publicada.

El objetivo de la investigación no es conseguir publicaciones, sino contribuir a una modificación del conocimiento previo. No tiene sentido emprender un proyecto sin tener en cuenta los estudios previamente realizados. De esta manera, evitaremos la realización de estudios redundantes o irrelevantes. Asimismo, la información obtenida de estudios previos facilitará el diseño metodológico de nuestro estudio y permitirá la realización de valoraciones comparativas de sus resultados. Si consultamos en trabajos previos los criterios y procedimientos de nuestro proyecto y las incidencias habituales a las que tendremos que enfrentarnos, podremos ahorrar esfuerzos y evitar errores.

Por lo tanto, una vez planteado el problema de investigación el primer paso es realizar una revisión sobre el tema. En ocasiones, la realización de esta revisión será la que dé

origen a la pregunta de investigación. Sorprendentemente, muchos proyectos se inician en la fase de diseño del cuaderno de recogida de datos, sin considerar que su elaboración será más fácil y fiable si antes hemos revisado proyectos similares.

No es necesario hacer revisiones extensas. En este sentido, algunos formularios para proyectos de investigación de agencias de financiación tienen limitaciones de espacio para ello. La revisión debe sustentarse en referencias bibliográficas actualizadas, obtenidas idealmente a partir búsquedas detalladas y estructuradas. La mención de los términos, estrategias y amplitud de las búsquedas será un aval de nuestra documentación. Recomendamos para ello utilizar procedimientos propios de las revisiones sistemáticas.

Podemos distinguir varios pasos en una búsqueda bibliográfica:

- Identificar los conceptos sobre los que se desea obtener información.
- Elegir los términos de búsqueda.
- Definir la profundidad o exhaustividad de la búsqueda.
- Elegir las fuentes de información.
- Elegir los tipos de documentos.
- Preparar la sintaxis de búsqueda.
- Depurar la estrategia de búsqueda.

Identificar los conceptos sobre los que se desea obtener información

En una búsqueda para documentar un proyecto, estos conceptos serán los elementos de la pregunta de investigación, estructurada en cuanto a paciente, intervención o exposición, comparación y resultado.

Veamos un ejemplo: queremos hacer un estudio para evaluar la capacidad predictiva de la obesidad abdominal (índice de cintura/talla superior a 0,5) para hipertensión arterial en la infancia. Los conceptos básicos los extraeremos de los componentes de la pregunta de investigación:

- **Población:** niños, adolescentes.
- **Intervención/Exposición:** obesidad abdominal (índice de cintura/talla).
- **Comparación:** ausencia de obesidad abdominal.
- **Resultado:** hipertensión arterial.

Además de estos elementos básicos, con frecuencia, otros elementos pueden ser necesarios para perfilar adecuadamente la búsqueda, que tendrán que ver con el tipo de diseño epidemiológico de nuestro estudio (por ejemplo: transversal, cohortes, etc.), la población (por ejemplo: geográficos, cronológicos, etc.) y factores relacionados con la exposición (por ejemplo: perímetro de cintura, índice de masa corporal, etc.) o el resultado (por ejemplo: síndrome metabólico, resistencia a la insulina, etc.).

Elegir los términos de búsqueda

Cada uno de los componentes de la pregunta de investigación debe ser concretado en uno o varios términos de búsqueda. Aunque esta elección de términos se puede hacer a partir de nuestro conocimiento previo, lo ideal es acudir a las listas de términos estandarizados, conocidos como descriptores, de las principales fuentes de información. La lista más empleada es la del *Medical Subject Headings* (MeSH) de Medline (<http://www.ncbi.nlm.nih.gov/mesh>). Estos descriptores se han clasificado en estructuras de árbol con términos sucesivamente subordinados (figura 2.1). Podemos acudir directamente a la lista original en inglés o acceder a ella después de consultar los listados equivalentes en español (figura 2.2) de los Descriptores en Ciencias de la Salud (DeCS) del Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud, BIREME (<http://decs.bvs.br/E/homepagee.htm>). Asimismo, para cada descriptor podemos elegir subdescriptores que limiten la búsqueda a aspectos concretos, como etiología, diagnóstico, tratamiento, pronóstico, etc. Siempre conviene emplear los descriptores estandarizados, aunque con frecuencia incluiremos en la búsqueda otros términos equivalentes, seleccionados entre los sinónimos, variantes ortográficas y formas abreviadas, con los que conseguiremos búsquedas más sensibles.

Figura 2.1

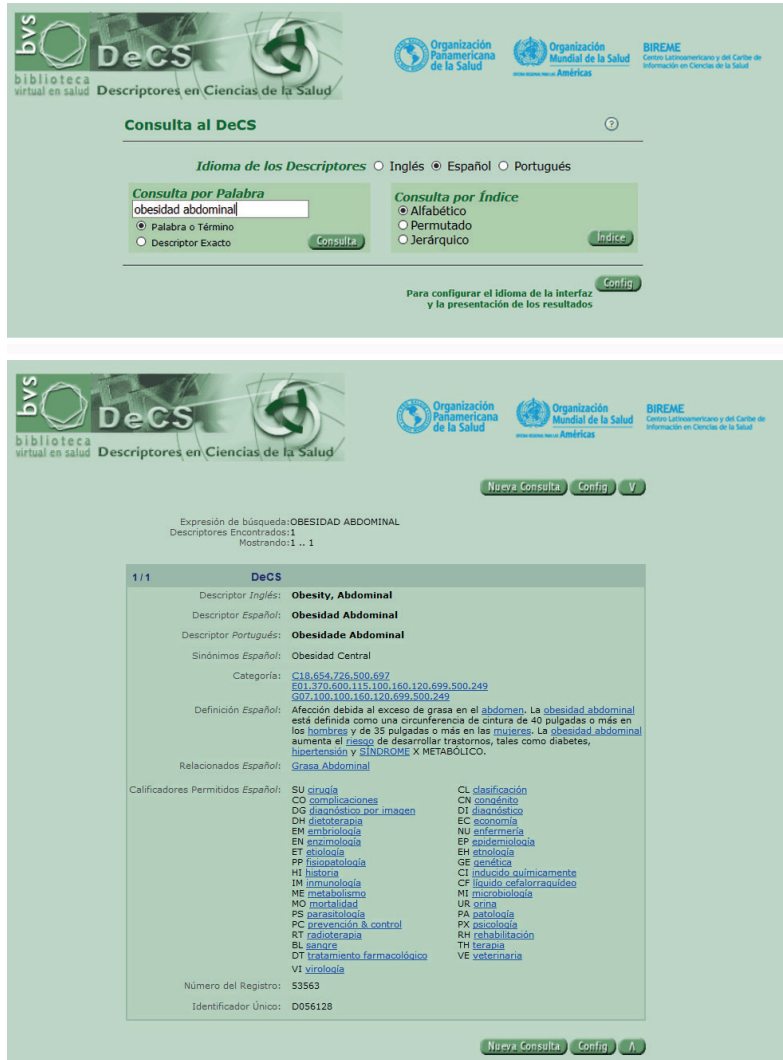
Pantalla de búsqueda de descriptores de PubMed

The screenshot displays the MeSH search interface. At the top, there is a search bar with 'MeSH' entered and a 'Search' button. Below the search bar, there are options for 'Limits' and 'Advanced'. The main content area is titled 'Obesity, Abdominal' and includes a definition: 'A condition of having excess fat in the abdomen. Abdominal obesity is typically defined as waist circumferences of 40 inches or more in men and 35 inches or more in women. Abdominal obesity raises the risk of developing disorders, such as diabetes, hypertension and METABOLIC SYNDROME. Year introduced: 2010'. Below the definition, there are 'PubMed search builder options' and a 'Subheadings' section with a grid of checkboxes for various categories. The 'Previous Indexing' section shows a hierarchical tree structure of MeSH terms, with 'Obesity, Abdominal' highlighted. Red boxes with numbers 1, 2, and 3 highlight the search bar, the subheadings grid, and the indexing tree respectively.

MeSH; accesible en <http://www.ncbi.nlm.nih.gov/mesh>. Descriptor estandarizado elegido (1), listado de subdescriptores (2) y estructura jerarquizada de descriptores (3)

Figura 2.2

Pantalla de búsqueda de Descriptores en Ciencias de la Salud



DeCS; accesible en <http://decs.bvs.br/E/homepagee.htm>.

En las figuras 2.1 y 2.2 podemos ver los resultados de las búsquedas realizadas en MeSH y DeCS con el término "Obesidad abdominal" y su correspondencia en inglés "Obesity, Abdominal". El descriptor de referencia seleccionado en MeSH es el que generalmente, aunque no siempre, tendrán asignados los artículos relacionados en Medline. El descriptor se puede precisar eligiendo uno o más subdescriptores (área marcada con un "2"). Si observamos el árbol de descriptores (área marcada con un "3") vemos los términos jerárquicamente relacionados, entre los que serán descriptores alternativos a seleccionar. Otras fuentes de información emplean listados y estrategias de indexación similares.

Para cada uno de los elementos de la pregunta de investigación tendremos que buscar los términos más apropiados, completando así el conjunto de elementos de nuestra estrategia de búsqueda.

Definir la profundidad o exhaustividad de la búsqueda

Si queremos estar seguros de que ningún trabajo relacionado haya quedado al margen de nuestra revisión, la búsqueda tendrá que ser muy amplia y sensible. Ello traerá como consecuencia la obtención de documentos intrascendentes que harán el trabajo más arduo, sin embargo, será la única forma de ser exhaustivos. Para hacer una búsqueda amplia y sensible, debemos emplear todos los términos relacionados existentes, términos comodín truncados (por ejemplo: bronqu* para bronquitis, bronquiolitis, etc.), no aplicar límites de tiempo o idioma, ni excluir fuentes de información.

Para ser exhaustivos, en vez de ceñirnos al descriptor específico (por ejemplo: "Obesity, Abdominal"), podemos asociar a la búsqueda términos relacionados o alternativos (por ejemplo: "Abdominal obesity", "Visceral Obesity", "Visceral obesities", "Waist-Height Ratio", "Waist-to-height", "waist to height", "waist height", "waist circumference to height", etc.). De igual manera, en vez de buscar exclusivamente en PubMed, haremos búsquedas en otras bases de datos (por ejemplo: EMBASE, LILACS, CINHALL, etc.).

Si, por el contrario, queremos hacer búsquedas reducidas y específicas, emplearemos exclusivamente los descriptores estándar, usaremos restricciones de tiempo o idioma, y exploraremos solo en las fuentes de información principales. Esta estrategia nos consumirá menos tiempo de trabajo y evitaremos revisar artículos intrascendentes, pero tendremos el riesgo de perder algún documento relevante.

Elegir las fuentes de información

Debemos elegir el tipo de fuente de información (primaria, secundaria o terciaria), el ámbito geográfico (nacional o internacional), el idioma y el tipo de documento a obtener. Las fuentes de información se pueden diferenciar en tres grandes grupos (tabla 2.1): fuentes de información primarias (libros de texto y revistas de biomedicina), secundarias (bases de datos bibliográficas, internacionales y nacionales) y terciarias (repertorios de revisiones sistemáticas y guías de práctica clínica y metabuscadores relacionados con la medicina basada en la evidencia, que incorporan una selección, análisis crítico y síntesis de información).

Tabla 2.1 Fuentes de información

Fuentes primarias	Fuentes secundarias	Fuentes terciarias
Revistas científicas Libros de texto	Internacionales: <ul style="list-style-type: none"> · Medline/PubMed · EMBASE · Web of Science · CINAHL · LILACC Nacionales: <ul style="list-style-type: none"> · IBECS · IME · CUIDEN 	Colaboración Cochrane Guías de práctica clínica: <ul style="list-style-type: none"> · GuiaSalud · National Guideline Clearinghouse Metabuscadores: <ul style="list-style-type: none"> · Tripdatabase · Epistemonikos

Habitualmente accederemos a las fuentes primarias a través de búsquedas en fuentes secundarias, que recopilan y clasifican los documentos primarios. Entre las fuentes secundarias más empleadas podemos destacar las generales, Medline (PubMed) y Embase, y otras más selectivas, como LILACS, de ámbito latinoamericano, CINAHL, de enfermería, y Web of Science (Institute for Scientific Information; Thomson Reuters), que incluye un repertorio selectivo de revistas y referencias de los trabajos indexados. Existen otras fuentes secundarias de alcance nacional, como el Índice Bibliográfico en Ciencias de la Salud (IBECS), el Índice Médico Español (IME, actualmente desactualizado) y CUIDEN (específico de enfermería).

La fuente secundaria más empleada, con diferencia, es Medline, a través de su motor de búsqueda PubMed. Su acceso gratuito y su eficiente sistema de clasificación e indexación le han hecho convertirse en un estándar de referencia. Aunque una búsqueda en PubMed permite obtener buenas búsquedas bibliográficas, éstas solo serán exhaustivas si las completamos con otras fuentes de información. EMBASE incluye un mayor repertorio de revistas que Medline, aunque gran parte de ellas son comunes a ambas fuentes. La principal limitación de EMBASE es su carácter “de pago”, por lo que muchas revisiones ignoran esta fuente de información, a pesar de que es la única vía de acceso a información que puede ser relevante.

Si nuestra revisión necesita recopilar ensayos clínicos, revisiones sistemáticas o guías de práctica clínica, tendremos que ampliar la búsqueda a fuentes de información terciarias. La Colaboración Cochrane, a través de la Biblioteca Cochrane Plus (<http://www.bibliotecacochrane.com/BCPMain.asp>), nos da acceso a revisiones sistemáticas y ensayos clínicos. Distintas instituciones facilitan el acceso a repertorios de guías de práctica clínica; sirvan como ejemplo Guíasalud (<http://portal.guiasalud.es/web/guest/guias-practica-clinica>) y National Guideline Clearinghouse (<https://www.guideline.gov/>). Asimismo, existen recursos que lanzan búsquedas automáticas en otras fuentes de información, que conocemos como metabuscadores; sirvan como ejemplo Tripdatabase (<https://www.tripdatabase.com/>) y Epistemonikos (<https://www.epistemonikos.org/es/>). Las búsquedas en estas fuentes de información tendrán que ser más sencillas, porque la información recopilada es muy seleccionada.

Elegir los tipos de documentos

Un aspecto importante a la hora de diseñar las búsquedas bibliográficas es establecer qué tipo de documentos queremos obtener en la revisión: libros, artículos originales, revisiones sistemáticas, guías de práctica clínica, etc. Para ello podemos incluir en la búsqueda términos propios de cada tipo de documento (por ejemplo: ensayos clínicos), o bien emplear los filtros metodológicos correspondientes, que algunas fuentes, como PubMed, incorporan en sus pantallas de búsqueda. También podemos hacer búsquedas selectivas directamente en repertorios especializados (por ejemplo: Cochrane Plus para revisiones sistemáticas y ensayos clínicos o GuíaSalud para guías de práctica clínica).

Preparar la sintaxis de búsqueda

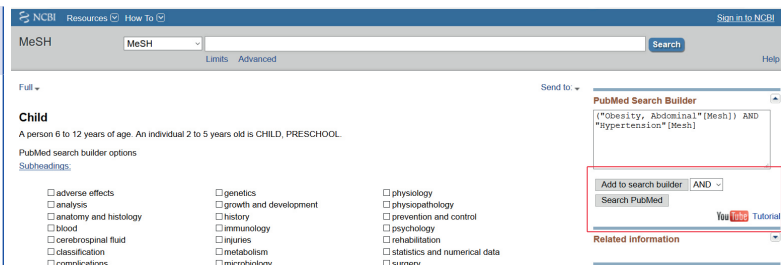
Una vez elegidos los términos y las fuentes de búsqueda, debemos elaborar secuencias de búsqueda adaptadas a cada fuente, en las que combinaremos los descriptores y términos relacionados empleando operadores booleanos (AND, OR, NOT). Podemos recurrir a paréntesis, truncamientos (bronqu*) y operadores de proximidad, cuando sea necesario. Habitualmente, los términos y operadores se encadenan en un texto lineal de longitud variable que se introduce en el campo de búsqueda. También se pueden fraccionar en varias secuencias, a las que se asignan números correlativos que luego se emplean para combinarlas. En algunas bases de datos se ofrecen múltiples campos con menús desplegables para asignar operadores booleanos de combinación.

Aunque podemos construir las búsquedas manualmente e introducirlas en los campos de búsqueda, la forma más precisa de construirlas es emplear los recursos asociados a los listados de términos estandarizados. Por ejemplo, en el motor de búsqueda PubMed, en vez de escribir la secuencia de búsqueda en el campo de la página de entrada, accederemos a la página MeSH de descriptores (figura 2.1). En esa página podemos buscar y elegir uno por uno los descriptores estandarizados. Cuando elijamos cada uno de ellos, podemos examinar su posición jerárquica en el árbol de descriptores, elegir el término concreto o expandir la búsqueda a todos los descriptores subordinados. Asimismo, podemos asignar a los descriptores los subdescriptores que los perfilen (por ejemplo: etiología, diagnóstico, tratamiento, etc.).

Una vez identificado el descriptor en la pantalla de MeSH (figura 2.1), con o sin algún subdescriptor asociado, lo incorporamos a la estrategia de búsqueda presionando el botón "Add to search builder" (figura 2.3). El término aparecerá junto al texto "[Mesh]", que indica al buscador que solo lo busque en el apartado de descriptores del artículo (no en otros apartados, como título, resumen, etc.). Si queremos quitar esta limitación, tendremos que borrar ese texto añadido.

Figura 2.3

Pantalla de búsqueda de descriptores (MeSH) en PubMed



Si hemos optado por construir la búsqueda en la pantalla de los descriptores (MeSH), repetiremos el mismo proceso para cada término de búsqueda, enlazándolos con las opciones "AND", "OR" o "NOT" disponibles junto al botón "Add to search builder". Una vez completada la secuencia de búsqueda, la lanzaremos pulsando sobre el botón "Search PubMed" (figura 2.3). La alternativa a este procedimiento asistido con MeSH

sería escribir la secuencia directamente en la caja de búsqueda de la página principal de PubMed o escribirla en un editor de texto y pegarla en dicha caja. El resultado de cualquiera de estos procedimientos será el mismo (figura 2.4).

En la tabla 2.2 presentamos dos estrategias de búsqueda para nuestro ejemplo de la obesidad abdominal, una búsqueda precisa y específica y otra amplia y sensible.

Tabla 2.2 Estrategias de búsqueda sobre obesidad abdominal e hipertensión

Búsqueda precisa y específica:

- ("Child"[MeSH Terms] OR "Adolescent"[MeSH Terms]) AND
- ("Obesity, Abdominal"[Mesh] OR "Waist-Height Ratio"[Mesh]) AND
- ("Hypertension"[Mesh])

Búsqueda amplia y sensible:

- ("child"[MeSH Terms] OR "adolescent"[MeSH Terms] OR child OR adolescent) AND
- ("Obesity, Abdominal"[Mesh] OR (Obesity AND (Abdominal OR Visceral)) OR "W-HtR" OR "WC/height" OR "WHtR" OR "waist:height ratio" OR "waist-to-height ratio" OR "waist to height ratio" OR "wthr" OR "Waist-Height Ratio"[Mesh] OR "Waist-to-height" OR "waist to height" OR "waist height" OR "waist circumference to height") AND
- ("Hypertension"[Mesh] OR "Blood Pressure"[Mesh] OR Hypertens* OR (elevat* AND "blood pressure") OR "high blood pressure" OR "Metabolic Syndrome X"[Mesh])

En la figura 2.4 se presenta el resultado de la búsqueda sencilla de la tabla 2.2.

Figura 2.4

Resultado de búsqueda en PubMed

The screenshot shows the PubMed search results page. The search query is: ("Child"[MeSH Terms] OR "Adolescent"[MeSH Terms]) AND ("Obesity, Abdominal"[Mesh] OR "Waist-Height Ratio"[Mesh]) AND "Hypertension"[Mesh]. The results are displayed in a list format, showing the first four results. The first result is "Prevalence of smoking, alcohol consumption and metabolic syndrome in patients with psoriasis" by Adigun E, Utzun S, Erduran F, Güler MA, published in An Bras Dermatol in 2018. The second result is "Differences in prevalence of hypertension and associated risk factors in urban and rural residents of the northeastern region of the People's Republic of China. A cross-sectional study" by Wang J, Sun W, Wells GA, Li Z, Li T, Wu J, Zhang Y, Liu Y, Li L, Yu Y, Liu Y, Qi C, Lu Y, Liu N, Yan Y, Liu L, Hui G, Liu B, published in PLoS One in 2018. The third result is "Prevalence of hypertension determined by ambulatory blood pressure monitoring (ABPM) and body composition in long-term survivors of childhood cancer" by Güler E, Col N, Buyukselik M, Balat A, published in Podiatr Hematol Oncol in 2018. The fourth result is "Cardio-metabolic and socio-environmental correlates of waist-to-height ratio in German primary schoolchildren: a cross-sectional evaluation" by Vorwieser E, Kelso A, Steinsacker JM, Kesztölyös D, URMEICE study group, published in BMC Public Health in 2018.

Depurar la estrategia de búsqueda

Cuando como resultado de la búsqueda se ha obtenido un número de documentos excesivo, estos pueden reducirse estableciendo limitaciones, como disminuir el período de búsqueda a menos años, o incluir en la búsqueda solo los descriptores principales (este sistema reduce extraordinariamente el número de trabajos recuperados, aumentando su relevancia). Por el contrario, si se han obtenido pocos documentos, debe replantearse de nuevo la búsqueda, aumentando el período cubierto a más años, realizando extensiones de descriptores a sus subordinados o utilizando palabras del lenguaje natural y todos los sinónimos posibles.

En nuestro ejemplo (figura 2.4) hemos encontrado 111 resultados (cambiará en función de la fecha de la búsqueda), un número accesible para la revisión detallada y selección de artículos. Si empleáramos la búsqueda más amplia (tabla 2.2), encontraríamos 1.010 referencias. La elección de una u otra secuencia de búsqueda va a depender de la exhaustividad con la que queramos hacer la revisión. Podemos examinar las referencias de las primeras páginas de resultados para ver la adecuación de la secuencia elegida; si en la búsqueda específica faltan artículos de referencia ya conocidos, tendremos que emplear una secuencia más sensible. Si al examinar los resultados de la búsqueda sensible encontramos muchas referencias no relacionadas, tendremos que revisar la secuencia de búsqueda.

En PubMed, al igual que en otras fuentes de información, contamos con recursos de filtrado de referencias muy útiles. Podemos ver en la figura 2.4 los distintos filtros disponibles en el área señalada a la izquierda de la imagen. Entre los filtros más usados podemos destacar el tipo de artículo, la disponibilidad de texto completo y la fecha de publicación; entre los otros filtros muy empleados que se pueden desplegar en posiciones inferiores, no visibles en la imagen, destacamos el filtro de grupos de edad.

Una vez elegida la secuencia de búsqueda definitiva, la búsqueda se completa con la lectura de títulos y resúmenes para poder identificar los artículos que nos interesa seleccionar para su examen detallado. En este paso del proceso de búsqueda resulta muy útil contar con gestores de referencias bibliográficas, programas desde los que podemos almacenar automáticamente y editar los artículos seleccionados. Estos programas nos facilitarán en gran manera la gestión de las búsquedas y el trabajo posterior de edición de informes y artículos. En un próximo apartado de este capítulo serán abordados.



Ejercicio práctico 2.1

Invitamos a los lectores a diseñar y depurar una búsqueda bibliográfica para documentar dos estudios. Recomendamos seguir los pasos descritos en este capítulo.

- a) Vamos a realizar un estudio para establecer si la asistencia a guardería se asocia a un mayor o menor riesgo de desarrollar sibilantes recurrentes en la primera infancia.
- b) Vamos a evaluar la eficacia del suero salino hipertónico nebulizado en la bronquiolitis aguda para reducir el riesgo de ingreso en servicios de Urgencias.



Gestores de referencias bibliográficas

Los gestores de referencias bibliográficas se han convertido en los últimos años en una herramienta esencial para recopilar, almacenar, organizar y, en definitiva, gestionar la información utilizada en cualquier trabajo de difusión y/o investigación.

El investigador tiene a su disposición múltiples fuentes de información en las que localizar, identificar y seleccionar los trabajos que necesita para documentar sus proyectos científicos, pero se encuentra con la dificultad de almacenar y gestionar estos trabajos. Para facilitar esta tarea, contamos con diferentes programas que ayudan al investigador a recopilar y manejar las referencias bibliográficas obtenidas en una base de datos. Estos programas se conocen como gestores de referencias bibliográficas.

La misión de un gestor de referencias es facilitar todas las tareas que se necesitan hacer con cada referencia bibliográfica, desde la captación automática de sus elementos de identificación y contenido (autor, título, tipo, editorial, año, número, páginas, resumen, etc.) hasta su asignación en forma de cita en los informes o artículos elaborados para el proyecto. El uso de un gestor de referencias reduce enormemente la carga de trabajo y el riesgo de cometer errores en este proceso.

Podemos definir como “cita” la mención abreviada o el enlace que se realiza desde el texto de un documento a la referencia bibliográfica completa al final del mismo. Tanto el formato de citación en el texto como de referencia al final del documento van a tener que adaptarse a las normas editoriales de la institución o revista que recibe el informe. Los gestores de referencias hacen esta adaptación de forma automática, pudiendo modificarse fácilmente desde una misma base de datos de referencias a diferentes normas editoriales. La forma más habitual de citación es la inserción de números correlativos intercalados en el texto, que se corresponden con el orden del listado de referencias al final del documento; los gestores de citas asignan automáticamente los números, actualizando el orden correcto con cualquier nueva cita, en el texto y en el listado de referencias. Cualquier investigador que haya tenido que elaborar informes con muchas referencias entiende la utilidad de esta función.

Podemos agrupar las funciones de los gestores de referencias bibliográficas en tres grandes áreas: la entrada, la organización y la salida de los datos. La **entrada de referencias** puede hacerse manualmente, escribiendo los elementos básicos de cada referencia, o bien, automáticamente, a partir de los resultados de las búsquedas realizadas en distintas fuentes de información. Esta captación automática puede ser directa, desde la pantalla de búsqueda, si el programa que usamos permite la exportación directa, o bien a través de salidas de resultados en ficheros de texto o con formatos específicos de gestores de citas, para lo que se requiere tener filtros específicos de importación.

Las referencias se recopilan en una base de datos independiente que puede guardarse en nuestro ordenador o, con algunos programas, en alojamientos *online* que se pueden

compartir. Los gestores tienen establecidas estructuras predefinidas que representan las diferentes tipologías documentales (artículos de revistas, libros, páginas web, comunicaciones a congresos, capítulos de libro, etc.). Cada estructura dispone de unos campos comunes básicos (autor, título, resumen, etc.) y unos campos específicos que dependerán del tipo de publicación. Considerando el volumen de información a incorporar a la base de datos, nadie debería incorporar manualmente una referencia si esta está accesible para su captación automática desde fuentes de información. Además, algunos de estos programas permiten la importación y/o el enlace de los documentos a texto completo.

Entre las funciones de **organización** podemos destacar: la clasificación o agrupación de referencias en carpetas específicas, la identificación de referencias duplicadas, la búsqueda de referencias y su revisión. Es habitual que las referencias obtenidas desde una fuente de información tengan que ser depuradas. Muchos de estos procesos de depuración tienen que ver con el hecho de que los títulos de los documentos se presentan en versión traducida al inglés; si el idioma original es otro, este debe ser reemplazado para su uso en la referencia. También es habitual tener que revisar ciertos nombres de autores que han sido almacenados con errores (por ejemplo: acentos, signos especiales, etc.).

Finalmente, la tercera área de funciones es la **salida de datos**. Podemos insertar citas en documentos, generar listas de referencia, adaptar las citas y referencias a las diferentes normas editoriales, y exportar referencias para compartirlas con otros usuarios o programas. La localización y selección de las citas a insertar se facilita con las herramientas de ordenación de referencias por campos específicos (autor, año, título) o la localización por carpetas o marcas temáticas.

Los gestores de referencias cuentan con información de los principales formatos de citación. Asimismo, en las páginas web de cada programa podemos encontrar formatos específicos para otras normas editoriales. Estas normas pueden ser incluso modificadas para crear formatos de citación personalizados. Cuando nuestro artículo debe ser modificado de una editorial a otra, la tarea de adaptación se hace automáticamente.

Contamos con numerosos programas de gestión de referencias, tanto de escritorio, programas que se instalan y usan desde ordenadores personales, como aplicaciones de entorno web, que se ejecutan desde un servidor *online*, donde se almacena la información. Algunos de ellos son gratuitos, como Zotero y Mendeley, y otros de pago, como RefWorks, Procite, EndNote y Reference Manager.

Los programas de escritorio deben ser descargados e instalados en el ordenador. Habitualmente también hay que instalar complementos de integración con buscadores *online*, para poder hacer exportaciones automáticas, y editores de texto, para poder insertar las referencias.

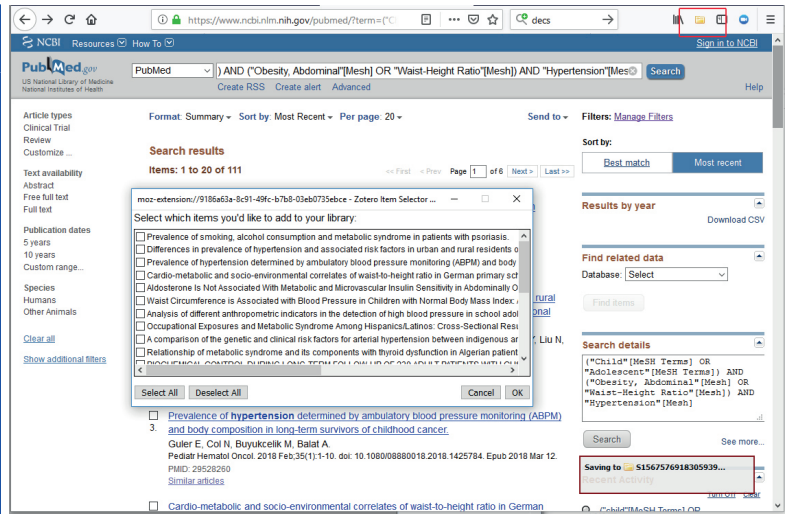
Mostraremos a continuación el uso de uno de estos programas. Hemos elegido Zotero por ser una aplicación gratuita bien integrada con buscadores como PubMed y fácil de integrar con editores de texto como Word. Veremos las dos funciones principales: cap-

tación automática de referencias e inserción de citas. Podemos descargar este programa en la URL <https://www.zotero.org/download/>. Pueden encontrarse tutoriales en la Web para aprender su instalación y manejo.

En la parte superior derecha de la figura 2.4 podemos ver un icono con forma de carpeta que aparece si tenemos instalado y activo Zotero en nuestro ordenador. Pulsando este icono con el ratón activamos la exportación de referencias. En la figura 2.5 se presenta la pantalla en la que se nos ofrece la opción de seleccionar las referencias presentes en la salida de resultados, para su exportación a Zotero.

Figura 2.5

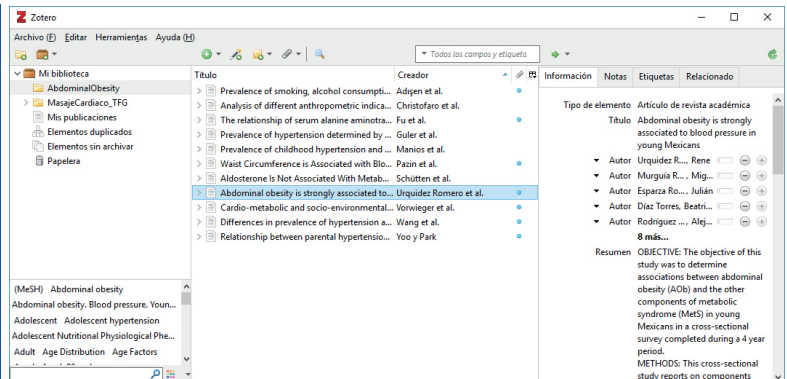
Pantalla de búsquedas de PubMed, con el icono de exportación de citas a Zotero integrado (icono con forma de carpeta señalado arriba a la derecha)



Podemos señalar las referencias a la vista en la pantalla de resultados y exportarlas a Zotero. Si optamos por examinar una única referencia, pulsando en su título, el icono de exportación de Zotero cambiará de aspecto, de una carpeta a una hoja; pulsando este icono se exportará la referencia a la vista. En la figura 2.6 podemos ver la pantalla de Zotero con algunas referencias exportadas.

Figura 2.6

Pantalla de Zotero con referencias importadas desde PubMed



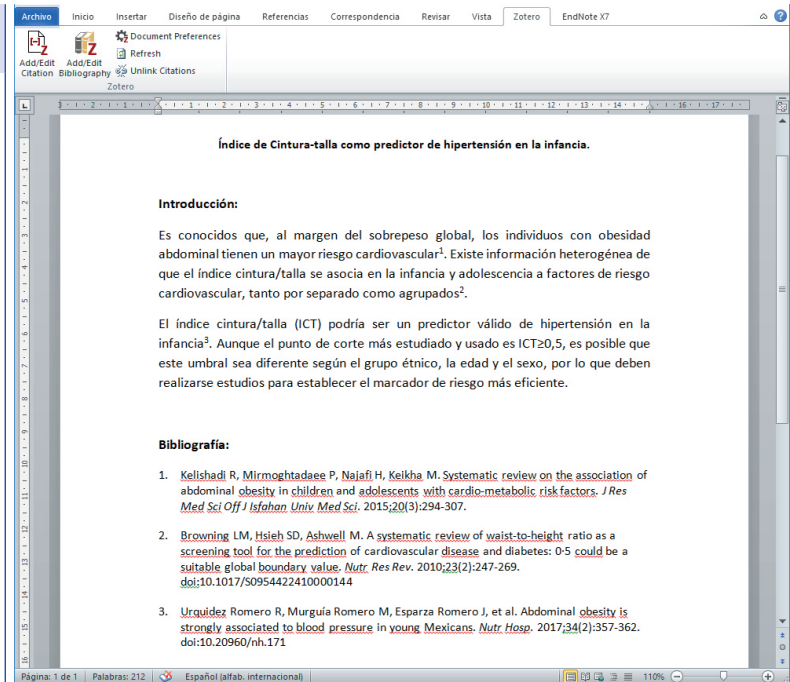
En la figura 2.6 vemos las tres áreas en las que se divide la aplicación. A la izquierda la referencia a la base de datos activa (Mi biblioteca), de la que cuelgan varias carpetas: la carpeta "Mis publicaciones", en la que podemos indicar los documentos creados con vinculación a la base de datos: una carpeta donde se guardan las referencias duplicadas; la carpeta con las referencias sin asignar a carpetas temáticas y la carpeta en la que se sitúan las referencias borradas. Podemos añadir cuantas carpetas queramos en las que clasificar las referencias por temas.

En el área intermedia vemos los títulos y autores de cada una de las referencias. Pulsando en las cabeceras correspondientes podemos ordenar las referencias, para localizarlas por título o autor. Si pulsamos el icono verde con el signo "+", disponible sobre esta área, se creará una nueva referencia en blanco. En el área de la derecha vemos los campos que comprenden los elementos de cada referencia, donde se pueden consultar o editar.

Una vez importadas y depuradas todas las referencias bibliográficas de nuestra búsqueda, podríamos pasar a la fase de citación en un documento. En la figura 2.7 vemos un documento de texto en el que hemos insertado varias citas intercaladas en el texto (con números correlativos entre paréntesis) y al final del mismo el listado de bibliografía en el que se detallan todas ellas ordenadas. En la parte superior izquierda de la imagen podemos ver los iconos del complemento de Zotero que permiten insertar citas (Add/Edit Citation) o insertar el listado de bibliografía (Add/Edit Bibliography). Para insertar una cita debe situarse el cursor en el punto del texto que se desee y pulsar el icono "Add/Edit Citation". Se abrirá un menú en el que nos deja elegir el formato de citación y a

Figura 2.7

Pantalla de un editor de texto en el que se han insertados citas y listado de referencias con Zotero



continuación nos mostrará las referencias disponibles para que elijamos la referencia a insertar. Cualquier nueva cita insertada entre las ya existentes recibirá el número que le corresponda, ajustándose automáticamente los números del resto de citas. Para ver la lista de referencias tendremos que situarnos al final del texto y pulsar el icono “Add/Edit Bibliography”; aparecerán todas las referencias ordenadas de las citas previamente insertadas.

Invitamos al lector a consultar los tutoriales sobre Zotero existentes en la Web y a aprender a manejarlo con los ejemplos prácticos propuestos en los ejercicios previos de este capítulo.

BIBLIOGRAFÍA RECOMENDADA

- Aleixandre-Benavent R, González Alcaide G, González de Dios J, Alonso-Arroyo A. Fuentes de información bibliográfica (I). Fundamentos para la realización de búsquedas bibliográficas. *Acta Pediatr Esp.* 2011;69(3):131-6.
- Aleixandre-Benavent R, Bolaños-Pizarro M, González de Dios J, Navarro Molina C. Fuentes de información bibliográfica (II). Bases de datos bibliográficas españolas en ciencias de la salud de interés en pediatría: IME, IBECs y MEDES. *Acta Pediatr Esp.* 2011;69(4):177-82.
- Aleixandre-Benavent R, González Alcaide G, González de Dios J, Alonso-Arroyo A, Navarro Molina C. Fuentes de información bibliográfica (III). Bases de datos bibliográficas extranjeras en ciencias de la salud de interés en pediatría: MEDLINE, Embase y LILACS. *Acta Pediatr Esp.* 2011;69(5):223-34.
- Alonso-Arroyo A, González de Dios J, Navarro Molina C, Vidal-Infer A, Aleixandre-Benavent R. Fuentes de información bibliográfica (XII). Gestores de referencias bibliográficas: generalidades. *Acta Pediatr Esp.* 2012;70(5):211-6.
- Contandriopoulos AP, Champagne F, Potvin L, Denis JL, Boyle P. Preparar un proyecto de investigación. 1.ª ed. Barcelona: SG Editores S.A.; 1991.

3

Diseños epidemiológicos Muestreo y aleatorización

Objetivos docentes:

- Conocer los principales tipos de diseños epidemiológicos.
- Aprender a seleccionar el diseño más adecuado para cada pregunta de investigación.
- Conocer las características de los estudios transversales, estudios de casos y controles, estudios de cohortes y ensayos clínicos.
- Diferenciar muestreo de aleatorización.
- Conocer los principales tipos de muestreo.
- Conocer los pasos del procedimiento de aleatorización.
- Reconocer la importancia de la ocultación de la secuencia de aleatorización.



Tipos de diseño de los estudios epidemiológicos

La epidemiología estudia la distribución de las enfermedades o acontecimientos relacionados con la salud y trata de identificar sus causas, su historia natural y el papel que desempeñan determinados factores de riesgo o protección. Para ello emplea una metodología basada en el método científico; el punto de partida es la identificación de una pregunta de investigación inspirada en el conocimiento actual sobre un determinado problema de salud. La contestación a esta pregunta requiere la formulación de una hipótesis de trabajo, que ha de contrastarse en un estudio llevado a cabo en una muestra de población. Generalmente, la pregunta de investigación de un estudio epidemiológico trata de identificar la relación entre un factor de estudio o exposición, que puede ser de riesgo o protección, y una variable de efecto o respuesta, habitualmente curación o enfermedad. La estrategia que sigamos para medir ambos parámetros, especialmente la secuencia temporal en que son valorados, va a condicionar el grado de evidencia que finalmente obtendremos sobre dicha relación.

Podemos clasificar los estudios epidemiológicos en función de distintos criterios (tabla 3.1): el objetivo de la pregunta de investigación (descriptivos o analíticos), las características del diseño (direccionalidad, tipo de muestreo y temporalidad), la existencia de intervención y las unidades de estudio.

Tabla 3.1 Criterios de clasificación de los estudios epidemiológicos

Pregunta de investigación:	Características del diseño:	Existencia de intervención:	Unidades de estudio:
<ul style="list-style-type: none"> · Descriptivos. · Analíticos (grupo de comparación). 	<ul style="list-style-type: none"> · Direccionalidad (anterógrada, retrógrada o simultánea). · Tipo de muestreo (exposición, efecto u otros). · Temporalidad (histórica, concurrente, mixta). 	<ul style="list-style-type: none"> · Observacionales. · Experimentales 	<ul style="list-style-type: none"> · Individuos. · Grupos.

Según el **objetivo** general de nuestra pregunta de investigación, distinguiremos dos grupos de estudios: los descriptivos y los analíticos. Los descriptivos tratan de contestar preguntas del tipo: ¿qué pasa?, ¿a quién?, ¿cuántos?, ¿dónde? o ¿cuándo?, mientras que los analíticos intentan responder a cuestiones como: ¿por qué?, ¿cómo evitarlo? o ¿sirvió para algo? La principal diferencia entre ambos radica en que los estudios analíticos requieren, para contestar sus preguntas, la existencia de grupos de comparación con los que contrastar la hipótesis formulada.

El diseño de los estudios epidemiológicos puede ser catalogado en relación a tres características: direccionalidad, tipo de muestreo y temporalidad. La **direccionalidad** se refiere al orden en el que la variable de exposición (factor de riesgo o protección) y la variable efecto o respuesta (enfermedad, curación) son investigados. Un estudio tendrá una direccionalidad anterógrada cuando la medición de la exposición anteceda a la del efecto; retrógrada cuando se mida primero el efecto y posteriormente se busque la exposición, y simultánea cuando ambas variables se determinen en el mismo momento. Solo la direccionalidad anterógrada es capaz de garantizar que la ocurrencia de la exposición preceda en el tiempo a la del efecto, lo que constituye uno de los criterios clásicos de causalidad.

También podemos diferenciar los estudios en función del **tipo de muestreo**; así, los sujetos de estudio podrán ser seleccionados por la presencia o ausencia de un factor de exposición (de riesgo o protección), de una variable de resultado o efecto (enfermedad), o bien por otros criterios de selección ajenos a ambos factores.

La tercera característica de diferenciación en cuanto al diseño de un estudio es la temporalidad. La **temporalidad** se refiere a la relación entre el tiempo en el que el investigador realiza el estudio y la ocurrencia temporal del factor de exposición y del efecto. La temporalidad es histórica cuando la exposición y el efecto han ocurrido antes de iniciarse el estudio, concurrente cuando ambos tienen lugar durante el proceso de investigación y mixta cuando el factor de exposición es histórico y el efecto concurrente. A la hora de garantizar la validez de la información recogida en un estudio, es la temporalidad concurrente la que ofrece un mayor grado de confianza.

La utilización de los términos **prospectivo y retrospectivo** para referirse al tipo de diseño de un estudio es un motivo frecuente de confusión. Algunos autores emplean estos términos para expresar direccionalidad, mientras que otros indican con ellos temporalidad. Asimismo, ambos términos se han empleado para denominar tipos concretos de estudios, como los estudios de cohortes (prospectivos) y de casos y controles (retrospectivos). También se han empleado para referirse a la estrategia de análisis estadístico, en función de que las pruebas de contraste de hipótesis se hayan definido antes de conocer los datos o durante el análisis. Considerando estas cuestiones, parece prudente evitar estos términos a la hora de describir el diseño de un estudio.

Como hemos visto, el principal interés de los estudios epidemiológicos es explorar la relación entre un factor de exposición y un efecto. Con frecuencia, establecer esta relación se ve mediatizada por otros factores que no podemos controlar. Por ello, la mejor manera de controlar la repercusión de una exposición es introducirla de forma controlada, de manera que desencadenemos una secuencia de acontecimientos que den lugar o no al efecto buscado. Si como investigadores somos meros espectadores y no modificamos el factor de exposición, estaremos realizando estudios observacionales; si, por el contrario, lo manipulamos, llevaremos a cabo estudios experimentales o de **intervención**. Aunque los estudios experimentales son los que más grado de evidencia aportan, determinadas preguntas de investigación no pueden ser contestadas con este diseño, ya que la intervención no sería ética o viable.

Finalmente, los estudios se diferencian en función de las **unidades de estudio** consideradas. Así, podemos distinguir los estudios realizados sobre sujetos individuales o sobre grupos de población. Los estudios realizados sobre grupos de población caracterizan una serie de diseños diferenciados; para los estudios descriptivos, los denominados estudios ecológicos, y para los de intervención, los conocidos como ensayos comunitarios.

En la figura 3.1 se representa un algoritmo que recoge, en función de sus características, los tipos de estudios epidemiológicos, representándose cada uno de ellos tanto más a la derecha y abajo cuanto mayor sea el grado de evidencia que aportan. Los estudios ecológicos y los ensayos comunitarios no han sido representados en el algoritmo. En la tabla 3.2 se presentan las características de los principales tipos de diseño.

Figura 3.1

Algoritmo de clasificación de los estudios epidemiológicos

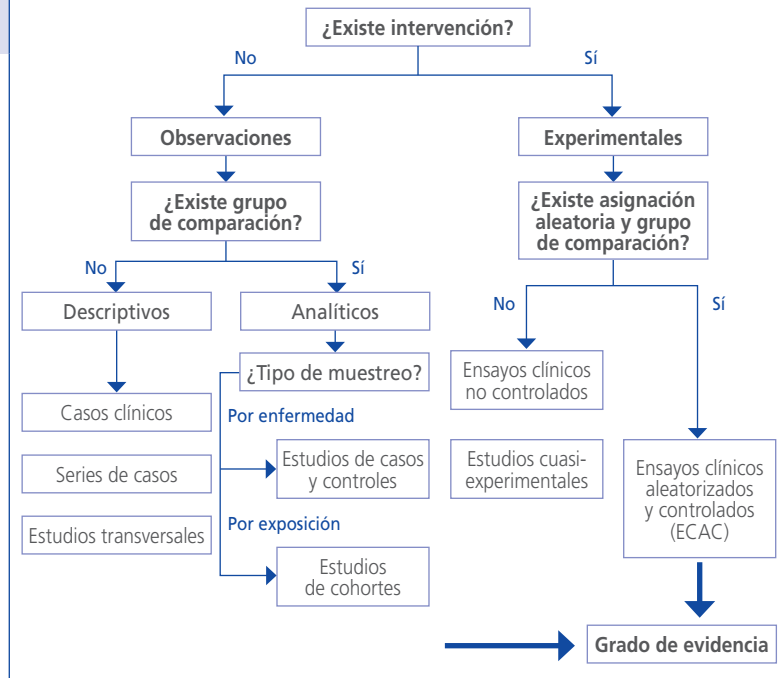


Tabla 3.2 Características de los principales tipos de estudio

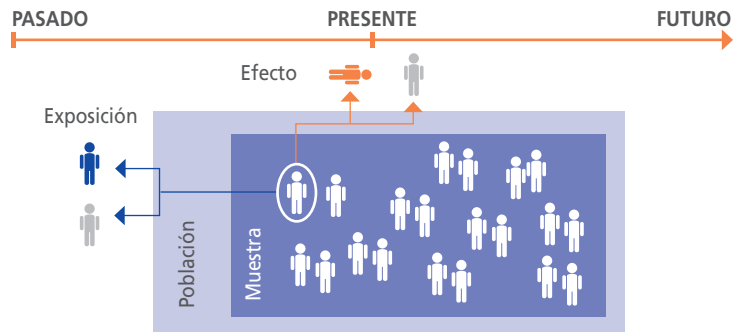
Tipo de estudio	Objeto	Muestreo	Temporalidad	Direccionalidad	Intervención
Transversal	Descriptivo	Otros criterios	Concurrente	Simultánea	No
Casos y controles	Analítico	Efecto	Mixta	Retrógrada	No
Cohortes	Analítico	Exposición	Histórica Concurrente Mixta	Anterógrada	No
Ensayo clínico	Analítico	Exposición controlada	Concurrente	Anterógrada	Sí

Estudios transversales

Los estudios transversales (figura 3.2) son estudios observacionales, descriptivos y simultáneos (no direccionales). Estudian la población total o una muestra significativa en un punto del tiempo o en un periodo corto, buscando la presencia o ausencia de factores de riesgo (exposición) y enfermedad (efecto). También son denominados estudios “de prevalencia”, ya que nos permiten cuantificar esta medida de enfermedad.

Figura 3.2

Esquema de estudio transversal



Los estudios transversales son muy útiles en planificación sanitaria, ya que permiten describir la distribución de las enfermedades y factores de riesgo y sugerir hipótesis que puedan comprobarse en otros estudios analíticos. Tienen como ventajas que son fáciles de ejecutar, relativamente poco costosos y fácilmente repetibles. Entre sus inconvenientes podemos destacar que son estudios sujetos a múltiples sesgos, sobre todo de selección (es primordial que la muestra elegida sea representativa), no sirven para investigación causal por carecer de direccionalidad, no permiten estimar incidencias ni riesgos y son poco útiles para estudiar enfermedades raras o de corta duración.

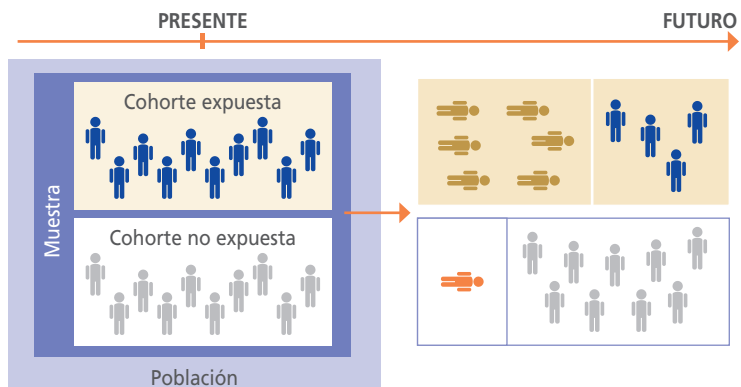
Estudios de cohortes

Los estudios de cohortes (figura 3.3) son estudios observacionales, analíticos, generalmente anterógrados, de temporalidad concurrente o mixta, en los que se muestrea en relación con la exposición. Se conocen también como estudios de "seguimiento", de "proyección", de "incidencia" o "prospectivos".

El punto de partida de estos estudios es la selección de dos grupos de sujetos, diferenciados según tengan o no un determinado factor de exposición, que se siguen a lo largo del tiempo para determinar la proporción de cada grupo que desarrolla la enfermedad objeto de estudio. Todos los sujetos tienen que estar libres de dicha enfermedad al inicio del estudio, pero ser susceptibles de poder padecerla.

Figura 3.3

Esquema de un estudio de cohortes



Las principales ventajas de los estudios de cohortes es que permiten calcular incidencias y estimar riesgos de forma más fiable, son eficientes para estudiar factores de exposición poco frecuentes, establecen la secuencia real de eventos (de exposición a enfermedad) y controlan mejor la existencia de sesgos. Como inconvenientes, podemos citar que son difíciles de ejecutar, de coste elevado y difíciles de reproducir, que no son factibles para enfermedades raras o con periodo de inducción largo (requerirían un gran tamaño muestral y un largo seguimiento) y que se ven repercutidos por cambios en las técnicas de medición a lo largo del tiempo o por desigualdades en la colaboración y seguimiento de las cohortes.

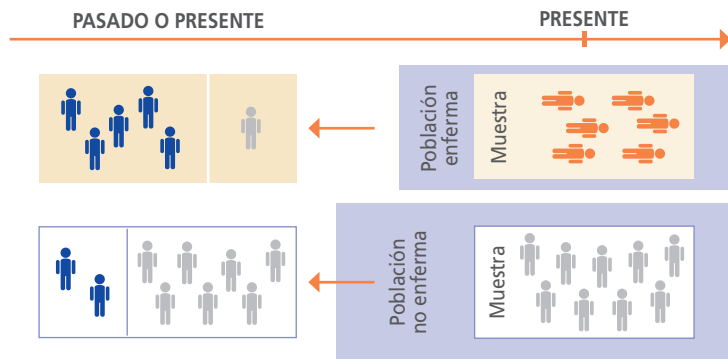
Estudios de casos y controles

Los estudios de casos y controles (figura 3.4) son estudios observacionales, analíticos, en los que el muestreo se hace en relación con la enfermedad o efecto, de direccionalidad retrógrada y temporalidad mixta. Se conocen también como estudios de “casos-testigos” o “retrospectivos”.

En estos estudios se parte de dos grupos de sujetos: unos que tienen la enfermedad (casos) y otros que carecen de ella (controles), e investigamos en ambos si en el pasado existen determinadas características (factores de exposición o riesgo) presumiblemente relacionadas con la enfermedad.

Figura 3.4

Esquema de estudio de casos y controles



Los estudios de casos y controles han proliferado en las últimas décadas por las indudables ventajas que aportan. Permiten estudiar enfermedades poco frecuentes, son de fácil ejecución, son muy eficientes (menor coste por unidad de estudio), tienen una buena capacidad para generar hipótesis y, si están bien diseñados y ejecutados, pueden alcanzar una validez cercana a la de los estudios de cohortes. Sin embargo, hay que tener en cuenta algunos de sus inconvenientes: con estos estudios no podemos calcular medidas de frecuencia (incidencia o prevalencia), ofrecen una peor estimación del riesgo que los estudios de cohortes, al no garantizar la secuencia temporal de la relación

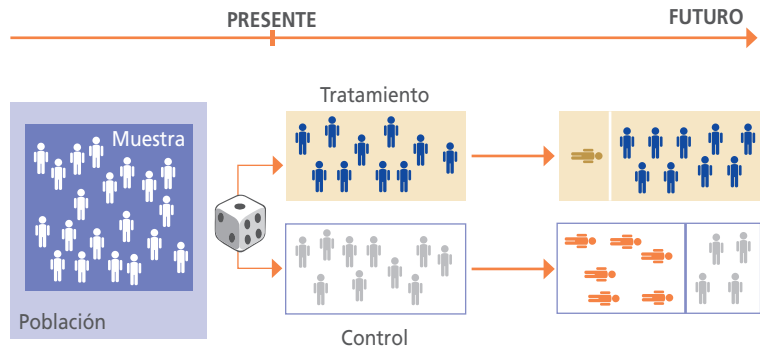
entre el factor de exposición y la enfermedad, presentan dificultades para el manejo de variables tiempo-dependientes (hábitos dietéticos y terapéuticos) y tienen una mayor susceptibilidad a los sesgos.

Ensayos clínicos aleatorizados y controlados

Los ensayos clínicos aleatorizados y controlados (ECAC) son estudios de intervención, analíticos, anterógrados, de temporalidad concurrente y muestreo de cohorte cerrada con control de la exposición (figura 3.5). Requieren la asignación aleatoria de la intervención que introduce el factor de estudio y la existencia de grupo de comparación. Los ECAC permiten evaluar los beneficios de los tratamientos o programas de intervención, ponderando sus riesgos y efectos indeseables. Con ellos podemos controlar los sesgos existentes en otros tipos de estudio y obtener información sólida sobre la relación causal entre una exposición (intervención) y un efecto (inferencia causal).

Figura 3.5

Esquema del ensayo clínico, aleatorizado y controlado



Para analizar sus resultados, debemos contrastar la comparabilidad de los grupos antes de la intervención. Los ECAC tienen importantes ventajas. Son los estudios que aportan mayor control sobre los distintos factores implicados en el diseño, por lo que son repetibles y comparables. A menudo ofrecen el único diseño posible para contestar a determinadas preguntas de investigación. Además, proporcionan la mejor evidencia sobre inferencia causal y generalización de los resultados. Como contrapunto presentan diversos inconvenientes: son estudios costosos en tiempo y dinero, pueden verse limitados por problemas éticos o de factibilidad y, en ocasiones, la aplicación de sus resultados se ve cuestionada porque la muestra de estudio no representa a la población donde quieren aplicarse.

A continuación veamos unos ejemplos con los que practicaremos la elección de un tipo de diseño. Recomendamos para cada supuesto las siguientes preguntas encadenadas (ver figura 3.1):

1. ¿Hacemos alguna intervención? Si la respuesta es sí, el estudio sería experimental. Si la asignación de la intervención experimental es aleatorizada y existe un grupo control, sería un ensayo clínico aleatorizado y controlado (ECAC). Si la respuesta es no, sería un estudio observacional y pasaríamos a la siguiente pregunta.
2. ¿Se comparan dos o más grupos de sujetos? Si la respuesta es sí, sería un diseño analítico; en el caso contrario, descriptivo.
3. Si es un estudio analítico, ¿por qué criterio se clasifican los grupos comparados? Si los grupos se diferencian en función de la presencia/ausencia de una enfermedad o efecto, sería un estudio de casos y controles (por ejemplo: recién nacidos con o sin malformación congénita); si se diferencian en función de la presencia/ausencia de un factor de exposición, sería un estudio de cohortes (por ejemplo: gestantes fumadoras o no fumadoras).
4. Si el estudio es descriptivo y se selecciona una muestra de sujetos en una población, el diseño es un estudio transversal.

Hay otros tipos de diseño no incluidos en el esquema de la figura 3.1, como son los estudios en los que las unidades de análisis son grupos de sujetos, tanto observacionales (por ejemplo: estudios ecológicos) como experimentales (por ejemplo: ensayos comunitarios). También hay otros tipos de diseño más complejos y mixtos que no abordaremos en este libro.



Supuesto 3.1

Queremos saber la frecuencia de desnutrición en pacientes pediátricos que ingresan por problemas no digestivos.



Supuesto 3.2

Queremos saber si los pacientes pediátricos ingresados para intervención quirúrgica programada con desnutrición al ingreso presentan mayor riesgo de infección de la herida quirúrgica que los pacientes sin desnutrición.



Supuesto 3.3

Queremos saber si en prematuros de menos de 32 semanas de gestación es eficaz la administración de probióticos para reducir el riesgo de enterocolitis necrotizante.



Supuesto 3.4

Queremos saber si la administración de aminofilina intravenosa en prematuros de menos de 32 semanas de gestación es un factor de riesgo de desarrollar enterocolitis necrotizante. Para ello comparamos la frecuencia de administración de aminofilina en prematuros con y sin enterocolitis necrotizante.

Pregunta 3.1

Qué tipo de diseño epidemiológico es el más idóneo para la pregunta de investigación del supuesto 3.1: ¿Cuál es la frecuencia de desnutrición en pacientes pediátricos que ingresan por problemas no digestivos?:

- a. Estudio de cohortes.
- b. Estudio transversal.
- c. Ensayo clínico.
- d. Estudio de casos y controles.

Pregunta 3.2

Qué tipo de diseño epidemiológico es el más idóneo para la pregunta de investigación del supuesto 3.2: ¿Influye el estado de nutrición de los pacientes ingresados por cirugía programada en el riesgo de infección de la herida quirúrgica?:

- a. Estudio de cohortes.
- b. Estudio transversal.
- c. Ensayo clínico.
- d. Estudio de casos y controles.

Pregunta 3.3

Qué tipo de diseño epidemiológico es el más idóneo para la pregunta de investigación del supuesto 3.3: ¿Es eficaz la administración de probióticos en prematuros de menos de 32 semanas de gestación para reducir el riesgo de enterocolitis necrotizante?:

- a. Estudio de cohortes.
- b. Estudio transversal.
- c. Ensayo clínico.
- d. Estudio de casos y controles.

Pregunta 3.4

Qué tipo de diseño epidemiológico es el más idóneo para la pregunta de investigación del supuesto 3.4: ¿El uso de aminofilina intravenosa en prematuros de menos de 32 semanas de gestación aumenta el riesgo de enterocolitis necrotizante?:

- a. Estudio de cohortes.
- b. Estudio transversal.
- c. Ensayo clínico.
- d. Estudio de casos y controles.



Muestreo

Es habitual en epidemiología que los estudios no puedan abarcar toda la población disponible, fundamentalmente por razones de factibilidad. Por ello, los estudios se realizan en muestras obtenidas a partir de la población. Para obtener dichas muestras pueden emplearse distintas técnicas de muestreo. Estas técnicas se diferencian en probabilísticas

y no probabilísticas (tabla 3.3). Las técnicas probabilísticas tienen en común el hecho de que todos los sujetos del marco muestral o población de estudio (la población accesible para ser seleccionada) tienen una probabilidad conocida y distinta de cero de ser incluidos en la muestra.

Tabla 3.3 Técnicas de muestreo

Técnicas de muestreo probabilísticas:	Técnicas de muestreo no probabilísticas:
<ul style="list-style-type: none">· Muestreo aleatorio simple.· Muestreo aleatorio estratificado.· Muestreo por conglomerados (mono o polietápico).· Muestreo sistemático.	<ul style="list-style-type: none">· Muestreo consecutivo.· Muestro de voluntarios.· Muestreo por cuotas.

El más sencillo de los muestreos probabilísticos es el **muestreo aleatorio simple**, en el que todos los sujetos del marco muestral tienen la misma probabilidad de ser seleccionados. En este caso, la probabilidad de ser seleccionado es independiente de los sujetos, lo que asegura una mayor representatividad de la muestra y nos permite generalizar los resultados con mayor seguridad, tanto a la población de estudio como a la población diana (aquella a la que se quiere generalizar los resultados y que tiene las características de la población de estudio).

El muestreo aleatorio simple es, como hemos visto, una técnica equiprobabilística, pero en algunas ocasiones interesa que los elementos de muestreo del marco no tengan todos las mismas probabilidades de ser incluidos en la muestra, ya sea por características de la población o de determinadas variables de estudio.

En otras ocasiones, la población puede estar dividida en estratos que interesa tener en cuenta a la hora de realizar el estudio. Pensemos en un muestreo entre distintas zonas geográficas, diferentes en tamaño o densidad de población. Si realizamos un muestreo aleatorio simple, puede ocurrir que las zonas más pequeñas queden poco representadas en la muestra. En estos casos puede realizarse un **muestreo estratificado**, que intenta asegurar que la muestra tenga suficientes elementos de cada estrato como para describir y controlar su influencia en el fenómeno estudiado y realizar estimaciones de efecto ponderadas según la fracción de muestreo empleada en él.

Por último, consideremos el caso de que los sujetos pertenezcan de forma natural a grupos excluyentes. Por ejemplo, niños de diferentes aulas o escuelas o pacientes de diferentes hospitales o servicios. En estos casos puede realizarse un **muestreo por conglomerados**. Se listan los conglomerados y se hace un muestreo aleatorio simple de los conglomerados, en lugar de hacer directamente el muestreo aleatorio de todos los individuos de la población general. Una vez seleccionados los conglomerados, pueden entrar en el estudio todos los elementos del conglomerado (monoetápico) o un subgru-

po elegido dentro de cada conglomerado mediante muestreo (polietápico). Esta técnica de muestreo facilita el trabajo de campo, aunque tiene el problema de que obliga a tener en cuenta la variabilidad intra e intergrupos, por lo que es menos eficiente desde el punto de vista estadístico y suele precisar tamaños muestrales mayores.

Existen técnicas de muestreo probabilísticas más complejas, en las que se utilizan técnicas mixtas. Así, el muestreo estratificado por conglomerados, en el que la selección de conglomerados se hace por estratos (por ejemplo: estratos de colegios públicos y privados, y dentro de cada estrato un muestreo por conglomerados de aulas o colegios). Otra técnica cuasi probabilística es el **muestreo sistemático**, en la que se seleccionan sujetos siguiendo una fracción fija, por ejemplo, múltiplos de 5 (los que ocupan las posiciones 5, 10, 15, etc., de los que acuden a una consulta). Esta técnica permite obtener muestras representativas de la población de estudio, salvo que la posición elegida diferencie las características de la población.

Entre los muestreos no probabilísticos, el más sencillo es el **muestreo consecutivo**; por ejemplo, elegir los pacientes que acuden a nuestra consulta durante un periodo de tiempo. El problema de este tipo de muestreos es que la muestra puede no ser representativa de la población de estudio, lo cual compromete la generalización de las conclusiones obtenidas a partir de nuestros resultados.

En el **muestreo de voluntarios**, los participantes se ofrecen a participar en el estudio. Esto resulta útil en ciertos tipos de ensayos clínicos en fases iniciales y otros estudios que conllevan procedimientos prolongados no justificados por necesidades asistenciales. Podemos establecer criterios de inclusión de manera que garanticemos el perfil apropiado del sujeto investigado, aunque no podemos evitar que este sujeto tenga propiedades diferenciadas de la población general.

Por último, hablaremos del **muestreo por cuotas**. En este muestreo se eligen sujetos de forma que en la muestra se mantenga la misma proporción que en la población. Pensemos, por ejemplo, que algunas características como el peso, el sexo o la raza sean importantes para el estudio. En estos casos la selección se hace de forma que asegure la misma proporción dentro de la muestra que la que se observa en la población. El problema es que, al tener distinta probabilidad de ser seleccionado, podemos favorecer de forma inadvertida alguna otra característica, comprometiendo la representatividad de la muestra.

En cualquier caso, y sea cual sea la técnica de muestreo empleada, es preciso que se defina con precisión la población diana o de referencia sobre la que se extrapolarán los resultados, la población susceptible de pertenecer al estudio y sus criterios de inclusión. Todos estos aspectos, junto con la calidad de los datos recogidos y el número y motivo de los datos perdidos, condicionarán la calidad del estudio y su validez.

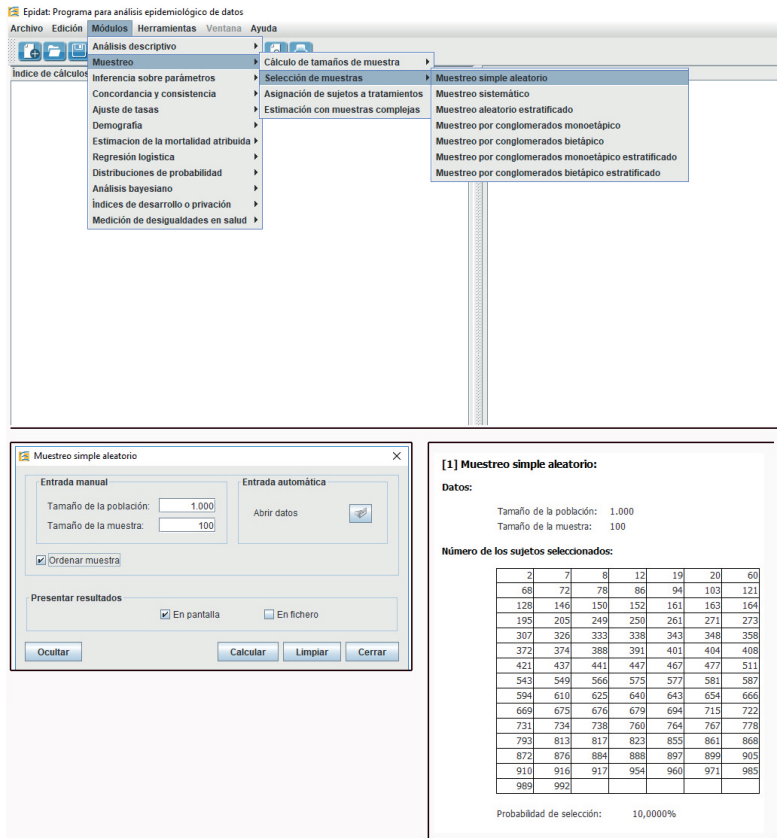
Existen diferentes herramientas informáticas que facilitan la realización de muestreos. La más simple y accesible es la utilización de la función "Aleatorio()" de la hoja de cálculo Excel, que genera números aleatorios, que podemos asignar a cada sujeto de la población y luego ordenarlos hasta alcanzar el tamaño muestral deseado. Para los que

no se manejen bien con hojas de cálculo, se pueden emplear diferentes programas. Pongamos como ejemplo el programa Epidat, *software* gratuito que puede ser descargado libremente (<https://www.sergas.es/Saude-publica/EPIDAT-4-2?idioma=es>), que no requiere instalación (el fichero descargado se puede ejecutar) y que emplearemos en varios capítulos de este libro (figura 3.6).

En el ejemplo de la figura 3.6 podemos ver un muestreo aleatorio simple para una población de 1.000 sujetos, de la que queremos seleccionar una muestra de estudio de 100. Una vez obtenidos los números correspondientes a los sujetos seleccionados, basta cruzar dichos números con los números de lista de los sujetos. El programa ofrece la opción de dar los números elegidos en un fichero Excel, si elegimos la opción "En fichero".

Figura 3.6

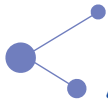
Pantalla de Epidat 4.2, desplegando el menú de muestreo, ventana de muestreo aleatorio simple y salida de resultados



P

Ejercicio práctico 3.1

Invitamos a los lectores a descargar e instalar el programa Epidat 4.2 (<https://www.sergas.es/Saude-publica/EPIDAT-4-2?idioma=es>) y realizar ejercicios de muestreo. Por ejemplo: seleccionar 50 sujetos en una población de 500.



Aleatorización

La aleatorización se emplea para asignar de forma no predecible a los participantes de un ensayo clínico a una de las alternativas de intervención. El objetivo fundamental de la aleatorización es balancear los grupos que intervienen en el ensayo, de forma que sean homogéneos en la distribución de todos aquellos factores, conocidos o desconocidos, que puedan sesgar los resultados del estudio. De esta forma, las diferencias que puedan existir entre los dos grupos serán probablemente debidas al efecto de la intervención en estudio.

Los objetivos de la aleatorización son asignar los tratamientos de forma equilibrada y generar grupos de comparación homogéneos, sin la interferencia voluntaria o involuntaria de los investigadores o de los pacientes. Insistimos en que es muy importante que no pueda predecirse a qué grupo será asignado el próximo participante como elemento fundamental para prevenir sesgos de selección. Este último objetivo se consigue mediante la **ocultación de la secuencia de aleatorización** a los investigadores, que será especialmente importante en estudios en los que no haya enmascaramiento de la intervención (ausencia de ciego). No debemos confundir la ocultación de la aleatorización con el ciego o enmascaramiento de la intervención (por ejemplo: uso de placebo) o de la evaluación del resultado (evaluadores del efecto del estudio ciegos a la intervención que ha recibido cada sujeto).

Lo habitual es elaborar una secuencia o código de aleatorización, que es una lista de los tratamientos a los que se asignarán los sucesivos participantes. Para ello se utiliza una tabla de números aleatorios o una secuencia aleatoria generada por ordenador y se realiza de forma oculta al investigador. Otros procedimientos, cada vez menos frecuentes, como la asignación por fecha de consulta o de nacimiento, por número de historia clínica, procedencia geográfica, etc., no son aceptables para generar el código de aleatorización, ya que pueden ser previsibles.

Existen diversas técnicas de aleatorización, de las que describiremos las cuatro más importantes: aleatorización simple, en bloques, estratificada y con técnicas de minimización.

Aleatorización simple

Se conoce también como aleatorización completa y es la técnica más sencilla y la utilizada con más frecuencia.

Esta técnica asigna cada participante a un grupo de intervención sin tener en cuenta la asignación de los participantes anteriores. Sería el equivalente de tirar una moneda al aire con cada participante. Es importante reconocer que este reparto aleatorio puede producir habitualmente grupos de diferente tamaño, por lo que debemos dudar de la calidad del procedimiento en aquellos ensayos en los que se utilice y los autores presenten grupos iguales, sobre todo si la muestra es de pequeño tamaño.

Este efecto de desigualdad de grupos puede producir pequeños desequilibrios entre las características basales de los dos grupos, especialmente si las muestras son pequeñas o existen varios grupos de intervención, que, dado el reducido tamaño, no serán estadísticamente significativas, aunque sean clínicamente importantes. En estos casos se preferirá el uso de otras técnicas de aleatorización.

Aleatorización por bloques

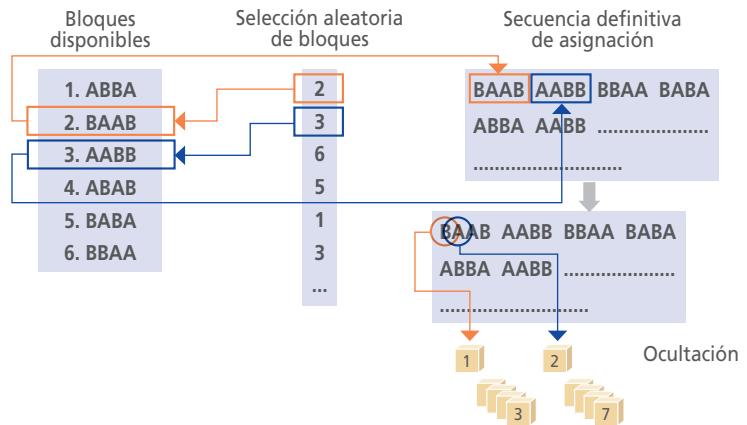
La aleatorización por bloques, también llamada por bloques permutados, tiene como objetivo el asegurar un balance periódico en el número de sujetos asignados a cada grupo de intervención.

Este procedimiento se realiza en tres pasos (figura 3.7):

1. Elegir el tamaño y número de bloques necesarios para el estudio. El tamaño debe ser múltiplo del número de grupos del ensayo.
2. Listar todas las permutaciones posibles en cada bloque.
3. Generar el código de aleatorización según el orden de selección de cada bloque. Después de completar las asignaciones de cada bloque nos aseguramos de tener el mismo número de individuos en cada grupo.

Figura 3.7

Esquema de aleatorización por bloques



Veamos un ejemplo: supongamos que queremos realizar un ensayo con 24 participantes y dos grupos de intervención (A y B). En este caso, un bloque de dos elementos sería predecible, mientras que uno de seis o mayor sería excesivo para nuestro tamaño muestral. Por tanto, elegimos seis bloques con un tamaño de cuatro elementos.

A continuación listamos las posibles permutaciones de dos elementos (el número de grupos) tomados de cuatro en cuatro (el tamaño del bloque): AABB, ABAB, ABBA, BAAB, BABA y BBAA.

Por último, se ordenan los bloques al azar del 1 al 6 y se van asignando los participantes de manera sucesiva a la opción que les corresponda, rellenando los bloques en el orden establecido por azar. De esta forma, cada cuatro participantes asignados se igualará el número de participantes de los dos grupos de intervención. En la práctica, se crean sobres cerrados con el orden de asignación, que quedan a disposición del investigador para realizar la aleatorización de los participantes en el ensayo.

Esta técnica de aleatorización es muy útil cuando el reclutamiento es lento, sobre todo con muestras pequeñas, y cuando hay posibilidad de que el ensayo se interrumpa prematuramente por cuestiones de eficacia o seguridad.

Su principal inconveniente deriva de la previsibilidad de asignación del tercero o cuarto de cada bloque, sobre todo en ensayos no enmascarados en los que se conoce la asignación de los participantes previos.

Aleatorización estratificada

La filosofía de este método es similar a la de los bloques, pero va un paso más allá.

En primer lugar, se establece una serie de grupos según un factor importante que se piense que puede influir en los resultados finales y se divide según puntos de corte basados generalmente en conocimientos de estudios previos.

Seguidamente, se determina el número de estratos dentro de cada nivel del factor escogido. Por ejemplo, en un estudio sobre tratamiento de la hipertensión podría elegirse la edad como primer factor (18-30 años, 31-50 años y más de 50 años) y, dentro de cada grupo, establecer un estrato sobre exposición al tabaco (fumador o no fumador). Quedarían así seis grupos (18-30 y fumador, 18-30 y no fumador, 31-50 y fumador, etc.). Finalmente, se genera una secuencia de aleatorización para cada uno de estos seis grupos, lo que puede hacerse mediante aleatorización simple o, más frecuentemente, mediante asignación por bloques.

El proceso de estratificación puede tener un mayor número de niveles. En nuestro ejemplo, cada uno de los grupos podría haberse estratificado, por ejemplo, por sexo (hombre o mujer), dando lugar a 12 estratos.

Esta técnica consigue mantener un balance en el número de participantes teniendo en cuenta todos los factores que se consideren de interés. Es preciso advertir que, si alguno de los factores considerados en la estratificación tiene un fuerte efecto en el resultado analizado, podrían reducirse las diferencias entre grupos por sobreajuste.

Aleatorización mediante minimización

También llamada aleatorización adaptativa, intenta también que las diferencias entre los distintos grupos sean las mínimas posibles.

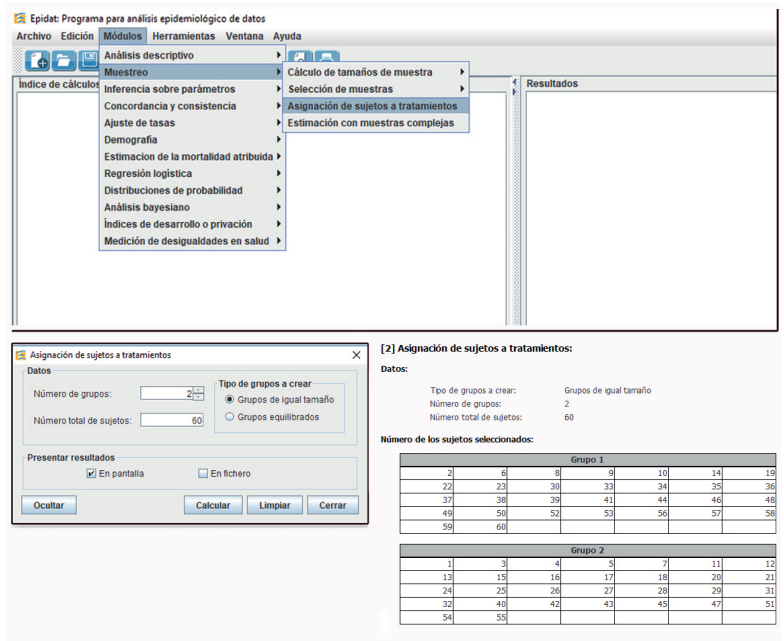
Los primeros sujetos, hasta un número previamente acordado, son aleatorizados de forma simple. A partir de ahí se va ajustando la probabilidad de asignación a cada grupo basándose en los desequilibrios que se hayan podido generar entre los distintos grupos de intervención o entre factores pronósticos que puedan influir en los resultados.

Es un método muy útil cuando existen numerosos factores pronósticos, especialmente con muestras pequeñas. El problema es que requiere recoger información y clasificar los participantes antes de la aleatorización, pero con la facilidad proporcionada por las aplicaciones informáticas y los sistemas de respuesta automática este método va ganando popularidad día a día.

Al igual que vimos con las técnicas de muestreo, contamos con diferentes herramientas informáticas para hacer la aleatorización. Una de ellas es el programa gratuito Epidat 4.2, empleado anteriormente para el muestreo. En la figura 3.8 podemos ver la pantalla de Epidat, con los menús desplegados para la aleatorización simple.

Figura 3.8

Pantalla de Epidat 4.2, con los menús desplegados para la aleatorización simple



Ejercicio práctico 3.2

Invitamos a los lectores a descargar e instalar el programa Epidat 4.2 (<https://www.sergas.es/Saude-publica/EPIDAT-4-2?idioma=es>) y realizar ejercicios de asignación aleatoria. Por ejemplo: asignar a 120 sujetos a recibir tres tipos de intervención (40 por grupo).

Respuestas y comentarios a las preguntas

Comentario a la pregunta 3.1

La respuesta más correcta es la "b", Estudio transversal.

En este estudio vamos a hacer una valoración nutricional al ingreso hospitalario a los pacientes pediátricos que ingresan por un problema no digestivo. No hay intervención (no es experimental), ni hay grupos comparados (no es analítico). Por ello es un estudio observacional, descriptivo y sin direccionalidad. En este caso nos interesa estudiar a una muestra de pacientes ingresados en un momento dado, por ello es un estudio transversal.

Comentario a la pregunta 3.2

La respuesta más correcta es la "a", Estudio de cohortes.

En este estudio vamos a clasificar el estado nutricional al ingreso de los pacientes ingresados para cirugía programada y vamos a valorar si hay diferencias en el riesgo de infección de la herida quirúrgica entre los pacientes normonutridos y los subnutridos. Es un estudio sin intervención (no es experimental), en el que sí hay grupos comparados (es analítico), en este caso en función de la presencia/ausencia de un factor de riesgo (subnutrición), por ello es un estudio de cohortes. Se diferencia del estudio de casos y controles en que los grupos se seleccionan en función de la exposición; en un estudio de casos y controles hubiéramos seleccionado niños con infección de la herida quirúrgica y otro grupo sin infección.

Comentario a la pregunta 3.3

La respuesta más correcta es la "c", Ensayo clínico.

En este estudio realizamos una intervención (probióticos), por lo que es un estudio experimental. Si la intervención es asignada a unos pacientes, siguiendo criterios aleatorios, y comparamos el resultado entre los que la reciben y el grupo que no la recibe, será un ensayo clínico aleatorizado y controlado. El ensayo clínico es un tipo de estudio experimental (se modifica la historia natural con una intervención controlada), analítico (se comparan al menos dos grupos, con intervenciones distintas), anterógrado (se evalúa primero el factor de exposición, la intervención, y luego el efecto, por ejemplo, ingreso hospitalario) y concurrente (el investigador asiste de forma prospectiva a la exposición y al efecto).

Comentario a la pregunta 3.4

La respuesta más correcta es la "d", Estudio de casos y controles.

Este estudio no es un estudio experimental. Aunque los pacientes hayan recibido una intervención, esta intervención no fue indicada por nosotros, sino que fue administrada siguiendo criterios clínicos ajenos a nuestro estudio. Es, por tanto, un estudio observacional. Aunque podríamos haber seleccionado sujetos con y sin aminofilina intravenosa y seguirlos en el tiempo, esta estrategia no es eficiente, porque tendríamos que reclutar y seguir a muchos pacientes para encontrar casos de enterocolitis necrotizante. Por ello seleccionamos sujetos con la enfermedad y por cada enfermo a un control sin enfermedad, mirando hacia atrás (retrógradamente) si recibieron la aminofilina con más o menos frecuencia. El estudio de casos y controles es un estudio observacional, analítico (se comparan dos o más grupos), en el que los grupos se comparan en función de la presencia/ausencia de enfermedad.

BIBLIOGRAFÍA RECOMENDADA

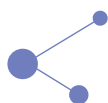
- Argimón JM, Jiménez J. Estudios experimentales I: el ensayo clínico aleatorio. En *Métodos de investigación clínica y epidemiológica*. Madrid: Elsevier España SA; 2004;33-48.
- Contandriopoulos AP, Champagne F, Potvin L, Denis JL, Boyle P. Preparar un proyecto de investigación. 1.ª ed. Barcelona: SG Editores S.A.; 1991.
- Hulley SB, Cummings SR (eds.). *Designing Clinical Research: An Epidemiologic Approach*. Baltimore: Williams and Wilkins; 1988.
- Jeehyoung K, Wonshik S. How to do random allocation (randomization). *Clin Orthop Surg*. 2014;6:103-9.
- Kelsey JL, Thompson WD, Evans AS. *Methods in observational epidemiology*. Nueva York: Oxford University Press; 1986.
- Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic research. Principles and quantitative methods*. Belmont, CA: Lifetime Learning Publications; 1982.
- Kramer MS, Boivin JF. Toward an "unconfounded" classification of epidemiological research design. *J Chron Dis*. 1987;40:683-8.
- Mahmoud S. An overview of randomization and minimization programs for randomized clinical trial. *J Med Signals Sens*. 2011;1:55-61.
- Wang D, Bakhai A. Randomization. En: Wang D, Bakhai A. *Clinical trials. A practical guide to design, analysis, and reporting*. Londres: Remedica; 2006;65-74.
- Xiao L, Huang Q, Yank V, Ma J. An easily accessible web-based minimization random allocation system for clinical trials. *J Med Internet Res*. 2013;15:e139.

4

Tipos de variables Diseño de base de datos

Objetivos docentes:

- Conocer el concepto de variable.
- Conocer los tipos de variables y las escalas de medida.
- Identificar la escala de medida correspondiente a cada variable.
- Saber seleccionar las variables de un estudio.
- Conocer los programas informáticos disponibles para la introducción de los datos de un estudio (bases de datos).
- Saber elegir el tipo de campo para cada tipo de variable.
- Saber diseñar tablas y formularios de bases de datos.
- Conocer los procedimientos de exportación e importación de bases de datos entre programas.



Variables

Recibe el nombre de variable toda característica medida en un estudio, se realice su medición en números (variables cuantitativas: edad o peso) o en categorías (variables cualitativas o categóricas). Se denomina variable porque, aunque podemos prever los valores posibles (espacio muestral), el valor observado en un momento dado en un individuo, grupo, comunidad o población es cambiante. Por ejemplo, podemos prever que los valores posibles del sexo de un nuevo recién nacido pueden ser masculino o femenino, pero hasta que no observamos al nuevo recién nacido no podemos asegurarlo. En este ejemplo, la unidad de observación es el recién nacido y el conjunto de recién nacidos examinados es la muestra de estudio.

Aunque la mayoría de las variables de uso frecuente pueden ser de fácil comprensión, existen otras que no lo son y, por lo tanto, requerirán de una definición operativa. Dicha definición incluirá el rango de valores posibles y la definición de cada uno de ellos. Veamos un ejemplo: evolución de una enfermedad. Podemos definir como valores posibles:

curación (desaparición de síntomas en ausencia de tratamiento), mejoría (desaparición de síntomas con tratamiento), persistencia o empeoramiento (persistencia o empeoramiento de síntomas incluso con tratamiento).

Las variables se diferencian en función de diversas características (tabla 4.1): su papel en la pregunta de investigación, el número de posibles valores y su escala de medición. Es importante conocer las características de las variables, ya que van a condicionar su recogida y análisis. Cuando presentemos los distintos test estadísticos comprenderemos la trascendencia del tipo de variables en la elección de los mismos.

Tabla 4.1 Tipos de variables

Según su papel en la pregunta de investigación:	Según el número de valores posibles:	Según su escala de medición:
<ul style="list-style-type: none">· Dependiente (respuesta).· Independiente (explicativa o exposición).	<ul style="list-style-type: none">· Continuas.· Discretas.	<ul style="list-style-type: none">· Nominal (dicotómica, politómica).· Ordinal.· De intervalos (continuas o de escala).· De razones (continuas o de escala).

Papel en la pregunta de investigación

Podemos diferenciar a las variables en función del papel que desempeñen en nuestra pregunta de investigación. Si estudiamos la asociación entre dos variables, las denominaremos como **“dependiente”** e **“independiente”**, respectivamente. La que recibe el nombre de dependiente (variable resultado o efecto) es aquella que pensamos viene condicionada por la otra variable, que denominamos independiente (variable explicativa o de exposición). Por ejemplo, si queremos estudiar la prevalencia de diabetes en una muestra de población y conocer las variaciones según la edad y el sexo de sus componentes, la presencia de diabetes sería la variable dependiente y la edad y el sexo las independientes. Por el contrario, si estudiamos la influencia de la diabetes sobre los niveles de presión arterial, esta sería la variable dependiente y la diabetes se consideraría variable independiente.

Número de valores posibles de su espacio muestral

Son variables **continuas** las que pueden adoptar un número teóricamente **infinito** de valores a lo largo de un continuo (por ejemplo: talla, peso). Son **discretas** cuando solo son posibles un número finito de valores (por ejemplo: número de hijos de una pareja; esta variable no puede tener fracciones). A la hora de identificar las variables continuas debemos tener en cuenta que, aunque los valores teóricamente posibles sean infinitos, muchas muestras contienen valores finitos, bien por tener escasas observaciones o por falta de precisión en la medición (por ejemplo: medición sin decimales).

Las variables **discretas** pueden medirse en cualquier **escala**, aunque habitualmente se medirán en escalas nominales u ordinales, mientras que las variables **continuas** solo pueden medirse en **escalas** de intervalos y de razones.

Escalas de medición de variables

Uno de los elementos fundamentales de la definición de una variable es el tipo de escala que utilizaremos para medirla. En función de la escala elegida decidiremos su codificación, tratamiento informático y estadístico.

Hay cuatro tipos de escalas de medición, que, ordenadas en orden creciente de potencia, según la proporción de información que contienen, son:

- Nominal.
- Ordinal.
- De intervalos.
- De razones o “ratios”.

Escala nominal

Consta de dos o más categorías **mutuamente excluyentes**. Si solo hay dos, se llama escala nominal **dicotómica**. A cada categoría se le suele asignar un número de código sin significado cuantitativo, lo que facilita su introducción en bases de datos. En cualquier situación, si se usa una codificación propia, debe tenerse claro lo que significa cada código para cada variable.

Veamos algunos ejemplos:

- Sexo: 1, masculino; 2, femenino.
- Fumar: 0, no; 1, sí.
- Estado civil: 1, casado; 2, soltero; 3, viudo; 4, divorciado.
- Procedencia del ingreso: 1, urgencias; 2, consultas; 3, otro hospital.

Dependiendo del programa que va a ser utilizado para el análisis, se prefiere codificar las variables nominales dicotómicas, en las que la presencia de enfermedad o del factor de exposición se suele codificar como uno (1), mientras que la ausencia de enfermedad o de exposición a algún factor se codifica como cero (0) o dos (2). Por ejemplo, el antecedente de hábito tabáquico puede codificarse como 1 y 0 (1, fumador; 0, no fumador) o como 1 y 2 (1, fumador; 2, no fumador). Aunque matemáticamente la presencia/ausencia de una característica se corresponde con la codificación 1-0, es frecuente usar la codificación 1-2, para evitar que variables vacías sean asignadas al 0 por error.

Escala ordinal

Las variables ordinales tienen la cualidad adicional, respecto a la escala nominal, de que **sus categorías están ordenadas** por rango; cada clase posee una misma relación posicional con la siguiente, es decir, la escala muestra situaciones escalonadas. Si se usan números, su única significación está en indicar la posición de las distintas categorías en la serie; sin embargo, no asumen que la distancia del primer escalón al segundo sea la misma que del segundo al tercero. Veamos algunos ejemplos:

- Clase social: 1 (baja), 2 (media), 3 (alta).
- Grados de reflujo vesicoureteral: grados 1, 2, 3, 4.
- Conformidad con una afirmación: 0 (completo desacuerdo), 1 (acuerdo parcial), 2 (acuerdo total).
- Fumar: 0 (no fumador), 1 (fumador leve; < 10/día), 2 (fumador moderado; 10-20/día) y 3 (gran fumador; > 20/día).

Existen escalas que serán mezcla de nominal y ordinal, porque solo algunas categorías estarán ordenadas por rango; esto ocurre en las escalas en las que un valor representa a una categoría inclasificable (por ejemplo: no sabe/no contesta o resultado indeterminado).

Escalas de intervalos

Las escalas de intervalos poseen la cualidad adicional de que **los intervalos entre sus clases son iguales**. Diferencias iguales entre cualquier par de números de la escala indican diferencias también iguales en el atributo sometido a medición. Veamos un ejemplo: la diferencia de temperatura entre una habitación a 22 grados centígrados y otra a 26 es la misma que la existente entre dos a 33 y 37 grados centígrados, respectivamente.

Sin embargo, la razón entre los números de la escala no es necesariamente la misma que la existente entre las cantidades del atributo. Por ejemplo: una habitación a 20 grados no está el doble de caliente que otra a 10. Ello se debe a que el cero de la escala no expresa el valor nulo o ausencia de atributo.

Escalas de razones

Su cualidad adicional es que el **cero sí indica ausencia de atributo**. En consecuencia, la razón entre dos números de la escala es igual a la existente entre las cantidades del atributo medido.

Ejemplos:

- Peso: medido en kilogramos.
- Concentración de glucosa en una muestra: medida en mg/dl.

- Tasa de mortalidad: muertes por 1.000 personas a riesgo.
- Ingresos: medidos en euros.

Las escalas de intervalos y razones se llaman también métricas o dimensionales. Las variables continuas van a medirse con escalas de razones o intervalos, por lo que es habitual que nos refiramos a ellas englobándolas como **escalas continuas**, ya que comparten estrategias de análisis, como la elección del test estadístico. Algunos paquetes estadísticos, como SPSS, las denominan simplemente **“escalas”**.

Cuando en la recogida de una variable continua (escala de razones o intervalos) no podemos recoger valores a partir de un límite (por ejemplo: cargas virales superiores a 100.000), si optamos por mantener esas observaciones en el análisis, la variable se comporta en ese rango como una variable ordinal.

En la tabla 4.2 presentamos un resumen de las características de las diferentes escalas de medida.

Tabla 4.2 Características de las diferentes escalas de medida de variables

Tipo de escala	Operatividad	Requisito
Nominal	Igualdad/desigualdad	Mutuamente excluyentes Exhaustividad
Ordinal	Jerarquía mayor/menor	Escalonamiento
De intervalos	Comparación de intervalos	Unidad constante
De razones	Comparación continua	Cero absoluto

Elección de las escalas

La selección de una escala estará determinada por la definición operativa de la variable que hayamos realizado y por los métodos disponibles para medirla. Para algunas variables solo existirá un tipo de escala elegible, mientras que para la mayoría tendremos varias alternativas.

En general, siempre que sea posible, interesa utilizar en la recogida de datos la escala más precisa, aunque no sea la que posteriormente se utilice en el tratamiento estadístico. Ello nos permitirá adecuar la decisión, una vez conocida la verdadera distribución de nuestros datos. Por ejemplo, siempre será preferible recoger el número de cigarrillos que fuma una persona al día que el dato categorizado como < 10 , $10-20$, > 20 ; la primera estrategia facilitará la búsqueda durante el análisis de puntos de corte distintos a los empleados *a priori*.

Cuando se categorizan algunas variables, como la edad o los ingresos económicos, pasan de ser variables continuas (de intervalos o de razones) a variables ordinales o nominales, según los puntos de corte elegidos.

Si una variable continua se limita en uno de sus extremos con una categoría abierta (por ejemplo: carga viral mayor de 100.000), forzamos que tenga que ser considerada como ordinal.

Las categorías de las variables deben estar clara y operativamente definidas, la escala debe ser exhaustiva (incluir todas las posibilidades) y las clases mutuamente excluyentes (cada observación solo puede ser asignada a una opción). A menudo tendremos que recurrir a categorías complementarias que cubran todo el espectro, como "otras", "no aplicable" o "desconocido", o establecer un orden de prioridades en caso de opciones intermedias.

Las escalas utilizadas para medir variables combinadas se llaman escalas compuestas, pueden basarse en combinaciones de categorías (por ejemplo: combinación de varios criterios diagnósticos simultáneos para asignar el estadio de un tumor), en el uso de fórmulas (por ejemplo: índice de masa corporal = peso/talla²) o de puntuaciones compuestas, obtenidas tras la suma de las asignadas a los distintos "ítems" implicados (por ejemplo: puntuación del test de Apgar).

Para poder diferenciar las escalas de medidas de una variable recomendamos hacer dos preguntas encadenadas:

1. *¿Los valores posibles de la variable son teóricamente infinitos?:*

Si los valores son teóricamente infinitos, la escala de medida será **continua**, bien de intervalos, bien de razones. Aunque la característica operativa de estas escalas es que la unidad de medida es constante, en la práctica la mayoría de las variables continuas cumplen este criterio. No tiene mucho interés diferenciar si la escala es de intervalos o de razones (existe el 0 absoluto), ya que ambas escalas comparten características de análisis.

Si la respuesta es negativa, pasaríamos a la siguiente pregunta.

2. *¿Están los posibles valores ordenados entre sí?:*

Si los posibles valores están ordenados entre sí, la escala es **ordinal**.

Si la respuesta es negativa, la escala es **nominal**.

A continuación haremos una serie de preguntas para que el lector practique la selección de la escala de medida para cada variable. Advertimos al lector que algunas de las variables de las preguntas podrían tener más de una posible escala de medida válida, en función de la precisión con que se recoja.

Pregunta 4.1

¿Qué escala de medida es la más apropiada para la variable "leucocituria"?:

- Nominal dicotómica.
- Nominal politómica.
- Ordinal.
- De intervalos o de razones (continua).

Pregunta 4.2

¿Qué escala de medida es la más apropiada para la variable "hepatomegalia"?:

- a. Nominal dicotómica.
- b. Nominal politómica.
- c. Ordinal.
- d. De intervalos o de razones (continua).

Pregunta 4.3

¿Qué escala de medida es la más apropiada para la variable "edad gestacional"?:

- a. Nominal dicotómica.
- b. Nominal politómica.
- c. Ordinal.
- d. De intervalos o de razones (continua).

Pregunta 4.4

¿Qué escala de medida es la más apropiada para la variable "número de crisis asmáticas en un paciente durante 1 año"?:

- a. Nominal dicotómica.
- b. Nominal politómica.
- c. Ordinal.
- d. De intervalos o de razones (continua).



Diseño de bases de datos

Las bases de datos (**BD**) son conjuntos de variables (datos u observaciones) organizadas bajo una estructura común. Cada elemento de la estructura se denomina CAMPO ("field") de la BD (que recoge la información correspondiente a una variable) y al conjunto de campos de una unidad de estudio lo denominamos REGISTRO. Las BD son, pues, una serie de registros con una estructura común y en las que la información se guarda en campos.

Utilizando un símil de la vida real, un archivador sería la BD, cada una de las fichas que contiene sería un registro y cada uno de los apuntes que aparece en la ficha, un campo.

Los **pasos** sucesivos en el diseño y manejo de una BD son los siguientes:

1. Selección y definición de variables.
2. Diseño de la estructura y creación de la BD (elección del programa).
3. Organización de la BD: creación de variables índice y grupos de variables (tablas) relacionados.
4. Creación de formularios de edición de datos.
5. Introducción de datos.

6. Depuración y transformación de variables.
7. Exportación e importación de datos entre programas.

Selección y definición de variables

La selección de las variables a incluir en una BD va a depender del objetivo del estudio. A la hora de seleccionar las variables tendremos en cuenta la pregunta de investigación y de ella extraeremos las variables fundamentales relacionadas, como la población o problema, la exposición o intervención, la comparación y el resultado. Las variables principales serán la de exposición o intervención y la de resultado o efecto.

Además de las variables principales, tendremos que seleccionar otros grupos de variables que resultarán claves para identificar casos (variables índice), describir la población de estudio (variables universales), controlar factores relacionados con las variables principales (variables relacionadas) y variables de fecha que faciliten el cálculo de medidas de tiempo.

Podemos, pues, diferenciar cinco grupos de variables en todo estudio:

1. **Variables índice:** que servirán para localizar a los sujetos de estudio o para buscarlos (por ejemplo: número de historia, nombre, etc.).
2. **Variables principales:** la variable o variables independientes o explicativas (por ejemplo: en un estudio en el que queremos ver si el tabaco es un factor de riesgo de cáncer, la explicativa o independiente sería fumar sí/no) y la variable o variables dependientes o efecto (en el ejemplo, la dependiente sería cáncer sí/no).
3. **Variables relacionadas:** son las que tienen relación con las principales, que podrían ser de confusión o interacción (en el ejemplo de tabaco y cáncer, la relacionada sería el consumo de alcohol).
4. **Variables universales:** son aquellas que hay que tener en cuenta en cualquier estudio (edad, peso, sexo, nivel social, etc.).
5. **Variables de fecha:** son variables útiles en la mayoría de los estudios (por ejemplo: fecha de nacimiento, fecha de diagnóstico, fecha de ingreso, etc.).

Para elegir las variables interesa buscar estudios similares previamente publicados, porque en ellos veremos las que otros autores han seleccionado y cómo las han definido. En ausencia de información previa, tendremos que diseñarlas y definir las nosotros. Cada variable tiene que ser predefinida en cuanto a tipo de variable (cualitativa o cuantitativa, discreta o continua) y a la escala de medida que sigue (nominal, ordinal o continua).

Las variables cuantitativas se introducirán tal cual en la BD. Sin embargo, las variables cualitativas deben ser codificadas antes de su introducción, preferiblemente empleando números enteros naturales. Esta estrategia facilita la introducción de datos y evita errores. Si tenemos que introducir la variable "sexo" en una BD, lo recomendable es asignar un número a cada categoría; por ejemplo: 1, masculino; 2, femenino. Podemos

entender fácilmente que es más rápido teclear el número "1" que la palabra "masculino". Asimismo, utilizando números evitamos errores en el análisis; al hacer recuentos, el ordenador no identifica como el mismo valor "masculino" y "Masculino".

En general, recomendamos que toda variable que va a ser considerada en el análisis (esto es, todas menos las de identificación: nombre, direcciones, etc.) se recoja con códigos numéricos.

Además de definir cada variable, debemos predefinir los valores no permitidos para cada variable y qué código introduciremos cuando tenemos valores perdidos para una variable. El tratamiento de los valores perdidos (no especificados o *missing*) es importante; es preferible asignar a dicha categoría un código, para variables cualitativas, o un valor imposible, para variables cuantitativas, que dejar en blanco el campo correspondiente. Si dejamos el campo en blanco, no sabremos si nos hemos saltado involuntariamente ese campo hasta que no revisemos el cuaderno de recogida de datos o la historia.

También es importante configurar *a priori* las relaciones entre variables. Por ejemplo, establecer qué variables deben estar disponibles para su introducción en función de los valores de variables previamente recogidas en la BD (por ejemplo: si se contestó "sí" a una pregunta previa; "*branching logic*").

Diseño de la estructura y creación de la BD (elección del programa)

Los datos de la BD se almacenan en tablas de filas y columnas, de forma análoga a como se presenta una hoja de cálculo. Las casillas de la fila superior indican los nombres de los campos (variables) de su columna correspondiente, mientras que cada fila contiene los distintos campos de cada registro. Cada casilla de la tabla es un elemento de información y corresponde a un campo de un registro.

La estructura de la BD va a depender del número de variables y de las agrupaciones y relaciones existentes entre ellas. Estos mismos factores van a condicionar la elección del programa en el que introduciremos los datos. Cuando la magnitud de la información a recoger es pequeña, la estructura de la BD será sencilla, por lo que cualquier programa servirá para crear y editar la BD.

Entre las alternativas para la introducción de datos podemos destacar:

- Hojas de cálculo (por ejemplo: Excel).
- Gestores de BD (por ejemplo: Access, MySQL, Filemaker, etc.).
- Paquetes estadísticos (por ejemplo: SPSS).
- Aplicaciones *online* (por ejemplo: RedCap, Google Forms, etc.).

La forma más sencilla de introducir los datos es recurrir a una **hoja de cálculo** (por ejemplo: Excel). Basta con escribir el nombre de cada variable en la casilla de la fila su-

terior de la hoja y en las filas inferiores los valores de cada variable para cada unidad de estudio. Si el número de variables es elevado, el formato tabla no será cómodo para la introducción de datos. Asimismo, la hoja de cálculo no facilita la introducción de datos con listas de valores codificados o la organización de grupos de variables relacionados. Es preciso señalar que algunas versiones web de Excel tienen la opción de crear formularios de recogida de datos, aunque no tienen la potencialidad de los gestores de BD.

Por ello, si el número de variables es alto o la BD cuenta con variables con listas de valores codificados o agrupaciones de variables relacionadas, recomendamos emplear otros programas o herramientas informáticas específicamente diseñados para ello: los **gestores de BD**. En este libro hablaremos de una herramienta habitualmente disponible en la mayoría de los ordenadores: Access.

En la figura 4.1 podemos observar la tabla de una BD de un estudio sobre fallo de medro en la infancia diseñado en el gestor de bases de datos Access. Esta tabla es similar a la que observaríamos en una hoja de cálculo Excel. En la imagen observamos una de las funciones que tienen los gestores de BD como Access: el despliegue de listas de valores de variables codificadas para su elección (ver la variable "sexo"). Otra ventaja de estos programas es que facilitan el control de los valores introducidos en cada variable, evitando errores. Asimismo, estos programas permiten agrupar y relacionar grupos de variables; así, podemos separar los datos generales de un paciente de los datos repetidos relacionados con episodios o consultas sucesivas. Otra herramienta muy útil, que veremos más adelante, es la elaboración de formularios para facilitar la introducción de datos (ver más adelante la figura 4.2). Finalmente, destacar que la BD de Access puede ser alojada en un servidor y que distintos colaboradores puedan editarla simultáneamente desde terminales conectadas al servidor.

Otra opción para introducir una BD es recurrir directamente a programas o **paquetes estadísticos**, como **SPSS**. Aunque algunos de estos paquetes incorporan gestores específicos, no son tan potentes ni versátiles como los gestores de BD. Además, es habitual que los puntos de introducción de datos sean diferentes de aquellos en los que se realiza el análisis estadístico, no estando estos programas disponibles en todos los ordenadores de

Figura 4.1

Estructura de una BD en formato tabla

ID	G1_Caso_Cc	G2_sexo	G3_fecha_n	G4_fecha_inc	F5_edad_mi	F6_peso_mi	F7_talla_ma
1	1		06/12/2013	10/03/2014	42	54	154
2	2		25/11/2013	26/11/2014	33	80	160
3	1		29/01/2014	13/11/2014	37	57	155
4	2		06/02/2014	01/12/2014	39	58	160
5	1		06/12/2013	09/01/2015	28	57	165
6	2		06/12/2013	05/12/2014	25	54	165
7	1		12/05/2014	14/11/2014	37	60	162
8	2		15/05/2014	03/12/2014	28	80	155
9	1		01/05/2014	05/12/2014	27	66	165
10	2		19/03/2014	05/12/2014	30	68	168
11	1		14/02/2012	03/11/2014	38	74	165
12	1		15/10/2012	28/10/2014	40	55	162
13	1	masculino	07/2013	28/07/2014	37	55	170
14	1	femenino	11/2013	15/07/2014	29	62	156
15	1		18/02/2013	14/08/2014	40	55	165
16	1		22/09/2013	25/06/2014	32	70	162,5
17	1		07/05/2014	23/10/2014	36	59	1,69

una organización. No obstante, tal y como comentamos al hablar de las hojas de cálculo, si la magnitud de la información es pequeña y la recogida de datos ocasional, podríamos optar por introducir la BD directamente en el programa estadístico.

La última de las opciones que se mencionaron son las aplicaciones online, como la utilidad de formularios de Google. Para algunas encuestas sencillas y ocasionales puede ser una buena opción. Existen otras herramientas online, especializadas y potentes, que no vamos a entrar a analizar en este libro.

Organización de la base de datos: ordenación y grupos de variables (tablas) relacionados

Como anteriormente hemos comentado, uno de los grupos de variables de toda BD son las variables índice. Los datos son almacenados habitualmente en el orden en que se introducen, pero este orden no suele ser el más útil para la edición de datos. Aunque los gestores de BD permiten ordenar los registros y hacer búsquedas empleando cualquier variable, lo habitual es que las BD se presenten en forma ordenada, empleando variables índice (por ejemplo: historia, nombre, número del registro en el estudio, etc.). Estas variables serán empleadas también para relacionar grupos de variables de las BD.

En función de la cantidad de información que manejemos (número de registros y/o número de variables), tendremos que decidir si conviene introducir todas las variables en una sola tabla o en varias tablas relacionadas. Por ejemplo, si tenemos datos generales de un paciente y datos repetidos de consultas sucesivas, puede interesar crear tablas separadas para introducir dicha información, que aparecerá asociada, empleando como vínculo alguna de las variables índice. Los gestores de BD nos permiten distribuir la información en varios ficheros o tablas sin que lo apreciemos, gracias a la utilización de tablas o formularios que presentan los datos adecuadamente combinados. Su principal ventaja radica en un óptimo aprovechamiento y organización del espacio. Por el contrario, un fichero único, con todas las variables, puede tener un importante número de campos vacíos, que aumentan innecesariamente el tamaño y entorpecen las tareas de búsqueda. No obstante, fusionar la información en una tabla única puede ser necesario a la hora de realizar el análisis estadístico, por lo que en bases de datos que se construyan para un análisis estadístico circunstancial no convendrá fraccionar los ficheros.

Una vez diseñados los grupos de variables o tablas y la forma de relacionarlas, tendremos que llevar a cabo la creación física de la BD. En una hoja de cálculo bastaría con poner los nombres en la primera fila y debajo los valores para cada paciente o unidad de análisis. Sin embargo, en los gestores de BD y en los programas de estadística debemos ceñirnos a unos tipos de campos específicos en función de las características de cada variable.

Se debe crear una tabla para cada grupo de variables, accediendo a la vista específica ("vista diseño" en Access; "vista de variables" en SPSS). Para cada variable debemos indicar al programa su nombre y tipo de variable (numérica, texto, etc.). También debe-

mos indicar el tamaño del campo (número de espacios disponibles para texto, números enteros y decimales disponibles, etc.), el formato (formato de fecha, tipo de caracteres), los valores permitidos o reglas de validación (comprobar que el valor se encuentra en un rango), la lista de valores para las variables codificadas y el valor o valores para los datos perdidos.

Aunque los programas actuales permiten asignar a las variables nombres sin restricciones, conviene evitar nombres de variable largos, con espacios en blanco o con caracteres especiales (ñ, guiones, acentos, puntos, etc.). Cuando se exportan datos entre programas (por ejemplo: de Access o Excel a SPSS), los nombres de variable complejos pueden originar errores o ser modificados automáticamente. Se recomienda emplear subrayados bajos “_” para separar palabras empleadas en el nombre de las variables (por ejemplo: número_historia). Una regla útil es empezar el nombre de la variable con una letra asociada al apartado del cuestionario de recogida de datos donde está la variable y un número correlativo, seguido de una descripción del contenido de la variable con abreviaturas (por ejemplo: para introducir la variable “peso al ingreso”, que ocupará el tercer lugar de las variables demográficas de una BD, podemos usar el nombre de variable “D_3_Peso_Ing”).

Una vez asignado el nombre de la variable, debemos indicar el tipo y otras características de los campos (variables). En la tabla 4.3 se presentan los tipos de campos disponibles y su correspondencia entre un gestor de BD (Access) y un paquete estadístico (SPSS).

Tabla 4.3 Tipos de campos en Access y SPSS

Access	SPSS
Texto: Debe indicarse el tamaño (número de caracteres).	Cadena: Debe indicarse el tamaño (número de caracteres).
Número*: Debe especificarse el tamaño eligiendo entre: Byte, Decimal, Entero, Simple y Doble. Se recomienda Byte para variables codificadas y Simple para variables numéricas.	Numérico: Debe indicarse el tamaño de cifras enteras y decimales disponibles.
Fecha/hora: Se recomienda fecha corta (dd/mm/aaaa) para fechas.	Fecha: Se recomienda dd.mm.aaaa; A diferencia de otros programas la unidad de fechas son segundos.
Sí/No: No se recomienda su uso.	
Autonumeración: Útil como número de registro automático.	
Otros: Objeto, Memo (texto flexible), Moneda, Hipervínculo.	Otros: Notación, Puntos, Coma, Moneda personalizada.

* **Byte:** enteros entre 0 y 255; **Entero:** enteros entre -32.768 y 32.767; El requisito de almacenamiento es 2 bytes; **Entero largo:** enteros entre -2.147.483.648 y 2.147.483.647; **Simple:** valores numéricos de punto flotante (con decimales) entre $-3,4 \times 10^{38}$ y $3,4 \times 10^{38}$; **Doble:** valores numéricos de punto flotante (con decimales) entre $-1,797 \times 10^{308}$ y $1,797 \times 10^{308}$; **Decimal:** valores numéricos entre $-9,999 \times 10^{27}$ y $9,999 \times 10^{27}$.

Recomendamos emplear los campos Texto o Cadena solo para variables que no van a ser utilizadas en el análisis (por ejemplo: nombre, domicilio). Las variables cualitativas se introducirán como números (en Access, número tamaño byte), salvo que los códigos a emplear contengan caracteres alfanuméricos (por ejemplo: códigos de la Clasificación Internacional de Enfermedades). Algunas variables que no serán usadas en el análisis pueden ser numéricas si su valor es originalmente numérico (por ejemplo: número de historia). El tamaño recomendado en Access para variables intrínsecamente numéricas es número-simple. No se recomienda el tipo de campo Sí/No de Access, ya que por defecto el programa asigna un "No", pudiéndose cometer el error de pasar de largo porque el campo está relleno sin asignar el verdadero valor; además, este tipo de variable no permite introducir un código para valor perdido.

A las variables cualitativas codificadas con números se les asignará un número correlativo (por ejemplo, nivel de estudios: 1, elementales; 2, medios; 3, superiores), destinando un código extra para el valor perdido. Recomendamos el "9" para códigos para variables con menos de nueve categorías, "99" para variables con menos de 99 categorías, etc. Esta regla facilita identificar el código 9 como "no especificado" en variables codificadas con distinto número de opciones.

Creación de formularios de edición de datos

Una de las principales razones para usar gestores de BD para la introducción de datos es la facilidad para crear formularios de recogida de datos. Estos formularios permiten ver los datos en formato "ficha", esto es, en vez de ver la matriz de filas y columnas, solo vemos los campos correspondientes a un registro, pudiendo navegar registro a registro en sucesivas pantallas para ver todos los datos. No debemos olvidar que los datos siguen físicamente almacenados como una tabla, el formulario solo es una "máscara" de acceso a los mismos que facilita el teclado de datos, tanto si el formulario es el único soporte de trabajo como si reproduce la hoja de recogida de datos en papel del proyecto.

Access ofrece la opción de crear un formulario asistido (todas las variables aparecen correlativas en vertical o repartidas por la ventana) a partir de la tabla de datos, aunque la opción de diseño de formulario es la mejor para mejorar el aspecto y funcionalidad del formulario.

En el formulario vemos todos los campos de un registro o paciente, bien en una ventana única que se prolonga verticalmente, o bien, tal y como se ve en la figura 4.2, ventanas solapadas a las que se accede a través de pestañas, en las que se editan las variables del estudio agrupadas. Podemos situar cada campo en la posición que queramos, personalizar los nombres de las variables, introducir instrucciones, etc. Podemos ver, asimismo, en la parte inferior de la ventana el navegador de registros, que nos permite navegar secuencialmente entre registros, y una casilla de búsqueda por si queremos encontrar un paciente concreto y añadir o modificar información.

Figura 4.2

Formulario diseñado en Access de la BD de la figura 4.1

The screenshot shows an Access form window titled 'FalloMedro'. The form is divided into three tabs: 'Datos Generales Familiares', 'Datos Personales', and 'Somatometrias'. The 'Datos Generales Familiares' tab is active, displaying various data entry fields. The fields are organized as follows:

- Top Section:**
 - G1_Caso_Control: 1
 - G2_sexo: 1
 - G3_fecha_nacimiento: 06/12/2013
 - G4_fecha_inclusion: 10/03/2014
- Datos Familiares:**
 - F5_edad_madre: 42
 - F6_peso_madre_kg: 54
 - F7_talla_madre_cm: 154
 - F8_edad_padre: 44
 - F9_peso_padre_kg: 80
 - F10_talla_padre_cm: 171
 - F11_situacion_familiar: 1
 - F12_estudios_madre: 3
 - F13_estudios_padre: 3
 - F14_trabajo_madre: 1
 - F15_trabajo_padre: 1
 - F16_numero_convivien: 4
 - F17_padres_nacidos_fuera_espana: 3
 - F18_numero_hermanos: 1
 - F19_menores_cuatro_e: 1
 - F20_etnia_gitana: 2

At the bottom of the form, there is a status bar showing 'Registro: 1 de 139', navigation icons, and a search field labeled 'Buscar'.

Introducción de datos

Ya hemos mostrado anteriormente las ventajas de los formularios para la recogida de datos. Es por ello que recomendamos usar esta opción para el tecleado de datos, registro a registro, en una o sucesivas sesiones.

Una vez que terminamos de introducir los datos, podemos cerrar el formulario y comprobar cómo la tabla vinculada ha actualizado la información. Es importante saber que, a diferencia de otros programas en los que al salir de los mismos nos preguntan si queremos guardar los cambios, Access almacena continuamente cualquier modificación (de datos), por lo que cualquier cambio se incorpora sin preguntarnos nada al salir. Por ello, recomendamos que al terminar la introducción de datos, provisional o definitivamente, busquemos el fichero correspondiente y hagamos una copia de seguridad.

Depuración y transformación de variables

A menudo, antes de llevar a cabo el tratamiento estadístico de un fichero, es preciso realizar ciertos cambios y comprobaciones destinados a depurar la BD, dejándola libre de errores, o a preparar los datos para procesos estadísticos concretos.

Es frecuente que se cometan errores en la transcripción o tecleo de datos, tanto más abundantes cuanto mayor cantidad de información se maneje y mayor número de personas la manipulen. Existen distintas técnicas encaminadas a reducir dichos errores:

- **Duplicación de ficheros:** cuando la BD es importante, resulta muy útil introducir los datos por duplicado (completo o solapado), comparando *a posteriori* ambas BD.
- **Chequeos lógicos:** con algunas variables se pueden realizar comprobaciones lógicas, buscando valores imposibles, códigos inexistentes, valores fuera de rango (si no se han limitado *a priori* en el diseño de la BD). También podemos comprobar la compatibilidad entre valores de variables (por ejemplo: presión sistólica mayor que la diastólica).

Otras operaciones habituales en proceso de depuración de BD son:

Combinación de ficheros

Cuando la información de un mismo individuo se encuentra separada en varias tablas (unidades de análisis), podemos requerir la fusión de las mismas para su tratamiento estadístico. Por ejemplo, podemos necesitar fusionar los datos evolutivos de revisiones periódicas de un paciente para estudiar cambios en la presencia o ausencia de un determinado síntoma o variaciones cuantitativas de parámetros bioquímicos. La operación requiere, en primer lugar, que las tablas tengan un mismo campo de identificación por el que deben estar ordenados y que no haya coincidencias entre los nombres de los campos de las tablas. Cada programa tiene sus instrucciones concretas para realizar dicha tarea. Otra combinación común de ficheros es la unión de tablas de la misma estructura con información de distintos individuos. Esto suele ser necesario cuando varias personas recaban la información por separado o cuando se ha distribuido el trabajo de introducción de datos al ordenador.

Detección y manejo de valores ausentes (*missing*)

Ya hemos dicho anteriormente que al diseñar una BD debemos definir para cada variable un código de "no especificado" (ausente o perdido). La utilización de dicho código nos permite registrar aquellos campos de los que no dispongamos información, porque no se haya podido obtener o porque se haya perdido. De esta manera diferenciaremos los valores ausentes de los errores en la introducción de datos.

Casi todos los paquetes estadísticos permiten asignar códigos o valores de "no especificado" o "*missing*". Para los campos numéricos es frecuente utilizar "- 1" (para variables métricas > 0), "9" (para variables categóricas con menos de ocho opciones), "99.9" o "999" (para variables métricas con rango inferior a 99 o 998) o cualquier valor comprendido fuera de un rango de valores (por ejemplo: variable categórica, grado de reflujo vesicoureteral; rango, 1-5; código "no especificado", < 1 o > 5).

Es importante conocer lo que el programa que empleamos hace con los campos numéricos que no se rellenan, especialmente si los deja vacíos o les asigna automáticamente el valor "0" o el código de "no especificado".

Esta disciplina en el tratamiento de los "*missing values*" facilita las tareas de depuración y detección de errores en nuestros ficheros. En general, los valores ausentes o perdidos ocasionan problemas en la explotación estadística de las BD, especialmente si empleamos técnicas de análisis multivariante. Las rutinas estadísticas anulan los registros incompletos, de manera que los contrastes estadísticos podrían realizarse con subficheros distintos en función de los campos que hayamos incluido en el análisis. Suele recomendarse que las variables con códigos "no especificados" en una alta proporción de casos sean excluidas, y de igual manera aquellos registros con muchos valores ausentes.

Detección y manejo de valores fuera de rango (*outliers*)

Los “*outliers*” son observaciones aparentemente inconsistentes con el resto de los datos (por ejemplo: en un estudio sobre la concentración de colesterol total en un grupo de escolares un resultado de 660 mg/dl; por ejemplo: edad mayor de 18 años en una muestra pediátrica). Es importante distinguir si los valores fuera de rango corresponden a errores en la determinación o recogida de datos o a valores excepcionalmente raros, pero posibles. Cada una de esos “*outliers*” ha de tener un tratamiento distinto.

En el primer caso parece razonable mantener el valor en el análisis; no obstante, dicha decisión va a influir en las estimaciones que se hagan sobre el conjunto de datos. Es previsible que se modifiquen las medidas de tendencia y dispersión (aumentando el valor medio y la varianza) y que incluso cambie la distribución de los datos, impidiendo, por ejemplo, la asunción de normalidad que requieren algunos contrastes estadísticos. De hecho, la inclusión o no de los valores “*outlier*” puede alterar la significación estadística de nuestros resultados. Por ello, algunos autores sugieren que, en determinadas circunstancias, se haga un tratamiento estadístico duplicado, con y sin “*outliers*”, de manera que podamos evaluar si hay diferencias apreciables. Hemos de ser siempre muy cuidadosos al retirar observaciones fuera de rango, y en todo caso, la decisión debe ser claramente abordada en la metodología y discusión del estudio.

Los paquetes estadísticos disponen de distintas técnicas para la detección de valores fuera de rango en una BD (tablas de frecuencia, histogramas, diagramas de caja, etc.).

Transformación de variables

Las transformaciones de datos se realizan, habitualmente, para crear nuevas variables con un formato más conveniente para el análisis. Las transformaciones más frecuentes son:

- Normalización de variables continuas (por ejemplo: transformación logarítmica).
- Recodificación de variables cualitativas (por ejemplo: agrupación de diagnósticos).
- Cambios de escala o unidades (por ejemplo: transformación de gramos a kilogramos).
- Variables calculadas o combinadas (por ejemplo: estimación de tiempos a partir de fechas, estimación de índice de masa corporal a partir de peso y talla, etc.).
- Categorización de variables continuas (por ejemplo: peso al nacimiento en bajo peso sí/no).

Las transformaciones de variables pueden ser realizadas de forma simultánea durante la introducción de datos o posteriormente, constituyendo un paso previo al análisis. En la segunda opción la operación se hace de una vez con el fichero completo, reduciendo el número de instrucciones necesarias, así como la probabilidad de error.

Una información práctica a la hora de manipular tiempos en las BD es saber que, aunque las fechas se introducen y muestran en formato día/mes/año, las BD las almacenan como números, en referencia al tiempo transcurrido desde una fecha de referencia (por

ejemplo: 1 de enero de 1900). Los distintos programas almacenan ese tiempo con unidades distintas (días, segundos), lo que es importante conocer cuando vayamos a realizar cálculos de tiempo. Otra cuestión práctica es recordar la importancia de introducir los años con los cuatro dígitos, para evitar que el programa asigne automáticamente el siglo (no es lo mismo 1916 que 2016).

Exportación e importación de datos entre programas

La gestión y análisis de BD requiere a menudo de la utilización de diferentes programas. Si queremos aprovechar los mejores recursos de cada uno de ellos tendremos que recurrir a la importación o exportación de datos.

El proceso más habitual es el diseño e introducción de datos en un gestor de BD y la exportación posterior a un paquete estadístico para su análisis. En ocasiones necesitamos incorporar información que obtenemos de otras fuentes, como códigos diagnósticos, escalas de gravedad, resultados de pruebas complementarias, que se pueden incorporar a nuestra BD por procedimientos de importación.

Cuando elijamos un programa para manejar nuestros datos, tenemos que asegurarnos de que tenga la capacidad para importar y exportar datos. Existen distintos formatos de fichero que son empleados y son aceptados como vehículo de transferencia de información. Los más usados son el formato Excel, los formatos delimitados (CSV) y los ficheros DBF (procedente del programa DBase). Tendremos que exportar nuestros datos con este formato para que otros programas los puedan importar. Algunas versiones de SPSS permiten leer directamente datos de Access.

No pretendemos que la información aquí recogida sea suficiente para aprender a diseñar y depurar BD, ese conocimiento solo podrá ser adquirido con la práctica. Hemos querido recopilar los pasos y factores a tener en cuenta.

Respuestas y comentarios a las preguntas

Comentario a la pregunta 4.1

La respuesta más correcta es la "a", Nominal dicotómica.

Aunque la presencia/ausencia de leucocituria corresponde a una escala nominal (valores mutuamente excluyentes), podría ser medida también en una escala ordinal (por ejemplo: número de cruces de leucocituria en la tira urinaria) e incluso en una escala continua de razones e intervalos (por ejemplo: recuento medio de leucocitos por campo). A la hora de elegir la forma de recoger la variable, tenemos que juzgar criterios de factibilidad y eficiencia. Cuando la recogida de la opción más precisa sea factible, será la elección preferida. Si el estudio se centra en la determinación de la leucocituria, el esfuerzo de una recogida más precisa merecerá la pena. Si es una variable secundaria, bastaría con una recogida menos precisa.

Comentario a la pregunta 4.2

La respuesta más correcta es la "a", Nominal dicotómica.

Aunque la presencia/ausencia de hepatomegalia corresponde a una escala nominal (valores mutuamente excluyentes), podría ser medida también en una escala ordinal (por ejemplo: número de traveses de dedo bajo reborde costal) e incluso en una escala continua de razones e intervalos (por ejemplo: dimensiones del hígado medidas mediante ecografía). A la hora de elegir la forma de recoger la variable, tenemos que juzgar criterios de factibilidad y eficiencia. Cuando la recogida de la opción más precisa sea factible, será la elección preferida. Si el estudio se centra en la presencia de hepatomegalia, el esfuerzo de una recogida más precisa merecerá la pena. Si es una variable secundaria, bastaría con la recogida menos precisa.

Comentario a la pregunta 4.3

La respuesta más correcta es la "d", De intervalos o razones (continua).

Aunque estamos acostumbrados a recoger la edad gestacional con valores enteros de semanas y fracciones de semanas desde la fecundación estimada por la fecha de la última regla, la edad gestacional adopta valores teóricamente infinitos si la midiéramos en unidades de tiempo pequeñas (minutos, horas, etc.). Por ello, al igual que otras variables de tiempo, siguen escalas continuas de intervalos o razones. No obstante, podría optarse por categorizar la variable simplificándola como variable ordinal (por ejemplo: menos de 32 semanas, 33-34 semanas, 35-36 semanas, mayor o igual a 37 semanas) o incluso nominal (a término y pretérmino).

Comentario a la pregunta 4.4

La respuesta más correcta es la "c", Ordinal.

El número de crisis asmáticas en 1 año no tiene valores posibles infinitos y la magnitud de cada crisis puede ser diferente en duración e intensidad (unidad no constante), pero sí que están ordenados entre sí (cuatro crisis son más que tres), por ello la variable es ordinal. Podríamos simplificar la variable estableciendo un punto de corte. Por ejemplo, considerar crisis frecuentes a aquellos con más de dos crisis, convirtiéndose la variable en nominal.

BIBLIOGRAFÍA RECOMENDADA

- Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.
- Milton JS. Estadística para biología y ciencias de la Salud. México: McGraw-Hill; 2001.
- Norman GR, Streiner DL. Bioestadística. México: Mosby/Doyma Libros; 1996.
- Woodwark M. Epidemiology study design and data analysis. London: Chapman & Hall/CRC; 1999.

5

Estadística descriptiva

Medidas de frecuencia, riesgo e impacto en epidemiología

Objetivos docentes:

- Conocer los fundamentos de la estadística descriptiva e inferencial.
- Conocer las principales medidas de centralización y dispersión.
- Entender la utilidad de la mediana y los rangos intercuartílicos.
- Conocer los principales métodos de representación gráfica.
- Conocer las principales medidas de frecuencia, riesgo e impacto en epidemiología.
- Saber elegir la medida epidemiológica más apropiada a cada tipo de estudio.
- Saber interpretar las estimaciones de riesgo y las diferencias de porcentajes o medias.



Estadística descriptiva e inferencia estadística

Podemos definir la estadística como el conjunto de métodos científicos ligados con la toma, organización, recopilación, presentación y análisis de datos, tanto para la deducción de conclusiones como para tomar decisiones razonables de acuerdo con tales análisis. Una definición más clarificadora sería: "El arte de tomar decisiones en presencia de incertidumbre".

Dentro de la estadística podemos diferenciar dos tipos: la estadística descriptiva y la inferencial. La estadística descriptiva se limita a describir y analizar un conjunto de datos, limitándose a los mismos, mientras que la estadística inferencial (o inductiva) trata de sacar conclusiones sobre una "población" a partir del análisis de los datos extraídos de un subconjunto de la misma ("muestra"). La inferencia estadística es el objetivo principal de la estadística, ya que es la que nos permite cuantificar nuestra incertidumbre y tomar decisiones cuando la probabilidad de error es mínima. Dentro de la inferencia estadística, podemos diferenciar a su vez dos tipos de estrategias: la estimación de intervalos de confianza de nuestras estimaciones y el contraste de hipótesis en el que confrontamos

dos o más alternativas, cuantificando la probabilidad de que las diferencias entre ellas se deban al azar.

Veamos un ejemplo para ilustrar las distintas funciones de la estadística: imaginemos que en una muestra de 63 niños con jaquecas recurrentes probamos en episodios sucesivos dos analgésicos distintos (A y B). De ellos, 31 manifestaron preferencia por el tratamiento A, 15 por el B y 17 no mostraron preferencia entre ellos. La estadística descriptiva nos permite estimar que un 49,2 % prefirió el tratamiento A, un 23,8 % el B, con una diferencia entre ambos del 25,4 % a favor del tratamiento A.

La inferencia estadística, mediante estimación de intervalos, nos permite avanzar hasta estimar que, con un 95 % de confianza (5 % de error), la diferencia observada se situaría en un intervalo entre el 3,6 y el 47,2 % a favor del tratamiento A.

Por último, el contraste de hipótesis nos permite calcular (test McNemar) que la probabilidad de que las diferencias encontradas sean debidas al azar es 0,018 (1,8 %), por lo que asumimos que el tratamiento A parece mejor que el B.

Los fundamentos y procedimientos seguidos para realizar estos cálculos se verán en capítulos sucesivos de este libro.



Medidas descriptivas

Como hemos visto al inicio de este capítulo, el primer paso del análisis estadístico es el cálculo de medidas descriptivas de la muestra de estudio. Podemos diferenciar varios grupos de medidas: de masa, de tendencia (o centralización) y de dispersión.

Medidas de masa

Son medidas de masa el tamaño muestral (n), el sumatorio y las frecuencias absoluta y relativa.

- Tamaño muestral: el recuento del número de casos.
- Sumatorio ($\sum X_i$): suma aritmética del valor de una variable de todos los casos.
- Frecuencia absoluta: recuento del número de ocurrencias de cada valor de una variable.
- Frecuencia relativa: proporción respecto al total.

En la tabla 5.1 se puede ver la tabla de frecuencias de la variable “número de hijos” de 20 parejas, tal y como lo ofrecen la mayoría de los paquetes estadísticos.

Tabla 5.1 Tablas de frecuencias de la variable "número de hijos" de 20 parejas

$X = \{2, 1, 1, 4, 5, 3, 2, 2, 1, 2, 1, 3, 2, 2, 1, 1, 1, 1, 5, 1\}$	Frecuencia absoluta	Frecuencia relativa	% acumulado
1	9	45 %	45 %
2	6	30 %	75 %
3	2	10 %	85 %
4	1	5 %	90 %
5	2	10 %	100 %
Total (tamaño muestra) = Sumatorio ($\sum x_i$) = 41	20	100 %	

Medidas de posición o tendencia o centralización

Las principales medidas de tendencia son la media, la moda y la mediana. Cada una de ellas describe una característica de los datos que estamos analizando.

- **Media muestral:** si tenemos X_1, \dots, X_n datos, se llama media muestral a la media aritmética de ellos.

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n} = \frac{\sum X_i}{n}$$

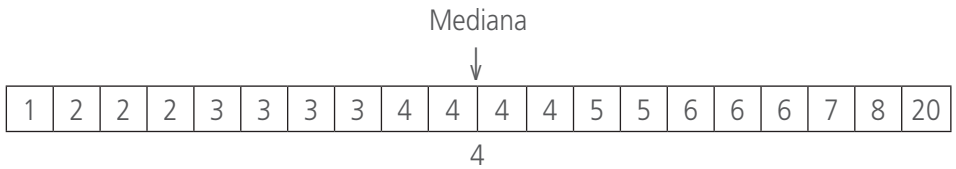
- **Moda muestral:** la moda es el valor que más se repite (puede no existir, y si existe, puede no ser única).
- **Mediana:** si ordenamos de menor a mayor los valores de una variable en una muestra (X_i), la mediana es el valor que está en el medio o la media de los valores que están en el medio (si la muestra es par):

$$\tilde{X} = \left\{ \begin{array}{l} X_{\frac{n+1}{2}} \text{ si } n \text{ impar} \\ \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} \text{ si } n \text{ par} \end{array} \right\} \{X_i\} \text{ ordenados}$$

Veamos de forma gráfica cómo localizar la mediana. En una muestra de 20 pacientes se recogieron las siguientes estancias hospitalarias:

$$X = \{2, 20, 3, 4, 5, 2, 3, 6, 7, 4, 2, 1, 3, 4, 6, 8, 6, 5, 4, 3\}$$

Si las ordenamos, la posición central la ocupan dos "4". La media de ambos es 4; esa es la mediana.



La medida más popular y empleada es la media; sin embargo, cuando los valores de una muestra no siguen una distribución normal, no es una buena medida de tendencia. En estas circunstancias recomendamos utilizar la mediana. Si la media y la mediana son muy diferentes, es poco probable que el valor medio describa la tendencia de los datos (probablemente no tengan una distribución de Gauss o normal), por lo que tendremos que dar la mediana o ambos.

Medidas de dispersión

Las medidas de tendencia no permiten describir los datos de una muestra, porque no informan de cuán alejados está cada uno de los valores respecto el valor central. Las principales medidas de dispersión son el rango, la varianza, la desviación típica, el coeficiente de variación y el rango intercuartílico.

- **Rango:** si ordenamos los valores de menor a mayor, es la diferencia entre los valores extremos (mínimo y máximo):

$$\{X_i\} \text{ ordenados} \quad X_n - X_1 \text{ (Máximo - Mínimo)}$$

- **Varianza:** la varianza es la media de las diferencias al cuadrado entre cada valor y la media. Se elevan al cuadrado para evitar que las diferencias negativas se anulen con las positivas. Se representa con s^2 .

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

- **Cuasivarianza:** la cuasivarianza es una fórmula de estimación corregida de la dispersión de los datos. Aunque la varianza describe fielmente la dispersión de los datos de la muestra, infraestima la dispersión de los datos en la población de la que procede la muestra si esta tiene pequeño tamaño muestral; por ello, la fórmula se corrige disminuyendo su denominador. La varianza que se emplea en inferencia estadística es la cuasivarianza, también conocida como varianza muestral o estimada, o simplemente varianza.

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1}$$

- **Desviación típica muestral o estimada:** como en el cálculo de la varianza las distancias entre cada valor y la media se elevan al cuadrado, la magnitud de la dispersión pierde sentido. Por ello, recurrimos a redimensionar la dispersión haciendo la raíz cuadrada de la varianza. De ahí resulta la desviación típica, representada por "s".

$$s = + \sqrt{s^2} = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}}$$

Al igual que con la varianza, la fórmula no corregida o desviación típica poblacional:

$$s = + \sqrt{s^2} = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n}}$$

Repasemos los pasos para el cálculo de la varianza y la desviación típica. No recomendamos realizar los cálculos manualmente, ya que estas medidas son estimadas automáticamente por cualquier calculadora, hoja de cálculo o paquete estadístico. La intención de conocer sus pasos es para entender su significado:

1. Anotar las observaciones y calcular la media.
2. Calcular la diferencia entre cada valor observado y la media.
3. Estas cifras se elevan al cuadrado y se suman, con lo que se obtiene la suma de los cuadrados.
4. La suma de los cuadrados se divide por una cifra que es igual al número de observaciones menos uno, con lo que se obtiene la media de la suma de los cuadrados o varianza.
5. Extrayendo la raíz cuadrada de la varianza, se obtiene la desviación estándar, que describe la dispersión a los lados de la media.

En la tabla 5.2 podemos ver el cálculo de la varianza y la desviación típica a partir de los datos de longitud de una serie de recién nacidos.

- **Coefficiente de variación:** el coeficiente de variación expresa la dispersión de los datos como medida ajustada. Al dividir la desviación típica por la media, nos indica el porcentaje de dispersión con respecto a la media. Generalmente se expresa como tantos por ciento. Resulta útil para comparar el grado de dispersión de variables con distintas unidades de medida o rango.

$$\text{C.V.} = \frac{s}{\bar{X}} \text{ a veces} \quad \text{C.V.} = \frac{s}{\bar{X}} \cdot 100$$

Tabla 5.2 Cálculo de la varianza y la desviación típica a partir de los datos de longitud de una serie de recién nacidos

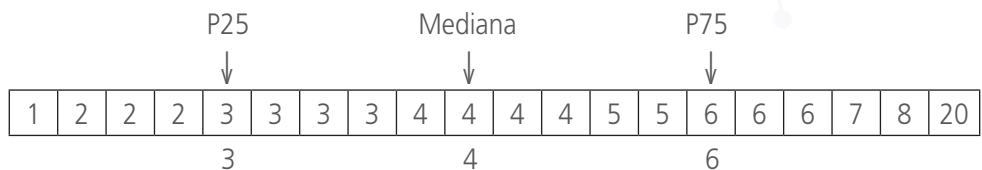
X_i (longitud RN)	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
48,5	-0,9	0,8
51,0	1,6	2,6
51,0	1,6	2,6
50,0	0,6	0,4
47,0	-2,4	5,8
47,5	-1,9	3,6
51,0	1,6	2,6
50,0	0,6	0,4
49,0	-0,4	0,2
50,0	0,6	0,4
51,0	1,6	2,6
51,0	1,6	2,8
48,0	-1,4	2,0
48,0	-1,4	2,0
50,0	0,6	0,4
49,0	-0,4	0,2
50,0	0,6	0,4
47,0	-2,4	5,8
51,0	1,6	2,6
48,0	-1,4	2,0
Sumatorio:	988	39,3
Media (\bar{X}):	49,4	Varianza: 1,96
Tamaño muestral:	20	Varianza muestral: 2,1
		Desviación típica: 1,43

- **Percentiles 25-75:** ordenando los X_i de menor a mayor el valor que deja a su izquierda el 25 % de los casos es el percentil 25 y el que deja por arriba a un 25 % de los casos es el percentil 75. El **rango intercuartílico** es el intervalo entre ambos percentiles. En muestras con distribución no normal es la mejor alternativa a la desviación estándar como medida de dispersión.

Veamos de forma gráfica cómo localizar los percentiles 25 y 75. En una muestra de 20 pacientes se recogieron las siguientes estancias hospitalarias:

$$X = \{2, 20, 3, 4, 5, 2, 3, 6, 7, 4, 2, 1, 3, 4, 6, 8, 6, 5, 4, 3\}$$

Si las ordenamos, los valores que corresponden a los percentiles 25 y 75 son "3" y "6". El rango intercuartílico sería $6 - 3 = 3$.



Métodos de representación gráfica

La representación gráfica de los resultados científicos es una herramienta de gran utilidad, pero a la vez enormemente traicionera. Tan perjudicial es una presentación pobre como excesivamente recargada. Como es más fácil trucar un gráfico que los datos numéricos de un estudio, es frecuente que se manipulen para aparentar resultados más vistosos.

En cuanto a las representaciones gráficas, se pueden hacer una serie de advertencias de interés:

1. En estos diagramas el convenio es que la frecuencia es proporcional al área (es proporcional a la altura solamente si tienen la misma anchura).
2. Para fines comparativos, es mejor usar frecuencias relativas.
3. El número de clases o grupos, cuando los datos no son categóricos, no debe de ser ni muy grande ni muy pequeño. Generalmente, se recomienda un número entre 5 y 20 categorías de igual tamaño. Otros autores recomiendan \sqrt{n} , siendo n = número de datos.
4. Los límites de las clases no deben coincidir con valores posibles de los datos (una cifra decimal más).

Según los datos de que dispongamos, podremos recurrir a distintos diagramas, que describimos a continuación:

- Diagramas de líneas y de áreas: describen recuentos, porcentajes o medidas de tendencia por grupos.
- Diagramas de barras: describen recuentos, porcentajes o medidas de tendencia por grupos (figura 5.1).
- Sectores (quesitos): describen frecuencias relativas de variables cualitativas (figura 5.2).
- Histogramas de frecuencias: describen recuentos de casos de variables cuantitativas por intervalos (figura 5.3).
- *Box-plots* (cajas de puntos): describen medidas de tendencia y dispersión de variables continuas de forma global o por grupos (figura 5.4).
- *Scatter plots* (nubes de puntos): describen la relación entre dos variables cuantitativas (figura 5.5).

Veamos a continuación los más importantes:

Figura 5.1

Diagrama de barras. Distribución de gastroenteritis bacterianas por estaciones

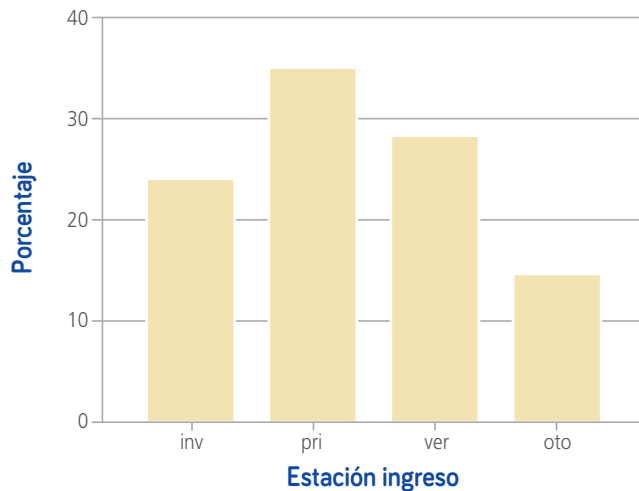


Figura 5.2

Diagrama de sectores. Distribución de gastroenteritis por sexos

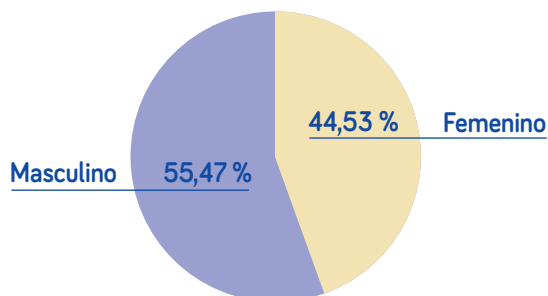
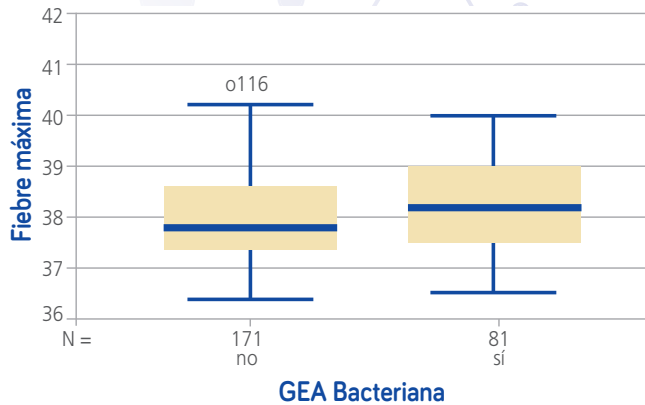


Figura 5.3

Diagramas de cajas.
Fiebre máxima de las gastroenteritis en función de la etiología bacteriana



Mediana: línea gruesa horizontal central; Percentiles 25 y 75: límites inferior y superior de la caja sombreada; Valores extremos (descartando valores atípicos): líneas horizontales finas inferior y superior; Valores atípicos: $< P25 - 1,5 \text{ RIC}$ o $> P75 + 1,5 \text{ RIC}$; RIC: rango intercuartílico; Valores atípicos señalados con su número de orden.

Figura 5.4

Histograma de frecuencias.
Recuento de leucocitos (miles) en pacientes con gastroenteritis

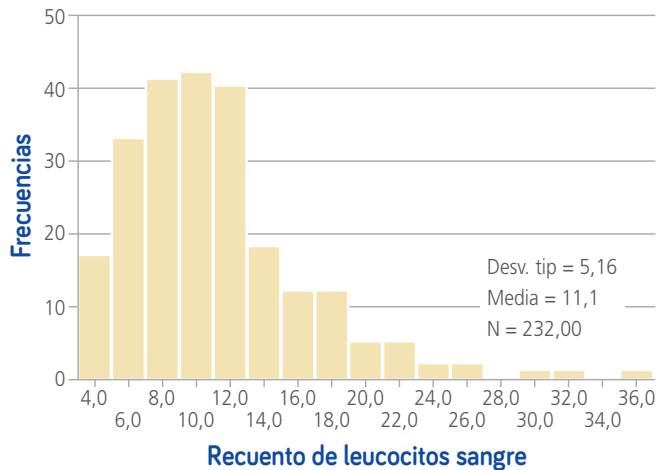
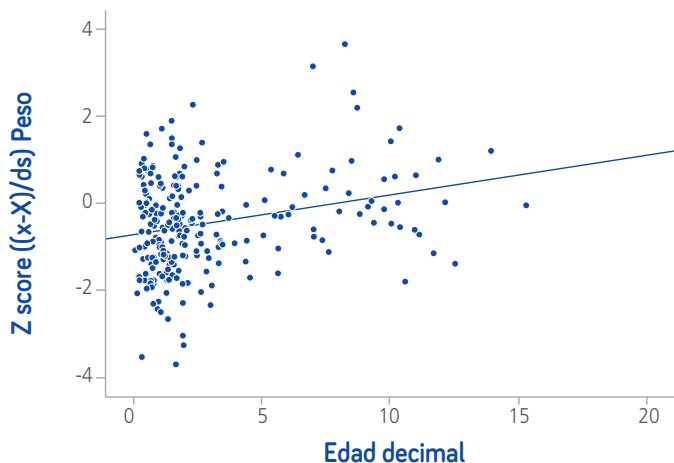


Figura 5.5

Gráfico de dispersión (Scatter Plot).
Correlación edad en años/estado de nutrición (estandarizado)



A continuación proponemos una serie de preguntas para repasar las medidas descriptivas presentadas hasta ahora.

Pregunta 5.1

¿Qué medida de tendencia o centralización presentaría usted para describir los datos de la figura 5.4 (recuentos de leucocitos)?:

- a. Moda.
- b. Media.
- c. Mediana.
- d. Media y mediana.

Pregunta 5.2

¿Qué medida de dispersión presentaría usted para describir los datos de la figura 5.4 (recuentos de leucocitos)?:

- a. Coeficiente de variación.
- b. Desviación típica.
- c. Varianza.
- d. Desviación típica y rango intercuartílico.

Pregunta 5.3

Si tuviera que comparar la dispersión de puntuaciones de dos parámetros biológicos de distribución normal con distinta unidad de medida, ¿qué medida de dispersión emplearía?:

- a. Coeficiente de variación.
- b. Desviación típica.
- c. Varianza.
- d. Rango intercuartílico.



Medidas de frecuencia, riesgo e impacto en epidemiología

Los resultados de los estudios epidemiológicos deben ser expresados en forma de medidas de salud o enfermedad, con las que indicamos frecuencias, diferencias, asociación, riesgo o impacto. La forma de presentar los resultados va a depender del tipo de estudio realizado, pero, sobre todo, de las características de la variable o variables de interés. En función del tipo o tipos de variables implicados podremos recurrir a distintas medidas epidemiológicas.

En epidemiología, el escenario más simple lo constituye el estudio de dos variables discretas dicotómicas. Este escenario corresponde al supuesto habitual de estudio de la asociación entre presencia/ausencia de un determinado factor de exposición y presencia/ausencia de enfermedad. Para analizar esta relación contamos con una serie de medidas de frecuencia, riesgo e impacto.

Otro escenario habitual es el que evalúa la asociación entre una variable discreta y otra continua. Este escenario corresponde a estudios que evalúan la repercusión de un factor

de exposición (por ejemplo: tratamiento versus placebo) sobre un efecto cuantificable en un rango continuo de valores (por ejemplo: presión arterial); en estos estudios la presentación de resultados se basará en las diferencias de medidas de tendencia (media, mediana) y dispersión de la variable continua entre grupos.

Incidencia y prevalencia

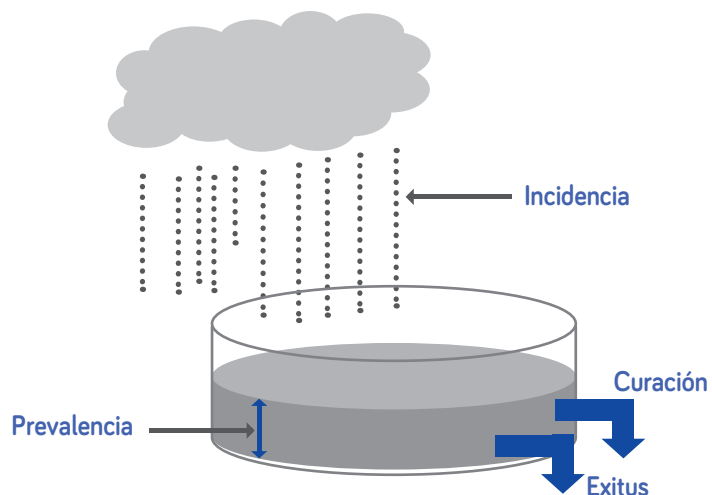
Incidencia y prevalencia son las expresiones de frecuencia de enfermedad más utilizadas en la literatura médica. Es importante distinguir entre ambas. La incidencia es el número de nuevos casos que han ocurrido durante un intervalo de tiempo dividido por el tamaño de la población en riesgo al comienzo del intervalo. Esta información es habitualmente obtenida de estudios de cohortes y expresada en forma de tasas.

La prevalencia es el número de individuos con una enfermedad o característica en un determinado punto en el tiempo dividido por la población en riesgo en ese momento. Se calcula habitualmente a partir de estudios transversales y se expresa como una proporción.

La prevalencia y la incidencia expresan información complementaria (figura 5.6). Una enfermedad con alta incidencia, pero con alta mortalidad o alta curación, tendrá una baja prevalencia en la población. Alternativamente, una enfermedad con una baja incidencia, pero con bajas mortalidad y curación (se cronifica), puede tener una alta prevalencia. El efecto de la mortalidad en la prevalencia puede repercutir en las características de las muestras seleccionadas para participar en un estudio, ya que la población susceptible de entrar en un estudio con casos prevalentes será una selección de pacientes con mejor pronóstico y menor presencia de factores de riesgo que la población identificada en un estudio con casos incidentes.

Figura 5.6

Representación gráfica de la relación entre incidencia y prevalencia



Podemos diferenciar dos tipos de incidencias: la incidencia acumulada y la densidad de incidencia. La incidencia acumulada corresponde al cociente entre el número de casos nuevos y la población en riesgo. Se utiliza cuando no se espera que la enfermedad se repita en un mismo sujeto y la población en riesgo sea estable (por ejemplo: frecuencia de meningitis en 1 año). La densidad de incidencia corresponde al cociente entre el número de casos nuevos y la suma de los periodos de tiempo que es seguida la población en riesgo (por ejemplo: casos-año). Se emplea cuando el riesgo de enfermedad es proporcional al tiempo de seguimiento, cada sujeto puede tener más de un caso y el tiempo de seguimiento de cada sujeto es diferente (por ejemplo: frecuencia de catarros en niños que acuden a una guardería a lo largo de un curso).

En la tabla 5.3 se presenta el cálculo de las medidas de incidencia en distintos tipos de estudios de cohortes.

Medidas de riesgo (asociación)

Mientras la incidencia y prevalencia expresan frecuencia de enfermedad en grupos concretos de pacientes, otras medidas permiten comparar las diferencias de frecuencia entre dos grupos distintos o estimar el incremento de riesgo asociado a una exposición concreta (tabla 5.3). Las principales medidas de riesgo son el riesgo relativo y la *odds ratio* (razón de ventajas). En las tablas 5.3 y 5.4 se presentan las fórmulas de cálculo.

El **riesgo relativo (RR)** se calcula dividiendo la incidencia en el grupo de sujetos expuestos a un determinado factor de riesgo o protección entre la incidencia en el grupo no expuesto. Solo puede ser calculado en estudios de seguimiento y mide la fuerza de la asociación entre exposición y enfermedad. Adopta valores entre 0 e infinito, menores de 1 para factores de protección y mayores para factores de riesgo; un RR de "1" supone el valor nulo (el riesgo en los dos grupos es igual), y cuanto más se aleje el valor de 1, por arriba o por abajo, mayor será la fuerza de la asociación. Un riesgo relativo mayor de 1 indica que el factor estudiado es un factor de riesgo. Un riesgo relativo menor de 1 indica que el factor estudiado es un factor de protección.

Cuando el seguimiento realizado a los sujetos incluidos en un estudio es heterogéneo, en vez de considerar la incidencia acumulada para el cálculo de riesgos, se recurre a la densidad de incidencia, en la que cada sujeto es contabilizado en el denominador en función del tiempo que es seguido. En este caso, el riesgo relativo es estimado a partir de la **razón de densidades de incidencia (RDI)** entre grupos.

En los estudios sin seguimiento longitudinal (estudios de casos y controles), como no es posible calcular la incidencia, no puede calcularse el riesgo relativo. Por ello, para estimar el riesgo se calcula la ***odds ratio (OR)***, que compara la *odds* de exposición (probabilidad de estar expuesto a un factor de riesgo dividida entre su complementario) en el grupo con enfermedad y la *odds* de exposición en el grupo control sin enfermedad, y que se calcula dividiendo ambas *odds*. La interpretación de la OR es

Tabla 5.3 Análisis de los estudios de cohortes. Medidas de riesgo e impacto

• **Estudios con seguimiento acumulativo.**

	Factor de exposición		
	Cohorte expuesta	Cohorte no expuesta	
Enfermos	a	b	N ₁
No enfermos	c	d	N ₀
	M ₁	M ₀	T

Incidencia acumulada en expuestos: $IA_1 = a / M_1$

Incidencia acumulada en no expuestos: $IA_0 = b / M_0$

$$\text{Riesgo relativo} = RR = \frac{IA_1 \text{ expuestos}}{IA_0 \text{ no expuestos}} = \frac{a / M_1}{b / M_0}$$

• **Estudios con datos Personas · tiempo.**

	Factor de exposición		
	Cohorte expuesta	Cohorte no expuesta	
Casos nuevos (enfermos)	a ₁	a ₀	
Total Personas · tiempo	L ₁	L ₀	T

Personas · tiempo = suma de personas por tiempo de seguimiento de cada una

Densidad de incidencia (DI) en expuestos: $DI_1 = a_1 / L_1$

Densidad de incidencia en no expuestos: $DI_0 = a_0 / L_0$

$$\text{Razón de densidades de incidencia} \text{ RDI} = \frac{DI_1 \text{ expuestos}}{DI_0 \text{ no expuestos}} = \frac{a_1 / L_1}{a_0 / L_0}$$

• **Medidas de impacto**

Diferencia de riesgos $DR = IA_1 - IA_0 = \frac{a}{M_1} - \frac{b}{M_0}$

Proporción atribuible (PA) $PA = \frac{IA_1 - IA_0}{IA_1} = 1 - \frac{IA_0}{IA_1} = 1 - \frac{1}{RR} = \frac{RR - 1}{RR}$

PA poblacional (PAP) $PAP = PA \times FE_{\text{casos}} = \frac{RR - 1}{RR} \times \frac{a}{N_1}$

FE_{casos} Fracción de exposición en los casos (proporción de casos expuestos)

Proporción evitable (PE) (si RR es < 1)

$$PE = \frac{IA_0 - IA_1}{IA_0} = 1 - RR$$

Tabla 5.4 Análisis de estudios de casos y controles. *Odds ratio*

Factor exposición (o riesgo)	Efecto (enfermedad)		
	Casos	Controles	
Exposición presente	a	b	N₁
Exposición ausente	c	d	N₀
	M ₁	M ₀	T

Odds de exposición en los casos = a/c

Odds de exposición en los controles = b/d

$$\text{Odds ratio} = \text{OR} = \frac{a/c}{b/d} = \frac{a \times d}{b \times c}$$

• **Estimación de RR a partir de OR**

$$\text{RR} = \frac{\text{OR}}{(1 - R_0) + (R_0 \cdot \text{OR})}$$

Donde R₀ es el riesgo de enfermedad en población no expuesta

similar a la del RR: “1” supone el valor nulo, valores menores de 1 indican disminución del riesgo y mayores de 1 indican aumento del riesgo. Hay que tener en cuenta que solo cuando la enfermedad estudiada es muy poco frecuente el RR y la OR ofrecen valores similares.

Otra medida de riesgo empleada en estudios de supervivencia es el **cociente de riesgos instantáneos**, más conocido por su término en inglés: **hazard ratio**. Se calcula en estudios de supervivencia (tiempo hasta mortalidad u otro evento concreto), que refleja el cociente de riesgos condicionados en los grupos comparados a lo largo de todo el tiempo de seguimiento. La interpretación del *hazard ratio* es similar a la del RR y la OR: el valor 1 indica ausencia de riesgo o asociación, valores mayores de 1, aumento del riesgo, y menores, disminución.

Medidas de impacto

Aunque con las medidas anteriores podemos estimar el riesgo que genera un factor de exposición sobre un efecto o enfermedad, esas medidas no nos informan del impacto que dicha exposición puede originar en el conjunto de casos existentes en una

población. Esta información podemos extraerla de otras medidas, como la diferencia de riesgos o la proporción atribuible (tabla 5.3).

Ambas medidas son estimadoras del efecto absoluto que ejerce la exposición sobre la incidencia de un suceso en el grupo expuesto o en la población total. Se usan para evaluar la importancia clínica o sanitaria de una exposición y nos informan del porcentaje de incidencia que se reduciría si se eliminara el referido factor de exposición. Son, por consiguiente, muy útiles, tanto en la clínica como en salud pública, para cuantificar el posible impacto de diferentes medidas de intervención.

La **diferencia de riesgos (DR)** se calcula restando de la incidencia en el grupo expuesto al factor de riesgo la incidencia en el grupo no expuesto. Adopta valores entre 0 y 1 (entre 0 y 100 si se expresa en tantos por ciento), constituyendo el "0" el valor nulo de ausencia de diferencias. La DR ofrece información independiente del riesgo relativo y puede variar, entre distintos grupos de pacientes, en función del riesgo propio de cada grupo. Así, podemos encontrarnos que factores que muestran un riesgo relativo muy alto apenas presenten modificaciones en la diferencia de riesgos porque el riesgo en la población (al margen de la contribución de dicho factor) sea muy bajo.

La **proporción o fracción atribuible (PA)**, también conocida como riesgo atribuible, fracción etiológica, fracción atribuible o población de riesgo atribuible, se define como la proporción de casos nuevos de la enfermedad, en el grupo de sujetos expuestos, que son atribuibles al factor de riesgo de interés. Se calcula dividiendo la diferencia de riesgos, anteriormente calculada, entre la incidencia en el grupo expuesto. Una extensión de esta medida es la **proporción o fracción atribuible poblacional (PAP)**, que extiende la proporción de casos nuevos a toda la población, esto es, al conjunto de sujetos tanto expuestos como no expuestos. Cuando el factor de riesgo es un factor de prevención, la medida de impacto estimada es la **proporción evitable**, que es la proporción de casos entre los expuestos que se evitarían aplicando el factor de prevención. Estas medidas de impacto no pueden ser directamente estimadas con los datos de un estudio de casos y controles, ya que no nos ofrece estimaciones de incidencia. No obstante, podemos emplear la OR como sustituto del RR para estimarlas, siempre que la enfermedad sea de baja prevalencia (en ese caso, la OR se aproxima al RR; para enfermedades más frecuentes existen fórmulas de estimación del RR a partir de la OR).

Los resultados de los ensayos clínicos suelen reflejar el efecto beneficioso de intervenciones terapéuticas que reducen el riesgo en el grupo expuesto (tabla 5.5). Por ello, la diferencia de riesgos en este supuesto, conocida como **reducción absoluta del riesgo (RAR)**, se calcula en sentido contrario, restando del riesgo en el grupo control el riesgo en el grupo de intervención. Otra medida de impacto aplicable a estos estudios, de gran interés clínico, es el **número necesario a tratar (NNT)**, que corresponde al inverso de la RAR ($1/\text{RAR}$), y que nos informa del número de pacientes que deberían ser tratados con la intervención terapéutica para que un caso se viera beneficiado, evitando un evento desfavorable.

Tabla 5.5 Análisis de los ensayos clínicos

Tasa de respuesta del grupo de intervención:	$P_i = \frac{\text{n}^\circ \text{ eventos grupo intervención}}{\text{total sujetos grupo intervención}}$
Tasa de respuesta del grupo control:	$P_c = \frac{\text{n}^\circ \text{ eventos grupo control}}{\text{total sujetos grupo control}}$
Reducción relativa del riesgo:	$RRR = \frac{P_c - P_i}{P_c}$
Reducción absoluta del riesgo:	$RAR = P_c - P_i$
Número necesario a tratar:	$NNT = \frac{1}{RAR}$

En la tabla 5.6 se presentan las medidas de frecuencia, riesgo e impacto más habituales en función del diseño del estudio, cuando el efecto se mide con una variable nominal dicotómica.

Quando la medida de efecto de un estudio es una variable cuantitativa, la estimación de la **diferencia de medias** entre los grupos de estudio constituye en sí misma una medida de asociación e impacto (por ejemplo: diferencia de medias de hemoglobina glicosilada en dos grupos de diabéticos con pautas de insulina distintas).

Tabla 5.6 Medidas de frecuencia, riesgo e impacto para variables nominales dicotómicas

Diseño	Frecuencia	Riesgo	Impacto
Transversal	Prevalencia	Razón de prevalencias	
Cohortes	Incidencia (acumulada o densidad incidencia)	Riesgo relativo (RR) Razón de densidades de incidencia	Proporción atribuible (PA) Proporción atribuible poblacional (PAP)
Casos-contróles		OR (<i>odds ratio</i>)	PA estimado, PAP estimado
Ensayo clínico	Incidencia	Riesgo relativo (RR) Razón de densidades de incidencia	Reducción absoluta del riesgo (RAR) Reducción relativa del riesgo (RRR) Número necesario a tratar (NNT)

En las figuras 5.7, 5.8 y 5.9 se presentan los cálculos de medidas realizados con una calculadora Excel accesible en goo.gl/ke8sUc. Este fichero puede abrirse *online* o descar-

garse. Dispone de tres hojas accesibles con tres pestañas en la parte inferior para hacer cálculos con estudios de cohortes, casos y controles y ensayos clínicos. Basta introducir los recuentos en la tabla marcada y el programa ofrece los cálculos.

Figura 5.7

Cálculos de medidas epidemiológicas de un estudio de cohortes con la hoja de cálculo accesible en goo.gl/ke8sUc

Estudios de Cohortes					
Introducir los datos en las celdas azules para hacer los cálculos					
Factor riesgo		Enfermedad			Incidencia
		+	-		
+		15	35	50	Expuestos 0,3000
-		15	85	100	No Expuestos 0,1500
		30	120	150	
Riesgo Relativo=		2,00	0,32	SD (ln RR)	
Intervalo de confianza (95%)					
Límite Inferior=		1,07			
Límite Superior=		3,76			
Proporción atribuible=		0,50			
P. atribuible poblacional=		0,25			

Con los datos de la figura 5.7 podemos interpretar que el factor estudiado es un factor de riesgo ($RR > 1$). En el capítulo 6 se explicará el fundamento, cálculo e interpretación del intervalo de confianza del 95 %. Ahora podemos adelantar que tenemos una confianza mayor del 95 % de que el factor estudiado es realmente un factor de riesgo, porque el intervalo de confianza no incluye en su interior el valor nulo para riesgos (el "1"). Además, podemos interpretar que un 50 % del riesgo en los sujetos expuestos al factor de riesgo (proporción atribuible 0,50) y un 25 % del riesgo en toda la población (proporción atribuible poblacional 0,25) se asocia al factor de riesgo.

Figura 5.8

Cálculos de medidas epidemiológicas de un estudio de casos y controles con la hoja de cálculo accesible en goo.gl/ke8sUc

Estudios de Casos y Controles					
Introducir los datos en las celdas azules para hacer los cálculos					
Factor riesgo		Casos	Controles		Razón de Exposición
		+	15		
-		15	21	36	Controles 0,43
		30	30	60	
Odds Ratio=		2,33	0,54	SD (ln OR)	
Intervalo de confianza (95%)					
Límite Inferior=		0,81			
Límite Superior=		6,73			

Con los datos de la figura 5.8 podemos interpretar que el factor de estudio podría ser un factor de riesgo ($OR > 1$); sin embargo, observando el intervalo de confianza, que incluye en su interior el valor nulo para riesgos (el "1"), con la muestra estudiada no tenemos suficiente confianza para declararlo. La calculadora no estima medidas de impacto, ya que para ello necesitamos saber el riesgo en la población no expuesta (aquí solo sabemos el riesgo en los casos y controles no expuestos, que podría no ser el mismo que el existente en la población).

Figura 5.9

Cálculos de medidas epidemiológicas de un ensayo clínico con la hoja de cálculo accesible en goo.gl/ke8sUc

Ensayo Clínico				
Grupo	Resultado de Interés		Total	Riesgo
	Sí	No		
Experimental	20	80	100	0,20
Control	40	60	100	0,40
Total	60	140	200	

	95% CI	
Reducción Absoluta Riesgo (RAR)	20,0%	7,6% a 32,4%
Riesgo Relativo (RR)	0,50	0,32 a 0,79
Reducción Relativa Riesgo (RRR)	50,0%	19,0% a 81,0%
Odds Ratio (OR)	0,38	0,20 a 0,71
Número Necesario a Tratar (NNT)	5	3 a 13

Con los datos de la figura 5.9 podemos interpretar que el tratamiento experimental reduce el riesgo del resultado de interés, con una reducción absoluta del riesgo del 20 %. Vemos que el intervalo de confianza no incluye en su interior el valor nulo para diferencias de proporciones o medias (el "0"). Por ello tenemos una confianza mayor del 95 % de que el tratamiento es eficaz. Además, podemos ver en términos relativos que el riesgo se reduce en un 50 % con respecto al grupo control (reducción relativa del riesgo). Por último, el programa nos calcula el número necesario a tratar, pudiendo interpretar que tenemos que tratar a 5 con el tratamiento experimental para que mejore un paciente con respecto al tratamiento control.

A continuación planteamos una serie de ejercicios prácticos para que realice el lector.



Ejercicio práctico 5.1

Queremos calcular medidas de riesgo e impacto de un estudio de cohortes en el que se han seguido a 100 diabéticos que realizan ejercicio físico intenso y a otros 100 que no. Se ha recogido en un periodo de 1 mes la frecuencia de hipoglucemia nocturna, encontrando que 30 sujetos del grupo con ejercicio intenso presentaron hipoglucemia, frente a 20 del grupo sin ejercicio. Estimen, utilizando sus propios medios o con la calculadora accesible en goo.gl/ke8sUc, las medidas de efecto.



Ejercicio práctico 5.2

Queremos calcular medidas de riesgo e impacto de un estudio de casos y controles que recoge 500 casos de muerte súbita y 1000 controles. Las madres de los casos fumaban 250/500 y de los controles (400/1000). Estimen, utilizando sus propios medios o con la calculadora accesible en goo.gl/ke8sUc, las medidas de efecto más apropiadas.



Ejercicio práctico 5.3

Queremos calcular medidas de eficacia de un ensayo clínico en el que se administró a 50 casos de bronquiolitis atendidos en urgencias suero salino hipertónico nebulizado y a otros 50 casos suero salino fisiológico. Ingresaron 10 casos de los tratados con suero hipertónico y 15 de los tratados con suero fisiológico. Estimen, utilizando sus propios medios o con la calculadora accesible en goo.gl/ke8sUc, las medidas de efecto más apropiadas.

Respuestas y comentarios a las preguntas

Comentario a la pregunta 5.1

La respuesta más correcta es la "d", Media y mediana.

La "c" también sería correcta.

Si observamos la distribución de los datos de recuentos de leucocitos, vemos que no adoptan una forma de campana (gaussiana), sino que los valores altos forman una cola asimétrica. Por ello es de esperar que la media esté sesgada respecto a la mediana. Aunque sería correcto dar solo la mediana, la información es más completa si damos ambas medidas; los lectores podrán interpretar por sí mismos cuál de las dos medidas elegir y no podrán reprocharnos que no demos la media. No obstante, para algunas variables las medias no son medidas de tendencia apropiadas (por ejemplo: media de puntuaciones del test de Apgar al nacimiento).

Comentario a la pregunta 5.2

La respuesta más correcta es la "d", Desviación típica y rango intercuartílico.

Si observamos la distribución de los datos de recuentos de leucocitos, vemos que no adoptan una forma de campana (gaussiana), sino que los valores altos forman una cola asimétrica. Por ello, es de esperar que la desviación típica no exprese convenientemente la dispersión de los datos, porque dicha dispersión será asimétrica a los lados de la media. Por lo tanto, la opción más prudente es calcular también el rango intercuartílico, que reflejará esa asimetría desde la mediana.

Comentario a la pregunta 5.3

La respuesta más correcta es la "a", Coeficiente de variación.

El coeficiente de variación refleja la dispersión de los datos respecto a la media en una escala relativa, como proporción o porcentaje respecto a la media. Por ello, si los dos parámetros tienen unidades de medidas distintas (por ejemplo: microgramos vs. miligramos), solo con ella podemos comparar la dispersión.

BIBLIOGRAFÍA RECOMENDADA

- Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.
- Argimón Pallás JM, Jiménez Villa J. Métodos de investigación clínica y epidemiológica. Barcelona: Elsevier; 2006.
- Milton JS. Estadística para biología y ciencias de la Salud. México: McGraw-Hill; 2001.
- Norman GR, Streiner DL. Bioestadística. México: Mosby/Doyma Libros; 1996.
- Rosner B. Fundamentals of Biostatistics, 7th Edition. Boston: Brooks/Cole, Cengage Learning; 2011.
- Rothman KJ. Epidemiología Moderna. Madrid: Díaz de Santos; 1987.

6

Estadística inferencial

Tamaño muestral e intervalos de confianza

Objetivos docentes:

- Repasar el concepto de inferencia estadística.
- Conocer el concepto de variable aleatoria y las distribuciones de probabilidad.
- Conocer las características de la distribución normal.
- Entender el concepto de error estándar.
- Entender el concepto de intervalo de confianza.
- Conocer el fundamento de la estimación del tamaño muestral.
- Aprender a manejar calculadoras epidemiológicas para calcular intervalos de confianza y estimar tamaños muestrales.



Inferencia estadística

En el capítulo anterior dijimos que la estadística es una herramienta que nos ayuda a tomar decisiones en presencia de incertidumbre. Nuestro objetivo es estimar parámetros de la población a partir de la información obtenida en muestras. Esta estimación va a estar siempre asociada a una mayor o menor incertidumbre, por muy grande que sea el tamaño de la muestra que estudiemos. También diferenciamos entre la estadística descriptiva y la inferencia estadística. Dijimos que la inferencia estadística es el objetivo principal de la estadística, ya que es la que nos permite cuantificar nuestra incertidumbre.

Dentro de la inferencia estadística diferenciamos dos tipos de estrategias:

- **La estimación de intervalos de confianza**, que nos informa del rango de valores entre los que se encontrará el parámetro poblacional a estimar.
- **El contraste de hipótesis**, con el que habitualmente confrontamos dos o más alternativas, cuantificando la probabilidad de que las diferencias entre ellas se deban al azar.

Veamos un ejemplo: en un ensayo clínico se compararon dos tratamientos, A y B, en dos grupos de 100 pacientes, para prevenir recaídas de una enfermedad. En el grupo A recayeron un 20 %, mientras que en el grupo B un 40 %. Podemos hacer una estimación de la eficacia de los tratamientos de dos maneras:

- Mediante estimación de intervalos de confianza: podemos decir que el tratamiento A es un 20 % mejor que el tratamiento B ($40\% - 20\%$) y que estimamos, con un 95 % de confianza (grado de error admitido de un 5 %), que dicha diferencia se encuentra entre un 7,6 y un 32,4 %. Todos los valores del intervalo son favorables al tratamiento A; el intervalo de confianza (no hay valores negativos) ya nos puede servir para tomar decisiones. A la hora de interpretar un intervalo de confianza debemos fijarnos en el valor nulo que implica la ausencia de efecto (el "0" para las diferencias o el "1" para los cocientes); en este caso vemos que el cero no está incluido en el intervalo, luego las diferencias serán estadísticamente significativas.
- Mediante contraste de hipótesis: podemos decir que el tratamiento A es mejor que el B, con una probabilidad de error de 0,002 (estimación realizada mediante comparación de proporciones con aproximación a la normal). Como el error es menor del 5 %, asumimos que el tratamiento A es mejor. Veremos más adelante que el contraste de hipótesis implica la formulación de hipótesis nula y alternativa, sobre las que se estima la probabilidad de error y, en función de la magnitud del error, podemos o no rechazar la hipótesis nula (no hay diferencias) y aceptar la alternativa (el tratamiento A es mejor que el B).



Probabilidad, variables aleatorias y distribuciones de probabilidad

Uno de los objetivos habituales de la investigación es estimar la frecuencia de una determinada característica, por ejemplo, la presencia o no de obesidad. Si nuestro estudio tiene suficientes sujetos (por ejemplo: estudiamos a una muestra representativa de 10.000 adolescentes de nuestra comunidad), la frecuencia relativa de obesidad (por ejemplo: $1.800/10.000 = 0,18$) es la mejor aproximación que tenemos a la probabilidad de estar obeso en nuestro medio. La probabilidad no solo se refiere a variables nominales dicotómicas, también se puede aplicar a cualquier otro tipo de variable; por ejemplo, puede interesarnos saber la probabilidad de que un diabético tenga una glucemia precomida entre 70 y 150 mg/dl.

Podemos definir **probabilidad** como un número entre 0 y 1, asociado con la verosimilitud de que ocurra un suceso. Asumiendo una muestra de tamaño suficientemente alta (infinita), esta probabilidad se estimaría con la frecuencia relativa.

A partir de esta estimación de probabilidad podemos decir que cada uno de nuestros adolescentes tiene una probabilidad de 0,18 (18 %) de ser obeso. Siguiendo terminolo-

gía propia de la teoría de la probabilidad, cada adolescente es un **experimento aleatorio**, del que antes de explorarlo sabemos que puede o no ser obeso (posibles resultados del experimento) y la probabilidad de cualquiera de ellos (obeso, $p = 0,18$), pero hasta que no lo exploramos no sabemos si lo es.

Las **variables** se caracterizan por ser fruto de observaciones repetidas de una misma característica, su información fundamental se puede resumir en un listado de los diversos resultados posibles y en la frecuencia (probabilidad) con que aparece cada uno de ellos. Lo habitual es que la probabilidad de cada uno de los valores de una variable siga algún tipo conocido de **distribución de probabilidad**. Existen muchas distribuciones de probabilidad, probablemente la más conocida sea la distribución normal, que siguen los valores de muchas variables continuas (por ejemplo: la talla de los adolescentes).

Cuando los distintos valores de una variable siguen una distribución de probabilidad la denominamos **variable aleatoria**. Para definir una variable aleatoria necesitamos conocer los valores posibles y la probabilidad de que ocurra cada uno de ellos. Veamos dos ejemplos. La variable aleatoria "curación de un tipo de tumor"; valores posibles: sí/no; distribución de probabilidad: binomial; probabilidad de curación: 0,75 (75 %). La variable aleatoria "longitud de los recién nacidos a término"; valores posibles: cualquier valor entre 40-60 cm; distribución de probabilidad: normal, con media de 50 cm y desviación típica de 2 cm.

Al elegir una variable asumimos un tipo concreto de distribución de probabilidad que condicionará las estimaciones y contrastes de hipótesis que queramos realizar con ella. A cada valor o rango de valores de una distribución de probabilidad le corresponde una probabilidad; esta relación se determina por lo que conocemos como **función de probabilidad**.

Existen múltiples distribuciones de probabilidad. Algunas ya han sido mencionadas, como la distribución binomial, para variables nominales dicotómicas, o la distribución normal, para variables continuas, pero hay muchas otras, como la distribución de Poisson (eventos raros que tienen lugar a lo largo de un periodo de tiempo o espacio), ji cuadrado (que siguen los valores observados y esperados de una tabla de contingencia), t de Student, F de Snedecor (que siguen los cocientes de varianzas), etc.

Invitamos al lector interesado a profundizar su conocimiento de las distribuciones de probabilidad en los textos mencionados en la bibliografía.

Por su interés, abordaremos las características de la distribución normal, ya que nos permitirá presentar la relación existente entre los valores de una distribución y su probabilidad.



La distribución normal

La distribución normal (o gaussiana o acampanada) es la distribución continua más ampliamente utilizada. Constituye la piedra angular de la mayor parte de los métodos

de estimación y contraste de hipótesis, por la asunción de que las variables aleatorias siguen una distribución normal.

La distribución normal puede ser aplicada no solo a variables con distribución esencialmente normal, sino también a variables de distribución no normal adecuadamente transformadas o a estimaciones de parámetros poblacionales de otras variables aleatorias (por ejemplo: proporción de obesos) realizadas a partir de los datos de muestras con un tamaño suficiente (en función del teorema central del límite $n \geq 30$).

La distribución normal viene caracterizada por su simetría, por su valor central (μ , **valor esperado**, media o esperanza matemática) y por su dispersión proporcional a la varianza (σ^2 , **varianza**). Nos basta con conocer la media y la varianza para estimar la probabilidad de cualquier rango de valores. Así, sabemos que a cada lado de la media se sitúa el 50 % de los valores, que entre la media ± 1 unidad de desviación típica (raíz cuadrada de la varianza) se encuentran el 68 % de los valores y que entre la media $\pm 1,96$ veces la desviación típica está el 95 % de los valores (figura 6.1).

Repasemos; la distribución normal es:

- Centrada a los dos lados de la media.
- Simétrica: el 50 % de los valores a cada lado de la media.
- Rango: media ± 1 desviación típica, el 68 % de los valores.
- Rango: media $\pm 1,96$ veces la desviación típica, el 95 % de los valores.

Además de esos valores de referencia, contamos con tablas de referencia detalladas de probabilidades para cada unidad de desviaciones típicas, no solo de forma simétrica, sino además asimétricas (figura 6.2). Así sabemos que el valor que corresponde a la media menos 1,65 veces la desviación típica deja a su izquierda el 5 % de los valores.

Figura 6.1

Distribución de probabilidad normal

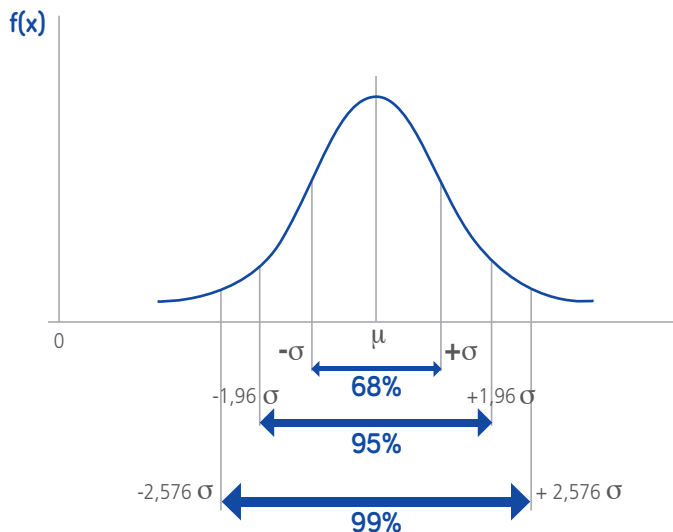
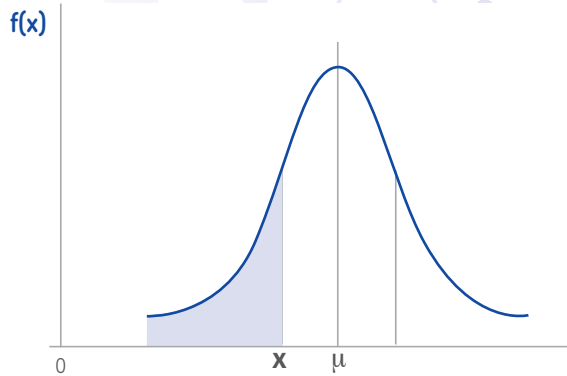


Figura 6.2

Rango de valores de una distribución normal



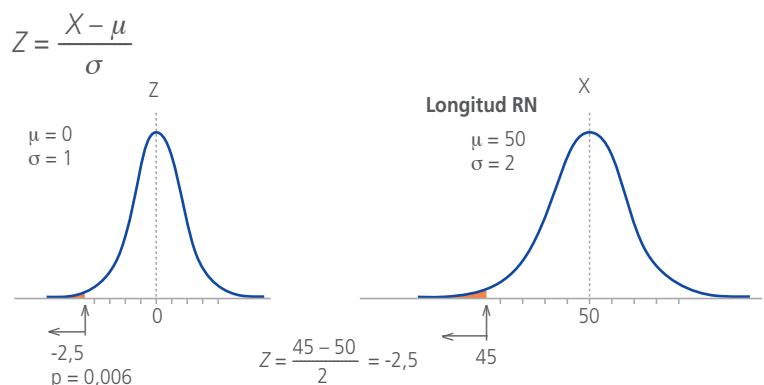
Hemos dicho que, conociendo la media y la desviación típica de una variable de distribución normal, podemos conocer la probabilidad de cualquier rango de valores; por ejemplo, en la distribución de longitudes de recién nacidos (media 50 cm, desviación típica 2 cm), podemos saber la probabilidad de que un recién nacido nazca con menos de 46 cm. Como 46 corresponde a la media menos dos veces (casi 1,96 veces) la desviación típica ($50 - 4$), podemos calcular la probabilidad de medir menos de 46 cm, que es, aproximadamente, 0,025 (un 2,5 %, ya que fuera del intervalo $\pm 1,96$ la desviación típica quedaba el 5% y aquí solo contamos un lado).

El cálculo exacto a partir de funciones de probabilidad requiere operaciones complejas, innecesarias, ya que tanto las hojas de cálculo como los paquetes estadísticos tienen memorizados los valores de referencia que corresponden a cada valor de una distribución de referencia que denominamos normal estandarizada o tipificada (Z), creada a partir de una transformación que consiste en restar a cada valor la media (centrar) y dividirlo por la desviación típica (estandarizar o tipificar). La distribución de referencia tiene una media 0 y una desviación típica 1 (figura 6.3).

$$Z = \frac{X - \mu}{\sigma}$$

Figura 6.3

Distribución normal (longitud de recién nacidos) de media 50 y desviación típica 2 y distribución normal estandarizada (Z) de media 0 y desviación típica 1



Con la centralización y estandarización podemos estimar la probabilidad de cualquier valor que siga una distribución normal, si sabemos la media y la desviación típica. También se puede saber a qué valor concreto le corresponde una probabilidad, buscando el valor de la normal estandarizada (Z) al que le corresponde dicha probabilidad y haciendo la transformación inversa (multiplicar por la desviación típica y sumar la media). Algunos de estos valores ya se han mencionado al describir la distribución normal, ya que son los factores que multiplicaban la desviación típica p (1; 1,65; 1,96).

$$P = \Phi(Z) \implies X = (Z \cdot \sigma) + \mu$$



Estimación por intervalos. El error estándar

Empezamos este capítulo diciendo que la inferencia estadística trata de sacar conclusiones sobre una “población” a partir del análisis de los datos extraídos de un subconjunto de la misma (“muestra”).

La primera estrategia de la inferencia estadística era la estimación por intervalos de confianza. Para ello contamos con las medidas descriptivas de la característica objeto de estudio en nuestra muestra. Habitualmente contaremos con una frecuencia relativa (proporción) para una variable nominal dicotómica o con una media (media muestral) para una variable continua. Nuestro objetivo es estimar las características correspondientes en la población: π (proporción) y μ (media). La mejor estimación que tenemos de ambos parámetros son nuestras medidas descriptivas. Pero como han sido estimadas en muestras, por prudencia, solo podemos decir que los parámetros a estimar, la π (proporción) o la μ (media) poblacionales, tendrán valores cercanos a los que hemos obtenido en nuestra muestra.

Pero, ¿cómo calculamos el intervalo de error alrededor de nuestras medidas muestrales? La aproximación más intuitiva es intentar saber si las proporciones o medias muestrales siguen algún tipo conocido de distribución de probabilidad.

Hagamos un ejercicio teórico a partir de los datos de una amplia muestra de casos reales de partos (12.000), de la que vamos a seleccionar muestras de creciente tamaño muestral. Nuestro objetivo es estimar la proporción de partos distócicos que hay en esa “población”. No olvidemos que incluso con muestras de gran tamaño muestral, las muestras solo son aproximaciones a la población. Asimismo, en la práctica este ejercicio teórico no es posible, ya que en nuestros estudios solo vamos a contar con una única muestra (que podría ser la que aquí vamos a usar como población).

Hemos seleccionado 120 muestras aleatorias de tamaño $n = 10$ (cada muestra 10 partos). En cada una de las 120 muestras estimamos la frecuencia relativa de parto distócico, con lo que obtendremos 120 proporciones, que utilizaremos como si fueran

observaciones individuales, que adoptarán valores entre 0 y 1, para confeccionar un histograma de frecuencias (figura 6.4). Como la proporción de partos distócicos en la muestra que hemos empleado como “población” es 0,15 (15 %), la mayoría de los valores van a ser 0,10 y 0,20; es importante advertir que esa información que sabemos aquí es siempre desconocida. El histograma (elaborado con el programa SPSS) nos calcula automáticamente la media y la desviación típica de esas observaciones, que han pasado a ser consideradas como variables continuas. Nos dice que la media es 0,15 y la desviación típica 0,115.

Ahora vamos a seleccionar 120 muestras aleatorias de tamaño $n = 20$ (cada muestra 20 partos). Haciendo el mismo procedimiento obtenemos un nuevo histograma de frecuencias, con estimaciones de media 0,15 y desviación típica 0,081.

Finalmente vamos a seleccionar 120 muestras aleatorias de tamaño $n = 100$ (cada muestra 100 partos). Haciendo el mismo procedimiento obtenemos un nuevo histograma de frecuencias, con estimaciones de media 0,15 y desviación típica 0,033.

En los histogramas (figura 6.4) podemos ver que, aunque el valor teórico poblacional era 0,15 (en el conjunto de los 12.000 partos), las proporciones de partos distócicos en las muestras individuales son muy variables, aunque las medias de todas las proporciones coinciden.

Podemos ver también cómo a mayor tamaño muestral los valores obtenidos son menos dispersos, lo que se traduce en un histograma más estrecho, con una apariencia cada vez más acampanada, y en una desviación típica progresivamente menor (0,115; 0,081; 0,033). Curiosamente, se puede deducir que el factor que pondera dicha dispersión es el tamaño muestral siguiendo una relación fija: raíz cuadrada del cociente de la proporción (0,15) por su complementario (0,85) y dividido por el tamaño de las muestras (10 o 20 o 100):

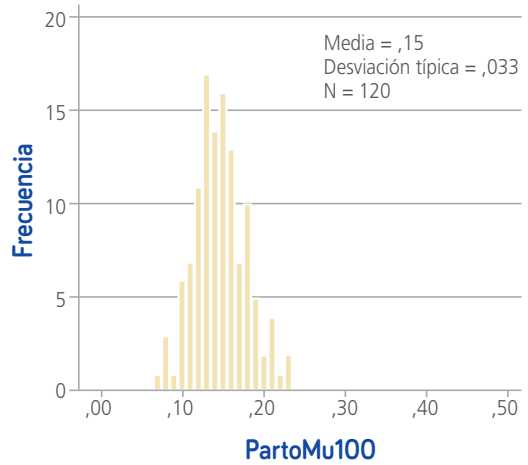
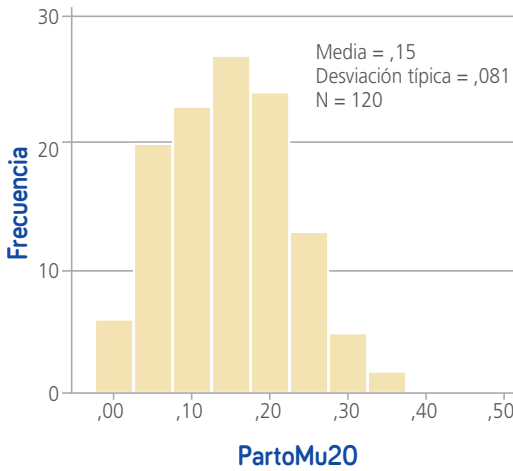
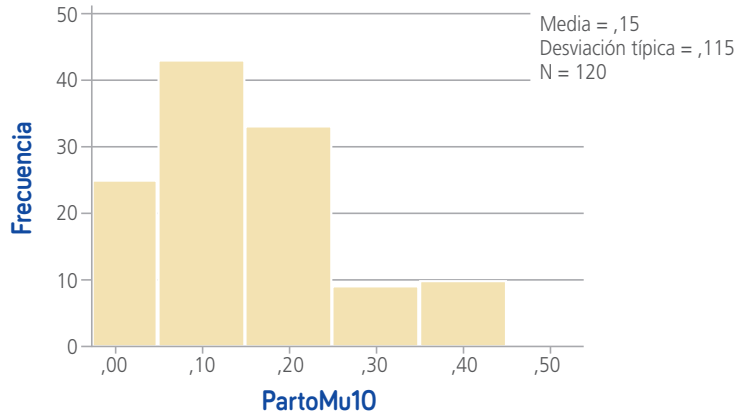
$$\sqrt{\frac{0,15 \cdot 0,85}{10}} = 0,112 \quad \sqrt{\frac{0,15 \cdot 0,85}{20}} = 0,080 \quad \sqrt{\frac{0,15 \cdot 0,85}{100}} = 0,035$$

Esta relación corresponde a lo que conocemos como **error estándar o error típico**, que se comporta como el factor de dispersión (desviación típica) de la distribución de proporciones muestrales, que, como otros parámetros, sigue una distribución normal (según el teorema central del límite para muestras $n \geq 30$). El error estándar se puede calcular a partir de la proporción que hemos encontrado en nuestra muestra, su complementario ($1 - p$) y el tamaño muestral (la aproximación a la normal es válida si el producto $[n \cdot p \cdot 1 - p]$ es mayor que 5).

$$\text{Error Estándar}_{\text{proporción}} = \sqrt{\frac{p \cdot (1 - p)}{n}}$$

Figura 6.4

Histogramas de frecuencia de las proporciones de partos distócicos en 120 muestras de tamaños $n = 10$, $n = 20$ y $n = 100$



Dijimos al comienzo de este apartado que nuestro objetivo era estimar un parámetro poblacional, pero no sabíamos cómo cuantificar nuestra incertidumbre para dar los valores entre los que es verosímil que se encuentre. Ahora ya tenemos los elementos que necesitamos para estimar el parámetro poblacional a partir de las medidas descriptivas de nuestra muestra. Como solo vamos a tener una muestra, asumiremos la proporción observada como el punto central de la estimación (**estimación puntual**). Utilizando esa proporción y el tamaño muestral calcularemos el error estándar, y utilizando las propiedades de la distribución normal podemos cuantificar entre qué rango de valores se encuentra.

Imaginemos que nuestra muestra contaba con 100 partos y obtuvimos una proporción de 0,17 (17 de los 100, 17 %, tuvieron parto distócico). Lo más probable es que hubiéramos encontrado un 15 %, pero por azar hemos encontrado otro resultado cercano. Asumimos ese 0,17 como proporción y calculamos el error estándar con la fórmula:

$$\text{Error Estándar}_{\text{proporción}} = \sqrt{\frac{0,17 \cdot (1 - 0,17)}{100}} = 0,0375$$

Cuando examinábamos la distribución de probabilidad normal vimos que en el rango entre 1,96 veces por debajo y por arriba de la media se encontraban el 95 % de los valores. Usemos esta propiedad para estimar el rango de valores entre los que tengo una confianza del 95 % de que se encuentre la proporción poblacional. Este cálculo corresponde a:

$$p \pm Z_{1-\alpha/2} \cdot \text{error estándar} \Rightarrow \text{Para IC 95\%} \Rightarrow 0,17 \pm 1,96 \cdot 0,0375$$

IC 95 % (proporción poblacional " π "): 0,0965 a 0,2435 (9,65 % a 24,35 %).

Habitualmente este rango se conoce como **intervalo de confianza** al 95 %. Se puede interpretar diciendo que tenemos un 95 % de confianza de que la proporción poblacional se encuentre entre un 9,65 y un 24,35 % (con un error menor del 5 %). Aunque esta es la interpretación más sencilla, la interpretación real corresponde a que, si hiciéramos 100 estudios con un tamaño muestral similar al nuestro, el verdadero parámetro poblacional estaría incluido en 95 de los 100 intervalos de confianza estimados.

Al igual que hemos deducido la fórmula del error estándar para la estimación de proporciones, el lector interesado puede consultar en textos especializados la deducción de las fórmulas de error estándar de otros parámetros. Lo importante es entender que ese error estándar es el factor de dispersión del conjunto teórico de estimaciones de dicho parámetro en sucesivas muestras (figura 6.5).

El procedimiento siempre es el mismo: a) calculamos la medida descriptiva o la medida de frecuencia, riesgo o impacto de nuestra muestra; b) estimamos a partir de dichos datos el error estándar; y c) usamos dicho error estándar y el valor Z correspondiente (para un intervalo de confianza del 95 %: 1,96) para estimar el intervalo de confianza.

Cuando manejemos variables continuas, definidas por sus medias y desviaciones típicas, no debemos confundir la desviación típica de nuestra muestra, que describe nuestros datos, con el error estándar, que nos permite estimar el rango de error en la estimación de la media poblacional.

Aunque el cálculo para proporciones, medias, diferencias de proporciones o diferencias de medias resulta sencillo, no recomendamos hacer dichos cálculos de forma manual. Para ello tenemos calculadoras epidemiológicas que realizan los cálculos fácilmente o, si tenemos los datos tabulados, podemos recurrir a paquetes estadísticos.

En la figura 6.6 se presenta un ejemplo de cálculo del intervalo de confianza para una proporción y en la figura 6.7 para una diferencia de proporciones. Ambos cálculos se han realizado con el programa Epidat 4.2, *software* gratuito que puede ser descargado libremente (<https://www.sergas.es/Saude-publica/EPIDAT-4-2?idioma=es>) y que no requiere instalación (el fichero descargado se puede ejecutar, aunque requiere tener instalada en el ordenador una versión actualizada de Java).

Figura 6.5

Fórmulas de estimación de errores estándar de algunos parámetros poblacionales

$$EE \text{ proporción} = \sqrt{\frac{p \cdot (1-p)}{n}}$$

$$EE \text{ diferencia proporciones} = \sqrt{\frac{\hat{p}_1 (1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2 (1 - \hat{p}_2)}{n_2}}$$

$$EE \text{ media} = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$

$$EE \text{ diferencia medias} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

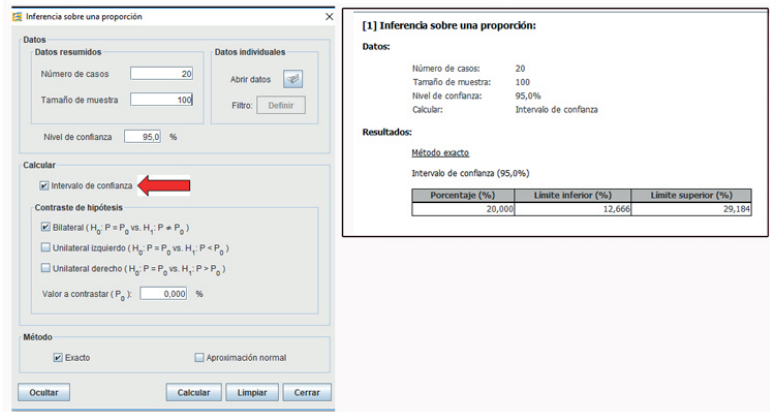
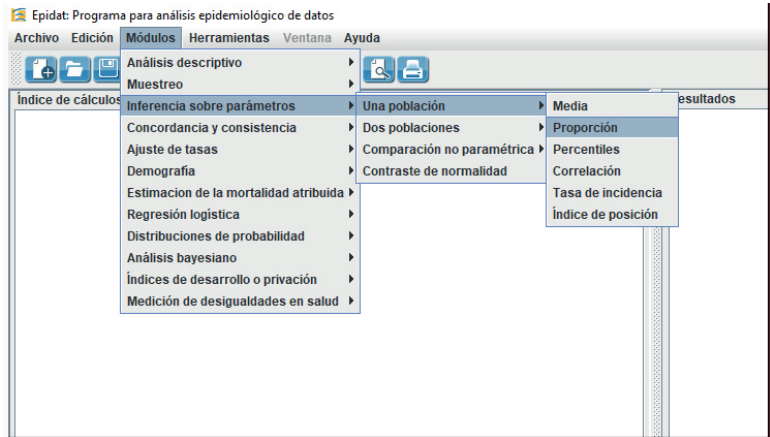
$$EE \text{ InOR} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

a	b
c	d

N: tamaño muestral; p: proporción; InOR: logaritmo neperiano de odds ratio; s: desviación típica muestral; σ : desviación típica poblacional (se estima a partir de s).

Figura 6.6

Cálculo de un intervalo de confianza para una proporción con Epidat 4.2. Se presenta el menú desplegado en el que se accede a la ventana correspondiente (en "Calcular" debe señalarse exclusivamente la opción de intervalo de confianza) y los resultados



El intervalo de confianza de la proporción 20 % (20/100) de la figura 6.6 se sitúa entre el 12,66 y el 29,18 %. Esto es, tenemos un 95 % de confianza de que la proporción poblacional se encuentre en este intervalo.

Figura 6.7

Cálculo de un intervalo de confianza para una diferencia de proporciones con Epidat 4.2. Se presenta el menú desplegado en el que se accede a la ventana correspondiente (en "Calcular" debe señalarse exclusivamente la opción de intervalo de confianza) y los resultados

[2] Inferencia. Comparación de proporciones independientes:

Datos:

	Muestra 1	Muestra 2
Número de casos	40	20
Tamaño de muestra	100	100

Nivel de confianza: 95,0%
Calcular: Intervalo de confianza

Resultados:

Población	Porcentaje (%)
1	40,000
2	20,000

Intervalo de confianza (95,0%)

Diferencia de proporciones	Límite inferior	Límite superior
	0,200	0,076 0,324

El intervalo de confianza de la diferencia de proporciones (40 % – 20 % = 20 %) de la figura 6.7 se sitúa entre 0,076 (7,6 %) y 0,324 (32,4 %). Esto es, tenemos un 95 % de confianza de que la diferencia de proporciones poblacional se encuentre en este intervalo. Aunque esto supone adelantarnos a lo que se abordará en el capítulo siguiente sobre contraste de hipótesis, la información que nos da el intervalo de confianza nos permite decir que la muestra 1 tiene mayor porcentaje que la muestra 2 (imaginemos que fuera éxito de tratamiento), porque en el intervalo de confianza no está incluido el 0 % (valor nulo que indica ausencia de diferencias), con un riesgo de error menor del 5 % ($p < 0,05$).



Ejercicio práctico 6.1

Invitamos a los lectores a descargar e instalar el programa Epidat 4.2 (<https://www.sergas.es/Saude-publica/EPIDAT-4-2?idioma=es>) y realizar ejercicios de cálculo de intervalos de confianza. Damos varios supuestos:

- a) Hemos estudiado a 200 lactantes sanos que acuden a revisión de los 6 meses y, de ellos, 60 mantenían lactancia materna. Calcule la prevalencia de lactancia materna a esa edad con su intervalo de confianza del 95 %.
- b) Hemos asignado a 70 recién nacidos con distrés respiratorio por aspiración meconial a recibir presión positiva continua en las vías respiratorias (CPAP) nasal y a otros 70 a recibir oxígeno en campana. La medida de efecto principal fue la necesidad de ventilación mecánica invasiva. Solo dos de los tratados con CPAP nasal requirieron ventilación invasiva, por 17 de los otros. Calcule la diferencia de proporciones con su intervalo de confianza del 95 %. Trate de interpretar este resultado; ¿podemos decir que la CPAP disminuyó la necesidad de ventilación invasiva?



Estimación del tamaño muestral

Hemos visto hasta ahora cómo cuantificamos nuestra incertidumbre para estimar parámetros poblacionales y el papel que tiene en ello el error estándar. Para calcular el error estándar utilizábamos distintas medidas descriptivas extraídas de la muestra de estudio. Podemos comprobar que en todas las fórmulas de los distintos errores estándar (figura 6.5) se encuentra el tamaño de la muestra o muestras de estudio. El camino recorrido hasta el momento ha sido del tamaño muestral a la precisión de nuestro intervalo.

Utilizando las mismas fórmulas de error estándar podemos recorrer el camino inverso. Si establecemos *a priori* un intervalo de precisión con el que queremos estimar nuestro parámetro poblacional, podemos estimar el tamaño muestral necesario para ello. Nos basta con despejar de la fórmula correspondiente el tamaño muestral. Así, para la estimación de un porcentaje se puede despejar “n”:

$$EE = \sqrt{\frac{p \cdot (1 - p)}{n}} \quad n = \frac{p \cdot (1 - p)}{EE^2}$$

Si incorporamos a la fórmula la precisión o diferencia (“d”) buscada (cada uno de los dos lados del intervalo de confianza), que corresponde a 1,96 veces ($Z_{\alpha/2}$) el error estándar, la fórmula quedaría:

$$n = \frac{Z_{\alpha/2}^2 \cdot p \cdot (1 - p)}{d^2}$$

Siendo “d” la precisión o diferencia ($d = Z_{\alpha/2} \cdot EE$)

Esta fórmula es aproximada, ya que para simplificarla hemos ignorado la corrección necesaria cuando la población de muestreo es finita. No obstante, la fórmula solo la pre-

sentamos para explicar el fundamento del cálculo del tamaño muestral. En la práctica recurriremos a calculadoras epidemiológicas para realizar las estimaciones.

Por lo tanto, para calcular el tamaño muestral de una proporción solo vamos a necesitar:

- La proporción esperada.
- La precisión de nuestra estimación o diferencia a estimar.

Cuando el parámetro a estimar es una media en vez de la proporción esperada, tendremos que saber la desviación típica de la característica medida, ya que aparece en la fórmula del error estándar correspondiente. Para otros parámetros, como la diferencia de proporciones, la diferencia de medias, *odds ratio*, etc., hay otros elementos a considerar que van a aparecer en las fórmulas.

Pero, ¿de dónde obtenemos la información necesaria? Veamos cómo hacerlo para la estimación de una proporción. Lo primero es establecer la proporción esperada. Para ello podemos recurrir a nuestro conocimiento previo sobre el objeto de estudio o buscar datos en la literatura publicada. Cuando no tenemos ninguna idea previa de la frecuencia esperada, la estrategia más conservadora es realizar el cálculo asumiendo que la prevalencia esperada es del 50 % (0,50). Para una misma precisión es la prevalencia que más tamaño muestral nos estimará.

El otro elemento a conocer es la precisión o diferencia a estimar. Esta precisión también requiere conocer el fenómeno estudiado, pero, generalmente, es el investigador el que establece la precisión en función de ciertos factores, de los que destacamos tres:

- La existencia de estimaciones de referencia con las que se quiere comparar.
- La estimación de la diferencia clínicamente importante.
- La disponibilidad o factibilidad del equipo investigador para reclutar sujetos.

Imaginemos que tratamos de estimar la prevalencia de obesidad en una población escolar de un área desfavorecida. Si tenemos la impresión de que esta población tiene una mayor prevalencia de obesidad que la población general, la precisión de nuestra estimación la obtendremos de la diferencia entre la prevalencia de la población general y la que nosotros creemos que va a tener la población desfavorecida. Si existe información publicada de que en la población general hay un 18 % de obesidad y esperamos encontrar una prevalencia entre la población desfavorecida de, al menos, un 25 %, la precisión requerida será la diferencia entre ambas cifras: 7 %. En la calculadora introduciremos el 25 % esperado y el ± 7 % de precisión que buscamos.

El segundo factor a considerar era la “diferencia clínicamente importante”. Usando el ejemplo anterior, si no podemos intuir cuál es la prevalencia en la población desfavorecida, tendremos que establecer qué diferencia de prevalencia consideraríamos suficiente como para interpretar la diferencia como “clínicamente importante”. En este ejemplo parece sensato asumir que una diferencia de prevalencia de obesidad de, al menos, el 5 % es clínicamente importante, ya que supone un aumento del 27 % de la prevalencia poblacional ($0,05/0,18 = 0,27$).

El tercer factor, la factibilidad, condiciona en la práctica muchas estimaciones de tamaño muestral. Con frecuencia el investigador tiene limitaciones de reclutamiento para alcanzar determinados tamaños muestrales, por lo que cae en la tentación de restringir sus objetivos de precisión. Si el cálculo del tamaño muestral supera su capacidad de reclutamiento, no es excepcional que se busquen objetivos menos precisos. Así, si una precisión de $\pm 5\%$, implica un tamaño muestral inaccesible, se explora el tamaño necesario para otros objetivos, por ejemplo, 7 o 10 %. Debemos advertir que esta actitud resulta poco rigurosa, ya que cuando el objetivo de precisión necesario es del 5 %, si no somos capaces de reclutar la muestra requerida, lo más razonable es renunciar a realizar el estudio o bien buscar la colaboración de otros grupos de investigación con los que ampliar el reclutamiento.

En proyectos que precisan la aprobación de un Comité Ético de Investigación o que optan a becas de investigación, la estimación del tamaño muestral es uno de los principales factores a considerar, ya que informa a los evaluadores de la factibilidad y relevancia del proyecto. Por ello, toda información que aportemos para justificar la estimación del tamaño muestral y la relevancia de la magnitud de los resultados a obtener es primordial.

Veamos cómo se realiza la estimación de tamaño muestral con una calculadora epidemiológica, disponible para su ejecución *online* (disponible en: <https://www.imim.cat/ofertadeserveis/software-public/granmo/>). Emplearemos los datos mencionados anteriormente (figura 6.8): prevalencia esperada de obesidad del 25 %, con una precisión de $\pm 7\%$. Hemos dejado en blanco la casilla destinada a la población de muestreo disponible (si la población del área desfavorecida es limitada, debe introducirse aquí, para ajustar el tamaño muestral a la población disponible) y a 0 el porcentaje de reemplazo (podríamos haber puesto aquí el porcentaje esperado de pérdidas para compensarlas en la estimación del tamaño muestral). Recordamos que este programa pide proporciones (de 0 a 1), no porcentajes. El programa nos indica que se necesitan 147 sujetos.

En la figura 6.9 se presenta el cálculo del tamaño muestral para encontrar una diferencia del 10 % en el riesgo de necesitar ventilación mecánica invasiva en dos grupos de

Figura 6.8

Cálculo del tamaño muestral para una estimación de una proporción con Granmo

Disponible en <https://www.imim.cat/ofertadeserveis/software-public/granmo/>.

Calculadora de Tamaño muestral GRANMO
 Versión 7.12 Abril 2012

Català Castellano English

Proporciones : Estimación poblacional

Nivel de confianza: 0.95 0.90 Otro

Población de referencia (Intro => Se asume una población infinita):

Estimación de la proporción en la población:

Precisión de la estimación para el nivel de confianza seleccionado:

Proporción estimada de reposiciones necesarias:

04/12/2018 10:03:41 Estimación poblacional (Proporciones)

Una muestra aleatoria de 147 individuos es suficiente para estimar, con una confianza del 95% y una precisión de ± 7.0000000000000001 unidades porcentuales, un porcentaje poblacional que previsiblemente será de alrededor del 25%. En porcentaje de reposiciones necesaria se ha previsto que será del 0%.

Proporciones

- Dos proporciones independientes
- Observada respecto a una de referencia
- Medidas apareadas (repetidas en un grupo)
- Bioequivalencia
- Estimación poblacional**
- Odds ratio (Estudios de Casos-Controles)
- Riesgo relativo (Estudios de Cohorte)
- Potencia de un contraste

Medias

Otras

recién nacidos con distrés respiratorio por aspiración meconial, tratados con CPAP nasal u oxígeno en campana (proporción esperada respectiva del 5 y 15 %). Además de las proporciones esperadas, debemos indicar la relación entre los grupos (si los grupos son de igual tamaño muestral, introduciremos "1") y la proporción esperada de pérdidas (si no prevemos pérdidas, introduciremos "0"). Por defecto, el programa asigna ciertos parámetros de los que no hemos hablado todavía y que se presentarán en el siguiente capítulo del libro:

- Riesgo alfa (error tipo I).
- Riesgo beta (error tipo II).
- Tipo de contraste (unilateral o bilateral).

Podemos adelantar que el riesgo alfa es la probabilidad de que encontremos diferencias cuando no las hay (error tipo I o de falso positivo) y el riesgo beta es la probabilidad de que no encontremos diferencias cuando sí las hay (error tipo II o de falso negativo). Se acepta que el máximo riesgo alfa asumible es 0,05 (5 %) y el máximo riesgo beta asumible es 0,20 (20 %). Cuanto más exigentes seamos en el umbral de error (riesgos menores), mayor será la muestra necesaria.

El tipo de contraste depende de la hipótesis que planteemos en el estudio; si nuestra hipótesis considera las dos colas de la distribución normal (hipótesis de diferencia) o solo una (hipótesis solo de superioridad o solo de inferioridad). El tipo de contraste elegido se traduce en que el factor que multiplicará el error estándar será 1,96 ($Z_{\alpha/2}$ bilateral) o 1,65 (Z_{α} unilateral), por lo que el contraste bilateral siempre implicará un mayor tamaño muestral.

Como vemos en la figura 6.9, el programa asigna, por defecto, un riesgo alfa de 0,05, un riesgo beta de 0,20 y un contraste bilateral. La estimación del tamaño muestral es de 133 sujetos en cada grupo.

Figura 6.9

Cálculo del tamaño muestral para una diferencia de proporciones con Granmo

Calculadora de Tamaño muestral GRANMO
Versión 7.12 Abril 2012

Català Castellano English

Proporciones : Dos proporciones independientes

Riesgo Alfa: 0.05 0.10 Otro

Tipo de contraste: unilateral bilateral

Riesgo Beta: 0.20 0.10 0.05 0.15 Otro

Proporción en el grupo 1:

Proporción en el grupo 2:

Razón entre el número de sujetos del grupo 2 respecto del grupo 1:

Proporción prevista de pérdidas de seguimiento:

calcula Limpia resultados Limpia todo Selecciona todo Imprimir

Proporciones

- Dos proporciones independientes
- Observada respecto a una de referencia
- Medidas apareadas (repetidas en un grupo)
- Bioequivalencia
- Estimación poblacional
- Odds ratio (Estudios de Casos-Controles)
- Riesgo relativo (Estudios de Cohorte)
- Potencia de un contraste

Medias

Otras

04/12/2018 10:05:10 Dos proporciones independientes (Proporciones)

Acceptando un riesgo alfa de 0.05 y un riesgo beta de 0.2 en un contraste bilateral, se precisan **133** sujetos en el primer grupo y **133** en el segundo para detectar como estadísticamente significativa la diferencia entre dos proporciones, que para el grupo 1 se espera sea de 0.05 y el grupo 2 de 0.15. Se ha estimado una tasa de pérdidas de seguimiento del 0%. Se ha utilizado la aproximación del ARCOSENDO.

Disponible en <https://www.imim.cat/ofertadeserveis/software-public/granmo/>.

P

Ejercicio práctico 6.2

Invitamos a los lectores a realizar estimaciones del tamaño muestral para dos proyectos con el programa Granmo (<https://www.imim.cat/ofertadeserveis/software-public/granmo/>):

- a) Queremos saber a cuántos lactantes sanos, que vemos en la revisión de los 3 meses de edad debemos estudiar para conocer la proporción de mantenimiento de lactancia materna, asumiendo un porcentaje aproximado del 60 %, con una precisión de ± 5 %. Repita el cálculo para una precisión de ± 10 %.
- b) Queremos saber el tamaño muestral de un estudio en el que comparamos el riesgo de reingreso por crisis asmática en dos grupos de niños con asma persistente que siguen dos tratamientos diferentes: un grupo con montelukast y otro con corticoides inhalados diarios. En el último año el riesgo de ingreso por asma en ese tipo de pacientes había sido del 15 %. Esperamos obtener una mejora del 5 % en el riesgo de ingreso con corticoides inhalados. Repita el cálculo para una mejora del 10 %.

BIBLIOGRAFÍA RECOMENDADA

- Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.
- Altman DG, Bland JM. Statistics notes: the normal distribution. BMJ. 1995;310:298.
- Altman DG, Bland JM. Statistics notes: variables and parameters. BMJ. 1999;318:1667.
- Milton JS. Estadística para biología y ciencias de la Salud. México: McGraw-Hill; 2001.
- Norman GR, Streiner DL. Bioestadística. México: Mosby/Doyma Libros; 1996.
- Rosner B. Fundamentals of Biostatistics, 7th Edition. Boston: Brooks/Cole, Cengage Learning; 2011.

7

Contrastes de hipótesis Elección del test estadístico

Objetivos docentes:

- Repasar el concepto de contraste de hipótesis.
- Conocer las principales pruebas de contraste de hipótesis.
- Conocer los factores a tener en cuenta en la elección del contraste de hipótesis.
- Saber elegir la prueba más apropiada para cada tipo de contraste de hipótesis.
- Aprender a interpretar los contrastes de hipótesis más comunes.



Contraste de hipótesis

En el capítulo anterior expusimos los fundamentos de la inferencia estadística. Diferenciamos en ella dos estrategias: la estimación por intervalos y el contraste de hipótesis. En este capítulo abordaremos el contraste de hipótesis, con él podemos comparar dos o más alternativas, cuantificando la probabilidad de que las diferencias entre ellas se deban al azar.

Recordemos un ejemplo presentado anteriormente: en un ensayo clínico se compararon dos tratamientos, A y B, a dos grupos de 100 pacientes, para prevenir recaídas de una enfermedad. En el contraste de hipótesis se plantean dos alternativas:

- Hipótesis nula: no hay diferencias de eficacia entre A y B. La diferencia de proporciones no es distinta de 0.
- Hipótesis alternativa; dos opciones: sí hay diferencias entre A y B (contraste bilateral) o A es más eficaz que B (contraste unilateral). La diferencia de proporciones es distinta/mayor que 0.

En el grupo A recayeron un 20 %, mientras que en el grupo B un 40 %. En el capítulo anterior calculamos para el mismo ejemplo que la diferencia de proporciones era del

20 %, con un intervalo de confianza del 95 % de 7,6 a 32,4 %. Como ese intervalo no incluye el valor nulo, que para una diferencia es "0", podíamos ya asumir la mayor eficacia del tratamiento A. Sin embargo, para resolver el contraste de hipótesis debemos cuantificar la probabilidad exacta de que la diferencia encontrada sea mayor que "0" por azar.

Contamos con varias pruebas con las que calcular esta probabilidad. Una de las opciones es la aproximación a la distribución normal de la diferencia de proporciones, cuyo error estándar era:

$$EE \text{ diferencia proporciones} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Podríamos realizar los cálculos por nosotros mismos, con el error estándar y nuestros conocimientos de la distribución normal, algo que no recomendamos. La mejor alternativa es usar alguna calculadora epidemiológica.

En la figura 7.1 se presenta el cálculo realizado con el programa gratuito Epidat 4.2 (<https://www.sergas.es/Saude-publica/EPIDAT-4-2?idioma=es>).

Figura 7.1

Contraste de hipótesis para una diferencia de proporciones mediante aproximación a la normal realizado con Epidat 4.2. Se presenta el menú desplegado en el que se accede a la ventana correspondiente (en "Calcular" debe señalarse el tipo de contraste; en este caso se ha optado por contraste bilateral)

The figure shows the Epidat 4.2 software interface. The top menu bar includes 'Archivo', 'Edición', 'Módulos', 'Herramientas', 'Ventana', and 'Ayuda'. The 'Módulos' menu is open, showing options like 'Análisis descriptivo', 'Muestreo', and 'Índice de cálculos'. Under 'Índice de cálculos', the path 'Inferencia sobre parámetros' > 'Dos poblaciones' > 'Comparación no paramétrica' > 'Contraste de normalidad' > 'Proporciones emparejadas' is highlighted. An orange arrow points to the 'Bilateral' option in the 'Contraste de hipótesis' section of the 'Inferencia. Comparación de proporciones independientes' window.

The 'Inferencia. Comparación de proporciones independientes' window shows the following data:

Datos resumidos	Muestra 1	Muestra 2
Número de casos	20	40
Tamaño de muestra	100	100

Level of confidence: 95.0 %

Contraste de hipótesis:

- Bilateral ($H_0: P_1 - P_2 = 0$ vs. $H_1: P_1 - P_2 \neq 0$)
- Unilateral izquierdo ($H_0: P_1 - P_2 = 0$ vs. $H_1: P_1 - P_2 < 0$)
- Unilateral derecho ($H_0: P_1 - P_2 = 0$ vs. $H_1: P_1 - P_2 > 0$)

Buttons: Ocultar, Calcular, Limpiar, Cerrar

The results window shows:

Datos:	Muestra 1	Muestra 2
Número de casos	20	40
Tamaño de muestra	100	100

Level of confidence: 95,0%
Calcular: Intervalo de confianza y contraste de hipótesis

Población	Porcentaje (%)
1	20,000
2	40,000

Intervalo de confianza (95,0%)

Diferencia de proporciones	Límite inferior	Límite superior
	-0,200	-0,324
		-0,070

Prueba de comparación de proporciones

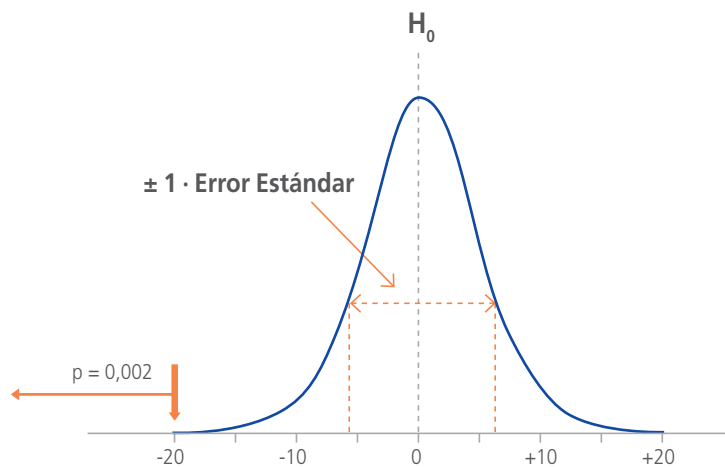
Contraste	Estadístico Z	Valor p
Bilateral	-3,088	0,002

La calculadora nos informa del valor Z (distribución normal estandarizada) correspondiente a una diferencia de 0,20 (20 %) en una distribución normal de media 0 y "desviación estándar" equivalente a nuestro error estándar (figura 7.2). El valor Z es 3,086 (valor más alejado que 1,96); aparece con signo negativo porque el grupo con menor recaídas lo hemos puesto primero, pero para el cálculo es irrelevante. Si cambiamos el orden, el valor Z sería positivo, pero el valor "p" sería el mismo, al que le corresponde una probabilidad (valor "p") de 0,002 (0,2 %). Como esta probabilidad es menor de 0,05 (5 %), podemos rechazar la hipótesis nula (no hay diferencias) y aceptar la alternativa (el tratamiento A es mejor que el B).

Rechazando la hipótesis nula y aceptando la alternativa asumimos un error de 0,002 (0,2 %). A este error lo denominamos error tipo I, o error de falso positivo (encontrar diferencias en la muestra cuando no las hay en la población), y a su probabilidad la llamamos "alfa". Es importante advertir que, aunque el error sea muy pequeño, siempre existe cierto riesgo de error.

Figura 7.2

Distribución normal de las diferencias de proporciones de media 0 (hipótesis nula [H_0]) y desviación típica equivalente a su error estándar

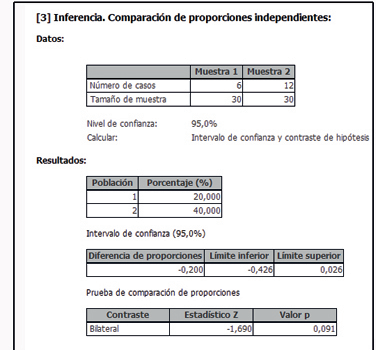
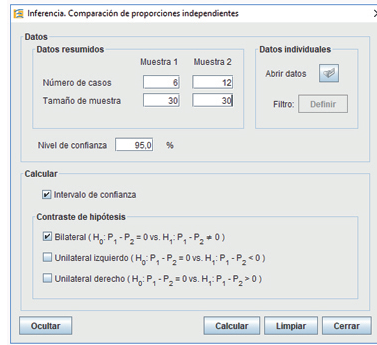
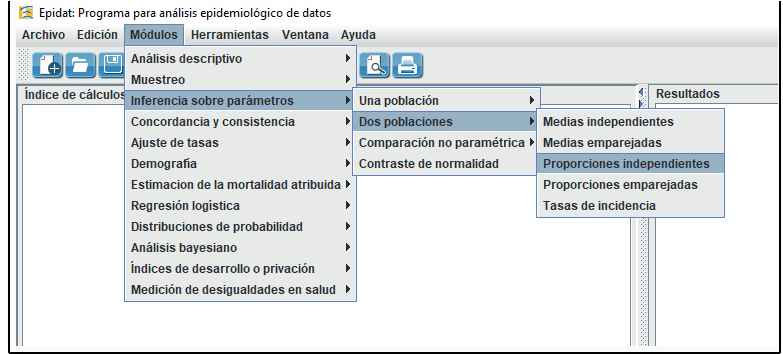


Veamos otro ejemplo: supongamos que en el estudio realizado anteriormente, en vez de contar con 100 sujetos en cada grupo, solo contáramos con 30 sujetos y que la proporción de recaídas fuera la misma: 20 % en el grupo A (6/30) y 40 % en el grupo B (12/30). En la figura 7.3 presentamos el nuevo cálculo.

Vemos cómo, aunque la diferencia porcentual entre tratamientos es la misma, al disminuir el tamaño muestral el error estándar aumenta y la probabilidad asociada a la diferencia encontrada cambia. La calculadora nos da un valor Z de 1,69 (menos alejado que 1,96), lo que para un contraste bilateral (figura 7.4a) le corresponde una probabilidad (valor "p") de 0,091 (9,1 %). Con este nuevo estudio no podríamos rechazar la hipótesis nula ni aceptar la alternativa, ya que el error tipo I (o de falso positivo) en el que incurriríamos sería mayor de 0,05 (5 %).

Figura 7.3

Contraste de hipótesis para una diferencia de proporciones mediante aproximación a la normal Epidat 4.2



¿Qué ha pasado? Que el nuevo estudio ha perdido potencia, aumentando el riesgo de error tipo II (riesgo beta) o de falso negativo (probabilidad de no encontrar diferencias en la muestra cuando sí las hay en la población). El tratamiento A podría ser más eficaz que el B, pero nosotros no hemos sido capaces de observarlo. Al aumentar el error estándar, la distribución normal es tan amplia que, aunque la diferencia sea grande, el valor nulo "0" es muy probable que quede dentro del intervalo de confianza (figura 7.4b).

Para que un resultado "negativo" (no hay diferencias estadísticamente significativas) sea aceptable, el estudio tiene que tener un error tipo II, cuantificado en el riesgo beta, menor de 0,20 (20 %). Al complementario del riesgo beta lo llamamos potencia (1 - beta). Por ello, un resultado "negativo" solo es aceptable si la potencia es mayor del 80 %.

Figura 7.4 a

Distribución normal de las diferencias de proporciones de media 0 (hipótesis nula H_0) y desviación típica equivalente a su error estándar

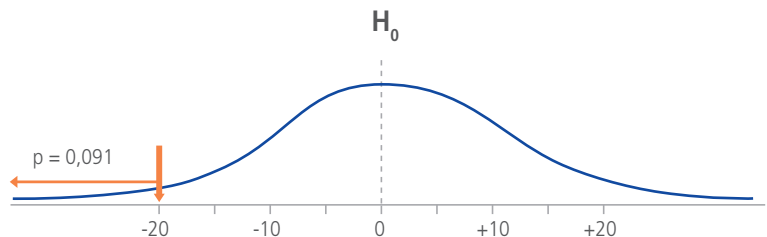
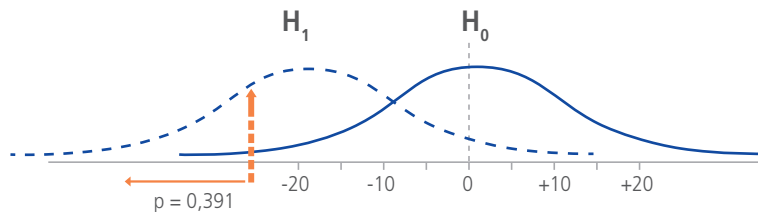


Figura 7.4 b

Distribuciones normales de medias 0 (H_0) y -20 (hipótesis alternativa [H_1])



Para calcular el error tipo II (riesgo beta) o la potencia ($1 - \beta$), recomendamos usar una calculadora epidemiológica. En la figura 7.4b se muestra el planteamiento en el que se sustenta el cálculo de la potencia. Si realmente existieran diferencias (H_1 cierta), existiría una distribución de diferencias de proporciones alternativa (H_1) a la hipótesis nula (H_0). En ese caso nuestro estudio podría haber encontrado cualquier valor comprendido en la distribución alternativa, pero solo los que quedan más alejados de la hipótesis nula darían una probabilidad menor de 0,05 en ella; podemos ver que ese valor es un valor más extremo que el que nosotros hemos encontrado (-0,20). En la figura 7.5 se muestra el cálculo para nuestro estudio.

Figura 7.5

Cálculo de la potencia de un contraste de hipótesis para una diferencia de proporciones con Epidat 4.2

[4] Tamaños de muestra. Comparación de proporciones independientes:

Datos:

- Proporción esperada en: 20,000%
- Población 1: 20,000%
- Población 2: 40,000%
- Razón entre tamaños muestrales: 1,00
- Nivel de confianza: 95,0%

Resultados:

Tamaño de la muestra*	Potencia (%)
60	39,1

*Tamaños de muestra para aplicar el test χ^2 sin corrección por continuidad.

La calculadora ha estimado que, con una muestra de 30 sujetos por grupo (60 en total), la potencia para estimar una diferencia del 20 % es 39,1 %. Como vemos, no alcanza la potencia mínima requerida del 80 %. Por ello, nuestro resultado "negativo" no sería aceptable; si estamos convencidos de que el tratamiento A es mejor que el B (así lo sugiere la diferencia encontrada), lo más razonable es plantear un estudio con mayor tamaño muestral.

Es importante destacar que en el cálculo de la potencia del estudio debemos introducir diferencias que consideremos clínicamente importantes, que no tienen por qué coincidir con las observadas en nuestro estudio. En nuestro supuesto hemos usado los datos del estudio, ya que un 20 % es aceptable como diferencia clínicamente importante. Si el estudio hubiera encontrado diferencias muy pequeñas (por ejemplo: 2 %), para el cálculo de la potencia deberíamos haber usado diferencias que consideremos clínicamente importantes, como, por ejemplo, un 10, un 15 o un 20 %. La potencia calculada sería interpretada como que, aunque el estudio no ha encontrado diferencias, tenía potencia suficiente para haber encontrado diferencias mayores de 10, 15 o 20 %. La elección de la diferencia requiere conocimientos del problema en estudio y no responde a criterios estadísticos.

Otra cuestión a advertir es que si en los cálculos de la potencia queremos usar riesgos alfa o beta alternativos (por ejemplo: riesgo alfa 0,01 o riesgo beta 0,10), los umbrales de cálculo de probabilidad cambiarán. Si disminuimos el riesgo alfa, aumentará el beta y viceversa; solo aumentando el tamaño muestral disminuirán los dos.

En la tabla 7.1 se resumen todas las situaciones posibles del contraste de hipótesis. Debemos tener en cuenta que sea cual sea la decisión de nuestro contraste, siempre existe un cierto riesgo de error, ya que la población es inaccesible. Recordemos que si el riesgo alfa es menor de 0,05, solo tendremos en cuenta la primera fila de la tabla. Cuando el riesgo alfa sea mayor, nos plantearemos el cálculo de la segunda fila, estimando el riesgo beta.

Tabla 7.1 Alternativas del contraste de hipótesis

Decisión	Realidad (¡¡Desconocida!!)	
	H ₀ Cierta	H ₀ Falsa
H ₀ Rechazada H ₁ Aceptada	Error tipo I (α) Falsos (+)	Decisión correcta
H ₀ NO Rechazada	Decisión correcta	Error tipo II (β) Falsos (-)

Error alfa: probabilidad de equivocarnos si rechazamos la hipótesis nula (H₀) cuando esta es cierta.

Error beta: probabilidad de equivocarnos si **no** rechazamos la hipótesis nula, a pesar de que sea falsa (H₁ cierta).

Potencia del test (1 – beta): probabilidad de rechazar la hipótesis nula cuando es falsa (encontrar diferencias cuando estas existen).



Ejercicio práctico 7.1

Invitamos a los lectores a descargar e instalar el programa Epidat 4.2 (<https://www.sergas.es/Saude-publica/EPIDAT-4-2?idioma=es>) y realizar contrastes de hipótesis para comparaciones de porcentajes. Damos varios supuestos:

- a) Hemos comparado el riesgo de ingreso por enfermedad infecciosa aguda de 100 lactantes que acudieron a guardería en el primer año de vida y 200 lactantes que fueron cuidados en su hogar. Ingresaron por enfermedad infecciosa aguda 10/100 lactantes del grupo de guardería y 10/200 del otro grupo. ¿Aumentó la asistencia a guardería el riesgo de ingreso?
- b) Queremos estimar la eficacia de los corticoides inhalados diarios en niños con asma persistente en comparación con montelukast. Para ello estudiamos a dos grupos de 200 pacientes que reciben uno de los dos tratamientos de estudio. Medimos la frecuencia de ingreso por asma en 1 año. Ingresaron 10/200 en el grupo con corticoides y 30/200 en el grupo con montelukast. ¿Fueron más eficaces los corticoides inhalados?



Pruebas de contraste de hipótesis

En el apartado anterior hemos empleado alguna prueba de contraste de hipótesis (aproximación a la normal de la diferencia de proporciones), pero existen muchas otras pruebas, entre las que tendremos que elegir la más apropiada para cada contraste.

En la elección del test estadístico tendremos que considerar los siguientes factores:

- a) Cuántas variables están implicadas: 1, 2 o más.
- b) Cuáles son las variables dependientes e independientes.
- c) Qué escalas de medida siguen las variables implicadas: nominal, ordinal, continua normal, continua no normal.
- d) Cuántos grupos de estudio hay: 1, 2 o más.
- e) Los grupos de estudio son independientes o están relacionados (o apareados).
- f) Queremos un contraste uni o bilateral.
- g) Qué umbrales de errores tipo I y II elegimos: 0,05 y 0,20, respectivamente, o inferiores.

En la tabla 7.2 se presenta un esquema simplificado para la elección de la prueba de contraste más apropiada. A continuación veremos unos ejemplos para las pruebas más habituales: test de ji cuadrado, test de la t de Student para muestras independientes y apareadas, ANOVA y coeficiente de correlación de Pearson.

Tabla 7.2 Esquema de elección del test de contraste de hipótesis más apropiado

Variable independiente	Variable dependiente		
	Nominal	Ordinal (continuas no normales)	Continua (razón o intervalos)
Nominal dicotómica (2 muestras)	Muestras independientes: · Test Z de comparación de proporciones · Test de ji cuadrado · Test exacto de Fisher	· Test U Mann Whitney (Wilcoxon suma rangos)	· Test t de Student de muestras independientes
	Muestras relacionadas: · Test McNemar · Test Z y método binomial	· T. Wilcoxon rangos con signo	· Test t de Student de muestras apareadas
Nominal politómica (> 2 muestras)	· Test de ji cuadrado · Método binomial	T. Kruskal - Wallis * M. apareadas: P. Friedman	ANOVA
Continua	Test t de Student	Coeficiente Correlación de Spearman ** ** También Ordinal/Ordinal	Coeficiente de correlación de Pearson Regresión lineal

- Análisis de supervivencia (tiempo hasta evento): método de Kaplan Meier y *Log-Rank*.
- Técnicas multivariantes: Variable dependiente nominal: regresión logística; Variable dependiente continua: regresión lineal múltiple; Supervivencia: regresión de Cox.

Variable nominal dicotómica frente a nominal dicotómica: comparación de proporciones entre dos grupos

En un estudio retrospectivo de cohortes se estudiaron los hábitos de introducción del gluten en lactantes. Se evaluó la introducción del gluten entre los 4 y 6 meses en función de distintas variables. Una de ellas fue el tener o no antecedentes de familiares celíacos. Solamente el 30,4 % de los 57 que tenían un familiar celíaco introdujeron el gluten entre los 4 y 6 meses, frente al 46 % de los 888 que no.

Veamos cómo elegimos la prueba más apropiada para este análisis. En primer lugar, valoraremos los factores a tener en cuenta en esta elección:

- Número de variables implicadas. Hay dos variables implicadas: "familiar celíaco" sí/no e "inicio de gluten entre 4 y 6 meses" sí/no.
- Variables dependientes e independientes. La variable independiente es la existencia de familiar celíaco y la dependiente es el inicio del gluten entre 4 y 6 meses.
- Escalas de medida de las variables implicadas. Ambas variables se miden en una escala nominal.

- d) Número de grupos de estudio. Hay dos grupos, como corresponde a una variable independiente nominal dicotómica.
- e) Grupos de estudio independientes o relacionados (o apareados). Los grupos son independientes, ya que unos casos no tienen nada que ver con los otros (serían relacionados o apareados si midiéramos una misma característica en el mismo grupo antes/después de una intervención).
- f) Contraste uni o bilateral. La elección de uno u otro contraste depende de la hipótesis alternativa que planteemos. Si nuestra hipótesis alternativa es que ambos grupos introducen el gluten de forma distinta, elegiremos un contraste bilateral. Si nuestra hipótesis alternativa solo contempla que los que tienen familiares con enfermedad celíaca introducen el gluten más tarde, elegiremos un contraste unilateral. La elección de un contraste unilateral se traduce en una reducción de la estimación de error tipo I ("p" más baja o significativa). Aunque el contraste unilateral es completamente ortodoxo, es habitual que se elijan contrastes bilaterales, ya que son más conservadores. Algunos test estadísticos son intrínsecamente bilaterales, por lo que cuando los usamos no tenemos otra opción (por ejemplo: test de ji cuadrado).
- g) Umbrales de errores tipo I y II. Habitualmente se eligen por defecto los umbrales 0,05 (error tipo I) y 0,20 (error tipo II). Algunos autores eligen umbrales de error más exigentes (por ejemplo: 0,01 y 0,10, respectivamente), intentando mostrar un mayor rigor en el contraste estadístico. Esta actitud podría estar justificada cuando en el estudio se realizan múltiples contrastes estadísticos, con variables independientes o dependientes alternativas, tratando de minimizar el error tipo I (a mayor número de contrastes, más riesgo de encontrar alguna diferencia significativa por azar).

El contraste estadístico de nuestro ejemplo podríamos describirlo como el análisis de la asociación entre dos variables nominales dicotómicas y también como la comparación de proporciones entre dos grupos (comparación de las proporciones de introducción de gluten entre los 4 y 6 meses entre los sujetos con y sin antecedentes familiares de enfermedad celíaca).

Aunque los factores a considerar en la elección de la prueba de contraste estadístico son los anteriormente detallados, se puede simplificar el procedimiento en tres pasos, siguiendo la estructura de la tabla 7.2:

1. Establecer las variables independiente y dependiente.
2. Elegir la fila en función de la escala de medida de la variable independiente. Hay cuatro opciones: nominal dicotómica con muestras independientes, nominal dicotómica con muestras apareadas, nominal politómica y continua.
3. Elegir la columna en función de la escala de medida de la variable dependiente. Hay tres opciones: nominal, ordinal y continua.

Podemos ver que las categorías de las columnas y filas no se corresponden. El esquema admite que si la variable dependiente o la independiente sigue una escala de medida que no aparece entre las opciones (por ejemplo: la variable independiente es ordinal), pueden intercambiarse las variables dependiente e independiente en el esquema.

Veamos la elección de nuestro ejemplo:

1. Las variables independiente y dependiente son: familiar celiaco e introducción del gluten entre los 4 y 6 meses.
2. La variable independiente es nominal dicotómica con muestras independientes (primera fila).
3. La variable dependiente es nominal (primera columna).

Las opciones que nos ofrece el esquema son tres: test Z de comparación de proporciones, test de ji cuadrado y test exacto de Fisher. La primera opción es la que empleamos en el apartado anterior de este capítulo, basada en la aproximación a la normal de la distribución de diferencias de proporciones. Mucho más popular y más utilizada es la segunda opción: el test de ji cuadrado (“ji” no “chi”, ya que la letra griega “ χ ” se lee “ji” en español). Es la que empleamos cuando buscamos asociación entre variables nominales, en general, tanto sean dicotómicas como politómicas, que podemos representar en una tabla de contingencia de dos o más filas por dos o más columnas. La tercera opción (test exacto de Fisher) se emplea cuando el tamaño muestral es pequeño y no se pueden utilizar los test anteriores, aunque puede ser empleada en cualquier circunstancia.

En la tabla 7.3 se presentan los datos de nuestro ejemplo en formato de tabla de contingencia 2 x 2 y analizados con un test de ji cuadrado.

Tabla 7.3 Tabla de contingencia 2 x 2 de asociación entre dos variables nominales dicotómicas y análisis mediante test de ji cuadrado y exacto de Fisher

		Introducción Gluten 4-6 meses			
		Sí	No	Total	
Familia Celiaco	Sí	Recuento % dentro de Familia Celiaco	17 30,4 %	39 69,6 %	56 100,0 %
	No	Recuento % dentro de Familia Celiaco	401 46,0 %	471 54,0 %	872 100,0 %
Total		Recuento % dentro de Familia Celiaco	418 45,0 %	510 55,0 %	928 100,0 %

Tabla 7.3 Tabla de contingencia 2 × 2 de asociación entre dos variables nominales dicotómicas y análisis mediante test de ji cuadrado y exacto de Fisher (continuación)

	Valor	gl	Significación asintótica (bilateral)	Significación exacta (bilateral)	Significación exacta (unilateral)
Chi-cuadrado de Pearson	5,192 ^a	1	,023		
Corrección de continuidad ^b	4,580	1	,032		
Razón de verosimilitud	5,370	1	,020		
Prueba exacta de Fisher				,026	,015
Asociación lineal por lineal	5,187	1	,023		
N de casos válidos	928				

a. 0 casillas (0,0%) han esperado un recuento menor que 5. El recuento mínimo esperado es 25,22.

b. Sólo se ha calculado para una tabla 2 × 2.

La prueba de ji cuadrado se basa en la comparación de los recuentos observados y esperados en cada casilla de la tabla de contingencia. Si las variables que se trata de asociar fueran independientes (no hubiera relación entre ellas), el recuento de cada casilla dependería exclusivamente del azar. Si fuera así, en la casilla superior izquierda (donde cruzan la columna "Sí" y la fila "Sí"), debería haber un recuento equivalente al producto de las probabilidades de tener un familiar celíaco ($56/928 = 0,018$) y de introducir el gluten entre los 4 y 6 meses ($418/928 = 0,450$); si multiplicamos dicho producto por el tamaño muestral ($0,018 \times 0,450 \times 928$), obtendremos un recuento esperado de 7,65. Sin embargo, en esa casilla hay un recuento observado de 17. Ese mismo cálculo se hace para cada casilla. Cuanto mayor sea la diferencia entre los valores esperados y observados en cada casilla, más baja será la probabilidad de que esa diferencia se deba al azar; si esa probabilidad es menor de 0,05, asumiremos que la asociación es estadísticamente significativa.

En la tabla 7.3 podemos ver los recuentos observados de cada casilla, los porcentajes respecto a las filas (30,4 % y 46 %) y en la parte inferior la estimación del error tipo I (significación) por varios métodos. Señalado con un sombreado naranja se destaca la "p" estimada mediante test de ji cuadrado (chi cuadrado en la figura) y en sombreado azul las dos estimaciones, bilateral y unilateral, mediante la prueba exacta de Fisher. La prueba de Fisher será obligatoria cuando en la nota que aparece en la parte inferior de la tabla nos indiquen un porcentaje de casillas con recuentos esperados menores de 5 en más del 20 % de ellas. Esta prueba se basa en la distribución de probabilidad binomial, por lo que no necesita cumplir los requerimientos de la aproximación a la normal ni de la distribución de ji cuadrado. En nuestro ejemplo no es necesario usar esta prueba, ya que el porcentaje de casillas es 0 % (señalado en negrita).

Si escogemos la “p” correspondiente a la prueba de ji cuadrado, vemos que al decir que las variables analizadas están asociadas, asumimos un error tipo I de 0,023 (2,3 %), menor de 0,05 (5 %). La existencia de asociación se puede interpretar igualmente diciendo que las proporciones o porcentajes comparados (30,4 % y 46 %) son diferentes y que esta diferencia es estadísticamente significativa.

Si hubiéramos elegido las estimaciones correspondientes al test exacto de Fisher, también hubiera dado estadísticamente significativo. Llamamos la atención de que la “p” para el contraste unilateral es más baja que para el bilateral (sombreado verde).

Variable nominal dicotómica frente a continua: comparación de medias entre dos grupos independientes

En un estudio transversal se midió la presión arterial en una muestra de 1.410 niños; queremos saber si los niños con obesidad abdominal (Índice cintura/talla $\geq 0,50$) tienen una mayor presión arterial sistólica que los niños sin obesidad abdominal.

Sigamos los pasos para elegir la prueba más adecuada en la tabla 7.2:

1. Las variables independiente y dependiente son: obesidad abdominal y presión arterial sistólica.
2. La variable independiente es nominal dicotómica con muestras independientes (primera fila).
3. La variable dependiente es continua (tercera columna).

La opción que nos ofrece el esquema es el test de la t de Student para muestras independientes. Debemos advertir que para poder elegir las pruebas que aparecen en la tercera columna, bajo la cabecera de “Continua”, la variable debe ser continua (con escala de intervalos o razones) y seguir una distribución normal. Los test correspondientes se basan en estimaciones de probabilidad que emplean medidas de centralización (media) y dispersión (varianza o desviación típica), por lo que si dichas medidas no tienen sentido en la variable analizada, la interpretación de los resultados se verá comprometida, aunque las estimaciones de probabilidad sean correctas (según el teorema central del límite para medias de variables continuas con tamaño muestral superior a 30). Así, mientras la comparación de medias de presión arterial entre grupos permite analizar las diferencias, podría no ser igualmente correcto comparar las medias de niveles de colesterol (variable que no suele seguir una distribución normal y cuya media podría estar sesgada). Si tenemos dudas sobre la validez de la comparación de medias, tendríamos que elegir uno de los test situados en la columna del medio, correspondiente a variables ordinales.

Aunque hay test estadísticos que permiten cuantificar la normalidad de una variable, la forma más simple de comprobar si una variable sigue una distribución normal es comparar la media y la mediana de los datos y observar su histograma de frecuencias. Si la

media y la mediana son muy diferentes y el histograma no adopta una forma acampanada simétrica, podría no ser razonable emplear estos test.

En la tabla 7.4 se presenta el análisis comparativo de medias mediante test de la t de Student para muestras independientes.

Tabla 7.4 Comparación de medias de muestras independientes mediante test de la t de Student para muestras independientes. Asociación entre una variable nominal dicotómica y otra continua

Estadísticos de grupo

Presión arterial sistólica		N	Media	Desviación típica	Error típico de la media
Obesidad abdominal ICT > 0,50	Sí	420	102,77	14,841	,724
	No	990	99,46	13,812	,439

Presión arterial sistólica	Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias					
	F	Sig.	t	gl	Sig. bilateral	Diferencia de medias	95 % intervalo de confianza para la diferencia	
							Inferior	Superior
Se han asumido varianzas iguales	1,918	0,166	4,030	1408	0,000	3,315	1,702	4,929
No se han asumido varianzas iguales			3,915	741,132	0,000	3,315	1,653	4,978

La salida de resultados nos muestra en la parte superior las medias y desviaciones típicas de presión arterial sistólica en cada grupo y en la parte inferior las estimaciones de probabilidad (Sig. Bilateral) y la diferencia de medias con su intervalo de confianza (en naranja). Como podemos ver, la tabla inferior contiene dos filas de resultados. Hemos escogido la superior porque las varianzas de las muestras comparadas parecen iguales; para saber esto, consideramos la significación de la prueba de Levene que aparece a la izquierda (en azul; $p > 0,05$). Si las varianzas fueran distintas, hubiéramos elegido la fila inferior. Concluimos diciendo que hay una diferencia de presión arterial significativa, con un error tipo I (alfa) menor de 0,001 (solemos sustituir el último "0" por un "1" poniendo delante el signo "<", ya que la salida de resultados no nos muestra más precisión en la estimación) y, por lo tanto, asumible, dado que es una " p " < de 0,05.

Además, podemos concluir que dicha diferencia es de 3,315 mmHg, con un intervalo de confianza del 95 % entre 1,702 y 4,929. A partir del propio intervalo podemos concluir lo mismo, dado que en su interior no se incluye el valor nulo de ausencia de diferencias "0".

Variable nominal dicotómica relacionada frente a continua: comparación de medias entre dos grupos relacionados

En un ensayo clínico se evaluó la eficacia de una intervención educativa para prevenir el desarrollo de obesidad. Se compararon los cambios de índice de masa corporal (IMC) de 648 niños de cuatro guarderías que recibieron la intervención con los de 336 niños de dos guarderías que no la recibieron. El grupo intervención mostró un aumento de puntuación Z estandarizada (por edad y sexo) de IMC de 0,24 (de + 0,12 a + 0,36) frente a un aumento de 0,41 (de + 0,56 a + 0,97) del grupo control.

Sigamos los pasos para elegir la prueba más adecuada en la tabla 7.2:

1. Las variables independiente y dependiente son: intervención educativa e IMC estandarizado (puntuación Z).
2. La variable independiente es nominal dicotómica con muestras relacionadas antes-después (segunda fila).
3. La variable dependiente es continua (tercera columna).

La opción que nos ofrece el esquema de la tabla 7.2 es el test de la t de Student para muestras relacionadas. En este caso asumimos que la puntuación estandarizada sigue una distribución normal. Debemos tener en cuenta que a la hora de realizar el procedimiento con el programa estadístico, en vez de introducir las variables independiente y dependiente tendremos que elegir dos variables dependientes, dado que habitualmente la variable dependiente estará recogida en dos campos: el IMC previo y el posterior a la intervención. No estamos analizando dos variables continuas, sino una misma variable continua en dos momentos.

Variable nominal politómica frente a continua: comparación de medias entre más de dos grupos

Queremos analizar la eficacia de tres tipos de intervenciones en una muestra de adolescentes obesos, aleatorizados a recibir una de tres posibles intervenciones diferentes: grupo A, solo dieta; grupo B, solo ejercicio, y grupo C, ambas intervenciones. Mediremos el efecto en base al IMC estandarizado al final del seguimiento.

Sigamos los pasos para elegir la prueba más adecuada en la tabla 7.2:

1. Las variables independiente y dependiente son: el tipo de intervención y el IMC estandarizado (puntuación Z).

2. La variable independiente es nominal politómica (tercera fila).
3. La variable dependiente es continua (tercera columna).

La opción que nos ofrece el esquema es ANOVA, análisis de la varianza. Este test permite comparar a la vez los tres grupos frente a la alternativa de comparar las intervenciones por pares (por ejemplo: dieta y ejercicio con respecto a solo dieta o solo ejercicio). Si hiciéramos comparaciones por pares, tendríamos que realizar más de un contraste estadístico, por lo que aumentaría el error tipo I. Por ello, la opción más correcta es emplear el ANOVA, que además tiene técnicas que facilitan contrastes entre pares sin incrementar el error.

Variable nominal politómica frente a nominal: comparación de porcentajes entre más de dos grupos

Usemos de nuevo el ejemplo del apartado anterior que analizaba la eficacia de tres tipos de intervenciones en una muestra de adolescentes obesos, aleatorizados a recibir una de tres posibles intervenciones diferentes: grupo A, solo dieta; grupo B, solo ejercicio, y grupo C, ambas intervenciones. En esta ocasión mediremos el efecto en base al porcentaje de casos que descienden el IMC por debajo del percentil 95 % para su edad y sexo.

El cambio que se produce afecta a la variable dependiente, que en vez de ser continua pasa a ser nominal (sí/no). Al cruzar en la tabla 7.2 la tercera fila (nominal politómica) con la primera columna (nominal), la opción que nos ofrece es el test de ji cuadrado, cuya realización e interpretación ya hemos visto. Solo cambia que en vez de analizar una tabla de 2×2 en este caso analizamos una tabla de 3×2 .

Variable continua frente a continua: correlación

Queremos saber cómo se correlacionan dos medidas de obesidad en la infancia. Para ello hemos medido el índice cintura/talla (IndCT) y el IMC estandarizado por edad y sexo (IMC-DS).

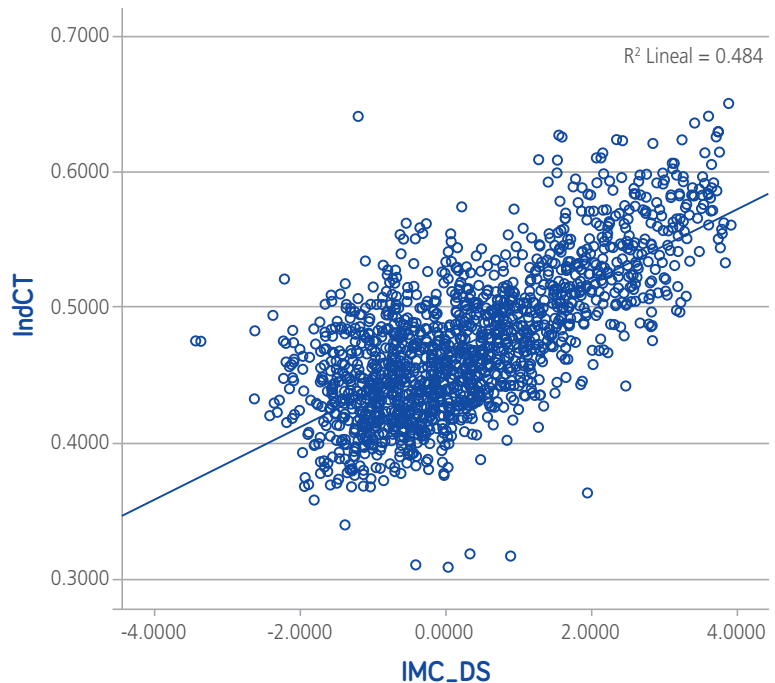
Sigamos los pasos para elegir la prueba más adecuada en la tabla 7.2:

1. Las variables independiente y dependiente son: el índice de cintura/talla y el IMC estandarizado (puntuación Z).
2. La variable independiente es continua (cuarta fila).
3. La variable dependiente es continua (tercera columna).

La opción que nos ofrece es la correlación de Pearson. En la figura 7.6 podemos ver la representación gráfica de la relación entre ambas variables y la recta de regresión que refleja la correlación.

Figura 7.6

Asociación entre dos variables continuas. Coeficiente de correlación



Una correlación perfecta directa (a mayor índice cintura/talla, mayor IMC) tendría un coeficiente de correlación de + 1 y una correlación perfecta inversa (a mayor índice de cintura/talla, menor IMC) de - 1. A la ausencia de correlación le corresponde un coeficiente de "0". Cuanto más se aleja el coeficiente de correlación de 0, mayor es la correlación. En este caso, el coeficiente de correlación de Pearson entre el índice de cintura/talla y el IMC estandarizado es de + 0,65 y significativamente distinto de "0" ($p < 0,001$). Por cada unidad de desviación estándar de IMC aumenta el índice cintura/talla 0,65 puntos.

Si alguna de las variables es ordinal o no sigue una distribución normal, tendríamos que emplear el test de correlación de Spearman, alternativa no paramétrica al de Pearson.

Otros contrastes

En el esquema de la tabla 7.2 hay otras pruebas de contraste de hipótesis que no vamos a abordar por separado, ya que ello nos obligaría a extender este documento. El lector interesado puede ver información al respecto en los textos de referencia.

Hay una serie de test de contraste de hipótesis que no se basan en parámetros que sigan las distribuciones de probabilidad habituales, por ello se conocen como pruebas no paramétricas. En vez de hacer las estimaciones en base a esos parámetros (media, varianza, etc.), comparan el conjunto de datos y hacen cálculos de probabilidad en función de la posición de cada uno de ellos en cada serie.

En la tabla 7.2 se detallan los más habituales. Para cada test paramétrico podemos encontrar uno o más test no paramétricos alternativos.

Al pie de la tabla se detallan otros procedimientos de contraste de hipótesis para análisis de supervivencia o análisis multivariante.

Finalizaremos este capítulo con una serie de preguntas para que el lector practique la elección del test estadístico. Nos ceñiremos a los test más habituales: ji cuadrado, t de Student para muestras independientes y relacionadas, ANOVA y coeficiente de correlación de Pearson.

Recomendamos seguir el esquema de la tabla 7.2, con el procedimiento simplificado en tres pasos:

1. Establecer las variables independiente y dependiente.
2. Elegir la fila en función de la escala de medida de la variable independiente. Hay cuatro opciones: nominal dicotómica con muestras independientes, nominal dicotómica con muestras apareadas, nominal politómica y continua.
3. Elegir la columna en función de la escala de medida de la variable dependiente. Hay tres opciones: nominal, ordinal y continua.

Pregunta 7.1

Queremos ver si los hábitos de consumo de zumos influyen en el riesgo de obesidad. Para ello diferenciamos a la población en función del consumo de zumos en cuatro grupos: consumidor habitual de zumos comercializados, consumidor de zumos naturales, consumidor habitual de ambos tipos de zumos y no consumidor habitual de zumos. A todos ellos les medimos el IMC. ¿Qué test de contraste de hipótesis elegiría?:

- a. Ji cuadrado.
- b. Correlación de Pearson.
- c. T de Student para muestras independientes.
- d. ANOVA.

Pregunta 7.2

Queremos ver si existen diferencias en los niveles de glucosa en ayunas entre diabéticos que usan pautas de insulina subcutánea intermitente y los que usan bomba de insulina. ¿Qué test de contraste de hipótesis elegiría?:

- a. Ji cuadrado.
- b. Correlación de Pearson.
- c. T de Student para muestras independientes.
- d. T de Student para muestras relacionadas.

Pregunta 7.3

Queremos ver si la educación materna (estudios elementales, medios o superiores) influye en la instauración de lactancia materna (lactancia materna exclusiva al mes de vida). ¿Qué test de contraste de hipótesis elegiría?:

- a. Ji cuadrado.
- b. Correlación de Pearson.
- c. T de Student para muestras independientes.
- d. T de Student para muestras relacionadas.

Pregunta 7.4

Queremos ver si hay asociación entre el tiempo de actividad física (medido en minutos/día mediante actimetría de muñeca durante 1 mes) y el IMC (medido al final del seguimiento). ¿Qué test de contraste de hipótesis elegiría?:

- a. Ji cuadrado.
- b. Correlación de Pearson.
- c. T de Student para muestras independientes.
- d. T de Student para muestras relacionadas.

Pregunta 7.5

Queremos ver si hay diferencias entre dos mediciones sucesivas realizadas de la presión arterial sistólica. ¿Qué test de contraste de hipótesis elegiría?:

- a. Ji cuadrado.
- b. Correlación de Pearson.
- c. T de Student para muestras independientes.
- d. T de Student para muestras relacionadas.

Respuestas y comentarios a las preguntas

Comentario a la pregunta 7.1

La respuesta más correcta es la "d", ANOVA.

La variable independiente es el consumo de frutas, que es una variable nominal politémica (tercera fila). La variable dependiente es el IMC, que es continua (tercera columna). Por lo tanto, el test más apropiado es ANOVA.

Comentario a la pregunta 7.2

La respuesta más correcta es la "c", t de Student para muestras independientes.

La variable independiente es el tipo de pauta de insulina, que es una variable nominal dicotómica que diferencia muestras independientes (primera fila). La variable dependiente es la glucosa, que es continua (tercera columna). Por lo tanto, el test más apropiado es la t de Student para muestras independientes.

Comentario a la pregunta 7.3

La respuesta más correcta es la "a", ji cuadrado.

La variable independiente es el nivel de estudios materno, que es una variable nominal politémica (tercera fila). La variable dependiente es la instauración de lactancia materna, que es nominal (primera columna). Por lo tanto, el test más apropiado es ji cuadrado.

Comentario a la pregunta 7.4

La respuesta más correcta es la "b", correlación de Pearson.

La variable independiente es el tiempo de actividad física, que es una variable continua (cuarta fila). La variable dependiente es el IMC, que es continua (tercera columna). Por lo tanto, el test más apropiado es el coeficiente de correlación de Pearson.

Comentario a la pregunta 7.5

La respuesta más correcta es la "d", t de Student para muestras relacionadas.

La variable independiente es la medición antes/después (advertimos que es una variable que no está como tal recogida en la base de datos), que es una variable nominal dicotómica que diferencia muestras relacionadas (segunda fila). La variable dependiente es la presión arterial sistólica, que es continua (tercera columna). Por lo tanto, el test más apropiado es la t de Student para muestras relacionadas.

BIBLIOGRAFÍA RECOMENDADA

- Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.
- Hu Y, He JR, Liu FH, Li WD, Lu JH, Xing YF, et al. Effectiveness of a Kindergarten-Based Intervention for Preventing Childhood Obesity. *Pediatrics*. 2017 Dec; 140(6). pii: e20171221.
- Marugán de Miguelsanz JM, Ochoa Sangrador C; Red de investigación de la SCCALP (Sociedad de Pediatría de Asturias, Cantabria, Castilla y León). Adecuación de los hábitos de introducción del gluten a las recomendaciones actuales. *An Pediatr (Barc)*. 2013;79(2):88-94.
- Milton JS. Estadística para biología y ciencias de la Salud. México: McGraw-Hill; 2001.
- Norman GR, Streiner DL. Bioestadística. México: Mosby/Doyma Libros; 1996.
- Riegelman RK, Hirsh RP. Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica. 2.ª ed. Washington, D.C.: Organización Panamericana de Salud; 1992;531.
- Rosner B. Fundamentals of Biostatistics, 7th Edition. Boston: Brooks/Cole, Cengage Learning; 2011.

8

Errores metodológicos

Objetivos docentes:

- Diferenciar entre errores aleatorios y sistemáticos.
- Conocer los tipos de errores sistemáticos.
- Conocer los tipos de sesgos de análisis.
- Aprender a identificar y controlar sesgos de confusión.
- Aprender a identificar y controlar sesgos de interacción.

Todo estudio epidemiológico ha de ser entendido como un ejercicio de medición en el que debe perseguirse la exactitud, esto es, la estimación del parámetro deseado con el mínimo error. Existen dos posibles fuentes de error: el error aleatorio y el error sistemático. Un estudio es preciso cuando carece de error aleatorio, es válido cuando carece de error sistemático.



Error aleatorio

Por error aleatorio (debido al azar) entendemos el error ligado al propio proceso de medición, esto es, a la falta de precisión. Esta falta de precisión o error aleatorio tendrá un valor distinto, e impredecible dentro de un rango, en cada una de las repeticiones de la medición. La variabilidad de una medición será, por tanto, un indicador de su precisión; cuanto más amplia sea dicha variabilidad, menor será su precisión. La variabilidad puede proceder de distintas fuentes: variabilidad biológica, imprecisión del instrumento de medida, inexperiencia del evaluador, etc.

Los errores aleatorios no suelen afectar a la validez interna de los estudios (no alteran la dirección de los resultados), pero sí limitan su potencia. En un estudio epidemiológico, la manera principal de reducir el error aleatorio consiste en aumentar el tamaño de la muestra. De esta manera minimizamos el riesgo de obtener resultados distorsionados

por azar, ya que la repetición de la medición tenderá a producir resultados distintos, pero cercanos, al valor verdadero del parámetro a medir.

En capítulos anteriores hemos explicado cómo podemos estimar el error aleatorio. Asumiendo que la repetición de un experimento con un mismo número de pacientes puede dar resultados diferentes por mero azar, la inferencia estadística nos permite cuantificar el rango de error a partir de medidas de dispersión de los resultados obtenidos y del tamaño muestral (error estándar). A menor dispersión de los resultados y mayor tamaño muestral, tendremos menor rango de error aleatorio.

Sea cual sea el parámetro elegido para expresar los resultados de un estudio, el rango de error de nuestra estimación puede expresarse como un **intervalo de confianza**, situado entre un valor por debajo y otro por encima del resultado obtenido. Cuanto más grado de confianza queramos atribuir a nuestra estimación de error, más amplio será el rango de dicho intervalo.

Habitualmente empleamos el intervalo de confianza al 95 %, cuya interpretación es: "si repetimos 100 veces un experimento en las mismas condiciones y con igual número de sujetos, el verdadero valor del parámetro poblacional que queremos estimar se encontrará incluido dentro de 95 intervalos de confianza de los 100 contruidos a partir de los resultados de cada experimento". Similar interpretación tendrían intervalos de confianza más exigentes.

La presentación de la incertidumbre, sobre el tamaño del efecto obtenido en un estudio, en forma de intervalos de confianza es el método preferido en las principales revistas científicas. A diferencia del nivel de significación obtenido en un contraste de hipótesis ("p"), que solo expresa la probabilidad de que la diferencia encontrada sea debida al azar (error tipo I o de falso positivo), el intervalo de confianza refleja la incertidumbre del resultado en las propias unidades de medida del efecto a evaluar, lo que facilita la interpretación de su relevancia clínica. Por otra parte, el intervalo de confianza permite juzgar la significación estadística del resultado, ya que solo será significativo cuando no esté comprendido en su rango el valor nulo de ausencia de efecto (por ejemplo: el "0" para diferencias de medias o riesgos, el "1" para medidas de riesgo).

También hemos visto que podemos medir el error aleatorio mediante el abordaje clásico que nos ofrece el **contraste de hipótesis**. El objetivo del contraste de hipótesis es permitir generalizaciones de los resultados de nuestra muestra a la población de la que procede. Cuando el resultado de un estudio en el que se comparan varias alternativas muestra diferencias, el siguiente paso es estimar si dichas diferencias corresponden a diferencias reales en la población o pueden ser explicadas por azar. Si la probabilidad de que el resultado obtenido debido al azar es muy baja, podremos asumir que corresponda a una diferencia real en la población, aunque siempre existirá una cierta probabilidad de error. Esta probabilidad será tanto más baja cuanto mayor sea la diferencia encontrada y más grande el tamaño muestral.

Por convención, si esta probabilidad es menor del 5 % ($p < 0,05$) se considera que no se debe al azar; en ese caso, se rechaza la hipótesis nula y se acepta la alternativa.

Esta decisión se tomará asumiendo siempre cierto margen de error, equivalente a la probabilidad anteriormente calculada (error tipo I, alfa) y expresada en el **valor "p" de significación**. No obstante, el valor "p" no ilustra la magnitud de los resultados obtenidos, como sí lo hace el intervalo de confianza; su nivel de significación, arbitrariamente elegido, va a verse muy influido por el tamaño de la muestra, al margen de cuál sea el efecto encontrado. Por ello va a ofrecer poca información sobre la relevancia clínica de los resultados.

En ocasiones, las diferencias encontradas en un estudio no alcanzan el nivel de significación del 5 %, porque la probabilidad de que los resultados sean debidos al azar es mayor ($p \geq 0,05$). En estos casos el contraste de hipótesis no permite rechazar la hipótesis nula. Sin embargo, la probabilidad de error tipo I puede ser $\geq 0,05$ (5 %) porque la diferencia real en la población sea pequeña, en cuyo caso no cometeremos ningún error, o bien porque hayamos obtenido una diferencia muy pequeña por azar, en cuyo caso estaremos cometiendo otro tipo de error: el error tipo II (beta). Este error será tanto más probable cuanto menor sea el tamaño muestral de nuestro estudio, lo que producirá un aumento del error aleatorio, que se traducirá en el cálculo de probabilidades. Como consecuencia, nuestro estudio tendrá poca **potencia** para encontrar diferencias, lo que se expresará como el complementario de la probabilidad de error tipo II (potencia = $1 - \text{beta}$). Si la potencia no alcanza al menos un 80 % (error tipo II menor del 20 %), nuestro estudio no tendrá la precisión suficiente para afirmar que no hay diferencias, debiéndonos plantear un ampliación de la muestra.



Errores sistemáticos

El error sistemático se corresponde con la falta de validez de un estudio. Un estudio es válido si "mide lo que realmente se quiere medir". La validez es, por tanto, la carencia de error sistemático. El error sistemático o sesgo (*bias*) es atribuible a errores metodológicos que, a diferencia del error aleatorio, no se reducen aumentando el tamaño muestral.

Resulta fundamental, al diseñar o analizar un estudio, valorar los sesgos que se hayan podido cometer: qué tipo de sesgos son, cuál es su magnitud y cómo pueden haber influido en los resultados. Tenemos que ser capaces, al menos, de identificarlos y estimar su posible repercusión. No siempre tendremos información como para estimar su magnitud, pero sí al menos tendremos que intuir su dirección, esto es, si los sesgos han podido aumentar o disminuir el efecto observado.

La validez de un estudio de investigación suele distinguir entre dos conceptos: validez interna y validez externa. Se entiende que hay **validez interna** cuando los resultados del estudio son válidos entre los propios sujetos del estudio. Por **validez externa** se

entiende la capacidad de generalización de los resultados, es decir, la capacidad de extender y aplicar las inferencias obtenidas a partir de ellos a una población diana (o población objetivo, también llamada externa) de referencia. Aunque una muestra mayor consigue una mayor precisión en las estimaciones, disminuyendo el error aleatorio, la representatividad de una muestra no puede definirse en un sentido estadístico, sino que requiere de un juicio racional sobre el parecido de la muestra a la población. Por otra parte, la existencia de validez interna es un requisito previo para que pueda darse la validez externa.

Podemos clasificar los sesgos en:

- Sesgos de selección.
- Sesgos de información (o clasificación).
- Sesgos de análisis (confusión e interacción).

Sesgos de selección

Los sesgos de selección se producen cuando la muestra de la que se obtienen los resultados no es representativa de la población, pero no por efecto del azar, sino por un error en los criterios de inclusión o exclusión, o en la sistemática de reclutamiento.

Estos sesgos afectan solo a la validez externa, ya que el diseño y análisis de los resultados pueden ser correctos, pero las conclusiones solo pueden aplicarse a la muestra estudiada. Este tipo de sesgos puede darse en diversas situaciones:

- **Selección incorrecta de los grupos de estudio:** si evaluamos el efecto de un factor de exposición o intervención, los grupos comparados solo deberían diferenciarse en el factor evaluado; sin embargo, estos grupos pueden presentar diferencias en otros factores implicados en el efecto. Esta comparabilidad se consigue en los estudios experimentales con la aleatorización, por lo que los estudios observacionales son más susceptibles a este tipo de sesgos.
- **Pérdidas durante el seguimiento:** cuando los que se pierden son sistemáticamente diferentes de los que no en variables relacionadas con el factor de estudio o la variable respuesta. Es lo que se denomina pérdidas informativas.
- **Ausencia de respuesta:** esto es propio de los estudios transversales. Hay que tratar de recoger información para saber si los que no contestan difieren en algo de los que sí lo hacen.
- **Supervivencia selectiva:** se produce cuando se seleccionan casos prevalentes en lugar de incidentes. Este tipo de sesgo es típico de los estudios de casos y controles, en los que los casos más graves o más leves están infrarrepresentados por *exitus* o curación.
- **Sesgo de autoselección por participación de voluntarios:** en general, existe riesgo de que estos individuos tengan características diferentes a los no voluntarios.

Sesgos de información o clasificación

Los sesgos de información o clasificación consisten en la distorsión del efecto estimado por errores en la medición y/o clasificación de los sujetos en una o más variables. Algunos ejemplos son: instrumentos de medición no válidos o no calibrados, criterios diagnósticos inapropiados o cambiantes, inconsistencia entre los evaluadores, ausencia de interpretación ciega de la exposición o del efecto, etc.

La influencia de estos errores en la clasificación del nivel de exposición o en el efecto puede distorsionar nuestros resultados, especialmente cuando los errores sean diferenciales. Entendemos por **error de clasificación diferencial** el producido cuando el criterio de clasificación en uno de los grupos que se comparan es diferente del criterio utilizado en el otro grupo. El efecto de este error suele traducirse en que los sujetos de un grupo sean clasificados incorrectamente y los del otro correctamente. La estimación del efecto puede verse aumentada o disminuida, produciéndose una distorsión en contra o a favor de la hipótesis nula.

El error de clasificación no diferencial es el cometido cuando se clasifica incorrectamente a los dos grupos por igual. La clasificación de exposición o de enfermedad es errónea para iguales proporciones de sujetos en los dos grupos de comparación. Este error tiende a producir una subestimación del efecto, esto es, una distorsión a favor de la hipótesis nula y, en consecuencia, genera resultados más conservadores. Por ello, el error no diferencial ocasiona menos problemas que el error diferencial.

Dentro de los sesgos de información podemos hacer referencia a dos muy típicos de los estudios de casos y controles (también de los de cohortes retrospectivas): sesgo de memoria (los que tienen un problema de salud recuerdan sus antecedentes de forma diferente a los que no lo tienen; por ejemplo: madres de recién nacidos malformados) y sesgo del entrevistador (la información se solicita o interpreta de distinta manera según el grupo al que pertenece el sujeto; por ejemplo: evaluadores no cegados al factor de estudio).

Sesgos de análisis

Mientras los sesgos de selección o clasificación solo pueden ser previstos y corregidos en la fase de diseño, otros errores pueden ser controlados en la fase de análisis. De estos **sesgos de análisis**, los principales son la confusión y la interacción.

Confusión

Cuando el efecto estimado de la relación exposición-enfermedad está distorsionado por la presencia de algún factor extraño en el análisis de los datos, decimos que existe confusión. El estimador del efecto de interés está distorsionado por estar mezclado con el efecto producido por un factor extraño que llamamos factor (o variable) de confusión. La presencia de tal factor de confusión altera los resultados y puede producir un aumento o disminución del efecto, o incluso cambiar la dirección del mismo.

Para que un factor sea de confusión necesita cumplir unos requisitos:

1. Debe ser factor de riesgo independiente de la enfermedad tanto para los expuestos como para los no expuestos en la población.
2. Debe estar asociado al factor de exposición que estudiamos en la población de donde provienen los casos.
3. No debe ser una consecuencia de dicha exposición, es decir, no debe ser un paso intermedio en la cadena o secuencia de causalidad entre la exposición y la enfermedad.

Veamos un ejemplo: en un estudio en el que queremos estimar el beneficio de la lactancia materna para evitar enfermedades infecciosas, tendremos que controlar la clase social de la familia, ya que podría comportarse como factor de confusión. La clase social cumpliría los requisitos de un factor de confusión, ya que se asocia a la enfermedad (riesgo de infección), al influir en los cuidados higiénicos; asimismo, se asocia a la exposición (tipo de lactancia) y no forma parte de la cadena causal entre exposición y enfermedad (el tipo de lactancia no produce cambios de clase social que puedan afectar a la incidencia de infección).

Para evitar sesgos de confusión podemos recurrir a una serie de técnicas. La primera y fundamental, solo disponible en estudios experimentales, es la aleatorización. Esta posibilidad que las variables de confusión conocidas y, lo mejor, desconocidas se distribuyan de manera homogénea entre los dos grupos.

Cuando no sea posible aleatorizar, podemos restringir los criterios de inclusión en función de la variable confusora o utilizar datos pareados en función de la misma. El problema, como es lógico, es que esto solo vale para las variables de confusión conocidas, pero no para las desconocidas.

Además, existen métodos para controlar el efecto de las variables de confusión durante la fase de análisis, como el uso de medidas de asociación ajustadas o los métodos de análisis estratificado o multivariante.

Si sospechamos que existe un factor de confusión en nuestro análisis, la forma más sencilla de comprobarlo es hacer un análisis estratificado, haciendo un análisis por separado de los sujetos con o sin dicho factor de confusión. Si la asociación entre exposición y enfermedad disminuye en cada estrato (por ejemplo: riesgos relativos u *odds ratios* que se acercan a 1) con respecto a la existente en el análisis global, existe un factor de confusión. Si hacemos una ponderación de las estimaciones de asociación de cada estrato, obtendremos la estimación verdadera de la asociación entre exposición y enfermedad. Aunque no es habitual, la confusión también podría producirse en el sentido contrario, con un aumento del efecto en cada estrato y, por lo tanto, en la estimación ajustada global.

Veamos otro ejemplo: en un estudio de cohortes se siguieron a 482 niños asmáticos, midiendo la incidencia acumulada de crisis asmática en un periodo en función de la exposición o no a tabaco en el hogar. Presentaron crisis un 62,5 % (60/96) de los ex-

puestos a tabaco en el hogar, frente a un 36,2 % (140/386) de los no expuestos. Ese mayor riesgo correspondía a un riesgo relativo (RR) de 1,72 (IC 95 %: 1,40 a 2,11), que supone un aumento estadísticamente significativo del 72 %. Para poder estar seguros de que esa asociación no es la consecuencia de la existencia de algún factor de confusión, se valoraron otras posibles variables. En la tabla 8.1 se presentan las frecuencias de crisis en función de la exposición a tabaco, estratificando por exposición alta o baja a contaminación ambiental.

Tabla 8.1 Frecuencias de crisis asmáticas en función de exposición a tabaco en el hogar estratificando por exposición a contaminación ambiental. Datos figurados

Contaminación ambiental	Exposición a tabaco en hogar	Crisis asmática		
		Sí	No	
Alta	Sí	50 (71,4 %)	20	70
	No	50 (62,5 %)	30	80
	Subtotal	100 (66,6 %)	50	150
Baja	Sí	10 (38,4 %)	16	26
	No	90 (29,4 %)	216	306
	Subtotal	100 (30,1 %)	232	332

Lo primero que llama la atención es que la diferencia de riesgo entre los expuestos a tabaco y los no expuestos ha descendido considerablemente en ambos estratos, aunque sigue siendo mayor en los expuestos. De hecho, en ninguno de los dos estratos la diferencia es estadísticamente significativa (en ambos, el valor nulo “1” está incluido en el intervalo): en el estrato de alta contaminación el RR del tabaco es 1,14 (IC 95 %: 0,91 a 1,43) y en el estrato de baja contaminación el RR es 1,30 (IC 95 %: 0,78 a 2,19). Haciendo el cálculo combinado de ambos estratos, el RR ajustado es 1,18 (IC 95 %: 0,95 a 1,45). En este ejemplo, la contaminación ambiental se comporta como factor de confusión en la relación entre tabaco y asma. Cuando el efecto observado en el análisis crudo (sin ajustar) se modifica en ambos estratos en la misma dirección, habitualmente atenuándose, respecto al valor nulo (en este caso al 1), será un factor de confusión.

Interacción

El otro sesgo de análisis que debemos controlar es el de modificación del efecto o interacción. Este sesgo ocurre cuando el efecto producido por una determinada variable se ve modificado en presencia de una tercera variable, produciéndose efectos de distinta intensidad en la misma o distinta dirección. Este fenómeno debe ser distinguido del de confusión y también controlado en el análisis.

Veamos un ejemplo: en una cohorte estudiamos el efecto de la lactancia materna exclusiva durante más de 4 meses sobre el riesgo de sibilantes recurrentes en los 3 primeros años de vida. El 13,8 % (80/579) de los que recibieron lactancia materna exclusiva menos de 4 meses tuvieron sibilantes recurrentes, frente a un 8,5 % (27/316) de los que la recibieron más de 4 meses. La lactancia materna menor de 4 meses se asoció a un aumento del riesgo del 61 % (RR: 1,61; IC 95 %: 1,06 a 2,44). Sin embargo, queríamos saber la influencia de otros factores en esa relación, en concreto el antecedente de asma en la madre.

En la tabla 8.2 se presentan las frecuencias de sibilantes recurrentes en función de la lactancia exclusiva menos o más de 4 meses, estratificando por antecedente de asma en la madre.

Tabla 8.2 Frecuencias de sibilantes recurrentes asmáticas en función de la lactancia materna, estratificando por asma materno

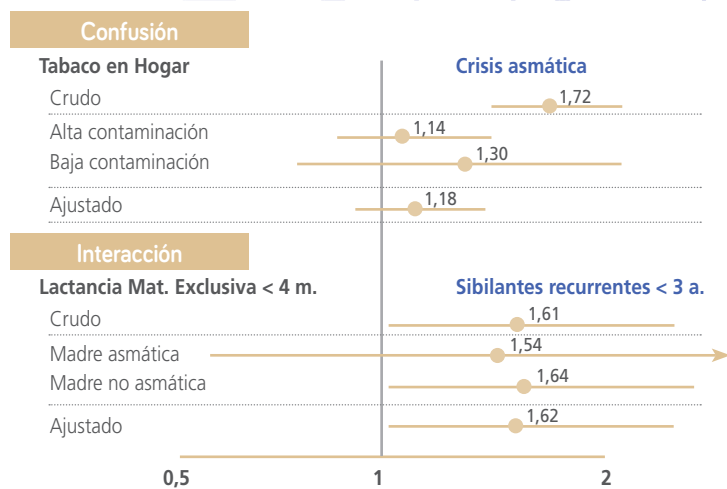
Asma materno	Lactancia materna exclusiva	Sibilantes recurrentes < 3 años		
		Sí	No	
Sí	< 4 meses	13 (20,6 %)	50	63
	≥ 4 meses	5 (13,5 %)	32	37
	Subtotal	18 (18,0 %)	82	100
No	< 4 meses	67 (13,0 %)	449	516
	≥ 4 meses	22 (7,9 %)	257	279
	Subtotal	89 (11,2 %)	706	795

Lo primero que observamos es que se mantiene en cada estrato la jerarquía de riesgos, aproximadamente un 6-7 % mayor en los que no tuvieron lactancia materna mayor de 4 meses, aunque en el estrato de asma materno los riesgos son mayores. Esto se traduce en estimaciones de riesgo con pequeñas diferencias: el RR de lactancia menor de 4 meses para sibilantes recurrentes es menor en el estrato de asma materno (RR: 1,52; IC 95 %: 0,59 a 3,94) que en el estrato sin asma materno (RR: 1,64; IC 95 %: 1,04 a 2,60). Por limitaciones en el tamaño muestral, en el primer estrato ese riesgo no es estadísticamente significativo (el intervalo de confianza incluye el 1). Haciendo el cálculo combinado de ambos estratos, el RR ajustado es 1,62 (IC 95 %: 1,07 a 2,45), similar al estimado crudo global anteriormente. En este ejemplo, el asma materno se comporta como factor de interacción o modificador del efecto.

En la figura 8.1 se presentan las estimaciones crudas, por estratos y ajustadas, para los ejemplos de confusión e interacción de las tablas 8.1 y 8.2. Si comparamos las estimaciones crudas y ajustadas con los factores de confusión, son diferentes, mientras que con los de interacción son similares. Además, podemos ver que las estimaciones de cada estrato se acercan ambas al valor nulo en el factor de confusión, mientras que con el de interacción una se aleja y la otra se acerca.

Figura 8.1

Ejemplos de estimaciones crudas, por estratos y ajustadas, de un factor de confusión y otro de interacción



Es muy importante tener en cuenta que confusión e interacción se interpretan de forma distinta. Siempre que encontremos una variable de confusión, nuestro objetivo será controlar su efecto y tratar de estimar una medida de asociación ajustada. Sin embargo, las variables modificadoras de efecto reflejan una característica de la relación entre exposición y efecto cuya intensidad depende de la variable modificadora. En estos casos, aunque podemos calcular una medida ajustada, como hacemos con los factores de confusión, esta describe peor la asociación entre exposición y efecto que las estimaciones por estratos. Por lo tanto, lo que debemos hacer cuando descubrimos una modificación de efecto es describirla y tratar de interpretarla.



Errores metodológicos en función del tipo de diseño

En función del tipo de diseño de un estudio podemos prever qué tipos de errores son más comunes. Conocer y anticipar estos errores nos va a facilitar la puesta en marcha de estrategias para evitarlos o controlarlos. Estos errores podrán ser de selección, de clasificación o de análisis.

En los estudios transversales el principal sesgo a evitar es el de no respuesta. Si elaboramos una encuesta para que un grupo de médicos la contesten y solo obtenemos respuesta de una parte de la muestra de estudio, incurriremos en un sesgo de selección. En este caso tendremos que tratar de explorar si la muestra final se diferencia mucho de la inicialmente proyectada o de la población de muestreo (por ejemplo: por edad y sexo). En todo caso, porcentajes de no respuesta muy altos pueden comprometer la validez externa del estudio.

En los estudios de cohortes el principal error que podemos tener es la pérdida de sujetos en el seguimiento. Si las pérdidas son altas, se compromete la representatividad de la muestra. Asimismo, en estos estudios estamos expuestos a sesgos de clasificación; así ocurre si no se utilizan procedimientos de cegamiento de la evaluación del efecto. Aunque en estos estudios la exposición no se controla y es conocida al inicio del seguimiento, se puede planificar un procedimiento por el cual el responsable de evaluar el efecto final desconozca el tipo de exposición.

En los estudios de casos y controles también se pueden cometer sesgos de selección, especialmente en la inclusión de controles. Si los controles presentan características diferentes a las de los sujetos de la población de la que proceden los casos, y estas características se comportan como factores de confusión, pueden distorsionarse los resultados. Por ello, en estos estudios resulta fundamental controlar los posibles factores de confusión en el análisis.

En los ensayos clínicos también se pueden cometer sesgos de selección en el proceso de aleatorización. Esto ocurrirá si no se utiliza un procedimiento correcto de aleatorización que garantice la homogeneidad de los grupos a comparar; para ello se recomienda emplear procedimientos de ocultación de la secuencia de aleatorización. Al igual que en los estudios de cohortes, otro error a controlar es el de las pérdidas en el seguimiento; se acepta que pérdidas superiores al 20 % comprometen la validez del estudio. Si la intervención no es ciega, también podremos incurrir en sesgos de clasificación al medir el efecto. Otros errores propios de los ensayos clínicos se producen en el momento del análisis, si se excluyen o se cambian de grupo los sujetos que no han seguido la intervención a la que han sido asignados. Para evitar estos errores se utilizan estrategias de análisis, como el análisis por intención de tratar, en el que cada sujeto es analizado en el grupo al que se asignó, independientemente de que hiciera el tratamiento o no.

A continuación haremos unos ejercicios para practicar la comprensión de los diferentes tipos de errores.

Pregunta 8.1

En un estudio para ver la frecuencia de hipertensión arterial en la infancia se emplea un dispositivo no calibrado de medición. ¿Qué tipo de error se puede estar cometiendo?:

- a. Error aleatorio.
- b. Sesgo de selección.
- c. Sesgo de clasificación o información.
- d. Sesgo de confusión.

Pregunta 8.2

En un estudio para ver la frecuencia de consumo de alcohol en la adolescencia se pasa un cuestionario a los niños que acuden a un Servicio de Urgencias. ¿Qué tipo de error se puede estar cometiendo?:

- a. Error aleatorio.
- b. Sesgo de selección.
- c. Sesgo de clasificación o información.
- d. Sesgo de confusión.

Pregunta 8.3

En un estudio transversal se envía una encuesta a la lista de correo de una sociedad científica formada por 500 miembros. Se obtiene respuesta *online* de 200 sujetos. ¿Qué tipo de error se puede estar cometiendo?:

- a. Error aleatorio.
- b. Sesgo de selección.
- c. Sesgo de clasificación o información.
- d. Sesgo de confusión.

Pregunta 8.4

En un estudio para estimar el riesgo de asma en función de la contaminación ambiental por tráfico se encontró que los expuestos a contaminación tenían un 50 % más de riesgo (RR: 1,50). Haciendo un análisis estratificado en función de la exposición a tabaco durante la gestación, se encontró que el riesgo asociado a la contaminación era mayor entre los expuestos a tabaco en la gestación (RR: 2,5) y menor entre los no expuestos a tabaco (RR: 1,20). ¿Qué tipo de error se puede estar cometiendo?:

- a. Sesgo de selección.
- b. Sesgo de clasificación o información.
- c. Sesgo de confusión.
- d. Sesgo de interacción o modificación del efecto.

Pregunta 8.5

En un estudio para estimar el riesgo de asma a los 7 años de edad se encontró que los niños que habían acudido a guardería en el primer año de vida presentaron el doble de riesgo (RR: 2,00) que los que no habían acudido a guardería. Se hizo un análisis estratificado en función de la exposición a tabaco en el hogar. En el análisis estratificado no se encontró asociación entre guardería y asma, ni entre los expuestos a tabaco (RR: 1,01) ni entre los no expuestos (RR: 0,97). ¿Qué tipo de error se puede estar cometiendo?:

- a. Sesgo de selección.
- b. Sesgo de clasificación o información.
- c. Sesgo de confusión.
- d. Sesgo de interacción o modificación del efecto.

Pregunta 8.6

En un estudio transversal en escolares se preguntó por la existencia de síntomas de asma y la exposición a tabaco en el hogar. Se entrevistaron a 100 sujetos. De los no expuestos a tabaco, tenían síntomas de asma el 5 % (3/60), mientras que de los expuestos a tabaco, el 15 % (6/40). Se estimó una razón de prevalencias de 3,00 (IC 95 %: 0,79 a 11,30). ¿Qué tipo de error se puede estar cometiendo?:

- a. Error aleatorio.
- b. Sesgo de selección.
- c. Sesgo de clasificación o información.
- d. Sesgo de confusión.

Pregunta 8.7

En un estudio se observó una asociación entre la toma de paracetamol en el primer año de vida con el riesgo de asma a los 3-5 años. En el análisis global se encontró un aumento del 21 % del riesgo (*odds ratio* [OR]: 1,21; IC 95 %: 1,04 a 1,41). Se hizo un análisis estratificado separando a los niños que tuvieron infecciones respiratorias frecuentes de los que no. La estimación ajustada de OR, controlando por infecciones respiratoria frecuentes, fue 1,03 (IC 95 %: 0,88 a 1,22). ¿Qué tipo de error se puede estar cometiendo?:

- a. Sesgo de selección.
- b. Sesgo de clasificación o información.
- c. Sesgo de confusión.
- d. Sesgo de interacción o modificación del efecto.

Respuestas y comentarios a las preguntas

Comentario a la pregunta 8.1

La respuesta más correcta es la "c", Sesgo de clasificación o información.

La utilización de un instrumento de medición de presión arterial no calibrado puede llevarnos a clasificar como hipertenso al que no lo es o viceversa. Por lo tanto, asistimos a un sesgo de clasificación o información.

Comentario a la pregunta 8.2

La respuesta más correcta es la "b", Sesgo de selección.

La población que acude a un Servicio de Urgencias puede no ser representativa de la población, y en este estudio especialmente, ya que la ingesta de alcohol puede estar sobrerrepresentada. Por ello podemos incurrir en un sesgo de selección.

Comentario a la pregunta 8.3

La respuesta más correcta es la "b", Sesgo de selección.

Este error es habitual en estudios transversales. Es posible que la predisposición a colaborar en la encuesta seleccione una muestra sesgada de profesionales con un perfil diferenciado. En sí mismo, la colaboración de encuestas *online* sesga el perfil del que responde.

Comentario a la pregunta 8.4

La respuesta más correcta es la "d", Sesgo de interacción o modificación del efecto.

Como hemos visto, la exposición a tabaco en la gestación se comporta como un tercer factor implicado en el riesgo de asma. En esta ocasión parece que la contaminación sí que se asocia a asma, pero este riesgo es mayor en los que estuvieron expuestos a tabaco que en los que no. Por ello es un sesgo de interacción. Este análisis concluiría dando los riesgos estimados por separado.

Comentario a la pregunta 8.5

La respuesta más correcta es la "c", Sesgo de confusión.

Como hemos visto, la exposición a tabaco en el hogar se comporta como un tercer factor implicado en la relación entre guardería y asma. Al hacer el análisis estratificado desaparece la asociación, probablemente porque la exposición a tabaco fuera en la muestra de estudio más frecuente entre los que acudieron a guardería que en los que no. Si combinamos las estimaciones por separado, el RR resultante estaría cercano a 1.

Comentario a la pregunta 8.6

La respuesta más correcta es la "a", Error aleatorio.

En concreto, un error tipo II. En este estudio hemos encontrado una diferencia importante (10 %) de la presencia de síntomas de asma en función de la exposición o no a tabaco en el hogar. Sin embargo, la diferencia observada no resulta significativa porque el intervalo de confianza de la razón de prevalencias comprende el valor nulo ("1"). Esto ocurre porque el tamaño muestral es pequeño. De hecho, la potencia del estudio para detectar una diferencia del 10 % es del 40,9 % (calculado con Epidat 4.2), que no alcanza el 80 % requerido.

Comentario a la pregunta 8.7

La respuesta más correcta es la "c", Sesgo de confusión.

En este estudio la variable "Infecciones respiratorias frecuentes" se comporta como un factor de confusión en la relación entre la toma de paracetamol y el riesgo de asma. Podemos comprobar que cumple los requisitos de un factor de confusión: se asocia a la enfermedad (las infecciones de vías respiratorias bajas predisponen a episodios de sibilantes o se manifiestan como tal). Asimismo, se asocia a la exposición (inducen la toma de paracetamol) y no forma parte de la cadena causal entre exposición y enfermedad (la toma de paracetamol es consecuencia, no causa, de la infección respiratoria).

BIBLIOGRAFÍA RECOMENDADA

- Arezina R, Duolao W. Source and control of bias. En: Duolao W, Bakhai A (eds.). Clinical trials. A practical guide to design, analysis and reporting. Londres: Remedica; 2006;55-64.
- Argimón Pallás JM, Jiménez Villa J (eds.). Bases metodológicas de la investigación clínica y epidemiológica. En: Métodos de investigación clínica y epidemiológica. Madrid: Elsevier; 2004;8-15.
- Argimón Pallás JM, Jiménez Villa J (eds.). Confusión y modificación del efecto. En: Métodos de investigación clínica y epidemiológica. Madrid: Elsevier; 2004;278-88.
- Molina Arias M, Ochoa Sangrador C. Errores en epidemiología. Errores sistemáticos. Factores de confusión y modificación del efecto. Evid Pediatr. 2016;12:16.
- Murray KW, Duggan A. Understanding confounding in research. Pediatr Rev. 2010 Mar;31(3):124-6.
- Sordillo JE, Scirica CV, Rifas-Shiman SL, Gillman MW, Bunyavanich S, Camargo CA Jr, et al. Prenatal and infant exposure to acetaminophen and ibuprofen and the risk for wheeze and asthma in children. J Allergy Clin Immunol. 2015 Feb;135(2):441-8.
- Wright AL, Holberg CJ, Taussig LM, Martínez FD. Factors influencing the relation of infant feeding to asthma and recurrent wheeze in childhood. Thorax. 2001 Mar;56(3):192-7.

Formulario para la memoria de un proyecto de investigación

	MEMORIA DEL PROYECTO DE INVESTIGACIÓN	FECHA:
	TÍTULO	PÁGINA XX DE

TÍTULO

Escribir el título: se recomienda que incluya de forma breve los componentes de la pregunta de investigación (población, exposición o intervención, comparación, efecto o resultado).

Tipo del Proyecto:

Individual / Coordinado / Multicéntrico

Duración: XX (años)

Investigador principal:

• Nombre, Apellidos	• Servicio e institución de trabajo
• NIF	• Correo electrónico

Investigadores colaboradores:

• Nombre, Apellidos, NIF	• Servicio e institución de trabajo
• Nombre, Apellidos, NIF	• Servicio e institución de trabajo
• Nombre, Apellidos, NIF	• Servicio e institución de trabajo
• Nombre, Apellidos, NIF	• Servicio e institución de trabajo
• Nombre, Apellidos, NIF	• Servicio e institución de trabajo
• ...	• ...

Añadir cuantas filas sean necesarias.
Anexo con hoja de firmas o Cartas de adhesión.

	MEMORIA DEL PROYECTO DE INVESTIGACIÓN	FECHA:
	TÍTULO	PÁGINA XX DE

RESUMEN (Objetivos y Metodología)

- Objetivos:
- Diseño: (Estudio transversal / Estudio ecológico / Estudio de casos y controles / Estudio de cohortes / Ensayo clínico; si es diseño mixto, detallar si descriptivo / analítico, concurrente / histórico, observacional / experimental).
- Ámbito de estudio: Hospital / Comunitario / Mixto, Unicéntrico / Multicéntrico.
- Población de estudio: Población diana, población de estudio.
- Muestra de estudio: Tamaño muestral, estrategia de muestreo.
- Instrumentación: Procedimientos, Variables de exposición y efecto.
- Determinación (Análisis): Medidas de frecuencia, riesgo e impacto / Medidas de validez o precisión; técnicas estadísticas.

ANTECEDENTES Y JUSTIFICACIÓN

- Antecedentes fundamentados en citas bibliográficas. Cuando sea apropiado, descripción de la búsqueda bibliográfica en la que se sustenta. Si existe experiencia previa por parte del grupo investigador, debe citarse.
- Este apartado no debe ser extenso (limitado en la mayoría de los formularios oficiales).
- Justificación de la pregunta de investigación: relevancia, interés, originalidad y factibilidad (ética y material). ¿Qué puedo aportar? (utilidad sobre el conocimiento, la práctica clínica, el paciente, el sistema sanitario o la sociedad).

BIBLIOGRAFÍA

Referencias bibliográficas siguiendo el estilo Vancouver.

HIPÓTESIS

- Formulación de la hipótesis de estudio siguiendo los componentes de la pregunta de investigación (PICoR) (PECOt): Población, Exposición o Intervención, (Comparación), Resultado (*outcome*), (tiempo).
- Ejemplos:
 - *Ejemplo A* (estudio observacional): en hermanos con enfermedad celíaca (E), respecto la población general (C) el riesgo de desarrollar la enfermedad es superior (R).
 - *Ejemplo B* (estudios experimental): en pacientes con bronquiolitis aguda leve-moderada (P), el tratamiento con aerosoles con suero salino hipertónico (I), en comparación con suero salino fisiológico (C), reduce el riesgo de precisar suplemento de oxígeno (R).
- En ocasiones, en estudios descriptivos exploratorios no se puede plantear una hipótesis previa. Puede plantearse un contraste teórico con la información disponible en la literatura.

	MEMORIA DEL PROYECTO DE INVESTIGACIÓN	FECHA:
		TÍTULO
		PÁGINA XX DE

OBJETIVOS

- Desarrollar la hipótesis de estudio en forma de objetivos: determinar frecuencia / riesgo / pronóstico / eficacia / efectividad / validez / precisión.
- Opciones:
 - Frecuencia: Incidencia / Densidad de incidencia / Prevalencia.
 - Riesgo: Riesgo relativo / *Odds ratio* / etc.
 - Impacto: Riesgo o proporción atribuible o prevenible.
 - Pronóstico: Supervivencia.
 - Eficacia / Efectividad de procedimientos terapéuticos: Reducción absoluta del riesgo, Número necesario a tratar.
 - Coste: Coste-efectividad, coste-utilidad, coste-beneficio.
 - Validez: Sensibilidad, especificidad, cocientes de probabilidades.
 - Precisión: Concordancia (kappa, coeficiente de correlación intraclase, etc.).
- Establecer el objetivo principal (medida de efecto principal) y los objetivos secundarios, enumerándolos.
- Los objetivos deben ser operativos concretando cada componente de la pregunta de investigación.

Ejemplo A: el riesgo de enfermedad celíaca es, al menos, el “doble” del de la población general a los 10 años de seguimiento. Esta operativización permitirá posteriormente estimar el tamaño muestral.

MÉTODOS

Diseño:

- Tipo de estudio: Descriptivo / analítico (existe grupo de comparación), concurrente (prospectivo) / histórico (retrospectivo), observacional / experimental.
- Diseño: Estudio transversal / Estudio ecológico / Estudio de casos y controles / Estudio de cohortes / Ensayo clínico.
- Otros estudios: Evaluación de pruebas diagnósticas / Revisión sistemática.

Sujetos de estudio:

- Población de estudio: descripción de la población accesible y susceptible de entrar en el estudio (diferente de población objeto del estudio en general o población diana).
- Muestra de estudio:
 - Tamaño muestral: justificación del tamaño muestral en función de la hipótesis y objetivos del estudio; para el cálculo del tamaño muestral se debe detallar (a partir de estudios previos) la medida principal de efecto, su precisión o la diferencia esperada entre grupos.

Ejemplo A: riesgo esperado de desarrollar enfermedad celíaca en hermanos de enfermos del 10 %. Precisión requerida $\pm 5\%$.

	MEMORIA DEL PROYECTO DE INVESTIGACIÓN	FECHA:
	TÍTULO	PÁGINA XX DE

Ejemplo B: diferencia esperada en el porcentaje de pacientes con bronquiolitis que requieren suplemento de oxígeno en función de que sean tratados con aerosoles de suero salino hipertónico frente a suero fisiológico del 15 % (riesgo basal en tratados con suero fisiológico del 30 %).

- Criterios de inclusión: deben describirse de tal manera que permitiera a una tercera persona reproducir el estudio o juzgar si un determinado paciente estaría incluido en él.
- Criterios de exclusión.
- Sistemática de muestreo: en los estudios en los que se haga una selección de sujetos entre la población de estudio debe detallarse el tipo de muestreo (muestreo sistemático o consecutivo / aleatorio simple / aleatorio estratificado / aleatorio por conglomerados, etc.) y los procedimientos operativos.

Variables de estudio:

- Definir todas las variables de estudio: nombre de la variable; tipo de variable (cualitativa o discreta / cuantitativa-continua); escala de medida (nominal / ordinal / continua de intervalos o razones), valores posibles (categorías o unidades de medida).
- Diferenciar las variables cuando corresponda en:
 - Variables de Exposición / Intervención (variable independiente). Podrían ser más de una.
 - Variables de Efecto / Resultado (variable dependiente):
 - Principal. Debe haber una variable de efecto o resultado principal.
 - Secundarias.

Ejemplo B:

- Variable de Exposición / Intervención: tipo de suero en el aerosol, variable cualitativa, discreta, de escala de medida nominal dicotómica (1, suero hipertónico; 2, suero fisiológico).
- Variable de Efecto / Resultado principal: necesidad de oxígeno, variable cualitativa, discreta, de escala de medida nominal dicotómica (1, precisa oxígeno; 2, no precisa).
- Variable de Efecto / Resultado secundaria: escala de gravedad de trabajo respiratorio (escala RDAI), variable cuantitativa, discreta, de escala de medida ordinal (valores enteros entre 0 y 17).
- En los estudios experimentales debe haber un apartado en el que se detallen los procedimientos de asignación aleatoria de tratamientos (ocultación de la secuencia de aleatorización) y de ocultación o enmascaramiento (ciego) de su aplicación.

Recogida de datos:

- Sistemática de recogida de datos: indicar quién recoge los datos (investigador, colaborador, autocumplimentado por el paciente, etc.), por qué medio (revisión de historias, registros informatizados, llamada telefónica, etc.).
- Indicar si se emplea algún método de enmascaramiento en la recogida de datos. Esto es especialmente importante en los estudios analíticos y, sobre todo, en los ensayos clínicos.

	MEMORIA DEL PROYECTO DE INVESTIGACIÓN	FECHA:
		PÁGINA XX DE

Análisis estadístico:

- **Descriptivo:** medidas de frecuencia de variables discretas y medidas de centralización y dispersión de variables continuas (media / mediana; desviación típica / rango intercuartílico):
 - Medidas de frecuencia: Prevalencia, Incidencia, Densidad de incidencia; Intervalos de confianza del 95 %.
 - Pronóstico: Supervivencia.
 - Coste: Coste-efectividad, coste-utilidad, coste-beneficio.
- **Analítico:**
 - Medidas de riesgo / impacto / validez / precisión:
 - Medidas de riesgo: Riesgo relativo / *Odds ratio* / etc.
 - Medidas de impacto: Riesgo o proporción atribuible o prevenible.
 - Eficacia / Efectividad de procedimientos terapéuticos: Reducción absoluta del riesgo, Número necesario a tratar.
 - Validez: Sensibilidad, especificidad, cocientes de probabilidades.
 - Precisión: Concordancia (kappa, coeficiente de correlación intraclase, etc.).
 - Contraste de hipótesis: opciones más frecuentes.
 - Dos variables discretas: test de ji cuadrado o pruebas exactas (Fisher) (para grupos relacionados test de McNemar).
 - Contraste de medias: test de la t de Student para muestras no relacionadas o relacionadas, análisis de la varianza o pruebas no paramétricas.
 - Supervivencia: *Log-rank* test.
 - Correlación entre variables cuantitativas: coeficiente de correlación de Pearson o Spearman.
 - Métodos de ajuste multivariante: regresión lineal múltiple (variable dependiente continua), regresión logística (variable dependiente discreta) o regresión de Cox (supervivencia).

UTILIDAD ESPERADA DEL ESTUDIO

- Si no se ha mencionado en el apartado de justificación, debe hacerse una reflexión sobre lo que puede aportar el estudio en cuanto al conocimiento previo, la práctica clínica, el paciente, el sistema sanitario o la sociedad.
- Las agencias de financiación de la investigación valoran especialmente la potencial utilidad económica de los resultados de investigación: patentes, reducción de costes, etc.

LIMITACIONES DEL ESTUDIO

- Comentar, si corresponde, las limitaciones posibles en relación con el tamaño muestral (falta de potencia estadística), la sistemática de muestreo (sesgos de selección), las variables medidas (sesgos de clasificación) o el análisis (sesgos de confusión o interacción).

	MEMORIA DEL PROYECTO DE INVESTIGACIÓN	FECHA:
	TÍTULO	PÁGINA XX DE

ASPECTOS ÉTICOS Y LEGALES

- Adecuación ética de los procedimientos.
- Consentimiento informado (Anexo).
- Aprobación por Comité Ético de Investigación Clínica (cuando corresponda; Anexo).
- Confidencialidad de datos personales.

CRONOGRAMA Y REPARTO DE TAREAS

- Cronograma de cada fase del estudio. Debe incluir la fase de análisis y explotación de resultados.
- Reparto de tareas.

EXPERIENCIA PREVIA DEL EQUIPO INVESTIGADOR

- Descripción de la experiencia previa de los investigadores en el tema: experiencia profesional, estudios realizados y su producción científica.

MEDIOS DISPONIBLES PARA LA REALIZACIÓN DEL TRABAJO

- Medios personales y materiales disponibles y su adecuación a los objetivos propuestos.
- Fuentes de financiación complementarias.

PRESUPUESTO DEL ESTUDIO

- En los trabajos en los que se solicita ayuda económica, justificación detallada de cada partida solicitada:
 - Gastos de personal: becarios y contratación de servicios. Antes de solicitar dinero para becarios, asegurarse de que la institución tiene recursos para su contratación (el personal del centro no puede recibir dinero en concepto de beca).
 - Viajes y dietas: se incluye habitualmente alguna cantidad para la asistencia a reuniones científicas.
 - Material inventariable (equipos): las ayudas para proyectos no deben usarse como vía de financiación de infraestructura; debe existir un equilibrio en la cantidad solicitada en este capítulo y otros gastos.
 - Material fungible: debe ser acorde con el tamaño muestral y los procedimientos previstos.
 - Otros gastos.
- En los trabajos en los que no se solicita ayuda económica, justificación de los costes estimados en carga de trabajo y material y fuente de financiación concertada. Si no se estiman costes por encima de la asistencia habitual, debe justificarse.
- Si el proyecto se ha diseñado para más de 1 año, debe detallarse por separado la financiación de cada año.

	MEMORIA DEL PROYECTO DE INVESTIGACIÓN	FECHA:
	TÍTULO	PÁGINA XX DE

Gastos de personal:	
Subtotal:	XXX.XXX,XX
Viajes y dietas:	
Subtotal:	XXX.XXX,XX
Material inventariable:	
Subtotal:	XXX.XXX,XX
Material fungible:	
Subtotal:	XXX.XXX,XX
Otros gastos:	
Subtotal:	XXX.XXX,XX
Total:	XXX.XXX,XX

ANEXOS

- Cuestionario de recogida de datos.
- Escalas auxiliares de recogida de datos.
- Consentimiento informado.
- Hoja de firmas o Cartas de adhesión de los colaboradores.
- Permisos de responsables de instituciones.
- Aprobación por Comité Ético de Investigación Clínica (cuando corresponda).
- Currículum de los investigadores.





Nutribén®

Especialistas en alimentación infantil

www.nutriben.es