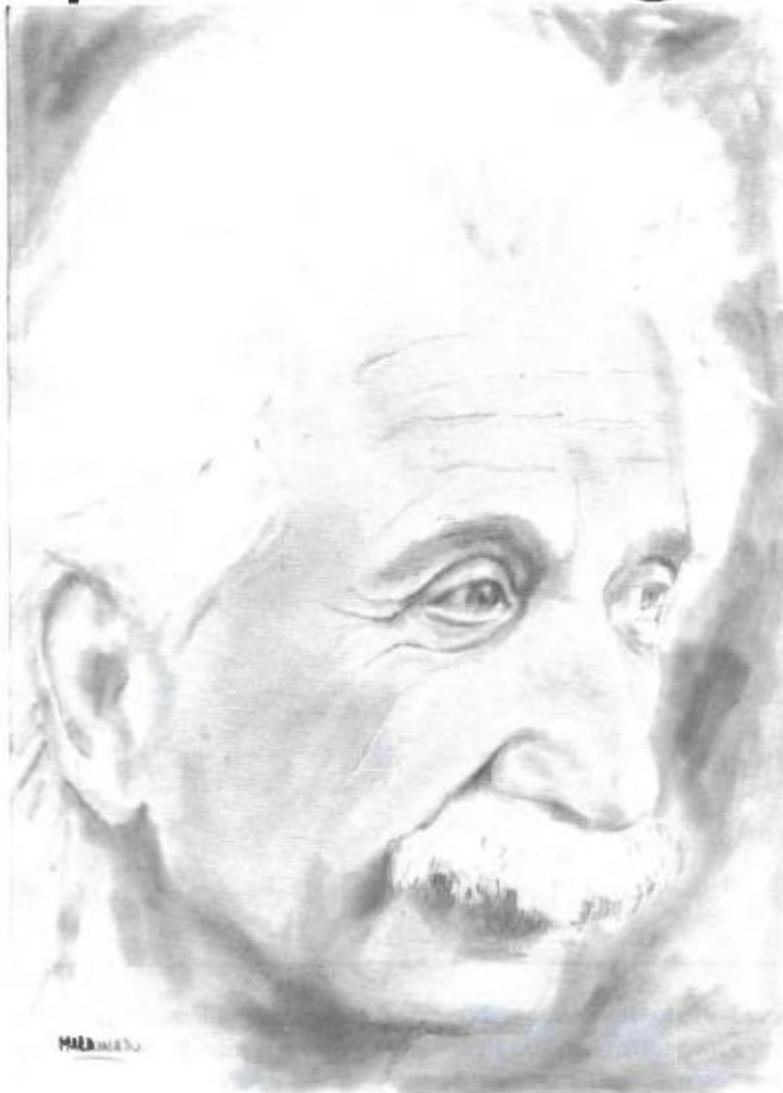


AAU

AMERICAN ANDRAGOGY
UNIVERSITY



Medicina Preventiva Bioestadística & Epidemiología

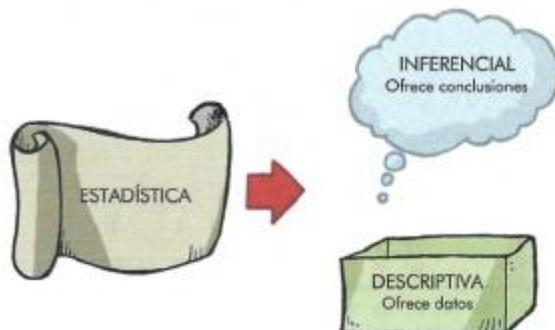


A lo largo de la historia el acto médico se ha basado en conocimientos que la mayor parte de las veces no han sido confirmados científicamente. La bioestadística es la herramienta que permite comprobar y demostrar la utilidad de un método diagnóstico o terapéutico.

1. Definición de estadística

La Estadística es el conjunto de métodos necesarios para recoger, clasificar, representar y resumir datos, así como para hacer inferencias —extraer conclusiones— objetivas y científicas a partir de ellos. Podemos dividirla en dos grandes apartados:

- Estadística Descriptiva.
- Estadística Inferencial.



Se describen dos tipos de estadística: descriptiva e inferencial

1.1. Estadística descriptiva

En primer lugar se precisa recoger información sobre el fenómeno estudiado y hacerla inteligible. Ello requiere sistematización, clasificación y representación visual de los datos, y su caracterización a través de una serie de medidas que resumen su comportamiento (básicamente, su tendencia central y su variabilidad).

1.2. Estadística inferencial

Con los datos ya caracterizados y sistematizados el siguiente objetivo es extraer conclusiones que proporcionen criterios para la toma de decisiones o sirvan para identificar nuevas preguntas a investigar.

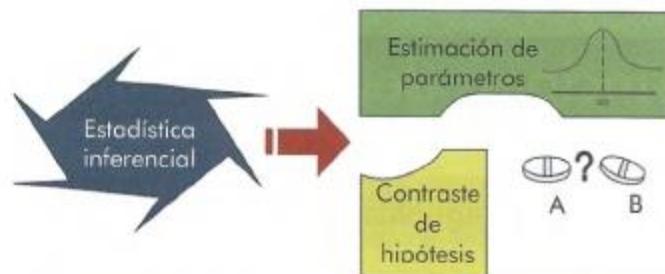
La estadística descriptiva nos da datos que apuntan a ciertas conclusiones y la inferencial nos dirá si esas conclusiones son correctas o no, calculando la probabilidad de acierto o error.

Por ejemplo, la decisión de indicar cierto tratamiento para determinada enfermedad se debe basar en la demostración de que ese tratamiento es efectivo y beneficioso frente a otras opciones o frente a la alternativa de no tratar. La demostración se basa en los datos obtenidos en un tipo de experimento denominado ensayo clínico aleatorio, en el se realizan las comparaciones entre el tratamiento propuesto y las demás alternativas. La decisión de determinar si el beneficio del nuevo tratamiento es significativo corresponde a la estadística inferencial.

Otro ejemplo de inferencia estadística consiste en determinar cuáles son los valores "normales" poblacionales de la concentración plasmática de un metabolito, de la tensión arterial, etc. medido por cierta prueba diagnóstica que se está evaluando.

Ambos ejemplos ilustran las dos aplicaciones fundamentales de la estadística:

- la estimación de medidas poblacionales, a partir de los datos obtenidos de una muestra, y
- el contraste de hipótesis, que se ocupa del estudio de comparaciones y asociaciones.



La estadística inferencial estima el valor poblacional o "verdadero" de un parámetro (la media de colesterolemia, un riesgo relativo, el tamaño de una diferencia) a partir de los datos medidos en una muestra. Para ello calcula un intervalo de confianza, que es un intervalo que contiene ese valor "verdadero" con una probabilidad alta. Por otra parte, proporciona criterios en el contraste de hipótesis para decidir acerca de la significación de una asociación o de una diferencia observada. Ambas aplicaciones se engranan, ya que utilizan algunos elementos comunes.

MIR 05 (8213): ¿Cuál de las siguientes afirmaciones es correcta en relación a la estadística?:

- La estadística tiene entre sus funciones la corrección de los errores y sesgos de un mal diseño.
- Ante un estudio con sesgos nos podemos fiar de los resultados si el valor de significación encontrado es razonablemente pequeño, por ejemplo $p < 0,0001$.
- La significación estadística es un criterio objetivo por lo que puede suplir la objetividad del juicio clínico.
- La mejor manera de investigar consiste en la obtención del mayor número p posibles, guiándonos por los resultados obtenidos.
- El valor de significación nos orienta sobre cual es la probabilidad de que la diferencia observada sea debida exclusivamente al azar.*

1.3. Variables

Una variable puede definirse como una característica observable y medible, que puede adquirir diferentes valores, y que representa el concepto que se quiere investigar.

En función de que utilicen o no valores numéricos se clasifican en:

a) Variables cuantitativas (adquieren valores numéricos):

- Cuantitativas discretas:** sólo pueden tomar ciertos valores, habitualmente números naturales o enteros, sin posibilidad de adoptar valores intermedios. Ejemplos: n° de hijos, n° ingresados en un hospital, etc.
- Cuantitativas continuas:** pueden tomar cualquier valor numérico, de modo que siempre cabe la posibilidad de infinitos valores dentro de un intervalo. Ejemplos: estatura, peso, etc.

b) Variables cualitativas (categorías no numéricas):

- Cualitativas nominales:** atributos que no mantienen entre sí ningún orden de sucesión o progresión. Cuando sólo existen dos posibilidades, hablaremos de variable dicotómica. Ejemplos: grupo sanguíneo (A, B, AB, O), resultado de un tratamiento (dicotómica: éxito, fracaso), etc.
- Cualitativas ordinales:** siguen algún tipo de gradación ordenada. Ejemplos: nivel socioeconómico (bajo, medio, alto), intensidad del dolor (sin dolor, dolor leve, moderado e intenso) (MIR), etc.

MIR 03 (7694): Se está realizando un estudio para conocer el efecto de un nuevo analgésico. La variable principal de respuesta es una escala de dolor con los siguientes valores: 1 (no dolor), 2 (dolor leve), 3 (dolor moderado) y 4 (dolor intenso). ¿De qué tipo de variable se trata?:

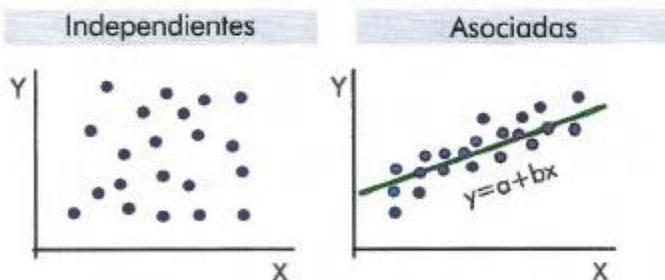
- Variable dicotómica.
- Variable continua.
- Variable discreta.
- Variable ordinal.*
- Variable cuantitativa.

MIR 08 (8992): Se realiza un estudio para determinar si se produce o no hemorragia digestiva con un determinado tipo de tratamiento. En este caso la variable principal del estudio es de tipo?:

1. Cualitativo continuo.
2. Categórico ordinal.
3. Categórico discreto.
4. Cualitativo binario.*
5. Categórico dependiente.

En numerosas ocasiones en bioestadística necesitaremos estudiar si dos (o más) variables están o no asociadas, es decir, si la variabilidad de una variable se sigue de una variabilidad en la otra variable, y esto no se debe al azar. El ejemplo típico es identificar factores de riesgo para el desarrollo de enfermedades. Y así clasificar la relación de nuestras dos variables del estudio en:

a) Variables asociadas (dependientes). En este caso las dos variables si estarían relacionadas de forma que una variable sería variable predictora de la otra, la variable desenlace. Ej: la variable fumar (predictora) se asocia con ca de pulmón (desenlace). Esta relación puede ser causal o no, pero están asociadas de forma significativamente distinta a la que se explicaría por el azar.



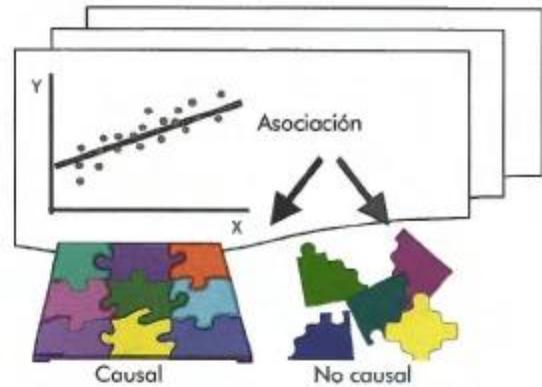
Dos o más variables pueden estar asociadas, depender una de otra, o ser independientes, y no tener relación entre sí

b) Variables no asociadas (independientes). Cuando dos variables son independientes asumimos que no están relacionadas. Sus valores varían de forma diferente.

2. Asociación y causalidad

2.1. Asociación ≠ causalidad

Es de gran importancia distinguir entre asociación y causalidad. La demostración de una asociación entre dos variables, no implica ni es sinónimo de una relación causa-efecto entre ambas. Existe la posibilidad de observar asociaciones sin que exista relación causal subyacente. Ello puede ocurrir en forma de asociaciones espúreas establecidas por sesgos, o bien en forma de asociaciones artificiales, por variables que evolucionan paralelamente a un factor causal.



El establecimiento de asociación entre variables no indica causalidad: puede ser una relación no causal, aspecto que se valorará con el estudio conjunto de los ciertos criterios de causalidad

Ej: ¿Es cierto que los dedos amarillos ocasionan cáncer de pulmón? No. Sin embargo, sí es cierta su asociación: el tabaco es causa de ambas, pero no hay relación causal entre dedos amarillos → cáncer de pulmón.



repeMIR

La asociación no implica ni es criterio de causalidad (2+)

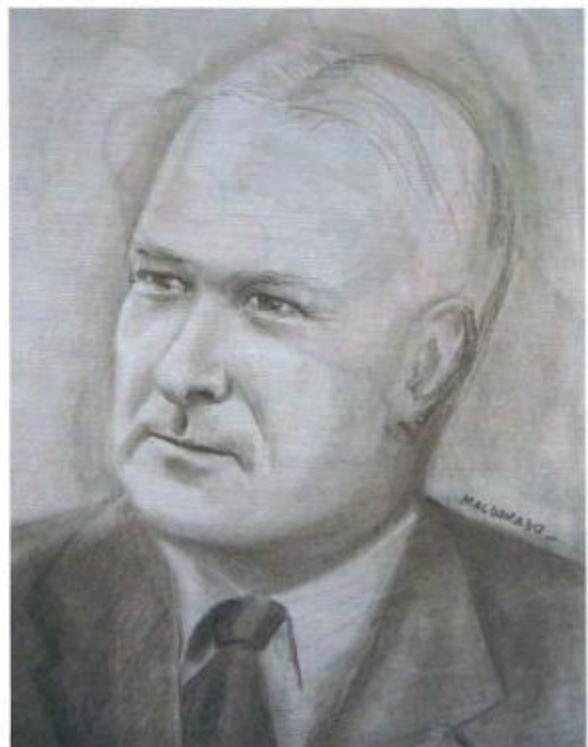
En general los fenómenos de la salud se consideran dentro de modelos causales multifactoriales. Es decir, un factor de riesgo identificado como factor causal debe interpretarse cuidadosamente dentro de una compleja trama de factores implicados.

Así para determinar una relación causal se acude a evaluar una serie de criterios epidemiológicos de causalidad, sobre los cuales se realiza una valoración conjunta.

2.2. Criterios de causalidad

¿Cómo reconocemos "la causa" de una enfermedad en bioestadística? No existen criterios únicos, ni necesarios ni suficientes (MIR) para dar respuesta a esta pregunta hoy en día. Existen varios modelos: determinista modificado, modelo de maraña o red, etc.

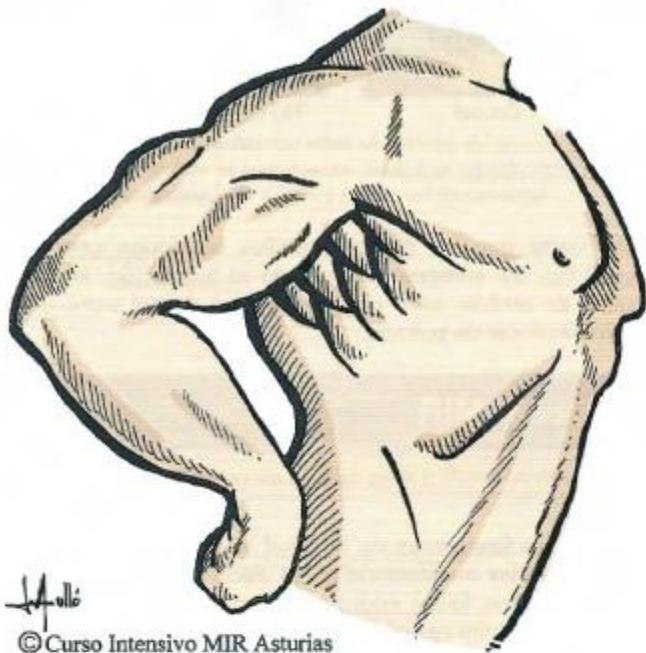
Los criterios más habitualmente utilizados son los propuestos por Bradford-Hill, que pueden clasificarse en criterios de validez "interna" (propios del estudio realizado) y en criterios de coherencia científica (la asociación hallada entre un factor causal y enfermedad no debe entrar en conflicto con lo que se sabe de la enfermedad).



Sir Austin Bradford Hill

A. FUERZA DE ASOCIACIÓN (5MIR) (MAGNITUD, INTENSIDAD DE LA ASOCIACIÓN O TAMAÑO DEL EFECTO):

A mayor intensidad de la relación entre dos variables, mayor es la probabilidad de que exista una relación causal entre ambas. Medidores de la fuerza de asociación: riesgo relativo (RR) y odds ratio (OR) (cuanto mayor sea el RR y la OR mayor será la fuerza de asociación). Otros medidores: coeficiente de correlación de Pearson, diferencia de riesgos, de proporciones o de medias...



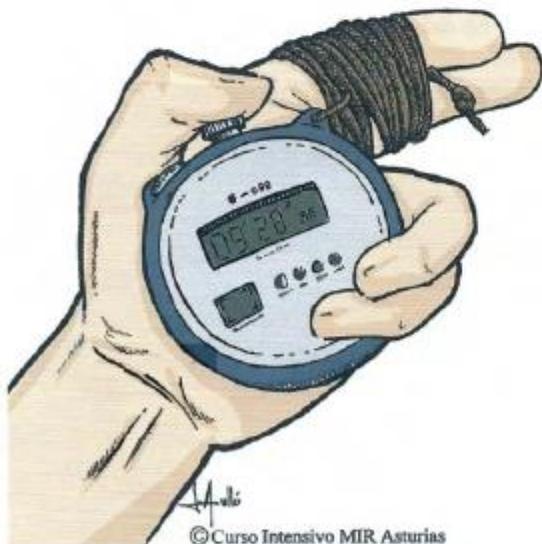
©Curso Intensivo MIR Asturias

Fuerza de la asociación (RR, OR)

No obstante, una asociación débil puede ser causal (ej consumo de cigarrillos-enfermedad cardiovascular, tabaquismo pasivo-cáncer de pulmón). Y una asociación muy fuerte puede deberse a un factor de confusión (ej dedos amarillos-cáncer de pulmón).

B. SECUENCIA TEMPORAL (4MIR) (TEMPORALIDAD):

La causa debe preceder al efecto, aunque puede ser difícil de establecer esta secuencia (¿la HTA es causa o consecuencia de la insuficiencia renal?). Es el único criterio considerado por algunos autores como necesario (condición sine qua non).



©Curso Intensivo MIR Asturias

Secuencia temporal

C. EFECTO DOSIS-RESPUESTA (3MIR) (GRADIENTE BIOLÓGICO):

Si un factor es causal, cuanto mayor es el tiempo y/o dosis de exposición al factor mayor es el riesgo de enfermedad (los fumadores moderados tienen tasas de incidencia de cáncer mayores que los no fumadores, y son aún mayores las de los grandes fumadores).



El término "dosis-respuesta", no se refiere a "respuesta al tratamiento" (MIR), esto no es ningún criterio de causalidad.

D. CONSISTENCIA (MIR):

Los resultados de un estudio deben mantenerse constantes y ser reproducibles por cualquier investigador y lugar. ¿Por qué es criterio de causalidad? Es poco probable que estudios diferentes con "errores" diferentes puedan llegar a una misma conclusión si ésta no fuera cierta.

E. PLAUSIBILIDAD BIOLÓGICA (5MIR):

La relación causal sugerida debe mantener la línea de los principios científicos aceptados en el momento. Por ejemplo, podría costarnos creer a priori que un microorganismo pueda originar daño no sólo a una célula, sino también a sus células hijas... hasta que descubrimos que es un virus con potencial carcinogénico.

Es decir creemos más en una relación causal si conocemos su mecanismo patogénico. Ej actual: colesterol y aterosclerosis; más verosímil tras teoría oxidativa de las LDL-colesterol.

F. ESPECIFICIDAD DE LA ASOCIACIÓN Y ANALOGÍA:

Si la relación entre un factor y un efecto es específica "B" sólo ocurre si ocurre "A", y la presencia de "A" sólo da lugar al efecto "B"- ello sugiere relación causal entre "A" y "B". La especificidad implica que una causa conduce a un único efecto, no a efectos múltiples; cierta especificidad aumenta la verosimilitud de la relación causal.



Especificidad de la asociación

La analogía postula que asociaciones causales similares pueden producir enfermedades similares. Ej: Si un fármaco puede ocasionar asma, quizá otro también pueda.

G. EVIDENCIA EXPERIMENTAL (EXPERIMENTO O CESACIÓN DE LA EXPOSICIÓN) (5MIR):

La demostración experimental es la prueba más sólida demostrativa de causalidad (4MIR), pero puede no ser posible realizar el tipo de estudio que se requeriría para ello.



©Curso Intensivo MIR Asturias



ATENCIÓN

¡Ojo!: que la causa sea vista posterior a una enfermedad no quiere decir que no pueda producir dicha enfermedad en los casos en que la precede (ej. un cardiópata tiene limitación de su actividad física y eso lo hace sedentario, pero eso no descarta que el sedentarismo vaya a ser un factor causal de la enfermedad cardiovascular).

La evidencia experimental es la prueba más sólida demostrativa de causalidad, fundamentalmente en forma de ensayo clínico aleatorio.

Una forma de salvar este problema es: si un factor produce un efecto determinado (enfermedad), este efecto debería cesar cuando desaparece el factor. Ej cáncer de mama-dieta grasa es una asociación que se postula como causal: apoyaría esta relación si, en un estudio experimental (ensayo clínico aleatorio), mujeres con dietas bajas en grasa, presentaron una menor incidencia de cáncer de mama.

Más que como criterio puede considerarse como un test de la propia hipótesis causal, dándole a la misma robustez.

repeMIR
El estudio clínico aleatorio evalúa fielmente la asociación causal entre un factor de riesgo y una enfermedad. (5+)

MIR 02 (7430): La mejor forma de verificar una hipótesis en epidemiología es a través de:

1. Un estudio descriptivo.
2. Un estudio experimental.*
3. Un estudio caso-control.
4. Un estudio de cohortes.
5. Un estudio transversal.

Criterios clásicos de causalidad de Bradford-Hill

De validez interna (propios del estudio):

- Fuerza de asociación: intensidad de relación.
- Secuencia temporal: causa precede al efecto.
- Efecto dosis-respuesta: mayor riesgo a mayor dosis y/o tiempo de exposición.

De coherencia científica (en sintonía con conocimientos científicos aceptados del momento):

- Consistencia: resultados reproducibles y constantes.
- Plausibilidad biológica: coherente con conocimientos actuales de la enfermedad.
- Especificidad de asociación (y analogía): A sólo si B, B sólo si A.
- Evidencia experimental/cesación de exposición: comprobación mediante ensayo clínico aleatorio.

repeMIR
No existen criterios únicos, ni necesarios ni suficientes (1+)

REGLA NEMOTÉCNICA
CRITERIOS DE VALIDEZ INTERNA

- **SE**cuencia **TE**mporal
- **FUE**rza de la asociación
- **E**fecto **DO**sis **R**espuesta

Sete fue un hedor (interno)



REGLA NEMOTÉCNICA

CRITERIOS DE COHERENCIA CIENTÍFICA

- **C**onsistencia, **CO**herencia
- **PLA**usibilidad biológica (credibilidad)
- **ESP**ecificidad de asociación, **A**nalogía
- Evidencia **experimental**

Co pla
española
experimental



2.3. Otros criterios de causalidad

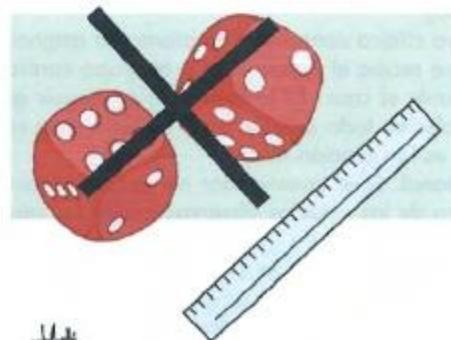
A. DESCARTE DE EXPLICACIONES ALTERNATIVAS:

Cuando se valora la relación entre un factor (presumiblemente causal) y un efecto (enfermedad) tenemos siempre que considerar la posibilidad de haber cometido algún tipo de sesgo, que dicha asociación se deba a algún factor de confusión o que se deba simplemente al azar (todas ellas explicaciones de la realidad alternativas de la causal), y por lo tanto que no exista esa asociación causal.

Por este motivo, algunos autores consideran primero la ausencia sesgos, factores de confusión y de azar (en este orden), en la evaluación de una hipótesis causal, y, si realmente no existe ninguno de estos criterios negativos, eso ya en sí mismo es un criterio importante de que nuestra asociación puede ser causal.

B. CRITERIO DE INDEPENDENCIA:

Es un criterio que se tiene en cuenta cuando se valoran factores de confusión: la asociación vista entre un factor de riesgo y una enfermedad no tiene por qué deberse enteramente al efecto de un factor de confusión, aunque se haya detectado en el estudio.



©Curso Intensivo MIR Asturias 2010
Establecer la ausencia de azar y sesgos en un estudio (incluyendo el sesgo de confusión), es un criterio de causalidad

Ej: Esto sucede así en el caso de la triple asociación de colesterol-tabaco-HTA en la cardiopatía coronaria. Los tres son factores causales y pueden comportarse como factores de confusión en un estudio, pero el colesterol será independiente si mantiene en cierto grado su efecto (sobre la HTA) tras haber controlado el efecto de la variable "tabaco" en el estudio.

C. CALIDAD DE LA EVIDENCIA (en la que apoyar recomendaciones en la práctica clínica):

Con dicho criterio jerarquizamos los diferentes tipos de estudios en función de su validez interna y la fuerza con que demuestran causalidad (grado en que controlan variables extrañas).

De más a menos: ensayo clínico > cohortes > casos y contro-

casos>informe de un caso... Es importante a la hora de valorar la aplicabilidad clínica de lo recomendado por un estudio (¿nos fiamos de él? ¿en qué grado?). Y así un estudio ecológico tiene menor validez que los estudios analíticos (casos y controles y cohortes) (MIR).

MIR 01 (7176): ¿Cuál de los siguientes estudio tiene MENOS validez para evaluar la existencia de una relación causal entre un posible factor de riesgo y una enfermedad?:

1. Estudio de casos y controles.
2. Estudios de casos y controles anidado en una cohorte.
3. Estudio ecológico o correlacional.*
4. Estudio de cohortes prospectivo.
5. Estudio de cohortes retrospectivo.

3. Tipos de estudios epidemiológicos

Los estudios epidemiológicos pueden clasificarse según varios criterios:

A. SEGÚN SU FINALIDAD PUEDEN SER:

- **Descriptivos**, si tienen como finalidad recoger las características de las variables que se están estudiando en la población, bien de persona, lugar o tiempo.
- **Análítico**, si estudia una posible relación causa-efecto.

B. SEGÚN LA PERSPECTIVA:

- **Transversal**. Se estudian las variables de una población en un momento determinado (realizando un "corte temporal"). Lo que se mide en este tipo de estudios es la prevalencia (número de casos en un momento determinado), y dado que se estudia un único momento de la población, es imposible establecer secuencias temporales. Es una "instantánea" de la situación.
- **Longitudinal**. Se analiza un periodo de tiempo. Puede servir para establecer una secuencia temporal, y lo que se mide es la incidencia (casos nuevos en un cierto tiempo). Es una "película" de la situación.

C. SEGÚN LA ASIGNACIÓN DE FACTORES DE ESTUDIO:

- **Experimental (ensayo clínico)**. Si el investigador asigna el factor que se estudia (por ejemplo, un tratamiento) a los distintos grupos. Pueden darse distintas situaciones:

1. **Ensayo clínico no controlado**. No hay grupo de control (p. ej. se estudia un grupo antes y después de un tratamiento).
2. **Ensayo clínico controlado**. Existe un grupo de control, sin el factor que se estudia (por ejemplo, un grupo sin tratamiento).
3. **Ensayo clínico controlado aleatorio**. La asignación al grupo que recibe el tratamiento o al grupo control se realiza mediante el azar. Es la forma de conseguir grupos comparables en todo excepto en el factor que se estudia y por tanto es la situación ideal.

- **Observacional**. Si el investigador no asigna el factor de estudio. Dentro de los estudios observacionales los grupos del estudio se pueden crear:

1. Teniendo en cuenta la exposición a un factor de riesgo (estudio de cohortes). Se investiga en cada grupo (expuestos y no expuestos) la frecuencia con que se desarrolla un determinado efecto.
2. Teniendo en cuenta la presencia del "efecto" (por ejemplo la enfermedad; estudio de casos y controles). Se investiga cada grupo (enfermos y no enfermos) la frecuencia con que se presenta un determinado factor de riesgo.

sarrollo de enfermedad o "efecto" después del inicio del estudio. Pero existen estudios de cohortes retrospectivas: un estudio con un grupo sometido a la radiación de la bomba de Hiroshima (1945) y otro no sometido a la radiación, donde se investiga el desarrollo de leucemias en los 20 años siguientes (1965) pero que se ha iniciado en 1995 con datos de un registro histórico.

- **Retrospectivo**: El inicio del estudio es posterior a la recogida de los datos (el investigador "mira hacia atrás"). Los estudios de casos y controles son retrospectivos: tenemos un grupo de casos y otro de controles e investigamos la presencia de un factor de riesgo en el pasado.

MIR 01 (7178): Si identificamos en la práctica habitual dos grupos de sujetos, uno expuesto a un fármaco o agente etiológico (por ejemplo, usuarias de contraceptivos orales) y otro no expuesto a este agente (por ejemplo, mujeres que utilizan otros métodos de contracepción) y los seguimos a intervalos regulares de tiempo para comparar la incidencia de tromboembolismo venoso entre uno y otro grupo, estamos planteando un:

1. Estudio de cohortes.*
2. Ensayo clínico secuencial.
3. Estudio caso-control.
4. Ensayo clínico pragmático.
5. Estudio ecológico.

MIR 01 (7175): Si identificamos en la práctica clínica habitual pacientes con agranulocitosis y comparamos su exposición a fármacos con la de un grupo de individuos de características similares pero que no padecen la enfermedad, estamos planteando un:

1. Estudio de cohortes.
2. Ensayo clínico secuencial.
3. Estudio caso-control.*
4. Ensayo clínico pragmático.
5. Estudio ecológico.

MIR 03 (7696): Identificamos en la práctica clínica habitual una población definida de pacientes con artritis reumatoide que no han respondido a un tratamiento con metotrexate, y que según criterio clínico reciben tratamiento o bien con etanercept o bien con infliximab. Seguimos la evolución de la respuesta clínica de la enfermedad a largo plazo con el objetivo de realizar una comparación entre ambos grupos. En este caso estamos planteando un:

1. Estudio de cohortes.*
2. Estudio coste-efectividad.
3. Estudio caso-control.
4. Ensayo clínico pragmático o naturalístico.
5. Estudio ecológico.

MIR 03 (7699): Señale la respuesta correcta. Un estudio en el que se seleccionan dos grupos de sujetos, uno formado por enfermos y el otro por personas libres de la enfermedad de interés, en el que se compara la exposición a posibles factores de riesgo en el pasado es un:

1. Estudio retrospectivo de doble cohorte.
2. Ensayo de campo.
3. Estudio retrospectivo de una serie de casos.
4. Estudio transversal.
5. Estudio de casos y controles.*

4. Medicina basada en pruebas

En los últimos años se ha intentado que las decisiones y actos médicos tengan una base científica. Con este fin se ha desarrollado la Medicina Basada en Pruebas (Evidence Based Medicine). Existe una escala que clasifica la "evidencia científica" que se puede extraer de un determinado estudio:

- Ia Pruebas obtenidas de meta-análisis de estudios controlados y aleatorizados
- Ib Pruebas de al menos un estudio controlado y aleatorizado
- Ila Pruebas de al menos un estudio controlado no aleatorizado
- Ilb Pruebas de al menos un estudio cuasiexperimental
- III Pruebas de un estudio no experimental descriptivo, como



ATENCIÓN

¡OJO! No es lo mismo estudio de cohortes que estudio prospectivo. Sigue leyendo...

D. SEGÚN EL MOMENTO DE INICIO DEL ESTUDIO:

- **Prospectivo**. El inicio del estudio es anterior a la recogida de los datos. (El investigador "mira hacia adelante"). Por ejemplo, la mayoría de estudios de cohortes son prospectivos (se crean dos grupos de expuestos y no expuestos y se comprueba el de-

estudios comparativos, estudios de correlación, y estudios de casos y controles

- IV Pruebas de informes u opiniones de comités de expertos, o experiencia clínica de autoridades respetables.

Evidencia fuerte

- Meta-análisis
- Ensayo clínico controlado aleatorio
- Ensayo clínico controlado no aleatorio
- Ensayo clínico no controlado
- Estudio de cohortes
- Estudios de casos y controles
- Estudios de corte transversal
- Series de casos
- Estudio de un caso

Evidencia débil

MIR 10 (9470): Para determinar la efectividad de un tratamiento quirúrgico comparando dos técnicas diferentes, el tipo de diseño que ofrece la mayor "evidencia científica" es:

1. Un estudio controlado aleatorizado de 50 pacientes.*
2. Una serie de casos de un único centro con 475 pacientes.
3. Un estudio de cohortes de 50 pacientes.
4. Un estudio con casos y controles de 100 pacientes.
5. Los consejos dados por un panel de expertos mundiales en base a su experiencia personal.

5. Principios éticos

Los principios éticos que has de regir en un estudio científico son el de autonomía, el de beneficencia y el de justicia.

5.1. Autonomía

La autonomía es la capacidad de una persona para decidir por ella misma. Está disminuida en caso de inmadurez o incapacidad psíquica e ignorancia.

El consentimiento informado es el documento en el que el paciente reconoce haber recibido y comprendido la información referente a un procedimiento diagnóstico o terapéutico.

En un estudio científico debe firmarlo cualquier individuo que participe, independientemente del grupo en que se incluya (MIR). Los individuos que no lo firmen no deben participar en el estudio (MIR), aunque limiten la generalización de resultados (MIR). Para que sea válido debe cumplir las siguientes características:

1. **Informado:** Se debe informar al paciente sobre todos los aspectos del estudio antes de incluirlo en él (MIR).
2. **Comprendido:** Se debe informar en términos comprensibles (MIR) y dejando un periodo de reflexión.
3. **Competente legalmente:** En caso de menores de edad o incapaces se recabará el consentimiento de su representante legal.
4. **Voluntario.** (MIR) Es decir, sin coerción.

MIR 03 (7629): El tratamiento antirretroviral en un paciente VIH+ con carga viral de 575.000/mm³ y una cifra de linfocitos de CD4 de 450/mm³:

1. Es obligado.
2. Sólo estaría indicado si la carga viral fuera superior a 1 millón de copias/ml.
3. En ningún caso se prescribiría si los CD4+ son >200/ml.
4. Está indicado si el paciente lo desea.*
5. En estas condiciones sólo estaría indicado en el contexto de un estudio clínico prospectivo.

MIR 03 (7707): Si en un ensayo clínico se usa placebo ¿cuál es el proceder correcto?:

1. Por respeto a los principios bioéticos comunicárselo a los sujetos que vayan a recibirlo.
2. No comunicárselo a los sujetos ya que en caso contrario se pierde el efecto placebo.
3. Comunicárselo a los sujetos una vez concluido el estudio.
4. Comunicárselo a las autoridades sanitarias y comités éticos, y, si estos lo aprueban, los sujetos que participan en el estudio no deben conocer su uso para evitar la pérdida del efecto placebo.
5. Comunicárselo a los sujetos en el momento de solicitar su consentimiento para ser incluidos en el estudio.*

5.2. Beneficencia

Se debe tratar a los pacientes protegiéndolos de daños y asegurando su bienestar. Incluye el principio de, ante todo, no hacer daño (*Primum non nocere*), pero añade como objetivos obtener el máximo beneficio con los riesgos mínimos.

5.3. Justicia

Se debe tratar a todos los seres humanos con consideración y respeto, sin establecer diferencias y distribuyendo los beneficios y las cargas de la investigación equitativamente, buscando el beneficio de todos. Los criterios de inclusión y exclusión de los estudios no han de ser discriminatorios, y no se deben incluir poblaciones vulnerables a no ser que el estudio sólo pueda realizarse en ellas.



REPASO

No hay que confundir ...

con ...

Variable cuantitativa: Pueden ser discretas o continuas.

Variable cualitativa: Pueden ser nominales u ordinales.

Asociación: Relación entre variables *significativamente* distinta de la que se explicaría por azar. *No implica causalidad.*

Causalidad: Se determina mediante un conjunto de criterios, ninguno de ellos necesario ni suficiente. Si un estudio no detecta asociación, esto no excluye causalidad.

Efecto dosis-respuesta: Es un criterio de causalidad. El riesgo de enfermar aumenta con mayor nivel y/o tiempo de exposición.

Respuesta a un tratamiento: No es criterio de causalidad.

Consistencia: Mismos resultados por distintos investigadores

Plausibilidad: En sintonía con los conocimientos actuales de la enfermedad



RESUMEN DE PRINCIPIOS DEL RAZONAMIENTO ESTADÍSTICO

1. DEFINICIÓN DE ESTADÍSTICA

- La **Estadística** es el conjunto de métodos necesarios para recoger, clasificar, representar y resumir datos, así como para hacer inferencias –extraer conclusiones- objetivas y científicas a partir de ellos. Así podemos dividirla en dos grandes apartados: Estadística Descriptiva y Estadística Inferencial.
- La **Estadística Descriptiva recoge, sistematiza y representa los datos** observados de un fenómeno a estudio, caracterizando su comportamiento a través de una serie de medidas (básicamente, su tendencia central y su variabilidad). La **Estadística inferencial extrae conclusiones** que proporcionen criterios para la toma de decisiones clínicas o sirvan para identificar nuevas preguntas a investigar. De esta forma, la estadística descriptiva nos ofrece datos que apuntan a ciertas conclusiones y la inferencial nos dirá si esas conclusiones son correctas o no, calculando la probabilidad de acierto o error.
- Dos son las aplicaciones fundamentales de la inferencia estadística: a) Estimación de medidas poblacionales (glucemia media poblacional); y b) Contraste de hipótesis, que se ocupa del estudio de comparaciones (eficacia de tratamientos) y asociaciones (tabaco-cáncer).

2. VARIABLES

- Una variable puede definirse como una característica observable y medible, que puede adquirir diferentes valores, y que representa el concepto que se quiere investigar. Se clasifican en: a) variables cuantitativas: discretas o continuas; y b) variables cualitativas: nominales u ordinales.
- Las variables cuantitativas discretas sólo pueden tomar ciertos valores numéricos, habitualmente números naturales o enteros, sin posibilidad de adoptar valores intermedios (nº de hijos en una familia), mientras que las cuantitativas continuas pueden tomar cualquier valor numérico (estatura, peso).
- Las variables cualitativas ordinales adquieren atributos que siguen algún tipo de gradación ordenada (intensidad del dolor (sin dolor, dolor leve, moderado e intenso MIR). Sin embargo, las variables cualitativas nominales no mantienen entre sí ningún orden de sucesión o progresión (grupos sanguíneos A, B, AB, O). Cuando sólo existen dos posibilidades, hablaremos de variable dicotómica (éxito versus fracaso).
- En otras ocasiones nos interesa saber si dos (o más) variables están o no asociadas, es decir, si la variabilidad de una variable se sigue de una variabilidad en la otra variable, y esto no se debe al azar (identificar factores de riesgo). Se distingue entonces entre: a) variables asociadas o dependientes (existe alguna relación entre las variables estudiadas); y b) variables no asociadas o independientes (dicha relación no existe).

3. ASOCIACIÓN Y CAUSALIDAD

3.1. ASOCIACIÓN ≠ CAUSALIDAD

- La asociación no implica causalidad (2MIR). Existe la posibilidad de observar asociaciones sin que exista relación causal subyacente (dedos amarillos y carcinoma de pulmón son variables asociadas, y no por ello los dedos amarillos son causa de cáncer).

3.2. CRITERIOS DE CAUSALIDAD

- Para determinar una relación causal se acude a evaluar una serie de criterios epidemiológicos de causalidad, sobre los cuales se realiza una valoración conjunta. No existen criterios únicos, ni necesarios ni suficientes (MIR). Los criterios más habitualmente utilizados son los propuestos por Bradford-Hill:

A. FUERZA DE ASOCIACIÓN (5MIR) (MAGNITUD, INTENSIDAD DE LA ASOCIACIÓN O TAMAÑO DEL EFECTO):

- A mayor intensidad de la relación entre dos variables, mayor es la probabilidad de que exista una relación causal entre ambas. Miden fuerza de asociación el riesgo relativo (RR) y odds ratio (OR) (cuanto mayor sea el RR y la OR mayor será la fuerza de asociación). Sin embargo, una asociación débil puede ser causal y una asociación muy fuerte puede deberse a un factor de confusión.

B. SECUENCIA TEMPORAL (4MIR) (TEMPORALIDAD):

- La causa debe preceder al efecto (aunque es difícil demostrarlo). Es el único criterio considerado por algunos autores como necesario.

C. EFECTO DOSIS-RESPUESTA (3MIR) (GRADIENTE BIOLÓGICO):

- Cuanto mayor es el tiempo o dosis de exposición al factor, mayor es el riesgo de enfermedad. ¡Ojo! el término "dosis-respuesta", no se refiere a "respuesta al tratamiento", este no es ningún criterio de causalidad (MIR).

D. CONSISTENCIA (MIR):

- Los resultados de un estudio deben mantenerse constantes y ser reproducibles.

E. PLAUSIBILIDAD BIOLÓGICA (5MIR):

- La relación causal sugerida debe mantener la línea de los principios científicos aceptados en el momento.

F. ESPECIFICIDAD DE LA ASOCIACIÓN Y ANALOGÍA:

- Si la relación entre un factor A y un efecto B es específica, sucederá que "B" sólo ocurre si se da "A" y la "A" sólo da lugar al efecto "B". La especificidad implica que una causa conduce a un único efecto, no a efectos múltiples (raro, salvo enfermedades infecciosas). La analogía postula que asociaciones causales similares pueden producir enfermedades similares. Algunos autores describen otro criterio, la coherencia, como sumación de los criterios de analogía y plausibilidad.

G. EVIDENCIA EXPERIMENTAL (EXPERIMENTO O CESACIÓN DE LA EXPOSICIÓN) (6MIR):

- La demostración experimental es la prueba más sólida demostrativa de causalidad (5MIR), pero puede no ser posible realizar el tipo de estudio que se requeriría para ello. Si un factor tiene un efecto determinado (enfermedad), este efecto debería cesar cuando desaparece el factor.

3.3. OTROS CRITERIOS DE CAUSALIDAD

A. DESCARTE DE EXPLICACIONES ALTERNATIVAS:

- Ante toda asociación supuestamente causal es preciso descartar primero que dicha asociación se deba a algún sesgo, factor de confusión o que se deba simplemente al azar.

B. CRITERIO DE INDEPENDENCIA:

- Es un criterio que se tiene en cuenta cuando valoramos factores de confusión: la asociación observada no tiene porqué deberse enteramente al efecto de un factor de confusión cuando éste exista; podría haber algo más (una relación causal independiente del factor de confusión).



C. CALIDAD DE LA EVIDENCIA (en la que apoyar recomendaciones en la práctica clínica):

- Se jerarquiza los diferentes tipos de estudios en función de su validez interna y la fuerza con que demuestran causalidad: **ensayo clínico (el mejor)**>cohortes>casos y controles>transversal>de riesgo agregado>serie de casos>informe de un caso. Y así un estudio ecológico tiene menor validez que los estudios analíticos (casos y controles y cohortes) (MIR).

4. TIPOS DE ESTUDIOS EPIDEMIOLÓGICOS

- **Descriptivos**, si tienen como finalidad recoger las características de las variables que se están estudiando / **Análítico**, si estudia una posible relación causa-efecto.
- **Transversal**. Se estudian las variables de una población en un momento determinado. Se mide la prevalencia. / **Longitudinal**. Se analiza un periodo de tiempo. Se mide la incidencia (casos nuevos).
- **Experimental (ensayo clínico)**. Si el investigador asigna el factor que se estudia. Puede estar controlado o no. El grupo control puede elegirse de modo aleatorio o no.
- **Observacional**. Si el investigador no asigna el factor de estudio:
 - **Estudio de cohortes**. Se investiga en cada grupo (expuestos y no expuestos) la frecuencia con que se desarrolla un determinado efecto.
 - **Casos y controles**. Se investiga cada grupo (enfermos y no enfermos) la frecuencia con que se presenta un determinado factor de riesgo.
- **Prospectivo**. El inicio del estudio es anterior a la recogida de los datos. (El investigador "mira hacia adelante"). / **Retrospectivo**: El inicio del estudio es posterior a la recogida de los datos (el investigador "mira hacia atrás").

5. MEDICINA BASADA EN PRUEBAS

<i>Evidencia fuerte</i>
Meta-análisis
Ensayo clínico controlado aleatorio
Ensayo clínico controlado no aleatorio
Ensayo clínico no controlado
Estudio de cohortes
Estudios de casos y controles
Estudios de corte transversal
Series de casos
Estudio de un caso
<i>Evidencia débil</i>

6. PRINCIPIOS ÉTICOS

- **Autonomía**. El consentimiento informado debe firmarlo cualquier individuo que participe, independientemente del grupo en que se incluya (MIR). Los individuos que no lo firmen no deben participar en el estudio (MIR), aunque limiten la generalización de resultados (MIR). Debe ser informado, comprendido, competente legalmente y voluntario.
- **Beneficencia**. Amplía el principio de no maleficencia (*primum non nocere*), incluyendo obtener el máximo beneficio con los riesgos mínimos.
- **Justicia**.



Uno de los objetivos finales del análisis estadístico es la **extrapolación** de conclusiones que sean **aplicables a poblaciones**. Sin embargo dichas poblaciones son normalmente **demasiado grandes para poder trabajar con ellas** (ej "toda España"). Es **prácticamente imposible** estudiar todas las variables necesarias en grandes grupos de personas, por ello **recurrimos al estudio de grupos más reducidos**, denominados **muestras**.

Para ello se reúne **un grupo** (pequeño, en principio) de personas que sean "**parecidas**" a la población que nosotros querríamos estudiar **en las variables** que nos interesan. **Seleccionamos aquellos más representativos**. Toda muestra ha de representar fielmente la población de la que se extrae, de forma que **estudiando nuestra muestra "sea como" estudiar a toda la población** en sus variables epidemiológicas más importantes.

Definir **cuál es la población de estudio**, cuál es el **tamaño necesario para las muestras** y qué **técnica de muestreo** utilizaremos para su extracción son labores de la **bioestadística** que abordaremos a continuación.

1. Población y muestra

Ante cualquier estudio nos planteamos siempre tres niveles de población:

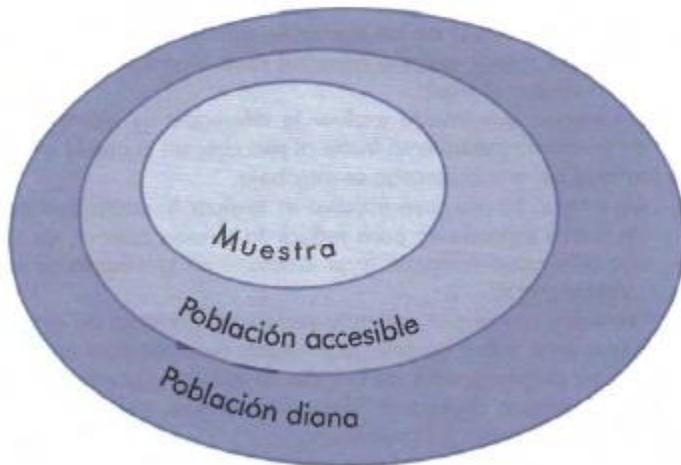
A. POBLACIÓN DIANA:

Se trata de aquella **población -grande-** a la que **se refiere el objetivo del estudio** (las mujeres españolas del ejemplo anterior), y a la que queremos **generalizar los resultados** del estudio. Es el término poblacional **más amplio**.

Pero, **¿tenemos fácil acceso a los datos de esta población?** ¿tenemos **registradas** a todas las mujeres españolas para seleccionar nuestra muestra? Probablemente **no**. Necesitamos **una población que conozcamos**; se trata de una **cuestión práctica**.

B. POBLACIÓN ACCESIBLE O DE ESTUDIO:

Es un subconjunto de la **población diana**, y tiene la **ventaja** de que **sus datos** son **accesibles**, fáciles de conocer. Existen diferentes **registros** o **circunstancias** que facilitan la colaboración, la calidad de los datos y el seguimiento. **Registros como** censos, listas electorales, padrón... y **circunstancias como** embarazadas que acuden a un hospital, drogadictos que acuden a un centro de rehabilitación, etc. **¿Sería fácil** conocer datos de todos los drogadictos o de todas las embarazadas de otra forma?



Existen tres niveles de población: muestra, población accesible y población diana

C. MUESTRA:

Es el grupo limitado de individuos que **realmente vamos a estudiar**.

De nuevo es un subconjunto del anterior. Seleccionamos sujetos con la finalidad de que este nuevo grupo de sujetos **represente a la originaria población diana** en sus características y variables.

Sin duda la **característica más importante de una muestra** es su **representatividad**, de la que depende la posibilidad de **generalizar a la población diana las conclusiones alcanzadas**. La representatividad depende de los **dos factores** más importantes que **caracterizan una muestra**:

- Tamaño muestral**, número de individuos que integran la muestra.
- Técnica de muestreo**, método que se utilizó para elegir a sus integrantes.

REPRESENTATIVIDAD



La representatividad de una muestra se apoya sobre dos aspectos fundamentales: el tamaño muestral y la técnica de muestreo

Si alguno de estos dos aspectos **se realiza incorrectamente**, la representatividad de la muestra **no se puede garantizar**, y la validez externa del estudio queda comprometida: es decir, los resultados obtenidos pueden ser válidos para la muestra, pero no son generalizables a la población diana ni a ninguna otra población.

2. Tamaño muestral

El tamaño muestral es el **número exacto de sujetos** que componen la muestra y debe ser **óptimo** tanto para los objetivos de nuestro estudio como para las **condiciones particulares** que se dan y se asumen en él:

- un número **insuficiente** de participantes puede **impedir encontrar diferencias o asociaciones buscadas**, y concluir **erróneamente** que no existen (estudio con **insuficiente potencia**);
- un número **excesivo** de sujetos **encarecerá** innecesariamente el estudio y **puede detectar como significativas** diferencias o asociaciones que son en realidad **irrelevantes** (y dar nuevamente un falso resultado a nuestro estudio).

El cálculo del tamaño muestral se realiza mediante **diversas fórmulas** existentes según el diseño del estudio. Básicamente son dos los **condicionantes globales** que determinan el tamaño de una muestra:

- variabilidad del fenómeno estudiado
- exigencia deseada en las estimaciones.

2.1. Factores determinantes

Desarrollaremos posteriormente los **dos aplicaciones básicas de la inferencia estadística**:

- estimación de parámetros poblacionales*** (ej. calcular cuál puede ser la **altura** media de los jóvenes españoles) y
- contraste de hipótesis** (**método estadístico** aplicable en muchas situaciones: búsqueda de **asociaciones** o estudios **comparativos, significación** estadística...).

Son distintos los **determinantes del tamaño muestral** según el caso:

A. EN ESTIMACIÓN DE PARÁMETROS POBLACIONALES:

- Variabilidad de la medida que se estima, expresada por su **desviación típica** o su **varianza**. **A mayor variabilidad, mayor tamaño** muestral necesario. Ej. si deseamos conocer la altura de los niños de 8 a 10 años y en las escuelas de las que extraemos nuestra muestra observamos **una gran diferencia** en las alturas de los distintos niños, **necesitaremos estudiar más niños que si la variabilidad fuera menor**.

* Una medida de tendencia central (ej. la media aritmética) o de dispersión (ej. la desviación típica) se denomina **estadístico** cuando se refiere a los datos de una distribución muestral, y **parámetro** cuando se refiere a la población. Es decir, el **estadístico** es a la muestra lo que el **parámetro** es a la población.



© Curso Intensivo MIR Asturias

La variabilidad del fenómeno a estudio determina, junto con otros factores, el tamaño muestral necesario

2. Nivel de confianza y error α deseados en la estimación. Veremos después que el nivel de confianza es la probabilidad de que el intervalo de confianza hallado en nuestro estudio para el parámetro que se estima contenga efectivamente en su interior el verdadero valor de dicho parámetro. A mayor nivel de confianza requerido, mayor tamaño muestral necesario. Por ejemplo: intervalo de confianza 160-180 mg/dl de colesterolemia y un nivel de confianza del 95% → "Podemos tener una confianza del 95% de que el verdadero valor poblacional de la colesterolemia está entre 160-180 mg/dl"; lo que nos deja un 5% de equivocarnos. El nivel de confianza es establecido a priori por el investigador (en el diseño del estudio). Con ello se establece también la probabilidad de error tipo α (MIR): probabilidad de que el verdadero valor del parámetro estimado "caiga" fuera del intervalo de confianza calculado, que nuestra medida sea errónea. Por ejemplo, un nivel de confianza del 95% supone asumir un riesgo α del 5%. Un mayor nivel de confianza lleva implícito un menor riesgo α asumido en la estimación (son conceptos complementarios). Por lo tanto a menor riesgo α que se desee asumir, mayor tamaño muestral requerido.
3. Precisión de la estimación (MIR). Un intervalo de confianza con un nivel de confianza dado resultará con mayor o menor amplitud (distancia entre sus extremos), dependiendo del tamaño muestral. A menor amplitud deseada, es decir, mayor precisión de la estimación, más sujetos deberán estudiarse. Ej anterior: 180 menos 160 hacen 20 mg/dl de amplitud del ejemplo anterior. Si quisiéramos estrechar más (175-165mg/dl) nuestro intervalo de confianza para el mismo nivel de confianza necesitaríamos aumentar el tamaño muestral.



© Curso Intensivo MIR Asturias 2003

4. Proporción esperada del fenómeno a estudiar (MIR). En el caso concreto de la estimación de una proporción poblacional (por ejemplo, la prevalencia de una enfermedad -¿cuál es la proporción de hipertensos en la población?), el cálculo del tamaño muestral requiere anticipar un valor esperado para la proporción poblacional que se estima. Tal valor puede tomarse de datos orientativos previos o, si no se dispone de información alguna, puede asignarse por defecto un valor de 0,5¹ (50%), en cuyo caso se obtendrá un tamaño muestral máximo.
5. Número de no respuestas (MIR) (no respondedores como pacientes que no siguen el programa, por ejemplo) debe ser pequeño, y no estar relacionado con la enfermedad que estudiamos. A mayor número esperado de pacientes que no respondan o se pierdan en el estudio, mayor habrá de ser el tamaño muestral.

MIR 04 (7970): Un investigador desea determinar la prevalencia de tabaquismo en niños de 12 años en una zona urbana mediante un estudio descriptivo. Para calcular el tamaño muestral que necesita para su estudio ya posee los siguientes datos: tamaño de la población de 12 años, porcentaje de pérdidas, la precisión con la que desea dar la prevalencia que obtendrá (por ejemplo 5%) y ha seleccionado un nivel de confianza (por ejemplo 95%). ¿Qué otro dato le falta?:

1. Estimar el error beta.
2. Estimar la proporción esperada de tabaquismo.*
3. Estimar la desviación estándar del tabaquismo.
4. Estimar el error alfa.
5. Estimar la media esperada del tabaquismo.

MIR 06 (8473): Para el cálculo del tamaño muestral en un estudio descriptivo en el que la variable de interés es dicotómica se precisa conocer los siguientes ingredientes, EXCEPTO uno:

1. Proporción esperada de la variable de interés en la población.
2. Precisión deseada del intervalo de confianza.
3. El error beta asumible por el investigador.*
4. Nivel de confianza del intervalo.
5. Fórmula para calcular el tamaño muestral para un estudio descriptivo con una variable dicotómica.

MIR 09 (9249): ¿Cuál de las siguientes situaciones sugiere que necesitará un menor tamaño muestral para alcanzar los objetivos de un ensayo clínico?:

1. Un ensayo que intenta evaluar la diferencia de mortalidad de un nuevo tratamiento frente al placebo, en el que la mortalidad del grupo placebo es muy baja.
2. Un ensayo clínico cuyo objetivo es evaluar la capacidad de un nuevo tratamiento para reducir la tensión arterial, en el que se producirá mucho error aleatorio en la medida de la tensión arterial.
3. Un ensayo en el que pretende evaluar el efecto de un inmunosupresor sobre la dosis acumulada de esteroides en pacientes diagnosticados de arteritis de células gigantes, en el que se intenta demostrar diferencias de dosis pequeñas entre ambos grupos de tratamiento.
4. Un ensayo clínico en el que usted asume una probabilidad de error alfa del 0.1 en vez del 0.05 habitual.*
5. Un ensayo clínico en el que usted pretende tener una potencia del 0.9 en vez del 0.8 habitual.

El error alfa es la probabilidad de encontrar un resultado significativo que realmente no existe. Si toleramos un error alfa mayor, nuestro estudio no es "tan bueno".

B. CONTRASTE DE HIPÓTESIS:

1. Variabilidad (3MIR) del fenómeno que se estudia, al igual que en la estimación de parámetros.
2. Riesgo α o de error tipo I (error de primera especie) (2MIR). Determina la probabilidad, asumida a priori por el investiga-

¹ Para el cálculo del tamaño muestral en este caso debemos multiplicar la proporción esperada del fenómeno a estudio por su complementario, esto es, si $P=0.3$ su complementario sería $q=0.7$. Si buscásemos el producto más alto entre todas las posibilidades (0.1 por 0.9; 0.2 por 0.8; 0.4 por 0.6...), el mayor sería 0.5 por 0.5. De ahí que ante falta de información sobre la proporción esperada obtenemos por 0.5.

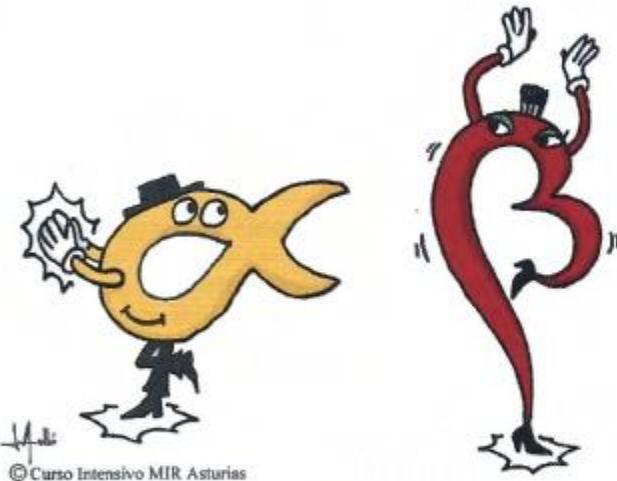
dor, de equivocarse en identificar como significativa una diferencia o una asociación que en realidad no existe (el error de "ver lo que no hay"). El riesgo α que el investigador establece, determina el nivel de confianza ($1 - \alpha$) del estudio², a partir del cual una diferencia o una asociación observadas se considerarán significativas, es decir, no atribuibles al azar. A menor riesgo α o mayor nivel de significación, mayor tamaño muestral requerido (6MIR).

Intervalo de confianza de...	Error α de...
95%	5%
99%	1%

3. **Riesgo β o de error tipo II (error de segunda especie) (MIR).** Es la probabilidad asumida de equivocarse en no detectar como significativa una diferencia o una asociación que en realidad sí existe (MIR). El riesgo β determina directamente la potencia del estudio (MIR) (su complementario: $1 - \beta$), que es la capacidad del estudio para detectar diferencias o asociaciones existentes. A menor riesgo β deseada o mayor potencia del estudio, mayor tamaño muestral necesario (6MIR).

Son complementarios...

α	Nivel de confianza (estimación de parámetros poblacionales)
β	Potencia



Recuerda que alfa y beta son errores aleatorios, no sistemáticos (sesgos)

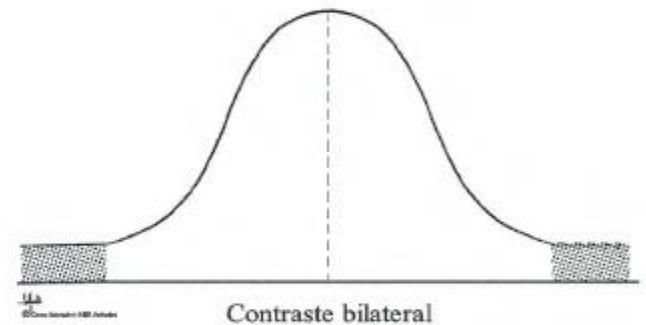
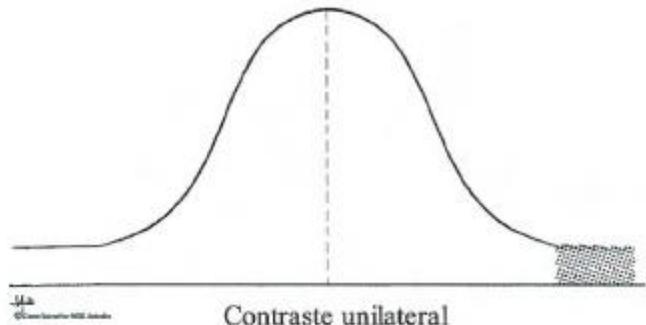
4. **Magnitud de la diferencia mínima que se desea detectar o mínima diferencia importante (MIR).** En una comparación, el investigador debe establecer la mínima diferencia que desea que el estudio detecte con el riesgo α establecido. A menor magnitud de la diferencia mínima detectable, mayor tamaño muestral necesario (5MIR). Es más fácil detectar diferencias de un 20% en las cifras de tensión arterial que una diferencia de 0.05% entre dos grupos de tratamiento antihipertensivo, por ejemplo. Requeriría mayor tamaño muestral la diferencia pequeña (0.05%).
5. **Estimación del porcentaje de pacientes que "se perderán" durante el estudio (3MIR).**

MIR 02 (7435): ¿Cuál es la forma correcta de manejar las pérdidas de pacientes en un ensayo clínico?

1. Reemplazarlos.
2. Tratar de evitarlas recordando a los pacientes su compromiso de participación en el estudio.
3. No hay un procedimiento adecuado, si se producen pérdidas dan lugar al fracaso del estudio.
4. Tener en cuenta las pérdidas previsibles al calcular el tamaño muestral.*
5. Reemplazarlos, pero asignándolos al azar a uno u otro grupo.

² Se considera en principio, en términos puristas, que el riesgo o error alfa es el complementario del nivel de confianza. Sin embargo, muchos autores utilizan indistintamente error alfa que nivel de confianza, al igual que aparece en muchas preguntas MIR.

6. En un ensayo que mide mortalidad: mortalidad encontrada en estudios previos en las diferentes alternativas (MIR).
7. **Contraste bilateral (dos colas) o unilateral (una cola) (MIR).** Los contrastes de hipótesis bilaterales o de dos colas requieren mayor tamaño muestral, porque son más exigentes. Es el caso de comparar, por ejemplo, si dos fármacos son iguales o diferentes en su eficacia considerando tanto que A pueda ser mejor que B, como que B pueda ser mejor que A (dos opciones). En el caso de un contraste unilateral o de una cola se compararía si los fármacos son iguales, o si uno de ellos en particular es mejor que el otro (A mejor que B, por ejemplo), sin valorar la opción contraria (B mejor). Por esta razón es el test de contraste unilateral requiere menor tamaño muestral, pues es menos exigente en sus hipótesis.



El contraste bilateral es más exigente que el unilateral, dado que tiene en cuenta dos posibilidades alternativas, en vez de una

MIR 05 (8222): Estamos planeando hacer un ensayo clínico para comparar la eficacia de dos antibióticos para el tratamiento de la neumonía. ¿Cuál de los siguientes datos NO es útil para calcular el tamaño de la muestra?:

1. El número de pacientes que ingresan con diagnóstico de neumonía en nuestro centro, que es de 5 a la semana.*
2. El porcentaje de pacientes que se curan con el fármaco de referencia, que en los estudios publicados es de un 85%.
3. La probabilidad de obtener un resultado falso positivo (error tipo 1), que la fijamos en un 5%.
4. La proporción de pérdidas de pacientes previstas, que en otros estudios similares de la literatura era de un 10%.
5. La diferencia mínima clínicamente relevante entre los dos fármacos, que la consideramos de un 10%.

MIR 13 (10190) (191): Un Pediatra desea estudiar el sobrepeso en los niños de 14 años, según los valores del IMC (índice de masa corporal). Para estimar el tamaño muestral necesario propone un nivel de confianza del 95% y una precisión de 1 unidad de IMC. ¿Qué más parámetros necesita conocer para determinar el tamaño muestral?

1. La media del IMC en la población.
2. La varianza del IMC.*
3. La media y la desviación típica del IMC.
4. El tamaño de la población y la media del IMC.
5. La desviación típica del IMC y el tamaño de población.

C. NO DETERMINAN EL TAMAÑO MUESTRAL...

Número de sujetos disponible (3MIR) (es diferente de n° o % de sujetos que se estima que se perderán), duración del seguimiento de pacientes (MIR) o los recursos disponibles no determinan el tamaño muestral (MIR). El tipo de enmascaramiento tampoco determina el tamaño muestral (2MIR).

Mayor tamaño muestral si...		
Estimación de parámetros	Variabilidad	↑
	Nivel de confianza	↑
	Error tipo α	↓
	Precisión	↑
	Proporción esperada	50% ¹
Contraste de hipótesis	Variabilidad	↑
	Nivel de confianza	↑
	Riesgo α (error tipo I)	↓
	Potencia (1 - β)	↑
	Riesgo β (error tipo II)	↓
	Menor mínima diferencia importante ²	↑
	Contraste bilateral o unilateral	Contraste bilateral
Mortalidad esperada en cada alternativa (estudios de mortalidad)	↑	

¹Menor tamaño muestral cuanto más se aleje de 50%, sea mayor o menor, dado que sería lo mismo multiplicar 0.3 por 0.7, que 0.7 por 0.3.

²Efectivamente cuanto más pequeña sea la diferencia que queramos ver (0.05% frente a 20%) más difícil va a ser, mayor tamaño muestral necesitaremos.

En general...

Mayor tamaño muestral necesario a mayor variabilidad del fenómeno a estudio y/o mayor exigencia deseada en el estudio (tanto en estimación de parámetros poblacionales como en contrastes de hipótesis).



MIR 00 Familia (6698): En un ensayo controlado, indique cual de los siguientes factores **NO** condiciona el tamaño de la muestra del estudio:

1. El error aleatorio de primera especie o de tipo I (α) que se acepta en su diseño.
2. El error aleatorio de segunda especie o de tipo II (β) que se acepta en su diseño.
3. El poder estadístico del estudio.
4. El nivel de enmascaramiento de los tratamientos en estudio.*
5. La variabilidad del parámetro de medida principal.

MIR 00 (6924): ¿Cuál de los siguientes aspectos **NO** ha de ser considerado a la hora de calcular el número de pacientes que hay que incluir en un ensayo clínico?:

1. Tipo de variable principal de respuesta.
 2. Grado deseado de protección frente a los errores de tipo I y II.
 3. Porcentaje de pacientes que previsiblemente se perderán.
 4. Tipo de enmascaramiento de los tratamientos.*
 5. Diferencia entre tratamientos que se desea detectar si existe.
- El tipo de enmascaramiento pretende evitar los sesgos de información.

MIR 01 (7197): De los siguientes factores, indique aquel que **NO** condiciona, a priori, el tamaño de la muestra en un ensayo clínico:

1. La magnitud de la diferencia que se intenta encontrar.
2. El nivel de enmascaramiento de los tratamientos en estudio.*
3. La variabilidad del parámetro de medida.
4. El nivel de protección frente al error de tipo I.
5. La potencia que se pretende alcanzar.

MIR 02 (7436): Queremos hacer un estudio para comprobar si un nuevo fármaco antiagregante es más eficaz que la aspirina para reducir la mortalidad cardiovascular. Para poder planificar el estudio necesitamos saber **cuantos** pacientes debemos incluir. ¿Cuál de los siguientes parámetros **NO** influye en el cálculo del tamaño de la muestra?:

1. La mortalidad cardiovascular encontrada en los pacientes tratados con aspirina en los estudios previos.
2. La proporción de pérdidas previstas.
3. La mortalidad cardiovascular esperada en los pacientes tratados con el nuevo antiagregante.
4. El número de centros que van a participar en la realización del estudio.*
5. La probabilidad de obtener un resultado falso positivo (error tipo I).

3. Técnicas de muestreo

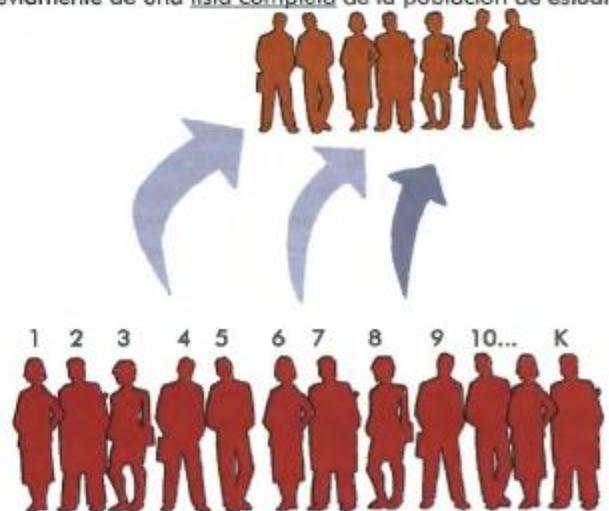
Junto con el tamaño muestral apropiado, la representatividad de una muestra depende de la aplicación de una técnica de muestreo correcta. Existen diferentes técnicas para seleccionar los sujetos que constituirán la muestra, divididas en función de que sean muestreos probabilísticos o no probabilísticos.

3.1. Muestreo probabilístico

Abarca una serie de técnicas que tienen en común utilizar algún procedimiento de selección aleatoria, de modo que todos los individuos de la población de estudio tienen una misma probabilidad de ser incluidos en la muestra, siendo ésta determinada por el azar. De este modo, es más probable que las muestras obtenidas sean representativas (MIR), y se previenen sesgos de selección (MIR).

A. MUESTREO ALEATORIO SIMPLE:

Cada individuo o unidad de muestreo tiene la misma probabilidad de ser escogido. Presenta la ventaja de ser una técnica simple, pero con el inconveniente de que requiere disponer previamente de una lista completa de la población de estudio.



En el muestreo aleatorio simple se extrae un grupo de la población aleatoriamente

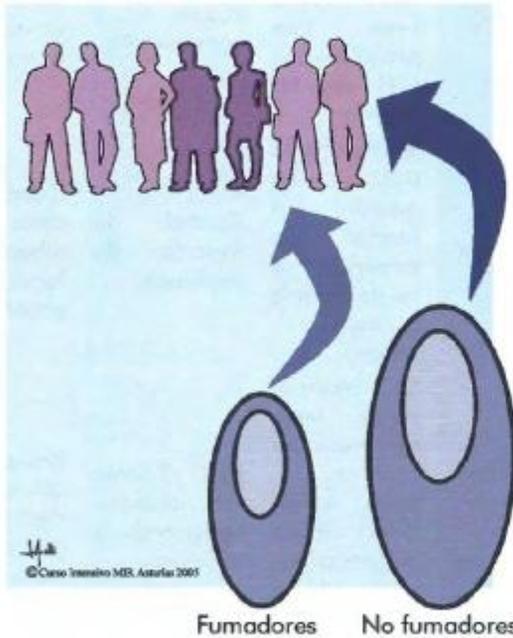
Ejemplo: extraer una muestra de cinco alumnos de una clase diciendo al azar cinco números de la lista de la clase. Todos los alumnos tienen la misma probabilidad de pertenecer a la muestra.

B. MUESTREO ESTRATIFICADO:

Es una variante del muestreo aleatorio simple. La población se divide en dos o más subgrupos (estratos) excluyentes entre sí, en función de cierta variable que interese controlar por poder ac-



tuar como factor de confusión (por ejemplo fumadores y no fumadores en estudio de cardiopatía coronaria). Seguidamente, se extrae aleatoriamente una muestra de tamaño proporcional en cada subgrupo o estrato.



Separamos a la población en función de un factor que deseamos controlar como factor de confusión, extrayendo posteriormente sujetos proporcionales al tamaño de cada grupo formado.

Es el método de aleatorización de elección cuando se pretende que un factor pronóstico se distribuya por igual en los dos grupos de tratamiento (MIR), y facilitará controlar el posible efecto del mismo como potencial factor de confusión.

MIR 02 (7433): Le encargan el diseño de un ensayo clínico en el que es muy importante que un factor pronóstico se distribuya por igual en los dos grupos de tratamiento. El método de aleatorización que usted utilizaría es:

1. Aleatorización simple.
2. Aleatorización por bloques.
3. **Aleatorización estratificada.***
4. Aleatorización centralizada.
5. Aleatorización ciega (ocultación de la secuencia de aleatorización).

Ejemplo: Supongamos que se realiza un estudio acerca de consumo de alcohol y desarrollo de cardiopatía coronaria. Se podría encontrar una asociación estadísticamente significativa entre ambos términos. Sin embargo, si estratificásemos a la población según fumadores y no fumadores, encontraríamos que el consumo de alcohol y tabaco se encuentran muy ligados entre sí, de modo que la gran mayoría de consumidores de alcohol se encontraría en el grupo de fumadores, con lo cual el tabaquismo ha actuado como factor de confusión, produciendo la observación de una aparente relación entre el consumo de alcohol y cardiopatía coronaria (falsa por otra parte). Mediante un muestreo estratificado se controlaría este fenómeno de confusión.

Esta técnica presenta el inconveniente de que se precisa conocer la distribución de la variable por la que se ha de estratificar.

Control de factores de confusión

Fase de diseño:

- Muestreo aleatorio (especialmente estratificado)
- Emparejamiento, matching
- Criterios de restricción

Fase de análisis:

- Análisis estratificado (y técnicas de ajuste o estandarización)
- Análisis multivariante (múltiples factores)

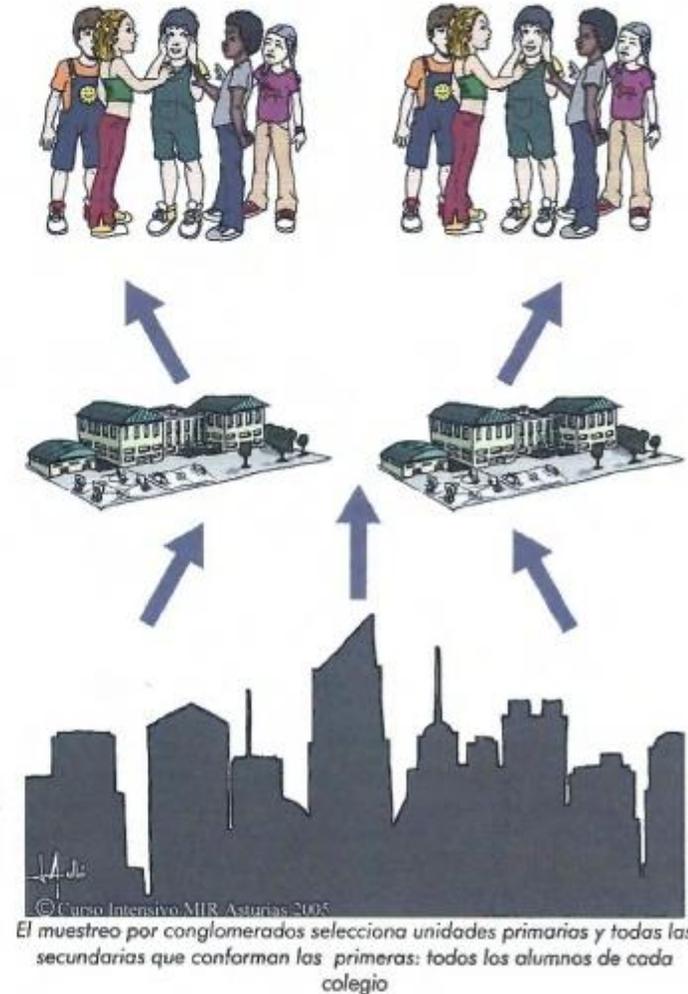
C. MUESTREO EN ETAPAS MÚLTIPLES (POLIETÁPICO):

Se seleccionan unidades primarias de muestreo de una población. Posteriormente se seleccionan, dentro de las unidades primarias, unidades secundarias, y así sucesivamente las veces que sea necesario hasta llegar a extraer los individuos de nues-

tra muestra (seleccionando el total o parte de la última unidad escogida).

Ejemplo: Si se desea estudiar la prevalencia de determinada enfermedad sobre niños en edad escolar, se pueden seleccionar escuelas (unidades primarias) y, en cada escuela, muestrear niños (unidades secundarias).

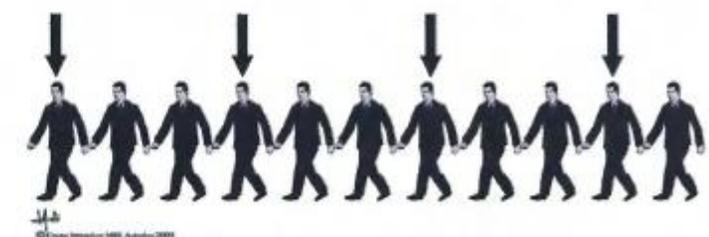
Cuando en un muestreo de dos etapas se incluyen todas las unidades secundarias (por ejemplo, si las unidades primarias son escuelas y escogemos a todos los niños en cada escuela), se denomina muestreo por conglomerados.



El muestreo en etapas múltiples tiene la ventaja de su utilidad cuando la población de estudio es grande y dispersa pero se encuentra agrupada por alguna característica, o cuando no se dispone de un listado de toda la población de estudio. Su desventaja es que complica el cálculo de parámetros poblacionales, y que puede originar un sesgo de selección si los individuos en cada unidad primaria fueran homogéneos respecto alguna de las variables estudiadas (ej: seleccionar para un estudio de altura media en población infantil un colegio con un buen equipo de baloncesto).

D. MUESTREO SISTEMÁTICO:

Se divide el total de la población de estudio entre el tamaño de muestra deseado, hallando así una constante k de muestreo. El primer individuo se elige al azar y después se elige sistemáticamente uno de cada k individuos, según algún factor que los ordene. [$k = n/N$, siendo n el tamaño muestral y N el número total de individuos de la población]



Ordenada una población, en el muestreo sistemático se selecciona uno de cada k individuos

Ejemplo: Seleccionar en una organización colegial a uno de cada 10 sujetos según su número de colegiado. Como ventajas este procedimiento ofrece la simpleza de su aplicación, el poder llevarse a cabo sin precisar un listado completo de la población de estudio y el hecho de que, en poblaciones ordenadas según una tendencia determinada (por ejemplo, antigüedad en una institución), asegura una cobertura de todos los tipos. La desventaja consiste en que, si existe alguna relación entre el fenómeno a estudio y la constante de muestreo, se pueden producir sesgos de selección.

3.2. Muestreo no probabilístico

Las unidades de muestreo (individuos) se seleccionan empleando métodos en los que no interviene el azar. La desventaja común de estos métodos es que, al no tener todos los individuos la misma probabilidad de ser incluidos en el estudio, la representatividad de la muestra no está asegurada.

A. MUESTREO CONSECUTIVO:

Se selecciona a los sujetos que cumplen los criterios de inclusión especificados en el protocolo del estudio (por ejemplo, a medida que ingresan en un servicio hospitalario pacientes con la enfermedad que estudiamos). Es la técnica más utilizada en los ensayos clínicos, ya que para aplicar la medida terapéutica que se somete a evaluación (efectividad de un fármaco nuevo), no es posible conocer con anticipación qué individuos desarrollarán la enfermedad o incurrirán en los criterios de inclusión. Por ello, en general, no es posible la aplicación de una técnica de muestreo aleatorio en los ensayos clínicos.

No debe confundirse esto con la asignación aleatoria a los grupos de estudio (grupo de casos o grupo control) que se hace sobre los participantes seleccionados para el ensayo clínico.

Ensayo clínico aleatorizado utiliza...

Muestreo consecutivo (no probabilístico) Muestreo consecutivo (no probabilístico)

B. INCLUSIÓN DE VOLUNTARIOS:

Es preferible evitar esta técnica, ya que las personas que se ofrecen voluntariamente a la realización de un estudio suelen estar influidas por algún tipo de característica que afectará a la representatividad de la muestra (ej: pacientes hipocondríacos, preocupados por su salud).

MIR 06 (8472): ¿Cuál de los siguientes muestreos es de tipo no probabilístico?:

1. Muestreo sistemático.
2. Muestreo estratificado.
3. Muestreo de casos consecutivos. *
4. Muestreo aleatorio simple.
5. Muestreo por conglomerados.

MIR 08 (9009): Hemos realizado un estudio de investigación en una determinada área de salud en el que se han incluido 500 pacientes con hipertensión arterial mediante un muestreo de casos consecutivos. Este tipo de muestreo consiste en:

1. Escoger "a dedo" a los individuos entre la población accesible que se suponen más apropiados para participar en el estudio.
2. Dividir a la población en subgrupos de acuerdo a ciertas características, y escoger una muestra al azar de cada uno de esos grupos.
3. Enumerar cada unidad de la población accesible y seleccionar una muestra al azar.
4. Elegir a cada individuo que cumpla con los criterios de selección del estudio dentro de un intervalo de tiempo específico o hasta alcanzar un determinado número. *
5. Extraer una muestra al azar a partir de agrupaciones naturales de individuos dentro de la población.

Muestreo probabilístico: misma probabilidad al azar, mayor representatividad.

Aleatorio simple	Muestra n de población N, cada sujeto tiene una probabilidad n/N de ser elegido.	Sencillo y fácil cálculo y comprensión.	Requiere conocer listado completo.
Estratificado	Se separa la población en grupos en función de presencia o no de factor a controlar (tabaco).	Control de factores de confusión.	Debe conocerse la distribución del factor en población.
Polietápico (por conglomerados si dos unidades)	Se realizan varias fases de muestreo sucesivas hasta seleccionar todos o parte de los sujetos de la última unidad seleccionada.	Muy eficiente en poblaciones grandes y dispersas.	Complica los cálculos. Posibles sesgos de selección.
Sistemático	Se selecciona un sujeto al azar y a partir de éste, uno de cada k individuos de la población N: $k = n/N$.	No necesario listado completo. Cubre todos los tipos (en población ordenada).	Posible sesgo de selección.

Muestreo no probabilístico: distinta probabilidad, no asegurada representatividad.

Consecutivo	Pacientes que cumplen criterios de inclusión del protocolo clínico.	Técnica más utilizada en ensayos clínicos.	No se asegura representatividad.
Voluntarios	Se incluyen voluntarios.	??	No recomendado.



No hay que confundir ... con ...

Tamaño muestral:

Número óptimo de individuos para un estudio (debe calcularse).

Se calcula según el tipo de estudio, variabilidad de las distribuciones, más varios factores que decide el investigador.

Número disponible de pacientes:

Circunstancia limitante que no interviene en el cálculo del tamaño muestral.

Determinantes del tamaño muestral en:

Estimación de parámetros poblacionales: variabilidad, nivel de confianza, error α , precisión de la estimación, proporción esperada del fenómeno a estudio (máxima 0.5), número de no respuestas.

Contraste de hipótesis: variabilidad, riesgo α o error tipo I (1ª especie), riesgo β o error tipo II (2ª especie), mínima diferencia que se desea detectar, pacientes que se perderán (no responden), que el contraste sea bilateral (ó unilateral).

No determinan el tamaño muestral: número de sujetos disponible, duración del seguimiento, recursos disponibles ni tipo de enmascaramiento.

Muestreo probabilístico:

Utiliza procedimientos aleatorios.

Muestreo aleatorio simple, muestreo estratificado, muestreo en etapas múltiples, muestreo por conglomerados, muestreo sistemático.

Muestreo sistemático:

Técnica de muestreo probabilístico. Se elige uno de cada k individuos.

Intervalo de confianza: intervalo de valores de la variable que incluye con alta probabilidad el valor verdadero estimado (media poblacional).

Muestreo no probabilístico: No interviene el azar. Muestreo consecutivo, inclusión de voluntarios.

Muestreo consecutivo:

Técnica de muestreo no probabilístico. Se eligen individuos según criterios de un protocolo.

Nivel de confianza: probabilidad que se tiene de que el intervalo de confianza calculado incluya el verdadero valor poblacional. Ej. 95% de nivel de confianza \rightarrow 95% de probabilidades de acierto.

RESUMEN DE MUESTREO

1. POBLACIÓN Y MUESTRA

- Se distinguen tres niveles de población:
 - **Población diana:** población grande a la que se pretende generalizar los resultados del estudio.
 - **Población accesible o de estudio:** subconjunto de la anterior en la cual tenemos datos sobre todos los individuos integrantes. Debemos seleccionar un subconjunto representativo de la misma: la muestra.
- Población muestral o muestra:** grupo limitado de individuos que realmente estudiamos. La característica más importante de una muestra es su representatividad, la cual nos permitirá generalizar los resultados muestrales a la población diana (validez externa), y dependerá fundamentalmente del tamaño muestral y la técnica de muestreo utilizados.

2. TAMAÑO MUESTRAL

- Número exacto de sujetos que componen la muestra. El tamaño muestral ha de calcularse en función de la variabilidad del fenómeno a estudio y el nivel de exigencia de nuestras estimaciones, y serán sus factores determinantes distintos según se trate de: a) estimación de parámetros poblacionales (media poblacional de colesterolemia) o b) contraste de hipótesis (búsqueda de asociación entre variables o diferencias entre tratamientos).

2.1. ESTIMACIÓN DE PARÁMETROS POBLACIONALES

- Variabilidad** (desviación típica y varianza de la variable estimada). A mayor variabilidad del fenómeno a estudio, mayor tamaño muestral necesario.
- Nivel de confianza** y error α (MIR) deseados en la estimación (nivel de confianza de 95% \rightarrow error α de 5%). Se establecen a priori. A mayor nivel de confianza o menor error α , mayor tamaño muestral.
- Precisión de la estimación (MIR) (mayor precisión, a menor amplitud del intervalo calculado). Si se desea mayor precisión, se requiere mayor tamaño muestral.
- Proporción esperada del fenómeno a estudio (MIR). Para calcular el tamaño muestral necesario se requiere adelantar la proporción de sujetos que van a presentar la enfermedad a estudio. Por defecto se toma la máxima: 0.5. Cuanta más se acerque a 0,5, mayor tamaño.
- Número de "no respuestas" esperado (MIR). Pacientes que se perderán en el estudio. A mayor número, mayor deberá ser la muestra.

2.2. CONTRASTE DE HIPÓTESIS

- Variabilidad (2MIR) (desviación típica y varianza del fenómeno a estudio).
- Error α , tipo I o de primera especie (nivel de significación) (2MIR). Se establece a priori y determina la probabilidad asumida de equivocarse en identificar como significativa una asociación o diferencia que en realidad no existen. A menor error α , mayor tamaño muestral necesario (5MIR).
- Error β , tipo II o de segunda especie (MIR). También se establece a priori y expresa la probabilidad asumida de equivocarse en no detectar como significativa una diferencia o asociación que existen (MIR). A menor error β , mayor tamaño muestral requerido (5MIR).
- Potencia (3MIR). Complementario del error β ($1 - \beta$). A mayor potencia, mayor tamaño muestral.
- Mínima diferencia que se desea detectar en el estudio (MIR). A menor diferencia importante detectable (más difícil), mayor tamaño muestral necesario.
- En un ensayo que mide mortalidad: mortalidad encontrada en estudios previos en las diferentes alternativas (MIR).
- Contraste bilateral (dos colas) o unilateral (una cola) (MIR). El contraste bilateral es más exigente (valora la igualdad frente a dos alternativas: ser mayor o menor), por ello requiere mayor tamaño muestral. El contraste unilateral valora únicamente la igualdad (\leq) frente a la alternativa, ser mayor.
- Pacientes que se perderán en el estudio ("no respuestas") (3MIR). A mayor número previsto, mayor deberá ser el tamaño muestral calculado.
- No influyen en el cálculo del tamaño muestral: número de sujetos disponibles (3MIR), duración del seguimiento (MIR), tipo de enmascaramiento (2MIR), recursos disponibles...

3. TÉCNICAS DE MUESTREO

- Se distinguen técnicas de muestreo probabilísticas y no probabilístico. El muestreo probabilístico utiliza algún procedimiento de selección aleatoria, siendo más probable que las muestras obtenidas sean representativas y evitando sesgos de selección (MIR).

3.1. MUESTREO PROBABILÍSTICO

A. MUESTREO ALEATORIO SIMPLE:

- Sencillo y de cálculo rápido. Presenta el inconveniente de que requiere el listado completo de la población.

B. MUESTREO ESTRATIFICADO:

- Se divide a la población en función de la presencia o no de un determinado factor. Seleccionamos posteriormente un grupo proporcional de entre ambos subgrupos (con y sin factor) para obtener la muestra. Es el método de aleatorización de elección cuando se pretende que un factor pronóstico se distribuya por igual en los dos grupos de tratamiento (MIR). Su ventaja fundamental es poder controlar factores de confusión, y su inconveniente la necesidad de conocer la distribución de dicho factor en la población.

C. MUESTREO EN ETAPAS MÚLTIPLES (POLIETÁPICO):

- Se seleccionan unidades primarias, dentro de las cuales se eligen una unidades secundarias y así sucesivamente, hasta obtener la muestra, incluyendo todos o parte de los individuos de una determinada unidad de muestreo. Técnica de muestreo muy útil ante poblaciones grandes y dispersas. Sin embargo el cálculo es complejo y se pueden producir sesgos de selección. Si se incluyen todas las unidades secundarias de las primarias seleccionadas, se denomina muestreo por conglomerados.

D. MUESTREO SISTEMÁTICO:

- Se eligen sistemáticamente uno de cada k individuos de una población hasta completar la muestra requerida, siendo k la constante de muestreo (n/N). No requiere un listado completo de la población y ofrece una buena cobertura de todos los tipos de una población ordenada (organización colegial). Pueden producirse sesgos de selección (si k tiene alguna relación con el fenómeno a estudio).

3.2. MUESTREOS NO PROBABILÍSTICOS

A. MUESTREO CONSECUTIVO:

- Se seleccionan pacientes con la enfermedad que estudiamos que cumplan unos criterios de inclusión establecidos previamente en un protocolo. Método más utilizado en los ensayos clínicos.

B. INCLUSIÓN DE VOLUNTARIOS:

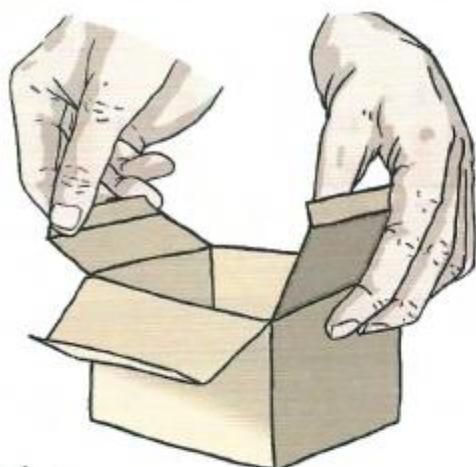
- No recomendada.

1. Validez interna y externa

La muestra debe representar a la población que se estudia, así como las variables elegidas deben representar el fenómeno objeto del estudio en esa misma población. Sólo de este modo las inferencias realizadas tienen validez. Se describen dos tipos: validez interna y externa.

A. VALIDEZ INTERNA:

Es el grado en que las conclusiones del estudio son extraídas correctamente (buena construcción y procedimiento), y soportadas con solidez por los datos obtenidos. Valida la verdad en el estudio. Es decir, los resultados son al menos ciertos en la muestra en la que hemos realizado el estudio. Es así como la validez interna permite hacer inferencias correctas sobre el efecto de un factor de riesgo en los participantes del estudio (MIR).



Julio

©Curso Intensivo MIR Asturias 2003

La validez interna es el grado en que se extraen adecuadamente los datos de un estudio, y son por ello extrapolables a los individuos participantes del mismo

MIR 00 Familia (6709): Si un estudio permite hacer inferencias correctas sobre el efecto de un factor de riesgo en los participantes del estudio, diremos que tienen una buena:

1. Sensibilidad.
2. Especificidad.
3. Valores predictores.
4. Validez interna.*
5. Validez externa.

MIR 11 (9710): Se dice que un estudio observacional tiene validez interna:

1. Cuando se obtiene un resultado estadísticamente significativo.
2. Cuando el investigador ha controlado razonablemente las posibles fuentes de sesgo.*
3. Cuando el investigador no ha cometido fraude y los resultados son auténticos.
4. Cuando se ha publicado en una revista con alto factor de impacto.
5. Cuando el intervalo de confianza de la principal medida de efecto no incluye el valor nulo.

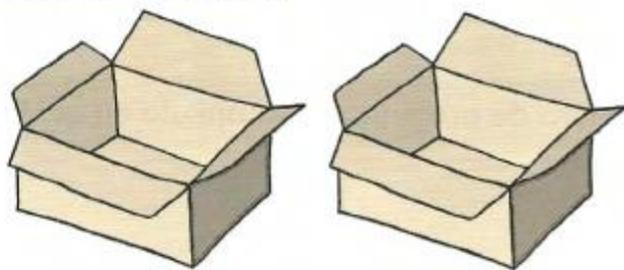
Todo fallo de procedimiento que impida que las conclusiones del estudio no sean extraídas de forma correcta afectará a la validez interna: sesgo de información (MIR) (sesgo de memoria (MIR)), utilización de grupos no comparables (MIR), sesgos por efecto de factores de confusión (MIR)...

Ejemplo: si en el estudio se debe medir la estatura de diferentes personas en centímetros, el utilizar una cinta métrica mal calibrada implica que los datos (las alturas) son ya de por sí falsos; la media y varianza serán incorrectas, por ejemplo.

B. VALIDEZ EXTERNA:

Es el grado en que las conclusiones del estudio son generalizables a la población diana. Valida las inferencias sobre la verdad

validez externa cuando se utiliza una muestra no representativa de la población general (MIR).



La validez externa valora el grado de comparabilidad y extrapolación de los resultados de un estudio a otras Poblaciones

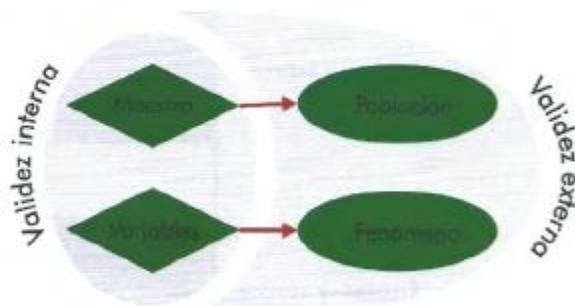
MIR 07 (8755): ¿Cuál es el principal inconveniente de las evaluaciones económicas realizadas dentro de los ensayos clínicos?:

1. La necesidad de utilizar suposiciones.
2. La baja validez interna.
3. La baja validez externa.*
4. La baja calidad de los datos de eficacia.
5. La falta de rigor científico.

Es decir, es difícil extrapolar los resultados de las evaluaciones económicas a otros estudios

Validez externa presupone validez interna, no así al contrario.

Ejemplo: si en el estudio anterior se utilizan jugadores de baloncesto para calcular la estatura media española -y si se ha utilizado una cinta métrica correctamente calibrada-, es más que probable que estos resultados, ciertos en la muestra, no sean extrapolables al resto de los españoles.



Estudiamos las variables de una muestra para poder aproximarnos al fenómeno de interés de la población. Para ello necesitamos validez interna (conclusiones del estudio bien extraídas) y validez externa (capacidad de extrapolarlas a la población u otras poblaciones)

No confundir...

Validez interna:

- Grado en que los resultados de un estudio son extraídos correctamente de sus participantes.
- Afecta a la validez interna: sesgo de memoria (sesgo de información), utilización de grupos no comparables, sesgo de confusión...

Con...

Validez externa:

- Grado en que las conclusiones de un estudio son generalizables a la población diana (u otras poblaciones).
- Afecta a la validez externa: toda alteración de la validez interna, sesgo de selección...

MIR 10 (9471): Unos investigadores realizan un estudio con pacientes diabéticos provenientes de una unidad de diabetes hospitalaria. La aplicación de los resultados del estudio a pacientes de una consulta de medicina de familia puede ser errónea por un problema de:

1. Validez interna.
2. Validez externa.*
3. Precisión.
4. Sesgo de clasificación.
5. Confusión.

C. RELEVANCIA CLÍNICA

Otro factor que interviene en que un estudio sea correcto es la relevancia clínica (que un estudio sea "estadísticamente" válido

no implica que su resultado sea clínicamente importante). En la relevancia clínica intervienen la magnitud de la diferencia, la gravedad del problema que se va a investigar, la frecuencia, la morbimortalidad que genera y su coste, entre otros factores. La relevancia clínica no es subjetiva, sino medible. Su medida es la reducción del riesgo relativo. Una reducción de más del 50% indica que el estudio realizado ha sido clínicamente relevante.

MIR 04 (7961): En la evaluación de un ensayo clínico que pretenda comparar la eficacia de un nuevo antiepiléptico en dosis crecientes individualizadas, frente a fenitoína en dosis fijas de 150 mg/día en adultos con crisis generalizadas, ¿cuál es el principal problema?:

1. Relevancia clínica.*
2. Validez externa.
3. Validez interna.
4. Reclutamiento.
5. Análisis de la respuesta.

2. Errores y sesgos

Error en estadística es un concepto amplio que incluye todo lo que falta a la verdad de la causa que fuere: por azar, por mal diseño, realización o interpretación del estudio, etc. Existen dos tipos: el error aleatorio y el error sistemático. Un sesgo es el resultado de un error sistemático.



El error aleatorio es producto del azar, mientras en los sesgos existe una tendencia hacia algún lado, por errores en la metodología del estudio

Los errores que pueden afectar a las inferencias de un estudio se pueden controlar en las fases de diseño y realización iniciales o en la fase de análisis de datos posterior.

2.1. Error aleatorio

El error aleatorio se produce al azar, y no afecta a la validez del estudio (ni interna ni externa). Pero sí reduce la precisión en las mediciones y la probabilidad de hallar una asociación verdadera entre una exposición (tabaco) y un efecto (cáncer de pulmón). El error aleatorio se debe a la variabilidad en el muestreo y a la variabilidad en las mediciones.

A. VARIABILIDAD EN EL MUESTREO:

Aunque a partir de una misma población extraigamos repetidamente muestras del mismo tamaño con la misma técnica de

muestreo, las muestras resultantes serán diferentes entre sí en algún grado. Si el muestreo ha sido correcto cualquiera de estas muestras serán representativas. Y, sin embargo, son distintas. Esto es variabilidad en el muestreo.

Ejemplo: realizamos un estudio para estimar la prevalencia de HTA en una población e imaginemos que ésta fuera, en realidad, del 20%. A nosotros nos da una prevalencia estimada de 19%, por ejemplo (buen resultado!). Si extrajéramos en varias ocasiones el mismo número de individuos de la misma población con la misma técnica de muestreo, seguramente obtendríamos valores de prevalencia ligeramente distintos (21%, 22%, 18%...), y no por ello mal calculados. Esto sería debido a la variabilidad en el muestreo (aunque todo lo hayamos hecho bien).

Para disminuir el efecto de esta variabilidad en el muestreo debemos estudiar más sujetos. El aumento del tamaño muestral disminuye el efecto del error aleatorio (MIR) y aumenta la precisión de las estimaciones.

B. VARIABILIDAD EN LAS MEDICIONES:

Además de que pueden salirnos muestras ligeramente diferentes, también aumentará el error aleatorio por la variabilidad en las mediciones del fenómeno a estudio en los diferentes individuos de una muestra. Esto se debe a que nuestras mediciones no serán perfectamente exactas, entendida la exactitud como el grado en que una medición representa el verdadero valor que está siendo medido.

Ejemplo: supongamos que un individuo mide "exactamente" 175,420 cms de altura y nosotros le tallamos tres veces consecutivas: 174 cms, 176 cms y 175 cms. Evidentemente son distintas, aunque hemos medido con esmero. Existe una variabilidad que aumentará nuestro error aleatorio.

Uno de los orígenes del error aleatorio es la precisión -siempre limitada- de cualquier instrumento de medida. La altura es una variable cuantitativa continua, pero nuestro tallímetro sólo mide centímetros (no milímetros). Por lo tanto, no es posible realizar con exactitud una medición de 175,420 cms; lo más que nos acercaremos es a 175 cms, pero nuestra variabilidad en la medición no hará error en ± 1 .

No confundir...	Con...
Variabilidad en el muestreo: <ul style="list-style-type: none"> • Las infinitas posibles muestras de una misma población, aunque todas fueran representativas, serán siempre ligeramente diferentes unas de otras. • Principal fuente del error aleatorio • Disminuye, aumentando el tamaño muestral. 	Variabilidad en las mediciones: <ul style="list-style-type: none"> • La medición de las variables tendrá una exactitud siempre limitada por los propios instrumentos de medida, y se producirán ligeras desviaciones del verdadero valor que se está midiendo. • Puede reducirse, mejorando la precisión de los instrumentos de medida.
Precisión de una estimación: es inversamente proporcional a la amplitud del intervalo de confianza; a mayor amplitud, menor precisión, y viceversa.	Exactitud: grado en que una medición representa el verdadero valor que está siendo medido.

Así pues, dada su naturaleza aleatoria, el error aleatorio no afecta a la validez interna ni externa de los estudios, pero:

- 1) reduce la precisión de las estimaciones
- 2) disminuye la probabilidad de detectar como significativas diferencias o asociaciones existentes.

El error tipo I o alfa es un error aleatorio consistente en rechazar la hipótesis nula cuando es cierta (pasarse de listo, ver diferencias entre dos tratamientos que realmente no existen). El error tipo II o beta es un error aleatorio consistente en no rechazar la hipótesis nula cuando es falsa (estar ciego).

2.2. Error sistemático (Sesgo)

Sesgo es un error sistemático en la recolección, análisis, interpretación, publicación o revisión de los datos.

El error sistemático (sesgo) compromete la validez (interna y externa) de los estudios (MIR). Y, dado que las variaciones que originan los sesgos son sistemáticas y no aleatorias, al aumentar

del tamaño muestral no modifica ni corrige los sesgos de un estudio (2MIR).

Error aleatorio	Error sistemático (sesgo)
No afecta a la validez	Afecta a la validez
Disminuye con el aumento del tamaño muestral	El tamaño muestral no lo modifica

Tres son los tipos de sesgo más importantes:

1. Sesgo de selección: la población de estudio no representa a la población diana. Para algunos autores incluye el sesgo de seguimiento.
2. Sesgo de información: resultado de un error en la medición de las variables.
3. Sesgo de confusión: producido por variables que actúan como factores de confusión, y que alteran la relación entre otras variables (exposición y enfermedad).

A su vez, pueden clasificarse en: a) sesgos positivos, cuando un parámetro alcanza valores superiores al real o se produce una asociación espúrea; b) sesgos negativos, cuando la estimación hallada de un parámetro es inferior al real o se anula una asociación que verdaderamente existe; o c) sesgos que cambian el sentido de una asociación (o diferencia), cuando de factor de riesgo de una enfermedad pasa a ser un factor de protección de la misma por error.

No confundir...	con...
<p>Error aleatorio:</p> <ul style="list-style-type: none"> • Se produce al azar. • No afecta a la validez del estudio. • Reduce la precisión de las estimaciones (intervalos de confianza más amplios) y reduce la probabilidad de interpretar como significativas asociaciones o diferencias existentes. • Fuentes: variabilidad en el muestreo y en las mediciones. • Puede reducirse el error aleatorio aumentando el tamaño muestral (aumentando también la precisión de las estimaciones) <p>Precisión es la <u>ausencia de error aleatorio</u></p>	<p>Error sistemático (sesgo):</p> <ul style="list-style-type: none"> • Se produce de forma sistemática (siempre que se realice de nuevo, si no es corregida su causa). • Afecta a la validez interna y/o externa (sesgo de selección sólo la externa; el resto los dos). • No se modifica aumentando ni disminuyendo el tamaño muestral. • Existen: sesgos positivos, negativos y que cambian el sentido de la asociación (o diferencia). • Clasificación: a) Sesgo de selección y seguimiento; b) Sesgo de información o medida; y c) Sesgo de confusión. <p>Validez es la <u>ausencia de error sistemático</u>.</p>

MIR 05 (8229): Respecto a los sesgos que pueden aparecer en un ensayo clínico, se pueden tomar distintas medidas para evitarlos. Señale la respuesta **INCORRECTA**:

1. Para evitar el sesgo de selección (diferencias sistemáticas en los dos grupos de comparación) se debe ocultar la secuencia de aleatorización.
2. Para evitar diferencias sistemáticas en los cuidados recibidos por los dos grupos de comparación (sesgo de cointervención) se debe comparar con un control histórico.*
3. Para evitar una distribución no homogénea de los principales factores predictores, conocidos o no, de la variable de resultado, se asigna aleatoriamente a los participantes a cada grupo de tratamiento.
4. Para evitar diferencias sistemáticas en los abandonos, pérdidas o cumplimiento terapéutico entre los dos grupos de comparación (sesgo de atrición), se realiza un análisis por intención de tratar.
5. Para evitar el sesgo de información (diferencias sistemáticas en la evaluación de las variables de resultado entre los dos grupos de comparación), se realiza un enmascaramiento de la intervención.

MIR 07 (8743): Una de las siguientes afirmaciones con respecto a la aleatorización es **INCORRECTA**:

1. El proceso de asignación aleatoria debe estar documentado en todos sus pasos.
2. La aleatorización permite que la asignación de cada paciente incluido en el ensayo clínico sea impredecible.
3. En el caso de un ensayo con aleatorización 1:1, todos los participantes tienen las mismas probabilidades de ser asignados a uno u otro de los grupos en estudio.
4. La aleatorización asegura totalmente la homogeneidad entre los grupos en comparación ya que este es su objetivo.*
5. Idealmente el proceso de aleatorización debe estar enmascarado para evitar sesgos en la asignación a los grupos en comparación.

2.3. Sesgo de selección

Se produce cuando algunos sujetos de la población tienen más probabilidades de ser seleccionados para la muestra, pudiendo presentar características basales diferentes en cuanto a la distribución de factores del estudio (MIR) a las encontradas en general en la población de la que son extraídos.

MIR 02 (7431): En un estudio de cohortes, cómo definiría el fenómeno que puede producir el hecho que las características basales de los distintos grupos sean distintas en cuanto a la distribución de factores de riesgo para la patología cuya aparición se quiere investigar:

1. Sesgo de selección.*
2. Sesgo de información.
3. Sesgo de observación.
4. Un error aleatorio.
5. Sesgo por presencia de factor de confusión.

Es consecuencia de un muestreo incorrecto donde la muestra seleccionada no es representativa de la población (MIR). Los sesgos de selección afectan a la validez externa de los estudios, aunque éstos tengan validez interna, dado que las conclusiones extraídas no son generalizables. En general, las técnicas de muestreo probabilístico ayudan a prevenir sesgos de selección (MIR).

El sesgo de selección se puede producir en dos momentos del estudio: a) en la formación de la población de estudio o b) por problemas durante el seguimiento de los sujetos. Algunos ejemplos:

A. EN LA FORMACIÓN DE LA POBLACIÓN DE ESTUDIO: a. POR MALA DEFINICIÓN DE LA POBLACIÓN ELEGIBLE (SESGO DE INCLUSIÓN):

Por ejemplo, la selección por vía telefónica produce sesgos porque hay personas que no tienen teléfono y otras que tienen varios números. Solución: utilizar criterios de elegibilidad adecuados.

El ejemplo clásico es la encuesta que se realizó para pronosticar los resultados de las elecciones presidenciales de Estados Unidos en 1948. La encuesta se realizó por teléfono. Por aquel entonces sólo tenían teléfono las clases acomodadas (que votaban mayoritariamente al candidato republicano, Thomas Dewey). Ganó el candidato demócrata, Harry Truman.



Thomas Dewey



Harry Truman



Handwritten signature

© Curso Intensivo MIR Asturias

b. POR ERRORES EN LA EXTRACCIÓN DE LA MUESTRA:

Por ejemplo, muchas veces los que acceden a entrar en el estudio (se requiere siempre consentimiento informado) no son similares a los que no lo hacen; esto constituye un sesgo de autoselección. *Solución*: optimizar la tasa de participación y de respuesta. Cuanto más bajas sean, mayor probabilidad de sesgo; deben ser superiores al 80%.

c. POR USO INAPROPIADO DE PRUEBAS DIAGNÓSTICAS (ESPECIALMENTE EN ESTUDIOS DE CASOS Y CONTROLES):

Por ejemplo, queremos estimar la relación entre estreñimiento y cáncer de colon en un estudio de casos y controles. Para ello seleccionamos personas con cáncer (casos) y sin él (con-

troles), y preguntamos acerca del hábito intestinal ¿cuándo decimos que no tiene cáncer para entrar en "control"? Un investigador opta por un test de sangre oculta en heces y otro investigador por un enema opaco y una colonoscopia. El primero tendrá mayores errores. Algunos autores denominan a éste, sesgo de detección.



Si para detectar cáncer de colon en un estudio de casos y controles, un investigador opta por un enema opaco y otro por enema opaco y una colonoscopia, el primero tendrá mayores errores diagnósticos que el segundo

Existen dos sesgos de selección importantes con nombre propio:

• **SESGO DE BERKSON:**

Se produce cuando se selecciona la muestra en un medio hospitalario, presentando estos sujetos características distintas precisamente por estar en el hospital (están más enfermos que la población general).



© Curso Intensivo MIR Asturias

• **FALACIA DE NEYMANN (MIR):**

Se produce cuando se seleccionan casos prevalentes, en vez de incidentes, en estudios de casos y controles, siempre y cuando la exposición a estudio sea, además de factor de riesgo, un factor pronóstico de la enfermedad, es decir, que su presencia en el enfermo presuponga una mayor o menor supervivencia de vida.

De tal suerte que si se estudia un factor pronóstico que aumenta la supervivencia, será mayor la probabilidad de encontrar enfermos con ese factor de riesgo, sobrevalorando entonces la asociación de la enfermedad con dicho factor. Y viceversa, si el factor pronóstico disminuye la supervivencia, será menos probable encontrar enfermos con él, infravalorando en este caso la asociación entre la enfermedad y el factor pronóstico. Se evita tomando casos incidentes.

Ejemplo: la miocardiopatía dilatada (MD) alcohólica tiene mejor pronóstico que las de causa no alcohólica. Si se pretende hacer un estudio de casos y controles sobre la MD y se seleccionan casos prevalentes, en vez de incidentes, es muy probable que encontremos muchos más enfermos de MD de causa alcohólica que de ninguna otra, sobrevalorando la asociación entre ambas (arrojando un riesgo relativo superior al verdadero).

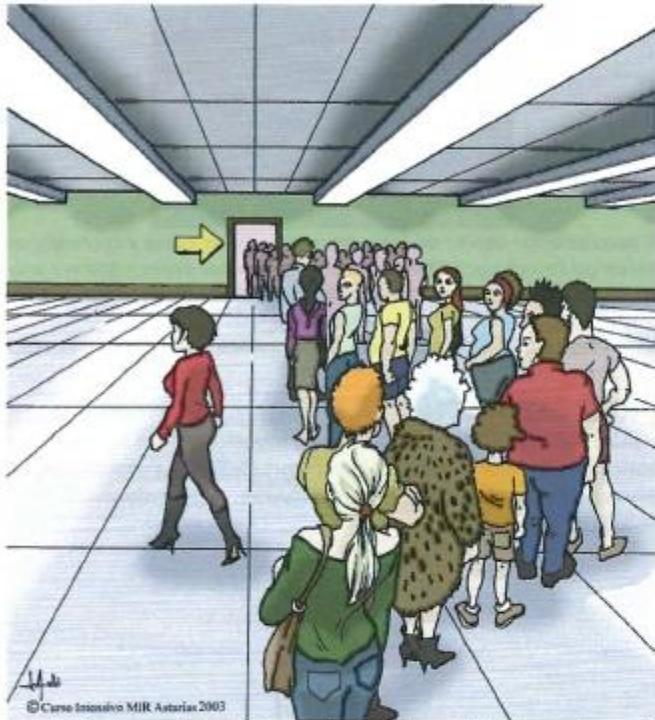
B. DURANTE EL SEGUIMIENTO (SESGO DE SEGUIMIENTO):

Puede clasificarse de forma separada, dentro del sesgo de selección o de información. No existe consenso en la clasificación de los sesgos. Pero debemos quedarnos con una idea clara sobre qué es el sesgo de seguimiento. Utilizaremos dos ejemplos:

a. POR PÉRDIDAS DURANTE EL SEGUIMIENTO:

Es decir, personas que dejan el estudio, donde la causa de este abandono puede estar relacionada precisamente con la exposición o el riesgo de enfermedad que valoramos. Por ejemplo, paciente que deja de acudir porque se encuentra mal (por la enfermedad). Solución: minimizar las pérdidas (¿incentivos?).

Las diferencias sistemáticas en los abandonos, pérdidas o cumplimiento terapéutico entre los dos grupos de comparación se denominan también sesgo de atrición (MIR).



Los individuos que dejan el estudio durante el seguimiento pueden producir sesgos y alterar los resultados del mismo

b. POR OBSERVACIÓN DESIGUAL EN CIERTOS SUJETOS:

Por ejemplo, realizar mayor número de pruebas diagnósticas a los sujetos expuestos que a los no expuestos en un estudio de cohortes. Parece evidente que en este caso habría mayor probabilidad de detectar enfermedad en los expuestos (mejor valorados) que en los no expuestos, incluso aunque hubiera igual o menor número de casos que en los no expuestos.

Sesgo de selección

- Selección de una muestra no representativa (algunos sujetos tienen más probabilidades de ser elegidos).
- Se corrige o reduce con muestreos probabilísticos.
- Sólo afecta a la validez externa (no a la interna).
- Clasificación: a) sesgo de selección propiamente (en la formación de la población): sesgo de inclusión, sesgo de autoselección, sesgo del trabajador sano...; b) sesgo de seguimiento (durante el seguimiento de participantes): sesgo de no-respuesta...
- Sesgo de Berkson: selección de una muestra en medio hospitalario.
- Falacia de Neyman: selección de casos prevalentes, en vez de incidentes, en estudios de casos y controles, cuando el factor a estudio es, además de factor de riesgo, un factor pronóstico de la enfermedad.
- Si este factor aumenta la supervivencia, se sobreestimarán su asociación (ej. miocardiopatía dilatada y alcohol). Si la disminuye, se infravalorará.

2.4. Sesgo de información

Todo error sistemático producido durante la medición de las variables del estudio, durante el proceso de recogida de datos. Así cuando se trata de mejorar los instrumentos de medida (para enfermedad o exposición) de un estudio se está evitando los sesgos de información o de medida (MIR).

Puede ser por: a) mala clasificación de las variables (expuesto/no-expuesto, o enfermo/no-enfermo), b) regresión a la media o c) falacia ecológica.

A. MALA CLASIFICACIÓN DE LAS VARIABLES:

Etiquetar erróneamente la variable "enfermedad" (enfermo/no-enfermo) o la variable "exposición" (expuesto/no-expuesto al factor de riesgo) en los participantes del estudio.



Etiquetar erróneamente a un enfermo como sano significa incluir en el estudio un sesgo de información por mala clasificación de la variable "enfermedad"

Ejemplo: si a muchos sujetos que han sido fumadores en el pasado se les etiqueta de "no-fumadores" (se les ha olvidado, por ejemplo), se estaría cometiendo un sesgo de información por mala clasificación (de la variable exposición). En este caso por un sesgo de memoria.

MIR 09 (9245): Se plantea un estudio de casos y controles para investigar la asociación entre la toma de antidepresivos durante el embarazo y el riesgo de malformaciones congénitas cardiacas. Con el fin de minimizar la posibilidad de un sesgo consistente en que las mujeres que han tenido un hijo con malformación recuerden mejor su posible exposición en el embarazo a medicamentos antidepresivos que las mujeres con niños no malformados, ¿cuál de los siguientes procedimientos sería el más adecuado?

1. Realizar un estudio de campo, con entrevistas directas a los casos y a los controles.
2. Realizar un estudio de corte transversal, para estimar la prevalencia de uso de antidepresivos en embarazadas.
3. Realizar un estudio en el que se obtenga de una base de datos la información sobre todas las prescripciones médicas realizadas durante el embarazo a los casos y a los controles.*
4. Seleccionar los controles mediante apareamiento (matching) con los casos.
5. Realizar un análisis por intención de tratar.

Se evitan sesgos por mala clasificación los denominados sesgos de observación, a los que son característicamente susceptibles los ensayos clínicos, debido a la subjetividad de las personas participantes (1º pacientes, 2º médicos, 3º analistas).

Se evitan mediante técnicas de enmascaramiento (MIR):

- Simple ciego (el paciente desconoce el tipo de intervención que se le administra: placebo o fármaco)
- Doble ciego (paciente y médico lo desconocen).
- Triple ciego (lo desconocen paciente, médico y analista de datos).

"EL ENMASCARAMIENTO evita los sesgos de INFORMACIÓN, (sesgo de observación)"



MIR 00 (6920): El propósito de emplear la técnica del doble ciego en un ensayo clínico es:

1. Evitar sesgos de medida.*
2. Asegurar la comparabilidad inicial de los grupos de tratamiento.
3. Evitar errores en la aleatorización.
4. Evitar sesgos de selección.
5. Asegurar la validez externa de los resultados.

Para valorar el sesgo de información se necesita un criterio de verdad, que no siempre existe. Los errores de clasificación del FR son más frecuentes que con la enfermedad.

Estos errores en la clasificación pueden producirse de forma diferencial o no diferencial entre los distintos grupos; y a su vez puede producirse también con varias variables.

a. NO DIFERENCIAL:

La proporción de veces que nos equivocamos al clasificar a los sujetos de enfermos/no-enfermos o expuestos/no-expuestos es similar en los diferentes grupos del estudio, dado que la sensibilidad y especificidad de las pruebas diagnósticas son iguales en unos y en otros.

b. DIFERENCIAL:

La proporción de veces que nos equivocamos al clasificar la enfermedad y exposición es, en este caso, distinta en los diferentes grupos del estudio, dado que la sensibilidad y especificidad de las pruebas diagnósticas son diferentes en unos y en otros (por ejemplo, mayor sensibilidad y especificidad en casos que en controles).

c. DE COVARIABLES:

En muchas ocasiones se estudian múltiples variables, donde unas influyen sobre otras (factores de confusión). Si cometemos errores en la clasificación de alguna o varias de ellas se puede dificultar la detección de variables que actúen como factores de confusión en el análisis.

Baste decir que un 10% de mala clasificación puede reducir a la mitad las posibilidades de detectar un sesgo de confusión; siendo virtualmente imposible ante un 30% (el ajuste es incapaz de detectar disminución del RR suficiente para sospechar que esa variable por la que hemos ajustado, sea factor de confusión).

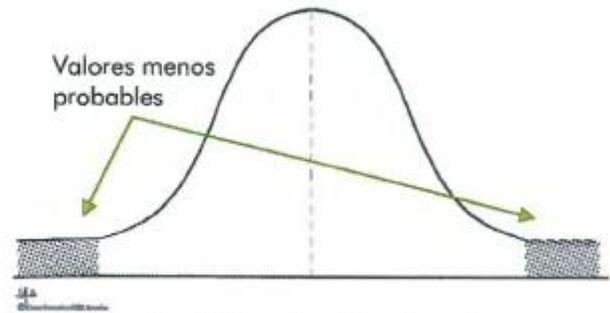
MIR 08 (9010): El sesgo amnésico que puede producirse en los estudios de casos y controles:

1. Es un sesgo de selección.
2. Es un sesgo de clasificación diferencial.*
3. Es un sesgo de clasificación no diferencial.
4. Es un sesgo de confusión.
5. No es realmente un sesgo sino un error que acontece en estudios con ancianos.

B. REGRESIÓN A LA MEDIA:

En distribuciones de variables cuantitativas continuas los valores extremos son menos probables que los valores próximos a la media, por lo que mediciones posteriores a un valor extremo inicial tienden a regresar a valores más centrales, y por ende, más probables. El efecto de la regresión a la media es mayor cuanto más extremos son los valores iniciales.

Valores más probables



En distribuciones de variables continuas los valores extremos son menos probables que los cercanos a la media

Esto es importante en la valoración de la eficacia de un tratamiento. Generalmente en medicina se tratan valores extremos de una variable continua (hipercolesterolemia, hipertensión...). Siguiendo el razonamiento, existe una tendencia natural (con o sin eficacia del tratamiento) a que las siguientes mediciones de la variable continua fueran menores, más cercanas a la media. De hecho podría postularse erróneamente que "el tratamiento es tanto más eficaz cuanto más grave sea la enfermedad (ya que los valores serían más extremos... pero también aumentaría la regresión a la media)".

C. FALACIA ECOLÓGICA:

Se produce por la propia agregación, que obliga a asumir que la exposición media de una comunidad es la misma que la que tiene cada sujeto de la misma, siendo esto una falacia. Cuanto mayor es la agregación mayor será la falacia ecológica.

Es la mayor limitación de los estudios ecológicos (en estos estudios se comparan comunidades de áreas geográficas diferentes).

Para evitarlo es recomendable comparar áreas que sean homogéneas, que tengan una distribución similar de las variables más importantes. Aún así es difícil de controlar, de forma que los resultados de estudios ecológicos no son válidos para establecer relaciones causales, aunque sí para generar hipótesis de trabajo (para otros estudios: cohortes).

Sesgo de información

- Error en la medición (recogida de datos).

2.5. Sesgo de memoria

El sesgo de memoria (recall bias, en inglés) es un sesgo típico de los estudios de casos y controles que consiste en que las personas del grupo de casos (con la enfermedad) recuerdan mucho mejor exposiciones a posibles factores de riesgo que las personas del grupo control (MIR).

MIR 13 (10197) (198): Queremos conocer si el consumo de café puede estar asociado a la malformación neonatal por lo que se diseña un estudio casos control. Se entrevista a un grupo de mujeres que han tenido niños con malformaciones y lo mismo en un grupo de madres sin hijos con malformaciones. Las entrevistas las realizaran dos entrevistadores entrenados mediante un cuestionario previamente validado. Además, los entrevistadores ignoran si la entrevistada es un caso o un control. ¿Que tipo de sesgo podemos introducir?

1. Sesgo de información.
2. Sesgo de memoria.*
3. Sesgo del entrevistador.
4. Sesgo de selección.
5. Falacia ecológica.

2.6. Sesgo del entrevistador

El entrevistador "fuerza" u orienta la respuesta del participante, consciente o inconscientemente. Para evitarlo es conveniente que el investigador desconozca la hipótesis que se desea estudiar.

2.7. Sesgo de obsequiosidad

Aquí el participante contesta lo que cree que el entrevistador desea oír. Para intentar corregir este sesgo se recomienda redactar las preguntas de un modo cuidadoso, y pedirle al entrevistado que explique o desarrolle sus respuestas.

2.8. Sesgo de respuesta invariable

El participante responde invariablemente lo mismo (por ejemplo responde siempre que sí). Típico de cuestionarios por ordenador, cuestionarios muy largos y preguntas dicotómicas.

2.9. Sesgo de falseamiento

Ocurre cuando las preguntas son consideradas por el participante como íntimas o traumatizantes, y llevan a dar respuestas falsas o incompletas.

2.10. Sesgo de cuestionario

Tiene lugar por un cuestionario confeccionado erróneamente (preguntas poco claras, dos preguntas en una, términos incomprensibles).

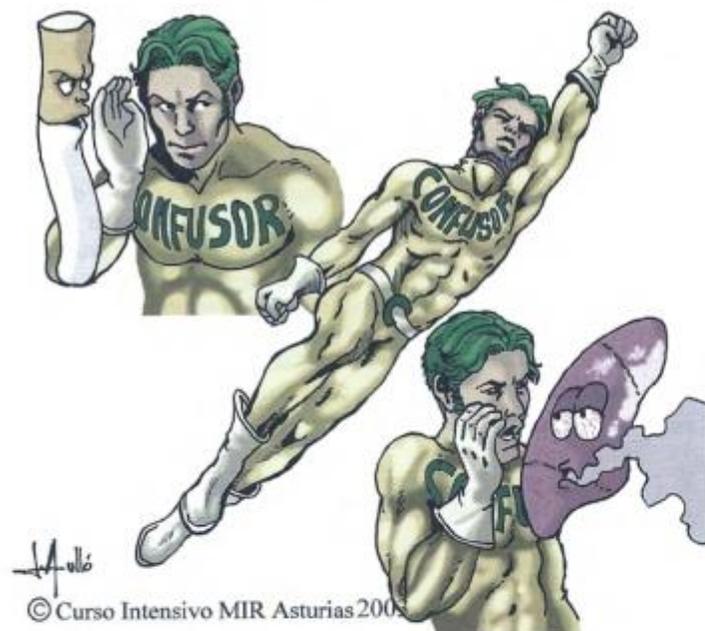
2.11. Sesgo o factor de confusión

Un sesgo o variable de confusión es una variable que influye sobre el resultado de las intervenciones y que no se distribuye homogéneamente entre grupos (MIR). Es un sesgo que se encuentra en la población de referencia y se produce por las relaciones que mantiene una variable, llamada fenómeno o sesgo de confusión, tanto con la variable predictora (factor de riesgo) como a la variable de efecto o desenlace (efecto o enfermedad). Los factores o variables de confusión, también conocidos como factores modificadores del efecto, pueden hacer que aparezca una asociación inexistente (sesgo positivo), que desaparezca o disminuya (sesgo negativo) o incluso que un factor de riesgo de enfermedad resulte, en el estudio, un factor protector (sesgo que cambia el sentido de la asociación).

A. REQUISITOS:

El factor de confusión debe:

- 1) ser factor de riesgo o marcador de riesgo del efecto (SMIR), independiente de la exposición, o una variable pronóstica para la enfermedad (RR u $OR \neq 1$)
- 2) asociarse a la exposición (SMIR) (RR u $OR \neq 1$)
- 3) no ser variable intermedia entre la exposición y el efecto (MIR), es decir, no ser un paso intermedio en la cadena causal entre la exposición y el efecto o enfermedad.



© Curso Intensivo MIR Asturias 2008

El sesgo de confusión se asocia tanto al factor de riesgo como a la enfermedad, sin ser un paso intermedio entre ambas

MIR 00 FAMILIA (6695): ¿Cuál de las siguientes afirmaciones es verdadera respecto a los factores de confusión en un estudio observacional?:

1. Se asocian a una posible causa y constituyen un factor de riesgo para el efecto estudiado.*
2. Son equivalentes a los sesgos de información no diferencial.
3. Se corrigen aumentando el tamaño de la muestra.
4. Se evitan con un buen entrenamiento de los investigadores que participan en el estudio.
5. Son uno de los tipos especiales de errores aleatorios.

MIR 00 (6927): ¿Qué es una variable de confusión en un ensayo clínico?:

1. Una variable para la que no hay un procedimiento validado de medición.
2. Una variable de respuesta poco sensible y específica.
3. Una variable basal que influye sobre el resultado de las intervenciones y que no se distribuye homogéneamente entre grupos.*
4. Una variable que, por error de diseño, se mide sin que en realidad aporte información sobre la diferencia entre las intervenciones.
5. La variable con cuya medición y mediante la aplicación de las técnicas estadísticas adecuadas se puede dilucidar si los tratamientos tienen distinta eficacia.

MIR 01 (7179): En un estudio epidemiológico observacional, ¿cómo definiría el fenómeno que puede producir la existencia de una variable que se relaciona a la vez con la exposición al factor etiológico y con la enfermedad estudiada?:

1. Sesgo de selección.
2. Sesgo de información.
3. Sesgo de observación.
4. Error aleatorio.
5. Sesgo por factor de confusión.*

MIR 03 (7695): En un estudio de casos y controles se estudia la relación entre cáncer de cavidad bucal (efecto) y consumo de bebidas alcohólicas (exposición). Como definiría el fenómeno que puede producir el hecho que el hábito de fumar se asocia a la aparición de cáncer bucal y también a un mayor consumo de bebidas alcohólicas?:

1. Sesgo de información.
2. Sesgo por factor de confusión.*
3. Un error aleatorio.
4. Sesgo de observación.
5. Sesgo de selección.

MIR 04 (7967): En un estudio de casos y controles se encontró una asociación entre la ingesta de alcohol y el desarrollo de cáncer esofágico. Dado que el consumo de tabaco se encontró asociado tanto a la ingesta de alcohol como al desarrollo de cáncer de esófago, el consumo de tabaco en este estudio debe considerarse como:

1. Un factor (sesgo) de confusión.*
2. Un sesgo de información.
3. Un sesgo de memoria.
4. Un sesgo de selección.
5. Un sesgo de falacia ecológica.

MIR 07 (8737): Se va a realizar un estudio para comprobar si la administración de determinado antipsicótico podría estar relacionado con un aumento de la incidencia de conductas suicidas en esquizofrénicos en los que se inicia el tratamiento de un brote agudo. Los pacientes más graves pueden ser más proclives a serles indicado este antipsicótico y también tienen mayor riesgo suicida. Teniendo en cuenta este posible factor de confusión "por indicación". ¿Cuál de los siguientes tipos de estudio minimiza mejor en el diseño la posibilidad de que se produzca este sesgo de confusión?:

1. Un ensayo clínico aleatorizado.*
2. Un estudio de casos y controles apareado.
3. Un estudio de cohortes de seguimiento prospectivo.
4. Un estudio de prevalencia.
5. Un estudio de casos y controles anidado en una cohorte.

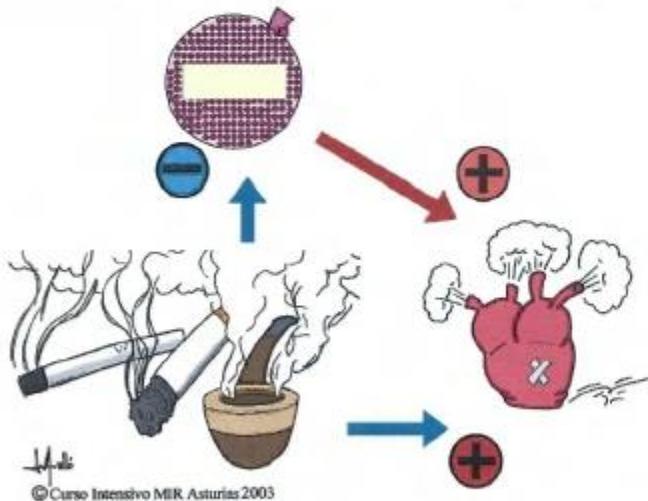
En definitiva, debe asociarse con el factor de riesgo y con la enfermedad independientemente, sin ser un paso intermedio en la relación entre ambos. Pero no siempre es fácil saber que una variable es un factor de riesgo y que no es un paso intermedio con otra variable que estudiamos.

Un equipo de investigadores estudia la relación entre niveles de HDL-colesterol y cardiopatía isquémica. Mediante un estudio de cohortes, valoran la fuerza de asociación entre los bajos niveles de HDLc y la incidencia de diagnósticos de cardiopatía. Esta fuerza de asociación es medida mediante el riesgo relativo (RR), que es la razón entre las incidencias de cardiopatía isquémica en fumadores respecto de no fumadores. Si un RR resulta superior a 1 y es estadísticamente significativo, se interpreta que los bajos niveles de HDLc son factor de riesgo para la cardiopatía isquémica, tanto más fuerte cuanto más elevado sea el valor del RR. En sus resultados, los investigadores observan un $RR=4$ y resulta estadísticamente significativo.

Se conoce que los niveles de HDL-colesterol son deprimidos por el tabaquismo, con lo cual se podría esperar que los sujetos del grupo de fumadores tuviesen una distribución de valores inferiores a los no fumadores. Por otra parte, se sabe también que el tabaquismo constituye un factor de riesgo independiente para el desarrollo de cardiopatía coronaria.

Para controlar este posible factor de confusión (tabaco), los investigadores separaron en fumadores y no fumadores a los sujetos participantes. Posteriormente, en la fase de análisis, realizaron un ajuste del RR según el factor tabaquismo valorado en los distintos sujetos, y obtuvieron un valor de $RR=2$ (HDLc-cardiopatía isquémica), inferior al observado a partir de los datos crudos.

Parte de la fuerza de asociación inicialmente observada entre bajos niveles de HDLc y cardiopatía isquémica era debida a una contribución independiente del tabaco. Es decir, el tabaquismo estaba actuando como factor de confusión, produciendo la observación de una asociación aparente entre bajos niveles HDLc y cardiopatía isquémica más intensa que la real.



El tabaco es factor de confusión en un estudio entre bajos niveles de HDL-colesterol y cardiopatía isquémica, por relacionarse con ambos y no ser un paso intermedio entre ambos

Pensamos en factor de confusión si el RR crudo y el RR ajustado difieren de forma significativa (3MIR). Y el factor de confusión será la variable por la que se ha ajustado en el estudio (tabaco).

2.12. Control de sesgo de confusión

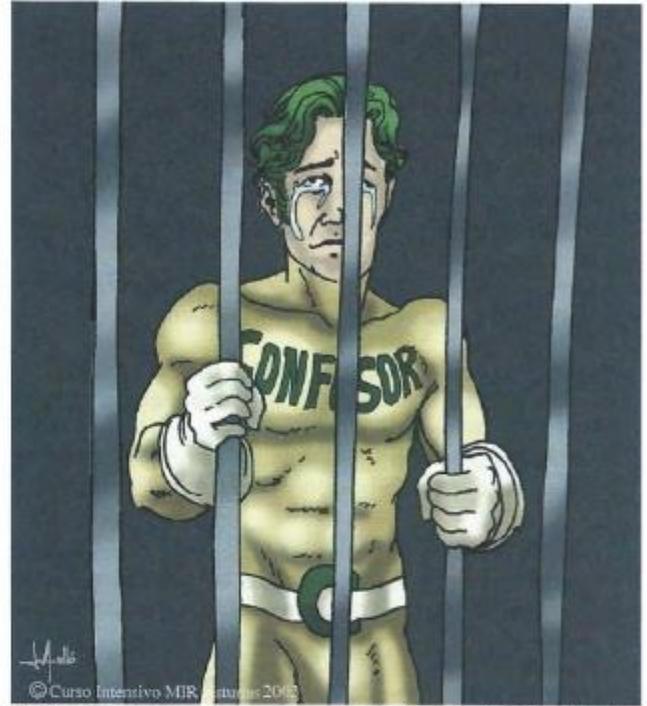
Los sesgos de confusión deben ser prevenidos por el investigador en el diseño del estudio o controlados en la fase de análisis. Para ello, se debe medir en los sujetos estudiados cualquier variable que se sospeche o se conozca que puede actuar como factor de confusión en la asociación considerada.

A. DURANTE EL DISEÑO:

1. Muestreo aleatorio (MIR), especialmente el muestreo estratificado (2MIR).
2. Emparejamiento (apareamiento, matching) (3MIR). Utilizada habitualmente en estudios de casos y controles para evitar que variables constitucionales (edad, sexo...) actúen como sesgos de confusión. Se selecciona, para cada caso, un control (o dos) que tenga el mismo valor de la posible variable de

confusión. Ejemplo: por cada mujer incorporada al grupo de casos, una mujer en el grupo de controles; por cada varón en los casos, un varón en los controles.

3. Criterios de restricción (2MIR). Se excluyen los sujetos que presenten la variable que queremos controlar.



Se puede controlar los factores de confusión durante el diseño del estudio y durante el análisis de los datos

B. DURANTE EL ANÁLISIS:

1. Análisis estratificado (2MIR) (análisis por separado de los datos según tenga o no la variable a controlar como posible factor de confusión) y técnicas de ajuste o estandarización (MIR) por factor de confusión (cálculo del riesgo relativo bruto y ajustado por factor de confusión).
2. Análisis multivariante (2MIR). Prueba de contraste de hipótesis que permite estudiar un conjunto de variables y las interacciones entre ellas. Sin embargo, el análisis multivariante no permite corregir otro tipo de sesgos (sesgos de selección ni sesgos de información) (MIR).

MIR 01 (7194): ¿Cuál de los siguientes **NO** es un método para controlar la confusión en los estudios de casos y controles?:

1. El apareamiento.
2. La restricción.
3. El ajuste multivariante.
4. La entrevista sistemática estructurada.*
5. El análisis estratificado.

MIR 10 (9472): ¿Cuál de las siguientes afirmaciones sobre los factores de confusión es correcta?:

1. Sólo hay que controlarlos cuando son una variable intermedia en la cadena causal.
2. Pueden controlarse mediante enmascaramiento.
3. Pueden contratarse mediante emparejamiento.*
4. No requieren control porque no modifican el efecto.
5. Sólo se pueden controlar recurriendo a técnicas estadísticas.

MIR 11 (9706): ¿Cuál de las siguientes afirmaciones acerca de la confusión en los estudios epidemiológicos es **FALSA**?:

1. Para "controlar" la confusión en un estudio las variables que la producen deben conocerse obligatoriamente de antemano.*
2. La confusión ocurre cuando una tercera variable se asocia tanto a la exposición como al desenlace.
3. Los análisis estadísticos multivariantes reducen la confusión.
4. El diseño de los ensayos controlados con distribución aleatoria minimiza la confusión.
5. La medición inexacta o sesgada de algunas variables puede generar confusión.

MIR 13 (10199) (200): Se quiere realizar un estudio para valorar el efecto de la exposición al consumo de alcohol sobre el cáncer de páncreas. Es posible que la relación pueda estar afectada por el efecto de la exposición al consumo de tabaco. Si sólo se quiere analizar el efecto del consumo de alcohol, ¿qué tipo de sesgo se puede producir?

1. Sesgo de exposición.
2. Sesgo de diagnóstico.
3. Sesgo de realización.
4. Sesgo por efecto vigilancia (o de Hawthorne).
5. Sesgo de confusión. *

Sesgo de confusión

- Variable basal que no se distribuye homogéneamente entre los grupos del estudio.
- Se asocia a la exposición, es factor de riesgo o pronóstico de la enfermedad y no es un paso intermedio entre ambas.

Se pueden:

- Prevenir durante la fase del diseño (muestreo aleatorio – especialmente de tipo estratificado-, emparejamiento o matching y criterios de restricción).
- Controlar durante el análisis (análisis estratificado con técnicas de ajuste o estandarización y análisis multivariante).

No confundir...

Validez interna: permite inferir conclusiones de los participantes del estudio.

Error aleatorio (debido al azar): reduce precisión y probabilidad de detectar diferencias o asociaciones ciertas. Se mejora con el aumento del tamaño muestral.

Sesgo de selección: muestra no representativa (falta de validez externa). Se corrige con muestreos probabilísticos. Pueden producirse en la formación o en el seguimiento de los grupos (sesgo de seguimiento).

Falacia de Neyman: selección de casos prevalentes, en vez de incidentes, en estudios de casos y controles, cuando el factor a estudio es, además de factor de riesgo, un factor pronóstico de la enfermedad.

Sesgo de observación: producido por subjetividad de participantes. Control con técnicas de enmascaramiento (simple, doble y triple ciego).

Con...

Validez externa: permite inferir conclusiones extrapolables a otras poblaciones.

Error sistemático (sesgos): afecta a la validez del estudio. No se modifica con el tamaño muestral (debe corregirse la causa).

Sesgo de información: errores en la medición de la variable. Puede ser diferencial (mala clasificación distinta según el grupo) o no diferencial (mala clasificación similar en todos los grupos).

Sesgo de Berkson: selección de muestra en medio hospitalario.

Sesgo de confusión: variables asociadas a factor de riesgo y enfermedad, sin ser paso intermedio entre ambas. Control con: muestreo probabilístico (estratificado), matching, criterios de restricción, análisis estratificado (ajuste) y análisis multivariante.



RESUMEN DE ERROR Y SESGO

1. VALIDEZ INTERNA Y EXTERNA

- La validez de los resultados de una muestra puede servirnos para realizar inferencias. Se distinguen dos tipos: validez interna y validez externa.

1.1. VALIDEZ INTERNA

- Grado en que los resultados de un estudio son extraídos correctamente de los participantes del mismo. La validez interna permite hacer inferencias correctas sobre el efecto de un factor de riesgo en los participantes del estudio (MIR). Afectan a la validez interna: sesgo de memoria (MIR) (sesgo de información (MIR)), sesgo de confusión (MIR), utilización de grupos no comparables (MIR)...

1.2. VALIDEZ EXTERNA

- Grado en que las conclusiones de un estudio son generalizables a la población diana (u otras poblaciones). La validez externa presupone validez interna. Afectan a la validez externa: sesgo de selección (muestra no representativa) (MIR), toda alteración de la validez interna...

2. ERROR ALEATORIO

- Error, incluye en Estadística, todo lo que falta a la verdad. Se distinguen dos tipos: error aleatorio y error sistemático, siendo el sesgo el resultado de un error sistemático.

- El error aleatorio se produce al azar y no afecta por ello a la validez del estudio (ni la interna ni la externa). Disminuye la precisión en las estimaciones (intervalos de confianza más anchos) y reduce la probabilidad de encontrar como significativas asociaciones o diferencias que realmente existan. Sus fuentes más importantes son: variabilidad en el muestreo y variabilidad en las mediciones.

2.1. VARIABILIDAD EN EL MUESTREO

- Resultado de las pequeñas diferencias que se producirían al muestrear infinitamente un población (existen infinitas muestras representativas ligeramente diferentes en una misma población). Principal fuente del error aleatorio. El aumento del tamaño muestral aumenta la precisión de las estimaciones (MIR), disminuye la variabilidad en el muestreo, y por ende, el error aleatorio.

2.2. VARIABILIDAD EN LAS MEDICIONES

- Cualquier instrumento de medida tiene un grado límite de precisión, a partir del cual no puede ser más exacto. La exactitud es el grado en que una medición representa el verdadero valor que se está midiendo. A mayor amplitud de un intervalo de confianza, menor precisión de la estimación.

3. ERROR SISTEMÁTICO (SESGO)

- Se produce de forma sistemática (siempre que se repite el proceso, si no se corrige la causa), por lo que afecta a la validez del estudio. El aumento o disminución del tamaño muestral no modifica los sesgos (2MIR). Existen sesgos positivos (asociaciones espúreas o sobrevaloradas), negativos (asociaciones anuladas o infravaloradas) y que cambian el sentido de la asociación (de factor de riesgo a factor protector).

- Clasificación: a) sesgo de selección y sesgo de seguimiento, b) sesgo de información o de medida, y c) sesgo o factor de confusión.

3.1. SESGO DE SELECCIÓN (Y DE SEGUIMIENTO)

- El sesgo de selección se produce cuando se elige una muestra no representativa (MIR): algunos sujetos tienen más probabilidades de ser elegidos para la muestra, pudiendo presentar características basales diferentes en cuanto a la distribución de factores del estudio (MIR) a las encontradas en general en la población de la que son extraídos. Se corrige o reduce con muestreos probabilísticos con asignación aleatoria (MIR) (todos los sujetos tienen la misma probabilidad de ser seleccionados). Sólo afecta a la validez externa (no a la interna (MIR)).

- Clasificación: a) sesgo de selección propiamente (en la formación de la población): sesgo de inclusión, sesgo de autoselección, sesgo del trabajador sano...; b) sesgo de seguimiento (durante el seguimiento de participantes): sesgo de no-respuesta...

- Dos sesgos de selección con nombre propio son:

- SESGO DE BERKSON: selección de una muestra en medio hospitalario.

- FALACIA DE NEYMAN: selección de casos prevalentes, en vez de incidentes, en estudios de casos y controles, cuando la exposición a estudio es, además de factor de riesgo, un factor pronóstico de la enfermedad. El estudio de casos prevalentes de una enfermedad de elevada letalidad puede producir sesgo de selección (MIR). Si este factor aumenta la supervivencia, se sobreestimará su asociación (ej. miocardiopatía dilatada y alcohol). Si la disminuye, se infravalorará. Se soluciona tomando casos incidentes de enfermedad.

3.2. SESGO DE INFORMACIÓN O MEDIDA

- Error en la medición de las variables del estudio (recogida de datos). La mejora de los instrumentos de medida pretenden evitar un sesgo de información (MIR).

- Clasificación:

- MALA CLASIFICACIÓN DE VARIABLES (exposición o enfermedad). Puede ser no-diferencial (sensibilidad y especificidad de pruebas diagnósticas es igual para todos los grupos del estudio) o diferencial (sensibilidad y especificidad distintas: más errores en unos grupos que en otros). -Sesgo de observación: debido a la subjetividad de los participantes (1° pacientes, 2° médicos, 3° analista/farmacéutico). Se controla mediante técnicas de enmascaramiento (2MIR) (simple, doble (MIR) o triple ciego).

- REGRESIÓN A LA MEDIA. Valores extremos regresan a valores centrales en siguientes mediciones de distribuciones cuantitativas continuas. Ojo en la interpretación de la eficacia de tratamiento.

- FALACIA ECOLÓGICA. Consiste en asumir que la exposición individual es igual a la exposición media poblacional, lo cual es falso, y más cuanto mayor sea el grado de hacinamiento. Por ello no son válidos para establecer relaciones causales (MIR).

3.3. SESGO DE MEMORIA

- Es un sesgo típico de los estudios de casos y controles que consiste en que las personas del grupo de casos (con la enfermedad) recuerdan mucho mejor exposiciones a posibles factores de riesgo que las personas del grupo control.

3.4. SESGO DE CONFUSIÓN

- Variable basal que no se distribuye homogéneamente entre los grupos del estudio (MIR). Requisitos de un factor de confusión (4MIR): a) se asocia a la exposición, b) es factor de riesgo o pronóstico de la enfermedad y c) no es un paso intermedio entre ambas.

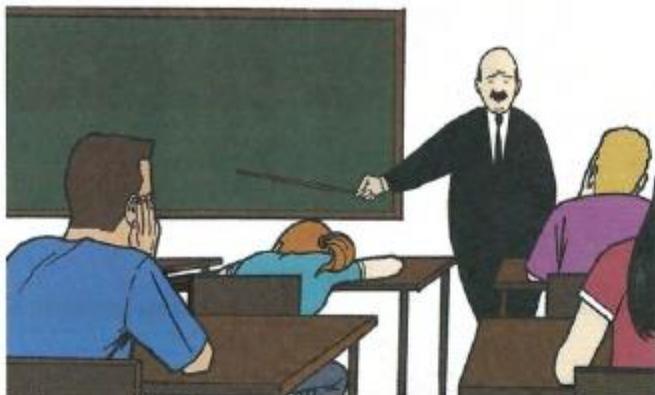
- Los sesgos de confusión se pueden:

- Prevenir durante la fase del diseño (muestreo aleatorio (MIR) –especialmente de tipo estratificado (2MIR)-, emparejamiento o matching (3MIR) y criterios de restricción (2MIR)).

- Controlar durante el análisis (análisis estratificado (2MIR) con técnicas de ajuste o estandarización –el riesgo relativo u odds ratio brutas y ajustadas difieren- (3MIR) y análisis multivariante (2MIR)). El análisis multivariante no controla otros sesgos (MIR) (selección, información)

1. Conceptos básicos

Una parte importante de la estadística descriptiva es la organización de los datos, permitiéndonos "ver" la información con mayor facilidad.



© Curso Intensivo MIR Asturias

La función de la estadística descriptiva es ayudarnos a ver de forma ordenada la información obtenida

- **DISTRIBUCIÓN:** conjunto de las observaciones realizadas sobre los distintos valores que toma una variable.
- **FRECUENCIAS:** número de veces o proporción del total que se observa cada valor determinado. Pueden ser absolutas o relativas (porcentual).
- **INTERVALOS DE CLASE (variable cuantitativas continuas):** segmentos de la variable definidos por sus respectivos límites de intervalo, siendo éstos los extremos de los valores contenidos en el intervalo. Ej: 1,6,4,5,6,7,8,10. Podemos hacer dos intervalos de clase: 1-5 (frecuencia: 3) y 6-10 (frecuencia: 5). Límites 1 y 5, y 6 y 10, respectivamente.
- **AMPLITUD:** distancia entre límites. Ej: 4 (5-1) y 4 (10-6), en el caso anterior.
- **MARCA DE CLASE:** valor representativo de todos los valores que abarca un intervalo de clase, y que suele ser el punto medio del intervalo. Ej: 3 (para 1-5) y 8 (para 6-10)

La segmentación de una variable cuantitativa en intervalos supone una pérdida de información, mayor cuanto menor sea el número de intervalos (agrupamos más los datos). Ej: antes partíamos de una distribución de 8 elementos y lo redujimos a sólo 2 intervalos, con sólo 2 valores representativos (marcas de clase).

2. Medidas de frecuencia

Los valores (x_i) de una variable y sus respectivas frecuencias se pueden ordenar y representar en filas y columnas, configurando así una tabla de frecuencias (v. tabla).

Las frecuencias pueden ser absolutas o relativas, y ambas pueden ser simples o acumuladas.

A. FRECUENCIA ABSOLUTA (n_i):

Número de veces que se repite (que se ha observado o medido) cada valor x_i de la variable. Ej: existen 38 elementos con valores comprendidos entre 130-150 en el primer intervalo

B. FRECUENCIA ABSOLUTA ACUMULADA (N_i):

Número de valores de la variable iguales o inferiores a uno dado, o a un intervalo determinado. Ej: en el tercer intervalo es de 188 (38+100+50). El último valor o intervalo será siempre el total de la población N .

C. FRECUENCIA RELATIVA (f_i):

Proporción de la frecuencia absoluta de cada valor o intervalo entre el total de observaciones (tamaño de la muestra o N).

$$f_i = \frac{n_i}{N}$$

Es una medida de probabilidad individual del valor o intervalo considerado. Ej: la probabilidad de encontrar un valor del intervalo 150-170 de la tabla es del 36,4% (0,364).

D. FRECUENCIAS RELATIVAS ACUMULADAS (F_i):

Proporción de la distribución con valores iguales o inferiores a uno dado. Sería la probabilidad de encontrar un valor igual o inferior a uno dado. La probabilidad del último valor o intervalo sería del 100% (ó 1).

Intervalo de clase	Marca de clase	n_i	N_i	f_i	F_i
$130 \leq x_i < 150$	140	38	38	13,8	13,8
$150 \leq x_i < 170$	160	100	138	36,4	50,2
$170 \leq x_i < 190$	180	50	188	18,2	68,4
$190 \leq x_i < 210$	200	20	208	7,3	75,7
$210 \leq x_i < 230$	220	18	226	6,5	82,2
$230 \leq x_i < 250$	240	15	241	5,5	87,7
$250 \leq x_i < 270$	260	12	253	4,4	92,1
$270 \leq x_i < 290$	280	9	262	3,3	95,4
$290 \leq x_i < 310$	300	8	270	2,9	98,3
$310 \leq x_i < 330$	320	3	273	1,1	99,4
$330 \leq x_i < 350$	340	2	275	0,6	100
Total		275		100	

Tabla de frecuencias de una distribución de colesterolemia en mg/dl en una muestra de 275 individuos, agrupada en intervalos de clase.

Existen tres medidas de frecuencia más importantes: proporción (MIR), razón (MIR) y tasa (MIR).

MIR 00 (6912): ¿Cuál de las siguientes **NO** es una medida de frecuencia?

1. Prevalencia.
2. Riesgo relativo.*
3. Tasa.
4. Razón.
5. Proporción.

El riesgo relativo es una medida de cuántas veces más enfermedad encontramos entre un grupo expuesto a un determinado factor de riesgo y un grupo no expuesto al mismo. Es la medida de fuerza de asociación propia de los estudios de cohortes (v. más adelante).

2.1. Proporción

Cociente en el que el numerador es una magnitud incluida en el denominador. Un ejemplo particular de proporción es la prevalencia (MIR). Ej: proporción de hipertensos en una población (HTA/ [HTA y no-HTA]; el numerador forma parte del denominador) es de 0,22 ó 22% (un porcentaje es una proporción multiplicada por 100).

$$P = \frac{a}{a+b}$$

Una proporción es una expresión de la probabilidad de un suceso. Ej: en el caso anterior la probabilidad de que un individuo elegido al azar sea hipertenso es $P(\text{hipertenso})=0,22$.



repeMIR

La prevalencia mide los casos existentes de una exposición o una enfermedad en un momento dado. (3+)

Medidas de probabilidad: frecuencia relativa (f_i) y proporción ($a/a+b$)

2.2. Razón

Cociente en el que el numerador no está incluido en el denominador. Es un término adimensional. Ej: razón de sexo (varones/mujeres).

$$R = \frac{a}{b}$$

Una **caso particular** es la **odds**, definida como una razón en la que el **numerador** expresa la probabilidad de que un determinado suceso ocurra y el **denominador** la probabilidad de que no ocurra. Es un término utilizado en los estudios de casos y controles para el cálculo de su medida de fuerza de asociación **odds ratio** (cociente entre **odds de exposición de casos** (casos expuestos/ casos no expuestos) y **odds de exposición de controles** (controles expuestos/ controles no expuestos).

2.3. Tasa

Es un **tipo especial de proporción** que **relaciona el cambio en una magnitud por** unidad de cambio en otra magnitud (generalmente **tiempo**). Las tasas epidemiológicas llevan habitualmente la dimensión **tiempo expresada en su denominador** como **personas-año (MIR)**, y contienen un **coeficiente múltiplo de 10** (normalmente 1000).

La **densidad de incidencia** de una enfermedad es una tasa que expresa los **casos nuevos aparecidos de una población por unidad de tiempo**. Ej: 50 casos nuevos en 5 años de cáncer de ovario en 1000 mujeres; su **densidad de incidencia** será de **10 nuevos casos por cada 1000 en cada año** ($50 / 5\% = 10\%$).

$$\frac{\text{casos nuevos}}{\text{total de personas} \cdot \text{tiempos en observación}}$$

No expresa probabilidad de tener (prevalencia) ni desarrollar (incidencia acumulada) una enfermedad (3MIR).

Proporción

Denominador contiene numerador. Ej. Prevalencia. Expresa probabilidad individual de tener enfermedad.

Razón

Denominador NO contiene numerador. Ej. Razón de sexo; odds (razón de probabilidades).

Tasa

Relaciona cambio de magnitud con tiempo (denominador habitual personas-año). Ej. Densidad de incidencia. NO expresa probabilidad individual de desarrollar ni de tener enfermedad.

3. Representación gráfica elemental

La **forma más correcta de representar** gráficamente una distribución **depende básicamente del tipo de variable** que se ha medido. Según éstas los **tipos de representación más comunes** son los siguientes:

3.1. Variables cualitativas y cuantitativas discretas

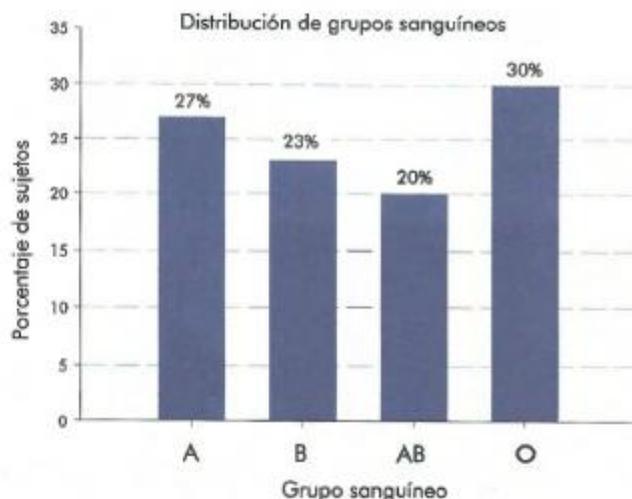
A. DIAGRAMA DE BARRAS:

Es el más habitual en este tipo de variables. Se representan los **valores de la variable en abscisas** (línea horizontal) y sus respectivas **frecuencias en ordenadas** (línea vertical, perpendicular a la anterior), resultando unas barras con las frecuencias de cada valor, pudiendo ser tanto absolutas como relativas.

Ej: trasladamos las frecuencias relativas de los distintos grupos sanguíneos a diagramas en el siguiente caso:

Grupo sanguíneo	Frecuencia (%)
A	27
B	23
AB	20
O	30
Total	100

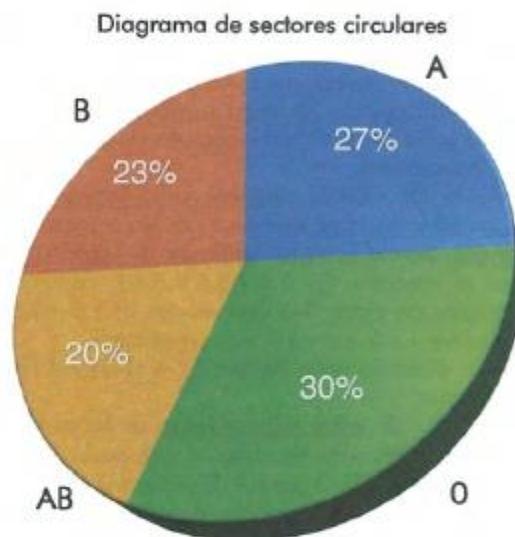
Distribución de grupos sanguíneos en una población. Se trata de una variable cualitativa nominal, y sus respectivas categorías distribuidas en proporciones.



El diagrama de barras es el tipo de gráfica más utilizada para variables cualitativas y cuantitativas discretas

B. SUPERFICIES REPRESENTATIVAS:

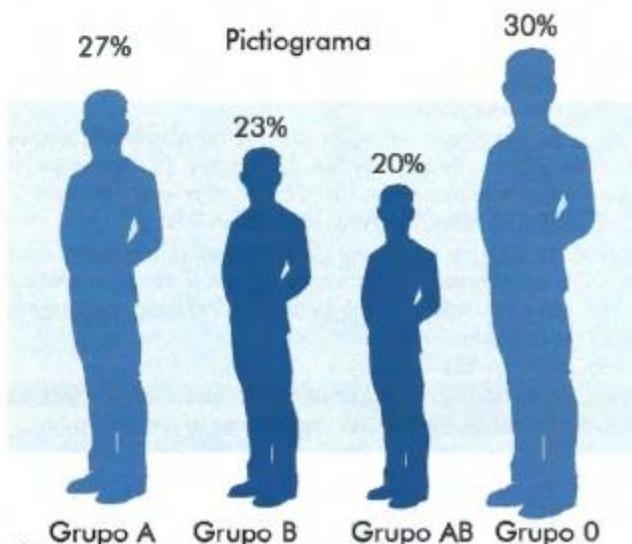
Áreas o volúmenes proporcionales a las frecuencias de cada valor. El tipo **más utilizado** es el diagrama de sectores circulares. Ej: grupos sanguíneos del caso anterior.



Distribución de grupos sanguíneos

Un diagrama de sectores circulares es un ejemplo de superficie representativa.

Otras superficies representativas son los pictogramas (**dibujos o símbolos**) y los cartogramas (**mapas geográficos**). Ej: pictograma de grupos sanguíneos.

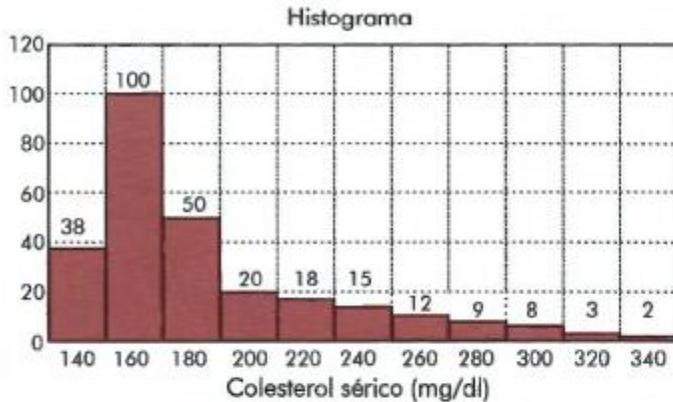


El pictograma representa mediante dibujos proporcionales, la frecuencia de los distintos valores

3.2. Variables cuantitativas continuas

A. HISTOGRAMA:

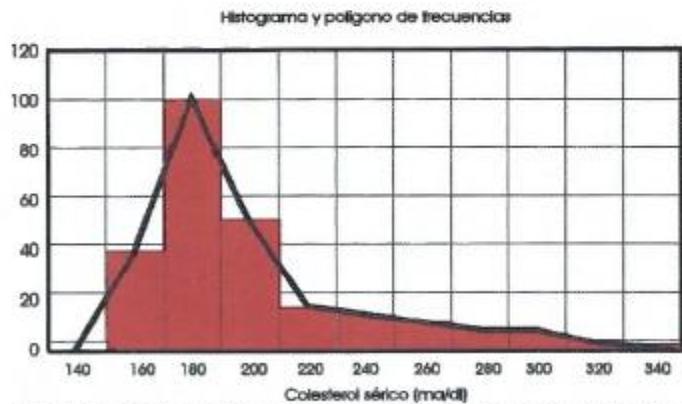
En el histograma de frecuencias utilizamos los valores de la variable continua ordenados y agrupados en intervalos de clase, de forma que en el eje de abscisas se coloquen los intervalos. La regla básica es: sobre los intervalos se levantan unos rectángulos cuya área es proporcional a la frecuencia del intervalo correspondiente, sean frecuencias relativas o absolutas. Dado que el área corresponde a altura (ordenadas) por base (abscisas = intervalo de clase), si todos los intervalos tienen la misma amplitud, la altura de cada rectángulo será proporcional a la frecuencia (lo habitual). En caso contrario calcularíamos la altura de cada intervalo como $\text{altura} = \text{frecuencia} / \text{base}$. Tomando como ejemplo la distribución de colesterolemia vista anteriormente, podemos construir el siguiente histograma:



El histograma es el tipo de gráfico más utilizado para variables cuantitativas continuas

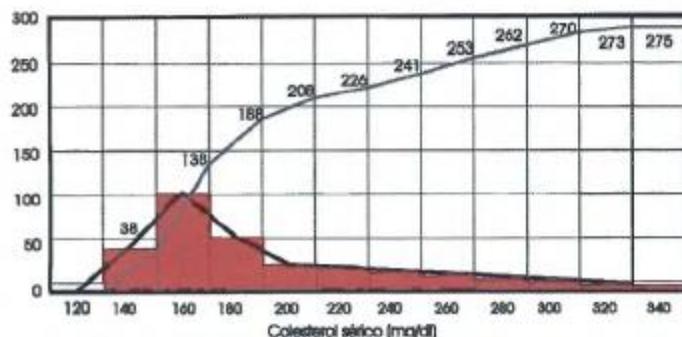
B. POLÍGONOS DE FRECUENCIAS:

El polígono de frecuencias (absolutas o relativas) se construye uniendo con una línea quebrada los puntos medios (marcas de clase) de la parte alta de rectángulos del histograma (de frecuencias absolutas o relativas), tal como vemos en el ejemplo:



El polígono de frecuencias se construye sobre el histograma, uniendo la parte alta de cada rectángulo.

En el polígono de frecuencias (absolutas o relativas) acumuladas se suman sucesivamente para su construcción las alturas de los rectángulos precedentes:



En el polígono de frecuencias acumuladas se van sucesivamente sumando la frecuencia acumulada de cada intervalo con los anteriores, hasta el total de la distribución.

Variable	Representación habitual
Cualitativa	<u>Diagrama de barras</u> : la altura expresa frecuencia de las variables (absoluta o relativa).
Cuantitativa discreta	<u>Diagrama de sectores, pictogramas o cartogramas</u> (superficies representativas): área o volúmenes expresan la frecuencia.
Cuantitativa continua	<u>Histograma</u> : el área (base x altura) expresa la frecuencia (absoluta o relativa). Si los intervalos de clase tienen la misma amplitud, la altura representa la frecuencia (igual base en cada rectángulo). <u>Polígono de frecuencias</u> (absolutas o relativas): une la parte alta de todos los rectángulos del histograma.

3.2. Variables cuantitativas continuas

El apuntamiento o curtosis de una curva indica lo "picuda" o "aplanada" que es. Una curva puede ser mesocúrtica, leptocúrtica (más apuntada de lo normal) o platicúrtica (menos apuntada y por tanto más aplanada de lo normal).

MIR 12 (9941):¿Cuál de los siguientes parámetros mide el apuntamiento de una distribución?:

1. El coeficiente de Fisher.
2. Los cuartiles.
3. La varianza.
4. La curtosis.*
5. La amplitud.

4. Representación gráfica avanzada

En los artículos de investigación biomédica es habitual que se ofrezcan ilustraciones que ayuden al lector a "ver" las respuestas a las preguntas investigadas (a diferencia de los gráficos del apartado anterior). Es decir, las inferencias realizadas sobre estimaciones de parámetros o sobre contrastes de hipótesis en el análisis de comparaciones entre grupos o de asociaciones entre variables.

Para ello se utilizan gráficos que representan, más que los datos medidos y sus frecuencias, las distribuciones resultantes con las relaciones establecidas entre distintas distribuciones y/o variables.

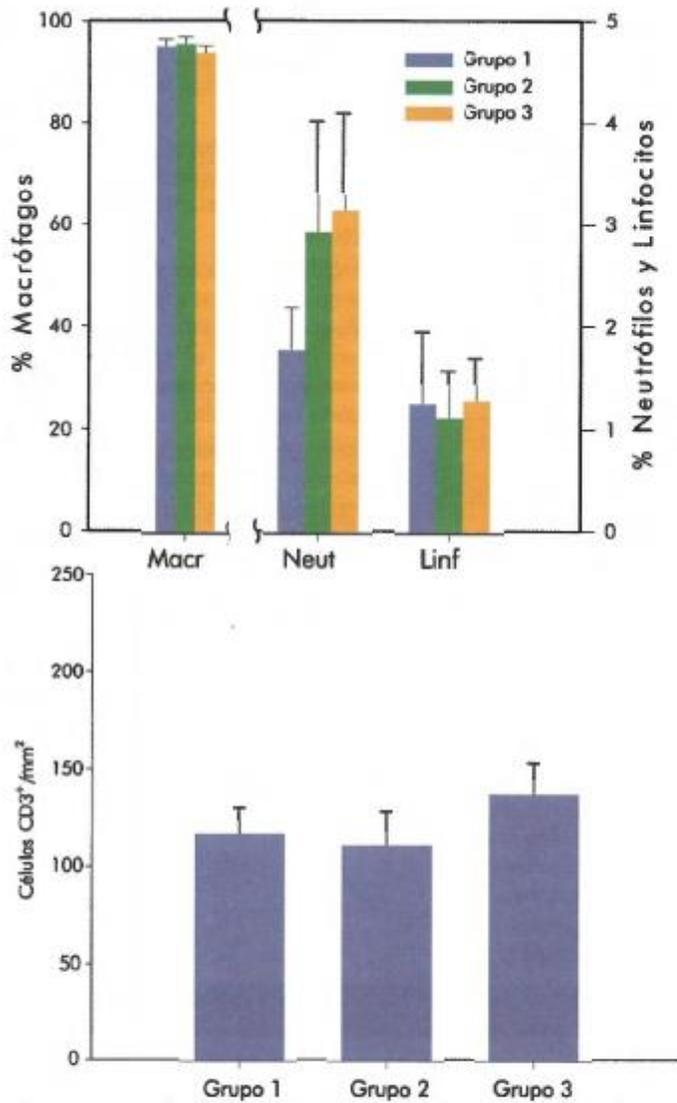
Las dos características fundamentales de una distribución son su tendencia central (expresada comúnmente por la media aritmética o la mediana) y su variabilidad (expresada mediante la desviación típica o la varianza, o utilizando percentiles en el caso de distribuciones irregulares).

En la estimación de un parámetro poblacional la medida de variabilidad considerada es su error estándar (por ejemplo, el error estándar de la media para la media aritmética). El error estándar se utiliza para definir los límites de intervalos de confianza (se desarrolla más adelante).

4.1. Comparación de grupos

Podemos representar el comportamiento de una variable en distintos grupos y utilizar esta representación con finés comparativos. Para ello debemos caracterizar adecuadamente la variable con una medida de su tendencia central, la media aritmética, y una medida de su variabilidad. Pero una variabilidad dentro del marco estadístico de la aproximación al verdadero valor del parámetro que estamos estimando. Esta medida de variabilidad suele ser el error estándar de la media, aunque pueden también utilizarse los límites de un intervalo de confianza.

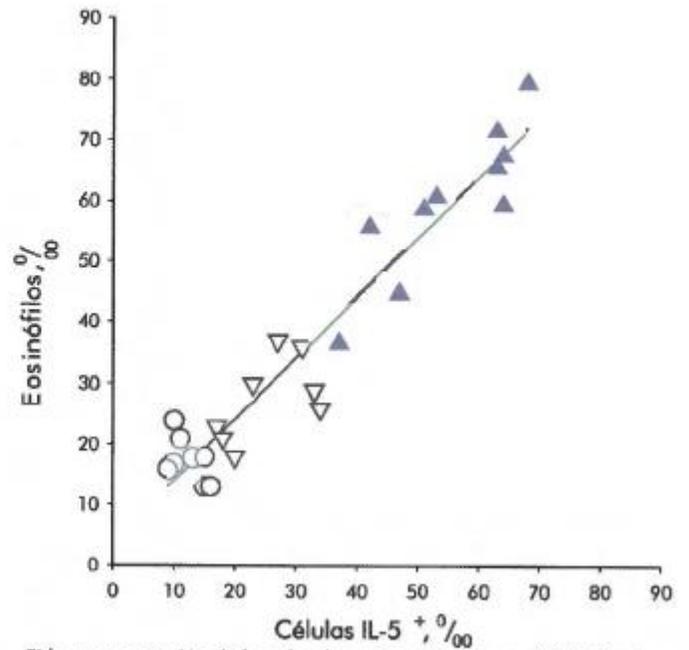
Así representaremos: a) una barra para la media aritmética y b) un segmento para el error estándar de la media. Ej: media y error estándar de la media en tres grupos de la variable células T (CD3⁺)/mm² (estos resultados se estudiarán después mediante análisis de la varianza de una vía)



En la presentación de resultados de estudios de investigación es frecuente expresar la media como un rectángulo, sobre el que se representa una medida de variabilidad (error estándar de la media o los intervalos de confianza para dicha media).

4.2. Asociación y regresión

Mediante unos ejes de coordenadas (x e y) se pueden representar en cada sujeto los valores que van adquiriendo dos variables y configurando una nube de puntos, valorando así si están asociadas. Cuando existe asociación, la recta o curva de regresión representará la relación entre una variable y otra, y se podrá predecir los valores de una de ellas (y o variable dependiente) en función de los de la otra (x o variable independiente), gracias a una fórmula matemática (relación lineal, exponencial...)



Si la representación de la nube de puntos entre dos variables tiende hacia una línea, se dice que ambas variables están asociadas según una relación lineal.



1. MEDIDAS DE FRECUENCIA

- El conjunto de observaciones de una variable (distribución) se puede dividir en intervalos de la variable (intervalo de clase) con un valor representativo para cada uno (marca de clase). Cada intervalo tendría un número de variables que correspondería a su frecuencia absoluta, y podríamos calcular la frecuencia relativa sin más que dividir la absoluta entre el total de observaciones o tamaño muestral N.
- Una frecuencia relativa expresa la probabilidad de encontrar en esa distribución un valor incluido en el intervalo (o suma de intervalos) considerado. Las frecuencias acumuladas (absolutas o relativas) expresan una frecuencia suma del intervalo correspondiente y los anteriores a él.
- Son medidas de frecuencia importantes: proporción (MIR) (ej. prevalencia MIR), razón (MIR) y tasa (MIR).

1.1. PROPORCIÓN

- Cociente en el que el numerador es una magnitud incluida en el denominador ($a/a+b$). La prevalencia es un ejemplo de proporción en el cual se expresa la probabilidad de que un sujeto esté enfermo (MIR) dentro de una población ($a+b$).

1.2. RAZÓN

- Cociente en el que el numerador NO está incluido en el denominador. En la razón de sexo se expresan cuántos hombres hay por cada mujer (ej. $3/1 \rightarrow 3$ a $1 \rightarrow 3$ hombres por cada mujer). Una odds es un ejemplo de razón en que se dividen dos probabilidades: la probabilidad de que algo suceda entre la probabilidad de que no suceda; por ello también se la denomina razón de probabilidades.

1.3. TASA

- Expresa la relación temporal que tiene el cambio de una magnitud. Su denominador habitual es personas-año (MIR). Un ejemplo es la densidad de incidencia: promedia el número de casos nuevos de enfermedad por unidad de tiempo (año). No expresa probabilidad individual de tener ni desarrollar una enfermedad (MIR).

2. REPRESENTACIÓN GRÁFICA ELEMENTAL

- El tipo de representación más habitual para cada distribución considerada se elige en función del tipo de variable de la misma, de forma que:

2.1. EN VARIABLES CUALITATIVAS O CUANTITATIVAS DISCRETAS:

- El gráfico más utilizado es el diagrama de barras. En él la altura de cada rectángulo será proporcional a su frecuencia, pudiendo ser frecuencias relativas o absolutas. Otras representaciones habituales son el diagrama de sectores circulares, pictograma y cartograma, en los cuales es el área el proporcional a la frecuencia (superficies representativas).

2.2. EN VARIABLES CUANTITATIVAS CONTINUAS:

- El gráfico más utilizado es el histograma, pudiendo ser los histogramas de frecuencias absolutas o relativas. Su área es proporcional a la frecuencia de cada intervalo, por lo que la frecuencia es resultado de multiplicar la altura del rectángulo por la base (el intervalo de clase). Si todos los intervalos de clase son iguales (iguales bases), las alturas serán también proporcionales a las frecuencias.
- Otras representaciones utilizadas son el polígono de frecuencias (resultante de la unión de la parte alta de todos los rectángulos en una línea quebrada) y el polígono de frecuencias acumuladas (resultante de ir sucesivamente sumando las alturas precedentes sobre cada nuevo rectángulo). Ambos se pueden construir sobre frecuencias absolutas o relativas (según sea el histograma).

3. REPRESENTACIÓN GRÁFICA AVANZADA

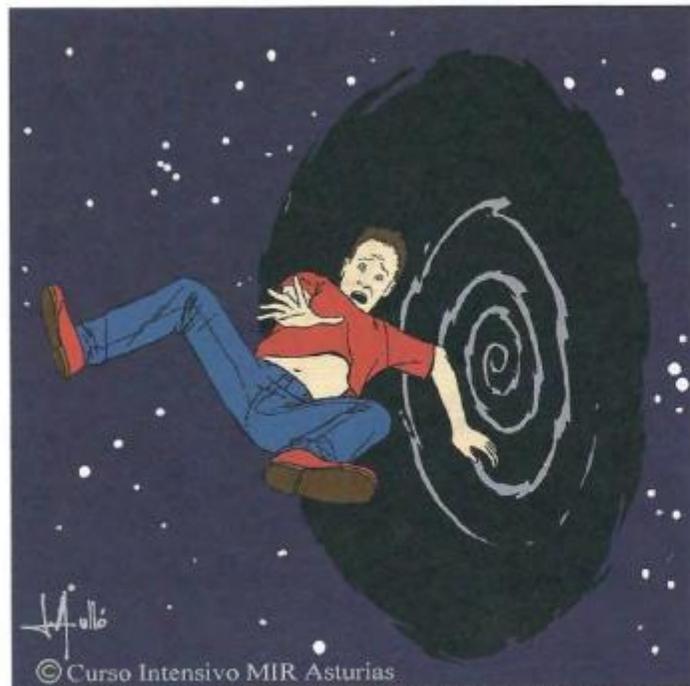
- En los artículos de investigación se suele representar las respuestas a las preguntas habituales, más que las diferentes observaciones. Para ello suele ser habitual la representación de medidas de tendencia central (media o mediana) y de dispersión (desviación estándar, o error estándar de la media) en comparación de grupos y la representación de una nube de puntos en asociación entre variables (regresión).

En estadística descriptiva utilizamos diferentes parámetros para caracterizar el comportamiento de la variable a estudio en cuanto a su tendencia central y variabilidad. Y para ello disponemos de:

- Medidas de tendencia central. Nos dan información acerca de los valores centrales hacia los que tiende la distribución; a saber, media (MIR), mediana (MIR) y moda (MIR).
- Medidas de posición. Nos localizan un dato determinado dentro de la serie, informándonos así acerca de la propia distribución. Son la mediana y los percentiles.
- Medidas de dispersión. Nos informan acerca de cómo de agrupados o dispersos se encuentran los valores de la muestra, es decir, su variabilidad. Las más importantes: rango (MIR) (recorrido o amplitud), desviación media, desviación típica o estándar, varianza y coeficiente de variación de Pearson.

Las medidas de tendencia central y de dispersión son frecuentemente utilizadas conjuntamente para establecer el comportamiento global de una serie de datos: sabremos hacia qué valor tienden a agruparse y cuánto se separan de él (es suficiente información inicial para hacernos una idea sobre la distribución).

1. Medidas de tendencia central



© Curso Intensivo MIR Asturias

Las medidas de tendencia central nos informan sobre los valores hacia los que tiende la distribución.

1.1. Media aritmética

Es la medida de tendencia central más simple (MIR) y utilizada.

A. DEFINICIÓN:

Promedio matemático de todos los valores que adquiere la variable. No deja necesariamente la mitad de los valores de la distribución a cada lado (MIR).

$$\bar{X} = \frac{\sum_{i=1}^N X_i \cdot N_i}{N}$$

B. CARACTERÍSTICAS:

Su cálculo únicamente es posible en variables cuantitativas (discretas o continuas), es decir, no válidas para variables cualitativas nominales (grupos sanguíneos: A, B, AB, 0) ni ordinales (dolor leve, moderado y severo).

No se puede calcular con clases abiertas, clases que no tengan límite superior o inferior. Ej: un intervalo ">80 años", haría imposible su cálculo.

No es recomendable su uso en distribuciones muy asimétricas (valores extremos muy altos o muy bajos), pues la media aritmética es "arrastrada" por la cola de valores extremos (2MIR), perdiendo calidad como medida de tendencia central. Ej: 1,2,3,4 y 140. La media es 30 (140+1+2+3+4/5); y no es muy representativo de la distribución.

C. OTROS TIPOS DE MEDIA:

a. Media ponderada: Se emplea cuando se desea dar distinto peso a diferentes valores. Para ello, cada observación se multiplica por un factor de ponderación.

b. Media geométrica: Raíz enésima del producto de los N valores de la variable. En distribuciones de datos con crecimiento exponencial es mejor estimador que la media aritmética.

c. Media armónica: Inverso de la media aritmética de los inversos de los valores de la variable.

d. Media cuadrática: Raíz cuadrada de la media aritmética de los cuadrados de los valores de la variable.

1.2. Mediana

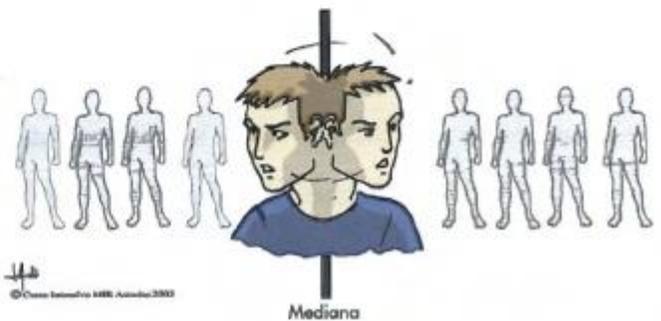
A. DEFINICIÓN:

Es el valor de la variable que deja el mismo número de observaciones por debajo y por encima de su valor (4MIR), es decir, deja el 50% de la muestra a cada lado. Ej anterior (1,2,3,4 y 140): su mediana es 3 (deja 1 y 2 a un lado, y 4 y 140 al otro). No se debe confundir con el punto medio del rango o recorrido de la variable.



repeMIR

Deja tantas observaciones por debajo como por encima, 50% de la muestra a cada lado. (4+)



La mediana es la medida de tendencia central que deja el mismo número de observaciones a ambos lados, coincidiendo con el centro de la distribución.

La mediana coincide con el percentil 50 (P50), siendo a la vez una medida de tendencia central y de posición.

B. CARACTERÍSTICAS:

Se puede calcular en variables cuantitativas (continuas o discretas) y en variables ordinales. No aplicable a variables cualitativas nominales.

Puede calcularse en una serie que tenga clases abiertas, a diferencia de la media.

No se influye por los valores extremos de una distribuciones muy asimétricas, siendo en estos casos mejor medida de tendencia central que la media (3MIR). Véase ejemplo expuesto (1,2,3,4 y 140): media 30, mediana 3 (¿cuál representa mejor la tendencia central de la mayor parte de las observaciones?). En las distribuciones simétricas unimodales (distribución normal), la media, la mediana y la moda coinciden (MIR).

MIR 04 (8015): Un estudio informa que la mediana de supervivencia de los pacientes sometidos a cierta intervención quirúrgica es de 7 años. Ello quiere decir que:

1. El valor esperado de tiempo de supervivencia es 7 años.
2. La mitad de los pacientes sobreviven más de 7 años.*
3. No hay ningún paciente que sobreviva menos de 7 años.
4. La mitad de los pacientes sobreviven aproximadamente 7 años.
5. 7 años es el tiempo de supervivencia más probable.

MIR 09 (9237): ¿Cuál de los siguientes índices **NO** es una medida de dispersión?

1. Desviación estándar.
2. Varianza.
3. Rango de amplitud.
4. Desviación media.
5. Mediana.*

MIR 10 (9467): Un estudio informa que la mediana de supervivencia de los pacientes después del diagnóstico de cierto tipo de cáncer es de 6 años. ¿Esto quiere decir que?

1. No hay ningún paciente que sobreviva menos de 6 años.
2. La mitad de los pacientes sobreviven aproximadamente 6 años.
3. El valor esperado del tiempo de supervivencia es 6 años.
4. No hay ningún paciente que sobreviva más de 6 años.
5. La mitad de los pacientes sobreviven más de 6 años.*

Físicamente divide en dos partes iguales el área de un histograma (50% a cada lado).

Media	Mediana
Cuantitativas (continuas y discretas)	Cuantitativas (continuas y discretas) y cualitativas ordinales
No válida si valores muy extremos (altos o bajos)	Válida aunque valores muy extremos

1.3. Moda

A. DEFINICIÓN:

Es el valor de presentación más frecuente de la distribución (MIR), el que más se repite (el que tiene mayor f_i). No es rigurosamente un valor de tendencia central, sino de tendencia central de posición.

B. CARACTERÍSTICAS:

Se puede determinar para cualquier tipo de variable (cualitativas o cuantitativas)

Puede ser un valor único (distribuciones unimodales) o múltiple (distribuciones multimodales: varios picos).

No se debe utilizar en series con pocos datos (fluctúa excesivamente al azar) y pierde representatividad si es una distribución muy dispersa (no hay grandes picos).

Aporta escasa información en variables cuantitativas, y prácticamente no se utiliza en inferencia estadística. Mayor utilidad en variable cualitativas.

Medidas de tendencia central: valor central hacia el que tiende la distribución.

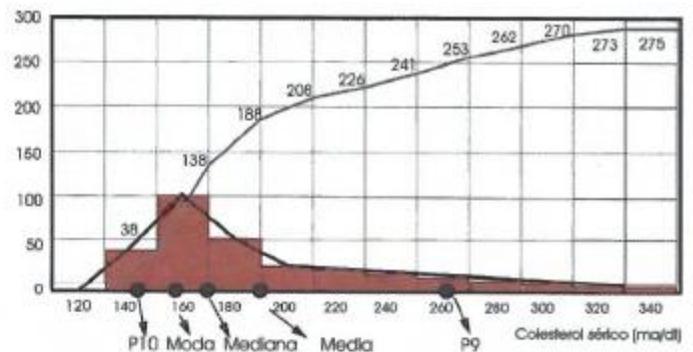
- **Media aritmética:** medida más simple y utilizada que promedia todos los valores de una distribución cuantitativa (continua o discreta). No recomendable en distribuciones muy asimétricas.
- **Mediana:** deja la mitad de la distribución a cada lado (coincide con el percentil 50). De elección en distribuciones asimétricas.
- **Moda:** valor más frecuente. Poco utilizada.

1.4. Ejemplo gráfico

Tomemos de nuevo la distribución de concentración de colesterol total sérico (mg/dl). Se trata de una variable cuantitativa continua, cuyas observaciones habíamos agrupado en intervalos de clase para configurar una muestra de $N=275$ individuos.

absolutas y polígono de frecuencias absolutas acumuladas. La distribución resultante es asimétrica, con una cola o sesgo de valores poco frecuentes hacia la derecha.

Observemos la situación en la gráfica del percentil 10, la moda, la mediana, la media aritmética y el percentil 90. La media es arrastrada hacia la derecha:



2. Medidas de posición

Son localizadores de un dato dentro de una distribución, ubicando un valor de la variable de acuerdo con particiones equiproportionales del total de las observaciones.



© Curso Intensivo MIR Asturias

Las medidas de posición localizan un dato mediante particiones equiproporcionales de la distribución.

2.1. Mediana

Es a la vez una medida de tendencia central y de posición (MIR), ya que es el punto de la distribución hasta el cual se ha acumulado la mitad de las observaciones, coincide con el percentil 50 (MIR).

2.2. Percentiles

A. DEFINICIÓN:

Son estadísticos de posición (MIR) que dividen el número de observaciones de la distribución en partes porcentuales acumulativamente, indicando el porcentaje de la distribución que se acumula hasta el valor del percentil (4MIR). Ej: el percentil 10 es aquel valor de la variable hasta el cual se ha acumulado el 10% de las observaciones (inclusive), de forma que queda el 90% por encima de dicho valor.

En la distribución de nuestro ejemplo el valor del percentil 90 (P_{90}) es 260,95. Esto quiere decir que el 90% de los individuos de la muestra tienen un nivel sérico de colesterol igual o menor a 260,95 mg/dl, y el 10% restante tiene más de 260,95 mg/dl.

mg/dl	260,95
90%	10%

Percentil "x" = "y" significa que hay un "x" % de individuos con menor o igual valor que "y" (4MIR).

$P_{80}=30$ cms \rightarrow el 80% tienen menor o igual que 30 cms; o el 20% tienen más que 30.

Los percentiles son estadísticos de posición que dividen el número de observaciones de la distribución en partes porcentuales acumulativamente. (4+)

B. CARACTERÍSTICAS:

Especialmente útiles en la descripción de distribuciones muy sesgadas con valores extremos asimétricos.

Muy adecuados para delimitar los valores raros. Ej: P97 y P3 de una distribución.

El percentil 50 coincide con la mediana (al igual que decil 50 y cuartil 50).

C. TIPOS:

a. **CENTILES:** dividen las observaciones de la distribución en 100 partes iguales (1,2,3,4...100).

b. **DECILES:** lo hacen en 10 partes iguales (10,20,30...100).

c. **CUARTILES:** lo hacen en 4 partes iguales (25,50,75,100).

Medidas de posición: localizan un dato determinado dentro de la distribución.

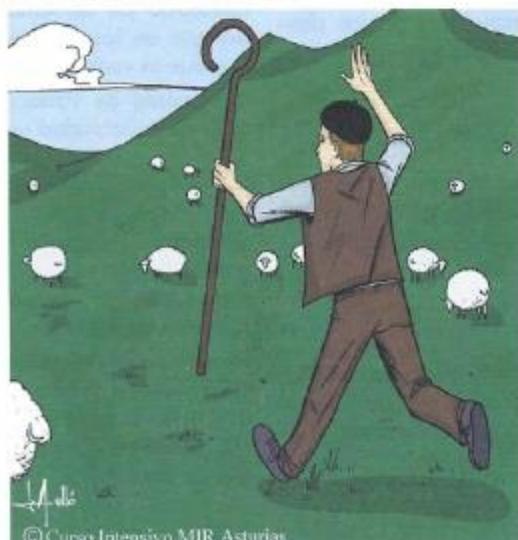
- Mediana: coincide con P50.
- Percentiles: dividen la distribución en partes porcentuales. Localizan valores raros, especialmente en distribuciones muy sesgadas.

3. Medidas de dispersión

Nos describen cómo de agrupados o dispersos se encuentran los datos de la muestra en torno a los valores centrales anteriores, siendo una expresión de la fluctuación del fenómeno estudiado.

MIR 13 (10189) (190): Con objeto de comprobar la eficacia de un tratamiento para dejar de fumar se compara un grupo control (con placebo) con un grupo tratado. Para que los grupos sean comparables es importante que no difieran mucho en la edad de los participantes. Nos informan que la media de edad en el grupo control es 52 años y que en el grupo tratado también es 52 años. A partir de dicha información podemos decir que:

1. Los grupos no difieren respecto a la distribución de la variable edad.
2. Si el estudio está bien diseñado, no nos puede dar la misma media de edad en ambos grupos.
3. Para comparar la distribución de la variable edad en ambos grupos sería conveniente conocer una medida de dispersión como la desviación típica además de la media. *
4. Para la conclusión final no importa la distribución de la edad en ambos grupos, sólo si el tratamiento es efectivo o no.
5. Si además de la media, coinciden la mediana y la moda podemos afirmar que los grupos no difieren respecto a la distribución de la variable edad.



Las medidas de dispersión estudian cómo de agrupados o separados están los datos entorno a los valores centrales de la distribución

3.1. Rango

A. DEFINICIÓN:

Diferencia entre los valores máximo y mínimo de la variable. Ej: en 2,5,6,7,8,120 sería 118 (120-2).

B. CARACTERÍSTICAS:

Sólo aplicable a variables cuantitativas.

Es la medida de dispersión que menos información proporciona (carece de utilidad en inferencia estadística).

3.2. Desviación cuartílica

Semidiferencia entre los cuartiles 3 y 1 (75% y 25%), por lo que refleja el promedio simple de datos incluidos en los dos cuartiles centrales.

$$DQ = \frac{1}{2} (Q_3 - Q_1)$$

Algunos autores definen otra medida, "recorrido intercuartílico", y se refieren a la diferencia entre tercer y primer cuartiles, abarcando el 50% central de los datos (sin dividir entre dos).

3.3. Desviación media

A. DEFINICIÓN:

Promedia las desviaciones de los datos de la distribución respecto a la media, hace la media de los valores absolutos de las desviaciones respecto a la media. Ej: 2,4,6,8. La media es 5. La desviación media es la media de las desviaciones (en valor absoluto: sin negativos). 2-5; 4-5; 6-5; 8-5 → (en valores absolutos): (3+1+1+3)/4 → Sería de 2.

La suma de las desviaciones a la media se anula (-1, -3, +1 y +3 al sumarlos se anularían), por eso utilizamos los valores absolutos.

$$DM = \frac{\sum_{i=1}^N |x_i - \mu|}{N}$$

B. CARACTERÍSTICAS:

Sólo aplicable a variables cuantitativas.

Escasa operatividad por el empleo de valores absolutos en estadística.

3.4. Varianza y desviación típica o estándar

A. DEFINICIÓN:

La varianza promedia los cuadrados de las desviaciones (hace la media de los cuadrados de las desviaciones a la media).

En el ejemplo anterior: $[3^2 + 1^2 + (-1)^2 + (-3)^2]/4$.

$$S^2 = \frac{\sum_{i=1}^N (x_i - \bar{X})^2 \cdot n_i}{n}$$

La desviación típica o estándar es la raíz cuadrada de la varianza.

$$S = \sqrt{S^2}$$

B. CARACTERÍSTICAS:

Sólo calculables en variables cuantitativas.

Son las medidas que mejor expresan la variabilidad (MIR) del fenómeno estudiado, y las más utilizadas. Cuanto mayores sean la varianza y la desviación típica de la distribución, mayor será su variabilidad.

La varianza presenta el inconveniente de estar expresada en unidades que son el cuadrado de las unidades de las observaciones originales, pero esto se obvia con la desviación típica o estándar (expresada en las mismas unidades que la media).

En casos en que no se utilice la media por no ser un valor representativo (distribuciones muy sesgadas), se recomienda no utilizar tampoco la varianza ni desviación típica.

3.5. Coeficiente de variación de Pearson

A. DEFINICIÓN:

Es *adimensional* (MIR), es decir, no tiene unidades de medida. Se calcula dividiendo la desviación típica entre la media de una distribución y multiplicando por 100 (MIR) (desviación típica "relativizada" a la media y expresada en porcentaje).

$$CV = \frac{S}{X} \cdot 100$$

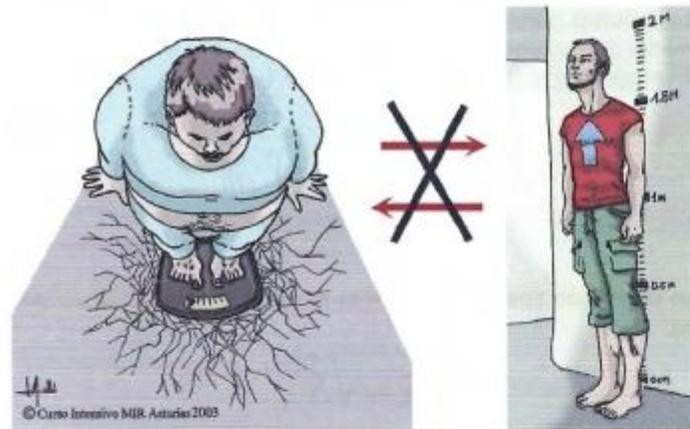
REGLA NEMOTÉCNICA



Concha Velasco es Superior a la Media

El Coeficiente de Variación de Pearson es un valor **adimensional**, es decir, sin unidades.

Son **adimensionales** la razón y el coeficiente de variación de Pearson.



El coeficiente de variación de Pearson se utiliza en la comparación de la variabilidad de distribuciones que utilicen distintas unidades de medida

B. CARACTERÍSTICAS:

Frecuentemente utilizado en la comparación de la dispersión o variabilidad de **métodos de medida diferentes** (3MIR) en situaciones en que las desviaciones típicas no son comparables directamente por estar referidas a distintas medias.

Medidas de dispersión: grado de variabilidad con que se separan o agrupan los valores de la distribución.

- **Rango (recorrido, amplitud):** diferencia entre los valores máximo y mínimo.
- **Desviación media:** promedia los valores absolutos de las desviaciones a la media.
- **Varianza S²:** promedia los cuadrados de las desviaciones a la media. Su unidad está elevada al cuadrado.
- **Desviación típica o estándar S:** raíz cuadrada de la varianza. Se expresa en las mismas unidades que la media. Medida de dispersión más utilizada.
- **Coeficiente de variación de Pearson:** cociente adimensional entre desviación típica y la media de la distribución (x100). Es la medida de dispersión de elección a la hora de comparar métodos de medida diferentes, que usen unidades de medida distintas.

No hay que confundir ... con ...

Variable: Función que puede adquirir distintos valores. Pueden ser cuantitativas (discretas, continuas) o cualitativas (nominales, ordinales).

Distribución: Conjunto de observaciones sobre los distintos valores que toma una variable.

Proporción: $\frac{a}{a+b}$

Razón: $\frac{a}{b}$

Casos nuevos

Tasa: $\frac{\text{Personas} \cdot \text{Tiempo}}$

Diagrama de barras: Gráfica para variables *cualitativas*. La altura de las barras es proporcional a las frecuencias.

Histograma: Gráfica para variables *cuantitativas continuas*. La superficie de los rectángulos es proporcional a las frecuencias.

Media aritmética: Promedio de todos los valores de la distribución. Es *arrastrada* por los valores extremos. Sólo variables cuantitativas.

Mediana: Valor que deja tantas observaciones por debajo como por encima. Mejor medida de centralización que la media en distribuciones sesgadas (asimétricas). Variables cuantitativas y cualitativas ordinales.

Moda: Valor más frecuente. Puede haber varias en una distribución.

Cualquier tipo de variable.

Percentil: valores que dividen la muestra en 100 partes iguales abarcando hasta el valor del percentil *inclusive*.

Deciles: percentiles 10, 20...
Cuartiles: percentil 25,50,75
La mediana es el percentil 50.

Desviación media: Medida de dispersión. Promedio de los valores absolutos de las desviaciones. Poco operativa.

Desviación típica o estándar: Medida de dispersión. Raíz cuadrada de la varianza. Se expresa en las mismas unidades que la variable.

Coeficiente de variación: Medida de variabilidad relativa.

Varianza: Medida de dispersión. Promedio de los cuadrados de las desviaciones.

$\frac{\text{Desviación típica}}{\text{Media}} \cdot 100$

Compara la dispersión de dos o más distribuciones. Comparación de la precisión de métodos de medida.

Análisis de la varianza: Prueba de significación estadística que compara las medias de 3 ó más distribuciones.

Análisis multivariante: Técnicas de inferencia estadística para el estudio simultáneo de diversas variables. Control de sesgos de confusión.



El coeficiente de variación de Pearson se emplea para comparar la variabilidad relativa de diferentes distribuciones (ej. métodos de medida). Es adimensional. (3+)



RESUMEN DE TENDENCIA CENTRAL Y VARIABILIDAD

En estadística descriptiva se utilizan distintas medidas para caracterizar el comportamiento de las variables a estudio. Existen tres grandes grupos: a) medidas de tendencia central, b) medidas de posición y c) medidas de dispersión (variabilidad).

1. MEDIDAS DE TENDENCIA CENTRAL

- Nos informan sobre los valores centrales hacia los que tiende la distribución de la variable. Las más importantes son: media (MIR), mediana (MIR) y moda (MIR).

1.1. MEDIA ARITMÉTICA

- Medida de tendencia central más utilizada y de fácil manejo matemático (MIR). Promedia los valores de la distribución. Sólo podemos calcularla con variables cuantitativas (continuas o discretas) y con intervalos de clase cerrados. No se recomienda en distribuciones muy asimétricas, por ser arrastrada por la cola de valores extremos (2MIR). Otras medias: ponderada, geométrica (para datos de crecimiento exponencial), armónica y cuadrática.

1.2. MEDIANA

- Deja la mitad de la distribución a cada lado (3MIR) (coincide con el percentil 50). Su cálculo es posible en variables cuantitativas y cualitativas ordinales, y no necesita que los intervalos de clase estén cerrados. De elección en distribuciones asimétricas (3MIR), ya que no se influye por los valores extremos. En distribuciones simétricas unimodales (distribución Normal) la media, mediana y moda coinciden (MIR) en el punto central.

1.3. MODA

- Valor más frecuente, pudiendo ser la distribución uni o multimodal. Poco utilizada, pero válida para cualquier tipo de variable.

2. MEDIDAS DE POSICIÓN

- Son localizadores de un dato dentro de la distribución, siendo ésta dividida en partes equipropocionales. Los más importantes son: mediana (coincide con el percentil 50) y percentiles.

2.1. PERCENTILES

- Dividen la distribución en partes porcentuales acumulativamente. Y así expresan el porcentaje (proporción) de la distribución con valores iguales o inferiores a uno dado (4MIR). Ejemplo: $P_{80} = 30$ cms, significa que el 80% de los valores de la distribución son iguales o inferiores a 30 cms, siendo el 20% superiores. Localizan valores raros, especialmente en distribuciones muy sesgadas. Tipos: percentiles o centiles (100 partes iguales), deciles (10 partes) y cuartiles (4 partes).

3. MEDIDAS DE DISPERSIÓN

- Nos informan acerca de cómo de dispersos o agrupados están los valores de la distribución en torno a su centro. Los más importantes son: rango, desviación media, desviación estándar, varianza y coeficiente de variación (Pearson).

3.1. RANGO (RECORRIDO, AMPLITUD)

- Diferencia entre los valores máximo y mínimo.

3.2. DESVIACIÓN MEDIA

- Promedia los valores absolutos de las desviaciones a la media. Poco operativa por utilizar valores absolutos.

3.3. VARIANZA (S²)

- Promedia los cuadrados de las desviaciones a la media. Sus unidades vienen expresadas al cuadrado (si la media está en "cms", la varianza vendrá expresada en "cms²").

3.4. DESVIACIÓN TÍPICA O ESTÁNDAR (S)

- Raíz cuadrada de la varianza, por lo que sus unidades son iguales a la media. Es la medida de dispersión más utilizada.

3.5. COEFICIENTE DE VARIACIÓN DE PEARSON

- Cociente entre desviación típica y la media de la distribución expresada en forma de porcentaje (x100) (MIR). Es un valor adimensional: se anulan las unidades de medida del numerador (desviación típica) y denominador (media), por ser las mismas.
- Es la medida de dispersión de elección a la hora de comparar métodos de medida diferentes (3MIR), que usen unidades distintas.

1. Conceptos básicos

1.1. Definición de probabilidad

La probabilidad es una medida de la verosimilitud de que un determinado suceso ocurra o no, con respecto a otros. Ej: si afirmamos que la probabilidad de que un fármaco cure a un enfermo es $P(\text{curación}) = 0,7$ (ó 70%), es como decir que al prescribir el fármaco a 100 enfermos esperaríamos que curase a 70 y fracasase en 30, como lo más probable.

Matemáticamente se define la probabilidad de un suceso $P(A)$ como el límite hacia el que tiende su frecuencia relativa (simple o acumulada) cuando el número de observaciones tiende a infinito.

$$P(A) = \lim_{n \rightarrow \infty} f_n(A)$$



La probabilidad del suceso aleatorio viene determinado por el límite al que tiende su frecuencia relativa cuando el número de observaciones tiende a infinito

Recordemos que tanto la frecuencia relativa como la proporción son medidas de probabilidad.

1.2. Experimento aleatorio

Proceso cuyo resultado no se puede predecir antes de realizarlo, depende del azar. Ejemplo: elegir al azar 10 individuos de la población y comprobar cuántos de ellos tienen grupo sanguíneo AB.

1.3. Espacio muestral

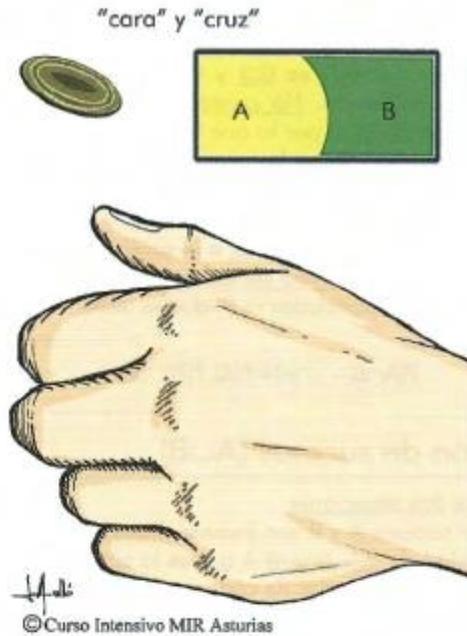
Conjunto de todos los resultados posibles de un experimento aleatorio. Se simboliza por Ω o E. Podemos considerar que es el fenómeno a estudio, pudiendo ser: a) finito (grupos sanguíneos A, B, AB, O) o b) infinito (números naturales 1, 2, 3...n). La suma de las probabilidades de todos los sucesos elementales de un espacio muestral, o probabilidad del espacio muestral, es igual a la unidad; es lo que se denomina suceso seguro. Es decir, $\sum P(e_i) = P(E) = 1$

1.4. Suceso

Un suceso es cualquier subconjunto de un espacio muestral. Puede ser elemental (un único elemento) o compuesto (ejemplo: elegir al azar diez individuos, y que dos de ellos tengan grupo sanguíneo AB).

A. SUCESO COMPLEMENTARIO:

Para un suceso A de un determinado espacio Ω , su suceso contrario o complementario (A^c) es el que ocurre siempre que no ocurre A, de forma que $P(A^c) = 1 - P(A)$. Ej: si la probabilidad de tener un grupo sanguíneo AB fuera 0,2 o 20%, la probabilidad del suceso complementario (no-AB: puede ser A, B ó O) sería de $1 - P(AB)$, es decir, 0,8 u 80%.

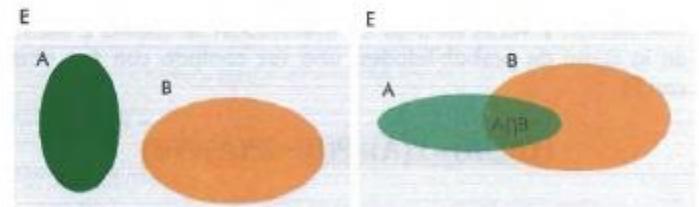


Sucesos complementarios: el suceso de A será B, si, siempre que no ocurra A, sucede B. Son incompatibles y la suma de ambos es el espacio muestral

Se cumple que dos sucesos complementarios son excluyentes mutuamente y su suma resulta 1 (probabilidad del espacio muestral, suceso seguro).

B. SUCESOS INCOMPATIBLES:

Dos sucesos A y B del espacio Ω son incompatibles si son mutuamente excluyentes, es decir, su realización simultánea es imposible: si A, no B, y viceversa. Ej: los sucesos "ojos azules" y "ojos castaños" son incompatibles, mutuamente excluyentes, y no complementarios (existen ojos de otros colores).



Dos sucesos son incompatibles si no pueden darse nunca a la vez

C. SUCESOS DEPENDIENTES E INDEPENDIENTES:

Dos sucesos A y B son independientes cuando la probabilidad de que ocurra uno de ellos no depende de la probabilidad de que ocurra el otro (no está condicionada al otro). Ej: un antiviral tiene una probabilidad del 85% de curar una viriasis y un marcapasos de curar una arritmia del 80%; y parecen independientes (no es lógico que un antiviral condicione al marcapasos ni viceversa).

Y serán dependientes si la probabilidad de que uno suceda está condicionada al hecho de que ocurra el otro, haciéndolo más o menos probable. Ej: supongamos que la cloxacilina cura el 30% de las estafilococias, pero si el estafilococo no es meticilín-resistente entonces sube al 95%; son sucesos dependientes. Evidentemente la eficacia de la cloxacilina viene condicionada por el hecho de que los estafilococos sean o no meticilín-resistentes.

Dos sucesos son incompatibles si son mutuamente excluyentes

Sucesos independientes: la probabilidad de que uno ocurra no influye en que suceda el otro.

2. Operaciones básicas

2.1. Intersección de sucesos ($A \cap B$)

Siempre que dos sucesos (A y B) sean independientes, la probabilidad de que ocurran los dos a la vez (A y B al mismo tiempo) es el producto de las probabilidades de ambos (MIR):

$$P(A \cap B) = P(A) \cdot P(B)$$

Ej: P("tener ojos verdes") es 0.3 y P("padecer Parkinson") es 0.02 en cierta población. No parecen influir el uno sobre el otro (son independientes) por lo que la probabilidad de que un enfermo de Parkinson tenga los ojos verdes será de 0.3×0.02 (0.006).

La misma ley se puede generalizar a N sucesos independientes, de modo que la probabilidad de que ocurran A y B... y n es el producto de las probabilidades individuales (MIR).

$$P(A \cap B \cap \dots \cap n) = P(A) \cdot P(B) \cdot \dots \cdot P(n)$$

2.2. Unión de sucesos ($A \cup B$)

Pueden darse dos situaciones:

1. Si dos sucesos A y B son incompatibles, entonces la probabilidad de que ocurra A o B es la suma de las probabilidades individuales de cada uno.

$$P(A \cup B) = P(A) + P(B)$$

2. Si A y B son compatibles (pero independientes), entonces la probabilidad de que ocurra A o B (al menos uno de los dos pudiendo ocurrir ambos) es la suma de las probabilidades de ambos menos su producto (al sumar A y B se cuenta dos veces el área de intersección, debe contarse sólo una vez):

$$P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B)$$

Si nos preguntaran la probabilidad de A o B, pero que no pudieran darse los dos a la vez, calculamos dicha probabilidad P como $P(A) + P(B) - 2P(A \cap B)$ (MIR), ya que debemos restar su intersección 2 veces (el área de intersección se cuenta 2 veces en la suma de probabilidades: una vez contada con A y otra con B).

$$P(A \cup B) = P(A) + P(B) - 2P(A) \cdot P(B)$$

2.3. Probabilidad condicionada [$P(A/B)$] (Teorema de Bayes)

Si dos sucesos A y B no son independientes y $P(B) \neq 0$, la probabilidad del suceso A condicionada a B [$P(A/B)$] expresa la probabilidad de que, habiendo ocurrido B, ocurra A:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Ojo porque, al no ser A y B independientes, la probabilidad de que ocurran A y B (numerador) no es el producto de sus probabilidades respectivas; esta probabilidad debe conocerse mediante su investigación directa.

Otra manera de escribirlo, teniendo en cuenta que para sucesos condicionados (no independientes), $P(A \cap B) = P(B/A)P(A)$:

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)}$$

El teorema de Bayes curiosamente calcula la probabilidad de que ocurra A condicionado a B utilizando la probabilidad de B condicionado a A (un poco lioso...). El siguiente ejemplo ilustra su verdadera utilidad práctica:

- Llamemos suceso A a la probabilidad de que llueva en Oviedo este fin de semana, y suceso B a la probabilidad de sufrir un accidente de tráfico en Oviedo.

- Supongamos que conocemos la probabilidad de que llueva en Oviedo este fin de semana (suceso A), y la probabilidad de sufrir un accidente de tráfico (suceso B). También conocemos la probabilidad de sufrir un accidente de tráfico si llueve (suceso B condicionado a A, es decir B/A).

Nos informan de que ha habido un accidente en Oviedo este fin de semana, pero no estábamos en la ciudad y desconocemos qué tiempo hacía. El Teorema de Bayes nos indica la probabilidad de que haya llovido una vez que sabemos que ha ocurrido un accidente, es decir $P(A/B)$.

Sucesos

- Sucesos complementarios: si A, no B, y viceversa (son por definición incompatibles, y su suma es el espacio muestral).
- Sucesos incompatibles: mutuamente excluyentes.
- Sucesos independientes: la probabilidad de uno no influye en la del otro.

Operaciones básicas

- Intersección: siendo independientes, la probabilidad de A y B (simultáneamente) es el producto de las probabilidades de ambos. Para más de dos, se procede de forma análoga (producto de sus probabilidades).
- Unión: probabilidad de A o B (uno, otro o ambos) es la suma de las probabilidades únicamente si son incompatibles; si son compatibles, se resta a dicha suma la probabilidad de la intersección.
- Probabilidad condicionada: la probabilidad de uno influye en la del otro.



1. CONCEPTOS BÁSICOS

1.1. DEFINICIÓN

- La probabilidad de un suceso es la verosimilitud de que éste ocurra. Matemáticamente es el límite al que tiende su frecuencia relativa cuando el número de observaciones tiende a infinito. Si la probabilidad de que un fármaco cure es de 0.7 (70%), esperaríamos que curase a 70 de cada 100, como lo más probable.

1.2. ESPACIO MUESTRAL

- Conjunto de todos los resultados posibles de un experimento aleatorio. Su probabilidad es 1: suceso seguro.

1.3. SUCESO COMPLEMENTARIO

- Si dos sucesos son complementarios, serán mutuamente excluyentes y la suma de sus probabilidades será de 1, ya que siempre que uno no se dé, sucederá el otro. Siempre que A, no B; y siempre que B, no A.

1.4. SUCESOS INCOMPATIBLES

- Dos sucesos son incompatibles si son mutuamente excluyentes, su realización simultánea es imposible.

1.5. SUCESOS INDEPENDIENTES

- Cuando dos sucesos son independientes la probabilidad de que uno suceda no influye (ni depende) en el hecho de que ocurra el otro.

2. OPERACIONES BÁSICAS

2.1. INTERSECCIÓN DE SUCESOS (A Y B)

- La probabilidad de que ocurran a la vez (intersección) dos sucesos A y B se puede calcular como el producto de sus probabilidades, siempre que A y B sean independientes (MIR). Si la probabilidad de A es 0.3 y la de B 0.5, ambos independientes, la probabilidad de A y B (simultáneamente) será de $0.3 \times 0.5 = 0.15$. De forma análoga se procede para varios sucesos: el producto de sus probabilidades será la probabilidad de que todos ellos ocurran simultáneamente, siempre que sean independientes (MIR).

2.2. UNIÓN DE SUCESOS (A O B)

- Dados dos sucesos A y B, la probabilidad de que al menos uno ocurra (A o B) se puede calcular como la suma de las probabilidades de ambos, siempre que sean incompatibles (no se pueden dar a la vez). En caso de que sean compatibles, la probabilidad de que al menos uno ocurra (A, B o los dos) será la suma de sus probabilidades, restando la probabilidad de su intersección: $P(A) + P(B) - P(A \cap B)$. La probabilidad de A o B, pero que no pudieran darse los dos a la vez, es igual a $P(A) + P(B) - 2P(A \cap B)$ (MIR).

2.3. PROBABILIDAD CONDICIONADA

- Si dos sucesos A y B no son independientes (son dependientes) y la probabilidad de B no es cero, la probabilidad de A condicionada a B será igual al cociente entre la probabilidad de su intersección y la probabilidad de B (¡ojo!, porque $P(A \cap B)$ no puede calcularse como el producto de probabilidades en este caso: son dependientes).



En capítulos previos hemos visto que las mediciones realizadas sobre los individuos de una muestra originan una distribución en la que a cada valor de la variable estudiada – colesterolemia– le corresponde una frecuencia con que ha sido observado –número de veces que medimos 163 mg/dl de coles-terolemia, por ejemplo. Esto es lo que se conoce como distribución muestral. Es decir, tenemos una distribución donde relacionamos el valor de la variable con el número de individuos de nuestra muestra que presentan ese valor.

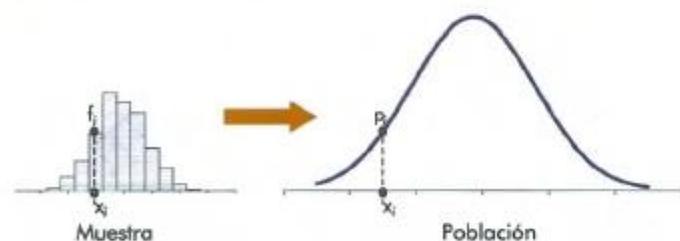
Una vez visto el significado de probabilidad, podemos ahora extender el concepto de distribución como paso previo a los capítulos de inferencia estadística.

1. Conceptos básicos

1.1. Frecuencia relativa y probabilidad

Habíamos dicho ya que la frecuencia relativa es una medida de probabilidad. Y recordemos que, si se hacen bien las cosas, una muestra debe representar a la población de la que se ha extraído. Por ello,

“Si a cada valor de una variable le corresponde una determi-nada frecuencia relativa en la distribución muestral, análoga-mente a cada valor le corresponderá una determinada proba-bilidad de aparición al extraer al azar un sujeto de la pobla-ción.”



Esta probabilidad dependerá de las propiedades de la distribu-ción de la variable en la población (distribución poblacional de la variable). Sin embargo, en función de una serie de datos de la muestra (media y desviación típica) y cumpliendo la variable en la población ciertos requisitos (distribución gaussiana o normal), se puede calcular la probabilidad de aparición de un valor de la variable en la distribución poblacional a partir de los datos obtenidos en la distribución muestral. Y es éste nuestro objetivo fundamental.

1.2. Variable aleatoria y función de probabilidad

A. VARIABLE ALEATORIA:

Todo fenómeno que toma diversos valores numéricos, depen-dientes de los resultados de un fenómeno aleatorio, con distin-tas probabilidades. Puede ser de tipo cuantitativo (discreto o continuo), siendo éste último caso el más frecuente en inferencia estadística.

B. FUNCIÓN DE DENSIDAD DE PROBABILIDAD¹:

Función matemática f(x) que predice la probabilidad de presen-tación de los distintos valores de una variable aleatoria en la distribución poblacional. Asigna a cada valor o intervalo de valores de la variable su probabilidad de aparición.

C. FUNCIÓN DE DISTRIBUCIÓN DE PROBABILIDAD:

Función matemática f(x) que expresa la probabilidad acumula-da (de manera análoga a las frecuencias acumuladas en las distribuciones muestrales). Esta herramienta matemática nos informa acerca de la probabilidad de encontrar valores iguales o inferiores a uno dado.

¹ Siendo puristas (no para el MIR), hablamos de función de probabili-dad en variables aleatorias discretas y función de densidad (de proba-bilidad) en variables aleatorias continuas.

Funciones de probabilidad

Función de densidad	Función de distribución
Frecuencia relativa de cada valor	Frecuencia relativa acumulada de cada valor

2. Modelos de distribuciones de probabilidad

El conocimiento de las funciones de probabilidad para cada una de las variables o fenómenos en particular puede parecer inalcanzable; habría que tener un montón de ecuaciones y tablas para cada variable médica: edad, talla, peso, colesterolemia, uricemia...

Afortunadamente, se han definido varios modelos con funcio-nes y propiedades definidas que nos sirven para el estudio de todas estas variables, evitándonos recurrir a una tabla (o función) para cada una de ellas. Estos modelos son las leyes o distribuciones de probabilidad teóricas. Las más importantes son:

- Distribución normal o gaussiana. Modelo de distribución más utilizada.
- Distribución binomial. Utilizada para variables cuantitativas discretas, definidas por el número de individuos que verifi-can una característica dicotómica (variables dicotómicas: cuántos hombres/ mujeres en una población, cuántos curar/ no curan con un fármaco).
- Distribución de Poisson. Usada en situaciones de gran ta-maño muestral (n grande) y probabilidad de aparición ex-tremadamente pequeña (p pequeña). Ej. número de oposito-res MIR tranquilos y relajados durante todo el tiempo de preparación.



REPASO

- En una distribución muestral cada valor de la variable tiene una frecuencia relativa, que expresa su probabilidad. Si el procedimiento de muestreo ha sido correcto, esta distribución muestral será representativa de la distribución poblacional de la variable.
- En la distribución poblacional la probabilidad de presentación de cada valor de la variable obedecerá a determinadas funcio-nes de densidad y distribución de probabilidad. A este efecto se han definido leyes o distribuciones de probabilidad teóricas, proporcionando modelos ajustables al comporta-miento de los fenómenos para su estudio. Entre las más utili-zadas se encuentran la distribución Normal y la distribución binomial.

2.1. Distribución normal

Es la más frecuentemente utilizada y sus propiedades son el fundamento de los procedimientos de inferencia más importan-tes. No es de extrañar: la mayor parte de las variables alea-torias en Biomedicina siguen una distribución de este tipo (altura, peso, colesterolemia, coeficiente intelectual...).

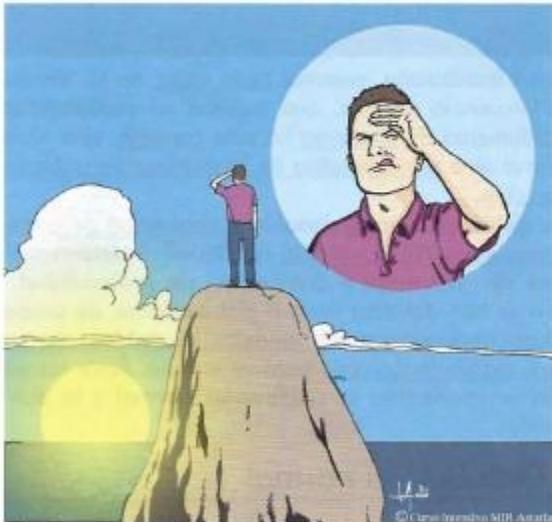
También se denomina campana de Gauss (tiene forma de campana y la descubrió Gauss).



Gauss, poco después de lo de la campana

A. CARACTERÍSTICAS:

1. Está definida por una función de probabilidad continua (2MIR) asintótica al eje de abscisas, por lo que es teóricamente posible cualquier valor de x entre $-\infty$ e ∞ (MIR) (nunca llega a haber un valor máximo o mínimo, siempre puede haber uno mayor o menor respectivamente).



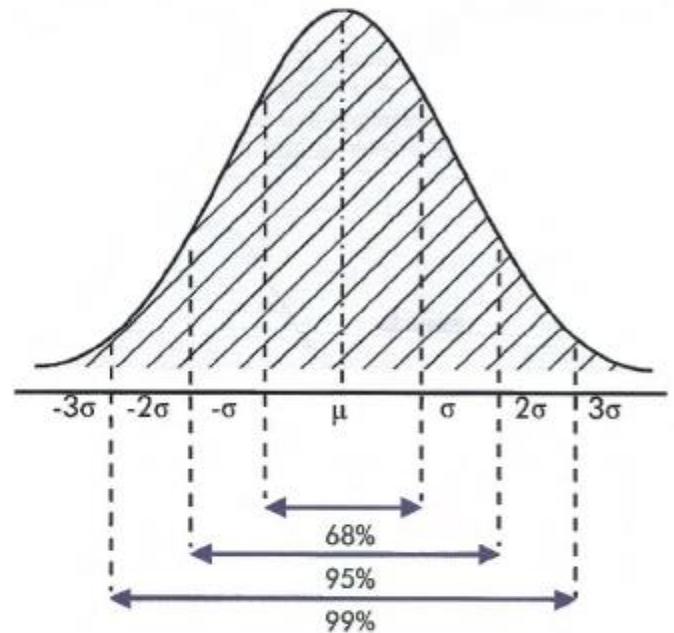
La distribución normal es asintótica con el eje de abscisas, lo que significa que cualquier valor es posible entre $-\infty$ e ∞

2. Es simétrica respecto al punto central (3MIR) de la campana donde coinciden media, mediana y moda (3MIR), lo que quiere decir que: a) es el valor medio (media o μ), b) deja el 50% de la distribución a cada lado (MIR) (mediana) y c) es el valor más frecuente (moda), disminuyendo hacia los lados, por lo que es además unimodal (MIR).
3. La distancia entre este centro y el punto de inflexión coincide con una desviación estándar o σ . De forma que la función queda suficientemente definida por la media aritmética (μ) y la desviación estándar (σ), o varianza (σ^2) (MIR). Cuanto más grande es σ , más se achata la campana (más plana). A esta característica se la conoce como curtosis: a) mesocúrtica (típica distribución normal); b) leptocúrtica (más puntiaguda); c) platocúrtica (más aplanada).
4. Su dispersión está perfectamente definida:
 - a) El intervalo $\text{media} \pm \text{una desviación estándar}$ [$\mu \pm \sigma$] agrupa aproximadamente el 68% de los valores centrales (MIR).

- b) El intervalo $\text{media} \pm \text{dos desviaciones estándar}$ [$\mu \pm 2\sigma$] agrupa aproximadamente el 95% de los valores centrales (5MIR).
- c) El intervalo $\text{media} \pm \text{tres desviaciones estándar}$ [$\mu \pm 3\sigma$] agrupa aproximadamente el 99% de los valores centrales.

repeMIR

- El intervalo [$\bar{X}_m \pm S_{\bar{X}}$] contendrá a μ con una probabilidad del 68%.
- El intervalo [$\bar{X}_m \pm 2S_{\bar{X}}$] contendrá a μ con una probabilidad del 95%.
- El intervalo [$\bar{X}_m \pm 3S_{\bar{X}}$] contendrá a μ con una probabilidad del 99%. (5+)



En una distribución normal. El 68% de los valores centrales se agrupan en el intervalo $\text{media} \pm \text{una desviación típica}$; 95% entre $\text{media} \pm \text{dos desviaciones típicas}$, y 99% entre $\text{media} \pm \text{tres desviaciones típicas}$.

EJEMPLO:

Si la media de una distribución normal es de 30 cms y su desviación estándar de 5 cms, podremos asegurar que:

- Entre 25 y 35 cms se encuentra aproximadamente el 68% de los valores centrales de la distribución [$\mu \pm \sigma$].
- Entre 20 y 40 cms se encuentra aproximadamente el 95% de los valores centrales de la distribución [$\mu \pm 2\sigma$].
- Entre 15 y 45 cms se encuentra aproximadamente el 99% de los valores centrales de la distribución [$\mu \pm 3\sigma$].

NOTA: Matemáticamente los intervalos correctos serían los siguientes: a) 68%: $\mu \pm 1$; b) 95%: $\mu \pm 1.96$; c) 99%: $\mu \pm 2.56$. Sin embargo, la aproximación que os recomendamos es la que hasta ahora han preguntado siempre en el MIR, y es además más sencilla de recordar.

Otro tipo de ejercicio no preguntado aún en el MIR: Media 50 años y desviación estándar de 10 años (de la pregunta anterior)...

- ¿Cuántos tienen más de 70 años y menos de 30? Respuesta: 5%.
- ¿Cuántos tienen más de 70 años? Respuesta: 2.5%.
- Y, ¿cuántos viven más de 80 años?...Respuesta... (debéis utilizar el intervalo con el 99%...).

5. Podemos utilizar la distribución normal (y aplicar sus propiedades) siempre que:

- La variable a estudio siga una distribución normal (MIR).
- El tamaño muestral sea lo suficientemente grande (MIR) (para que represente a la distribución de la que es extraída).

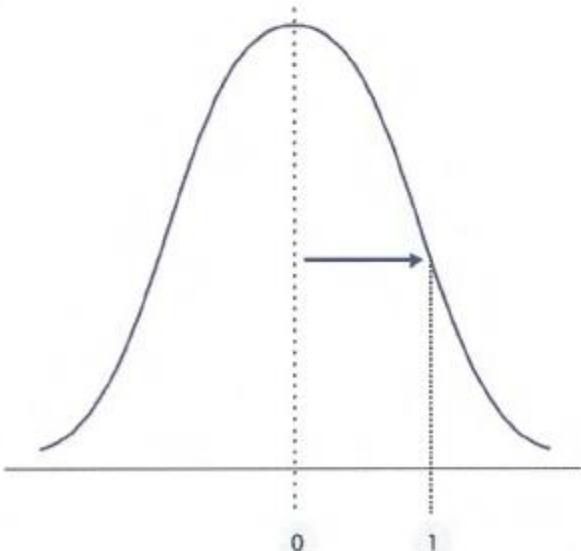
Propiedades de la Distribución Normal (9MIR)

- Función de probabilidad continua y asintótica con el eje de abscisas, donde cualquier valor de x es posible desde $-\infty$ e ∞ .
- Simétrica respecto a punto central donde coinciden media, mediana y moda:
 - es el valor promedio de la distribución
 - deja el 50% de los valores a cada lado
 - es el valor de presentación más frecuente; siendo la distribución unimodal.
- Queda suficientemente definida por la media aritmética (μ) y la desviación estándar (σ) (o su varianza: σ^2).
- Los valores centrales se agrupan como sigue:
 - el 68% en intervalo $[\mu \pm \sigma]$
 - el 95% en intervalo $[\mu \pm 2\sigma]$
 - el 99% en intervalo $[\mu \pm 3\sigma]$
- Ej: Si la media es 30 y la desviación típica es 5:
 - el 68% entre 25-35
 - el 95% entre 20-40
 - el 99% entre 15-45

B. DISTRIBUCIÓN NORMAL ESTANDARIZADA O TIPIFICADA (TRANSFORMACIÓN Z):

Cualquier distribución Normal se puede convertir en otra de media=0 y desviación típica=1, restándole a cada valor de la distribución original la media aritmética y dividiéndolo por la desviación típica ($z = (x - \text{media}) / \text{desviación típica}$).

Este procedimiento de estandarización interesa a veces desde el punto de vista práctico y se llama transformación Z. La distribución resultante se denomina distribución Normal estandarizada o tipificada, con $\mu=0$ y $\sigma=1$ (vale 1 tanto la desviación típica como la varianza). Esto hace que podamos utilizar esta distribución tipificada con todas las distribuciones normales que queramos (nos evitamos las medias y desviaciones típicas de cada una).



Distribución Normal estandarizada: media 0 y desviación típica 1

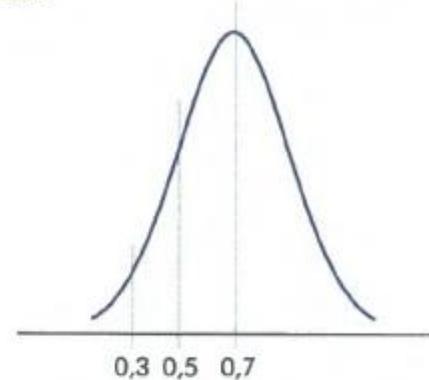
Un valor tipificado, abreviado como z , es un procedimiento que permite expresar cualquier valor inicial en términos de unidades de desviación típica, ya que la media es 0 (utilizando unas tablas).

2.2. Distribución binomial

Imaginemos que la probabilidad de que un determinado fármaco cure a un enfermo es $P=0,7$. Al aplicarlo sucesivamente a varios enfermos, cada uno de ellos presenta una dicotomía de posibilidades: curarse/ no-curarse, y lo que ocurre a

cada enfermo no está condicionado por lo que le sucede a los demás.

Si aplicamos el fármaco a una muestra de $n=100$ enfermos: ¿Cuál es la probabilidad de que se curen 50 pacientes y no se curen los 50 restantes, ó de que se curen 80 y no se curen 20? Las distintas posibilidades originan una variable discreta (proporción enfermos curados/ no-curados), que sigue una distribución binomial.



Distribución binomial de probabilidad $p=0,7$ y tamaño muestral n grande. Fijaos cómo la media de esta distribución binomial es 0,7, coincidiendo (como era de esperar) con la propia probabilidad de que el fármaco curase. Dado que en nuestro ejemplo el tamaño elegido es grande, se cumple que la distribución se aproxima a una distribución normal, coincidiendo media, mediana y moda en su centro.

A. CARACTERÍSTICAS:

- Se aplica a variables aleatorias discretas (número de individuos curados).
- La variable se define por el número de individuos, de entre los n de una muestra, que verifican una cierta característica dicotómica (el fármaco cura o no).
- La probabilidad p de presentación de dicha característica (suceso), debe ser constante e independiente en todos los individuos (la probabilidad de curación de un paciente con un fármaco no se influye por el hecho de que se cure otro).
- Los parámetros que definen una distribución binomial son la probabilidad del suceso independiente p (curación), la probabilidad del suceso complementario ($q=1-p$) y el tamaño de la muestra n .
- Al aumentar el tamaño de la muestra (n grande), la distribución tiende a aproximarse a la Normal (como en nuestro ejemplo).

Utilizamos la distribución binomial en variables dicotómicas (MIR)

2.3. Distribución de Poisson

A. CARACTERÍSTICAS:

- Podríamos considerarla un caso particular de la binomial.
- Utilizada en casos de tamaño muestral grande y variables con probabilidad de aparición pequeña. Se denomina también ley de los sucesos raros, al describir el comportamiento de variables de frecuencia muy baja. Ej. albinos en una población normal (en una población grande la probabilidad de encontrar un albino es muy pequeña), bacterias por cm^3 de aire (en una toda una "población grande de" aire la concentración de bacterias por cm^3 es escasa).

poisson

La distribución de Poisson es un caso particular de la binomial (SS), en la que la probabilidad p es muy pequeña y el tamaño n muy grande.

2.4. Otras distribuciones

Modelos especiales de distribuciones de probabilidad, denominadas **distribuciones en el muestreo**, que están asociados al proceso mismo de la inferencia estadística: **chi-cuadrado (χ^2)**, **t de Student-Fisher** y **F de Fisher-Snedecor**. Las veremos más adelante como pruebas de contraste de hipótesis.

No confundir...	con...
<p>Función de densidad de probabilidad: función matemática que asigna a cada valor una probabilidad de aparición. Análogo a <u>frecuencia relativa</u>.</p> <p>Distribución Normal: modelo teórico de distribución de probabilidad. Existen infinitas distribuciones normales, cada una de las cuales tendrá una media y desviación estándar diferentes.</p> <p>Distribución binomial: modelo teórico de distribución de probabilidad que se aplica a variables cuantitativas discretas, las cuales se definen por verificar o no una característica dicotómica. Ej. número de individuos que se curan (o no) con un fármaco.</p>	<p>Función de distribución de probabilidad: función matemática que asigna a cada valor la probabilidad de aparición de un valor igual o inferior al considerado. Análogo a <u>frecuencia relativa acumulada</u>.</p> <p>Distribución normal tipificada o estandarizada: modelo teórico de distribución de probabilidad, con valores fijos para la media -cero- y la desviación estándar -uno-. Cualquier distribución normal puede ser estandarizada mediante una transformación Z de sus valores.</p> <p>Distribución de Poisson: caso particular de la binomial, en que el tamaño muestral es muy grande y la probabilidad de aparición del fenómeno a estudio es muy pequeño. Ej. albinos en población general.</p>



RESUMEN DE PROBABILIDAD

1. CONCEPTOS BÁSICOS

- La frecuencia relativa de un valor expresa la probabilidad de aparición de dicho valor en su distribución. En inferencia estadística pretendemos calcular la probabilidad de aparición de un valor en su distribución poblacional a partir de los datos obtenidos de una distribución muestral.

2. MODELOS DE DISTRIBUCIÓN DE PROBABILIDAD

- Las más importantes son: a) distribución normal (campana de Gauss); b) distribución binomial; c) distribución de Poisson (ley de los sucesos raros).

2.1. DISTRIBUCIÓN NORMAL (CAMPANA DE GAUSS)

- Es el modelo más utilizado en inferencia estadística, siguiendo muchas variables biomédicas una distribución ajustable a ella (colesterolemia, uricemia, estatura...). Es una función de distribución de probabilidad con una propiedades muy importantes.
- Propiedades de la distribución normal (9MIR):
 - Función de probabilidad continua (2MIR) y asintótica con el eje de abscisas, donde cualquier valor de x es posible desde $-\infty$ e ∞ (MIR).
 - Simétrica respecto a punto central (3MIR) donde coinciden media, mediana y moda (3MIR):
 - es el valor promedio de la distribución
 - deja el 50% de los valores a cada lado (MIR)
 - es el valor de presentación más frecuente; siendo la distribución unimodal (MIR).
 - Queda suficientemente definida por la media aritmética (μ) y la desviación estándar (σ) (o su varianza: σ^2) (MIR)
 - Los valores centrales se agrupan como sigue:
 - el 68% en intervalo $[\mu \pm \sigma]$ (MIR)
 - el 95% en intervalo $[\mu \pm 2\sigma]$ (5MIR)
 - el 99% en intervalo $[\mu \pm 3\sigma]$
- Ej (6MIR): Si la media es 30 y la desviación típica es 5 los valores centrales estarían entre: 68% entre 25-35, 95% entre 20-40 y 99% entre 15-45.

2.2. DISTRIBUCIÓN BINOMIAL

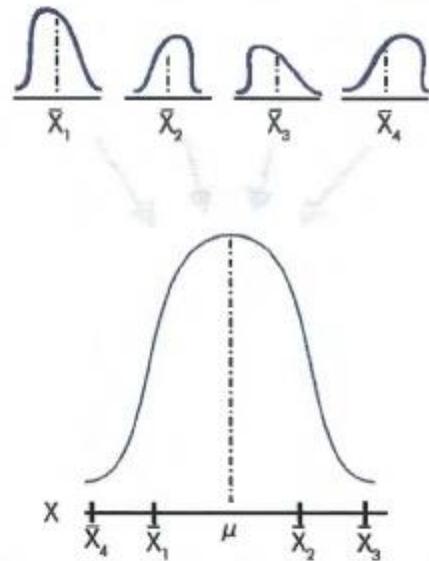
- La variable de la distribución es cuantitativa discreta, definiéndose dicha variable cómo "número de sujetos que verifican una característica dicotómica". Ej. número de sujetos que se curan con un fármaco: la variable de la distribución es discreta (número de sujetos: 1, 2, 3... ,n), pero la característica es una variable cualitativa nominal dicotómica (curación/ no curación).
- La distribución binomial se caracteriza por tres parámetros: a) la probabilidad p de aparición del suceso (curación); b) su complementario q ($q = 1-p$) (no curación); y c) el tamaño muestral n .

2.3. DISTRIBUCIÓN DE POISSON

- También denominada ley de los sucesos raros por aplicarse, como caso particular de la distribución binomial, a variables con probabilidad de aparición escasa (raro) y tamaños muestrales grandes. Ej. número de albinos en la población general

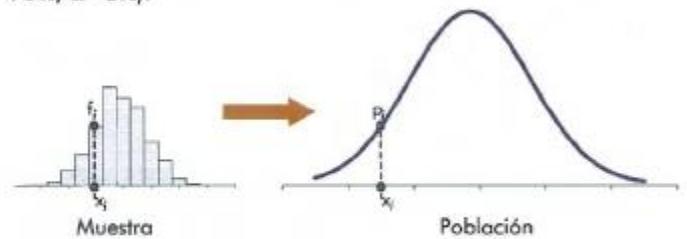
Dentro de la inferencia estadística vamos a considerar dos situaciones:

- Estimación de parámetros poblacionales.** Utilizamos medidas muestrales de un parámetro (colesterolemia) e intentamos inferir (estimar) cuál es el verdadero valor de la media de dicha variable en la población, basándonos, como veremos, en: a) en el teorema del límite central de la media, b) el error estándar de la media (de la distribución de las medias muestrales) y c) en las propiedades de la distribución Normal. Ej: muestra de 60 individuos de una población de 300 en donde detectamos que la media muestral de colesterolemia es de 150 mg/dl. Hacemos nuestros cálculos y nos resulta una media poblacional de colesterolemia para esa población de $160 \pm 2 \times 15$ mg/dl, es decir, intervalo de confianza del 95% de 130-190 mg/dl.
- Valoración de significación estadística.** Como veremos tanto el grado p de significación estadística como el cálculo de intervalos de confianza nos van a ayudar a saber si el resultado de un estudio ha sido visto porque realmente existe o simplemente por azar ("chiripa"), debido a la variabilidad de las diferentes posibles muestras. En dos ocasiones nos interesará la significación estadística:
 - Valorando si una asociación hallada entre dos variables (tabaco y cáncer de pulmón) es o no estadísticamente significativa.
 - Valorando si una diferencia entre dos grupos (% curados en los tratados con un fármaco y en los tratados con placebo) es o no estadísticamente significativa.



De una misma población se pueden extraer infinitas muestras, que siempre serán ligeramente diferentes unas a otras, aunque se utilice el mismo método de muestreo.

Podemos extraer infinitas muestras y hallar la "media de todas las medias", pero esto sería demasiado trabajo. Afortunadamente, la Estadística Inferencial nos proporciona un método para conocerla utilizando los datos de únicamente una muestra representativa cualquiera. La estrategia será dar a la pregunta una respuesta probabilística mediante la utilización de intervalos de confianza. Ej: el nivel de colesterolemia medio en la población es de 150-180 mg/dl (con un nivel de confianza del 95%, $\alpha=5\%$).



Partimos de datos provenientes de la muestra para extrapolar los resultados a la población.

Inferencia Estadística

Estimación

Contraste de Hipótesis

Dentro de la Inferencia Estadística distinguimos: estimación de parámetros poblacionales y contrastes de hipótesis. Utilizan conceptos y métodos similares, pero de forma didáctica vamos a considerarlos de forma independiente.

1. Estimación de parámetros poblacionales

1.1. Valor verdadero

Extraemos una muestra de una población. Calculamos la media de una variable aleatoria (ej: colesterolemia) en los individuos de la muestra. Y ahora queremos saber cuál es el verdadero valor de la media en la población, la media poblacional μ . Se trata de pasar de la verdad en el estudio (muestra) a la verdad en el universo (población). Nuestro problema es que queremos generalizar a la población los resultados de una muestra, a pesar de que las observaciones de la muestra no son directamente aplicables, en principio, a la población; aunque la muestra sea representativa.

Ello se debe a la inevitable variabilidad en el muestreo, una de las fuentes de error aleatorio. Para k muestras diferentes extraídas de una misma población, el valor de sus k medias será siempre ligeramente distinto en cada una de ellas. Obtendremos k valores ligeramente diferentes para media muestral: $\bar{X}_1, \bar{X}_2, \bar{X}_3, \dots, \bar{X}_k$, aún cuando todas las muestras sean representativas (v. dibujo). "No es un problema de representatividad, sino de variabilidad". Cada muestra nos da una media más o menos distinta a la anterior: ¿con qué valor nos quedamos? ¿cuál es el verdadero? ¿cuál es la media poblacional?

1.2. Intervalos de confianza

A. DEFINICIÓN DE INTERVALOS DE CONFIANZA:

Un intervalo de confianza es un intervalo de valores de la variable. En la práctica, cualquier intervalo de confianza contiene o no el valor verdadero. La correcta interpretación del intervalo de confianza implica decir que si se tomaron muchas pruebas aleatorias y se establece un intervalo de confianza del 95% de cada muestra, entonces el 95% de estos intervalos contendría el verdadero parámetro estimado.

B. ERROR α Y NIVEL DE CONFIANZA:

Para calcular un intervalo de confianza el investigador debe establecer a priori el error α que asume en la estimación, es decir, la probabilidad de que el valor verdadero del parámetro estudiado quede fuera de los límites del intervalo.

El nivel de confianza de la estimación o " $1-\alpha$ " (complementario de α) determina la probabilidad de que el intervalo de confianza contenga efectivamente el valor verdadero del parámetro poblacional. Es frecuente establecer el nivel de confianza en el 95%, es decir, $\alpha=5\%$ o 0.05.

Como hemos visto tanto el nivel de confianza deseado como el error α asumido son factores determinantes del tamaño muestral, que debe ser calculado en función de las necesidades del estudio.

C. APLICACIONES:

Los intervalos de confianza se aplican a la estimación de diversos tipos de parámetros:

- Medidas descriptivas de distribuciones: media poblacional, de una proporción (60-80% de curados con un fármaco).
- Medidas de fuerza de asociación: riesgo relativo (RR de 4-4.1) odds ratio (OR de 7.8)

- c) Medidas del tamaño de un efecto: la diferencia entre los curados con 2 fármacos diferentes (diferencias de 20-30% entre los tratados con A y con B).

1.3. Error estándar de la media

La amplitud de los límites del intervalo determina la precisión en la estimación (es más preciso el intervalo 150-160 mg/dl de colesterolemia que el intervalo 140-170 mg/dl, por ser éste más amplio). La falta de precisión se debe a la variabilidad en el muestreo, ya que esta variabilidad genera un error aleatorio propio de cualquier fenómeno al azar. Pues bien, la medida de esta variabilidad en el muestreo se denomina error estándar del parámetro estudiado; como es el caso del error estándar de la media.

El error estándar de la media es la medida de dispersión más adecuada para estimar el verdadero valor de una media poblacional (MIR)



En la construcción de los intervalos de confianza para la estimación de parámetros poblacionales, utilizamos el error estándar de la media.

El error estándar de un parámetro es una medida de la variabilidad impuesta por el error aleatorio del muestreo y depende de la dispersión de la variable estudiada, su desviación típica S, y del tamaño muestral, n (MIR). Tanto es así, que el error estándar de la media se calcula (MIR) con esta fórmula:

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

En consecuencia: a mayor tamaño muestral, menor error estándar del parámetro estudiado (media), y mayor precisión en la estimación (2MIR).

Podríamos decir, que el error estándar de una muestra equivale a la desviación típica del individuo en la población.

MIR 11 (9701): Si desea conocer la imprecisión esperada al estimar la media de una variable recurrirá a:

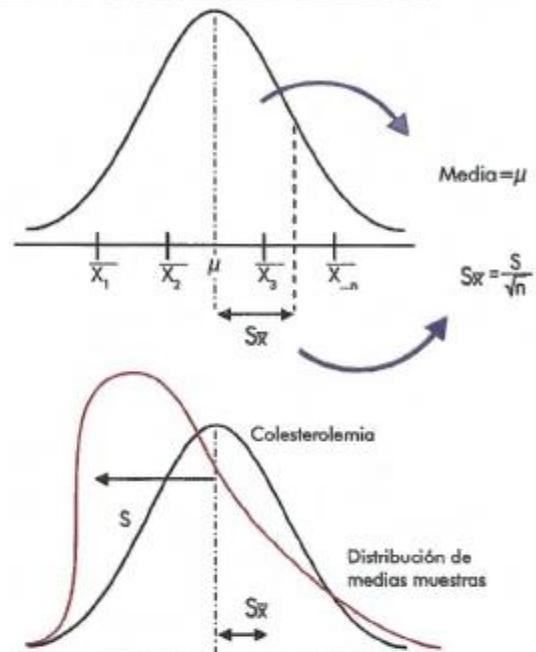
1. El error estándar (típico).*
2. La desviación estándar (típica).
3. La varianza de la variable.
4. El rango intercuartílico.
5. El coeficiente de variación.

Dado que el error estándar mide el error o imprecisión (variabilidad) a que está sujeta la estimación, utilizaremos el error estándar en el cálculo de los intervalos de confianza. Pero para ello necesitamos introducir el teorema del límite central de la media para su cálculo.

1.4. Teorema del límite central de la media

Imaginemos que de una población seleccionamos k muestras diferentes mediante muestreo repetido, y que a continuación realizamos una nueva distribución con las medias de esas k muestras, para una variable X que se ha medido. Obtenemos así la distribución de las medias muestrales, es decir, ahora la distribución no es de una variable sino de las medias de esa variable de todas las muestras extraídas.

El teorema del límite central de la media establece que, si reunimos muestras del mismo tamaño de una distribución no Normal, la distribución resultante de sus medias sí será Normal, siempre que las muestras sean suficientemente grandes. El tamaño necesario será pequeño en variables que ya de por sí sigan una distribución Normal y deberá ser mayor en variables con distribuciones muy alejadas de la Normal.



Aunque una distribución no se ajuste a la distribución Normal, el teorema del límite central de la media nos garantiza que la distribución de las medias de las infinitas muestras posibles de esa población, sí será una distribución Normal, siempre que se extraigan muestras representativas con tamaño muestral suficientemente grande.

ESTO EN LA PRÁCTICA SIGNIFICA...:

La distribución Normal tiene muchas ventajas estadísticas en su manejo: es simétrica, sus valores se pueden tipificar y utilizar tablas universales (con la media y su desviación típica ya caracterizamos la distribución). El que la distribución poblacional de una variable estudiada no se ajuste a ella supone, en teoría, que se nos van a complicar mucho todos los cálculos. Sin embargo, gracias a este teorema, podemos utilizar las ventajas de esta distribución teórica, cuando lo que buscamos es estimar un parámetro (coeficiente de inteligencia medio, colesterolemia, altura media, etc) mediante muestras a pesar de que dicho parámetro no siga una distribución normal en la población.

1.5. Cálculo del intervalo de confianza

Nos hemos quedado con una distribución, la de las medias muestrales, que, gracias al teorema del límite central de la media, podemos decir que se comporta como una distribución Normal, independientemente de las características de la distribución real de la variable estudiada (puede no seguir una distribución Normal), ¿de acuerdo?

Pues bien, habíamos dicho que podíamos caracterizar una distribución Normal conociendo simplemente su media y su desviación típica. En este caso, conociendo:

- a) μ (media poblacional, y por ende media de todas las medias muestrales).
- b) La medida de dispersión propia de esta distribución, que no puede ser otra que el error estándar de la media. Recordemos que el error estándar es la medida de la variabilidad en el muestreo, la medida de la imprecisión que nos obliga a estimar. Por lo tanto,

La distribución de las medias muestrales es una distribución Normal cuya media es μ (media poblacional) y cuya desviación típica es $S_{\bar{x}}$ (error estándar de la media) (MIR)

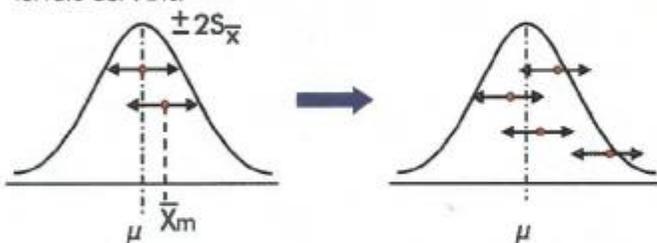
Ya que en esta distribución μ es la media y el error estándar de la media su desviación típica podemos decir, aplicando propiedades de la distribución Normal, que:

El intervalo $[\mu \pm 2S_{\bar{x}}]$ contiene el 95% central de las medias muestrales (MIR)

El intervalo $[\mu \pm 3S_{\bar{x}}]$ contiene el 99% central de las medias muestrales.

Entre esos "95% de valores" (por ejemplo) hay uno que nos interesa más: la media poblacional. Pues sólo nos queda un paso: si estos mismos intervalos (construidos con el error estándar de la media) los transportamos y los centramos sobre una media muestral (obtenida en nuestro estudio) entonces:

- La distribución de medias muestrales se forma por medias de las infinitas muestras posibles de cada población. El intervalo ± 2 errores estándares no deja de ser un área bajo la curva, que incluye el 95% de todas las medias muestrales. Si lo centramos en la media poblacional, será el 95% central del área; mientras que centrado ese intervalo sobre una media cualquiera incluirá un 95% de área no centrada, pero seguirá siendo el 95% de todo el área.
- Es decir, sólo un 5% de todas las medias muestrales se escapará del intervalo. Sólo existe un 5% de probabilidades de no coger la media poblacional, se centre donde se centre el intervalo del 95%.



Si centramos los intervalos de confianza sobre una muestra representativa cualquiera, y no sobre la media poblacional, tendremos unos intervalos que incluirán con un 68%, 95% y 99% respectivamente, a la media poblacional.

El intervalo $[\bar{X}_m \pm 2S_{\bar{x}}]$ contendrá a μ con una probabilidad del 95% (MIR)

El intervalo $[\bar{X}_m \pm 3S_{\bar{x}}]$ contendrá a μ con una probabilidad del 99%.

Tales intervalos se denominan intervalos de confianza de la media poblacional y sus límites, límites de confianza de la media. Para contestar preguntas MIR quedémonos con esto:

$[\bar{X}_m \pm S_{\bar{x}}] \Rightarrow$ Intervalo de confianza del 68%.

$[\bar{X}_m \pm 2S_{\bar{x}}] \Rightarrow$ Intervalo de confianza del 95% (7MIR)

$[\bar{X}_m \pm 3S_{\bar{x}}] \Rightarrow$ Intervalo de confianza del 99%

NOTA: Matemáticamente los intervalos correctos serían los siguientes: a) 68%: $\mu \pm 1$; b) 95%: $\mu \pm 1.96$; c) 99%: $\mu \pm 2.56$. Sin embargo, la aproximación que os recomendamos es la que hasta ahora han preguntado siempre en el MIR, y es además más sencilla de recordar.

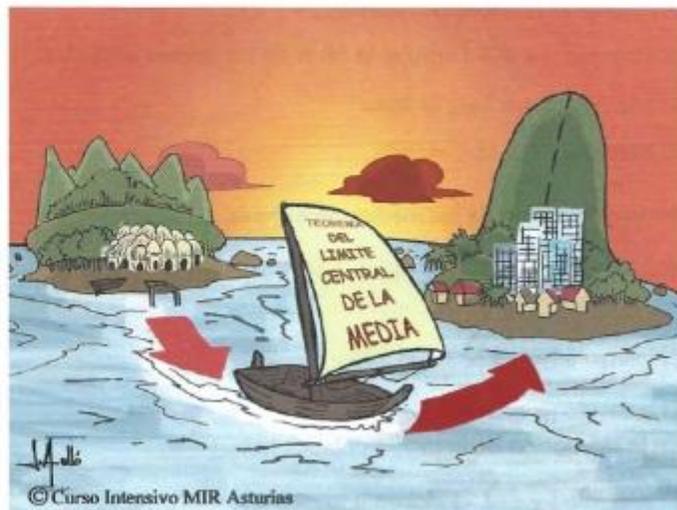
Una reflexión final: los intervalos de confianza serán más anchos a mayor dispersión de la variable medida (S), y más estrechos o precisos a mayor tamaño muestral (n).

SI EN UNA PREGUNTA MIR...

Nos hablan de...	Se refieren a...	Y buscaremos la opción del tipo...
Desviación estándar (o típica)	Medida de dispersión de una distribución Normal cualquiera (muestral, poblacional...)	El 95% de los valores centrales de... están entre... y ...
Error estándar	Medida de dispersión de la distribución de las medias muestrales.	Existe un 95% de confianza de que entre... y... esté el verdadero valor de...

MIR 06 (8480): Un artículo de una revista científica informa que el intervalo de confianza al 95% del nivel medio de colesterol en los adultos atendidos en un Centro de Salud es 192-208. Se aceptó que la variable tenía una distribución normal y el número de pacientes estudiados fue 100. ¿Cuál de las siguientes afirmaciones es FALSA?:

1. La probabilidad de que el nivel medio poblacional esté comprendido entre 192 y 208 es 0.95.
2. Si se repitiera el estudio muchas veces, en un 95% de ellas se obtendría una media muestral comprendida entre 192 y 208.
3. El 95% de los adultos de la población tiene un nivel de colesterol comprendido entre 192-208. *
4. La media muestral encontrada en el estudio es de 200.
5. La desviación típica muestral encontrada en el estudio es aproximadamente 40.



El Teorema del Límite Central de la Media nos lleva desde una distribución no necesariamente Normal (isla de perfil irregular) hasta una distribución Normal, la distribución de las medias muestrales (isla de perfil en campana de Gauss)

No confundir ...

Distribución de una variable: La que corresponde a los valores que toma esa variable:

- a) Distribución muestral: La obtenida de las observaciones sobre una muestra.
- b) Distribución poblacional: La que resultaría de medir a la población entera.

con ...

Distribución de las medias muestrales: Distribución teórica de las medias de todas las muestras que se pudiesen extraer de la población. Es una distribución Normal. Su media ("media de las medias") coincide con la media poblacional de la variable estudiada. Su desviación típica es el error estándar de la media.

No confundir ... con ...

Media poblacional: μ . Ej.: media de la colesterolemia en la población. Es la media que estimamos mediante los intervalos de confianza.

Coefficiente de variación: es adimensional y compara variabilidad relativa de diferentes métodos de medida (con diferentes unidades de medida)

$$CV = \frac{S}{\bar{x}}$$

Valores centrales de una distribución Normal cualquiera: Sus intervalos de probabilidad se construyen con la desviación típica de la distribución (S) (ej.: de colesterolemia) aplicando las propiedades de la distribución Normal.

En cualquier distribución Normal...

El intervalo $[\bar{x} \pm S]$ agrupa el 68 % de los valores centrales.

El intervalo, $[\bar{x} \pm 2S]$, el 95%.

El intervalo $[\bar{x} \pm 3S]$, el 99%.

En la distribución de las medias muestrales...

El intervalo $[\mu \pm S_{\bar{x}}]$, agrupa el 68 % central de las medias muestrales.

El intervalo $[\mu \pm 2S_{\bar{x}}]$, el 95%.

El intervalo $[\mu \pm 3S_{\bar{x}}]$, el 99%.

En la estimación de la media poblacional...

El intervalo $[\bar{x} \pm S_{\bar{x}}]$ incluye a la media poblacional con una probabilidad del 68%.

El intervalo $[\bar{x} \pm 2S_{\bar{x}}]$, con una probabilidad del 95%.

El intervalo $[\bar{x} \pm 3S_{\bar{x}}]$, con una probabilidad del 99%.

2. Valoración de significación estadística

Hemos estimado parámetros poblacionales (media poblacional) a partir de los datos de una muestra. Ahora lo que pretendemos es, hallado un valor, saber si éste es estadísticamente significativo, es decir, si según la estadística hemos abierto el eterno problema de la variabilidad, y nuestro valor es real, no debido al azar en el muestreo.

Para valorar la significación estadística utilizaremos test de contraste de hipótesis en dos situaciones:

- Calculando el grado de significación o valor p del estudio (comparándolo después con el error α).
- Calculando un intervalo de confianza (valorando si incluye el valor asignado a la hipótesis nula H_0).

Media muestral: \bar{x} . Ej.: media de colesterolemia en los sujetos de la muestra que estudiamos. Es sobre la que construimos los intervalos de confianza para estimar la media poblacional.

Error estándar de la media: Es la desviación típica de la distribución Normal de las medias muestrales.

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

Desviación estándar de una distribución: S . Corresponde a la variable medida. Ej.: desviación típica de la colesterolemia.

Intervalos de confianza de la media poblacional: Se construyen con el error estándar

de la media ($S_{\bar{x}}$), aplicado sobre la media muestral de nuestro estudio. Contienen a la media poblacional con una probabilidad determinada, que es el nivel de confianza.

MIR 03 (7692): El objetivo principal de contraste de hipótesis estadístico, aplicado a ciencias de la salud es:

- Confirmar la hipótesis alternativa y la hipótesis nula.
- Distinguir entre las diferencias debidas al azar y las diferencias clínicas.*
- Aumentar el valor predictivo del estudio para tener certeza en las conclusiones.
- Establecer el error alfa o error tipo I.
- Establecer el error beta o tipo II.

Y lo vamos a hacer en el estudio de dos situaciones muy comunes en biomedicina:

- Comparando proporciones de distintos grupos (2 grupos: comparando porcentaje de curados con A y con B).
- Asociaciones entre variables (tabaco y cáncer de pulmón).

Valorando la significación estadística en estas dos situaciones podría resultar que: a) un fármaco A cura "más que" un fármaco B, y b) las variables tabaco y cáncer de pulmón "están" asociadas (la Estadística lo no enunciaría así). O podría ocurrir que nuestros resultados no fueran significativos, por lo que la probabilidad de que el azar hubiera generado la aparente diferencia entre A y B o la aparente asociación entre las dos variables sería muy alta, y pensaríamos que no había ni esa diferencia ni esa asociación.

2.1. Significación estadística mediante el grado p

A. DEFINICIÓN DEL GRADO DE SIGNIFICACIÓN p:

El objetivo de un estudio en el que se plantean contrastes de hipótesis es extraer conclusiones sobre el "significado real" de un efecto observado (una diferencia o una asociación). Donde, El grado de significación p (MIR) es la probabilidad calculada de que el azar haya generado un efecto observado (diferencia o asociación) u otro más alejado (MIR), no existiendo realmente ese efecto (MIR). El grado de significación p depende del tamaño muestral (MIR) y de la magnitud del efecto observado, pero no es ni mide tal magnitud de la diferencia o asociación (MIR).

Ej: si el grado de significación $p=0.06$, significa que la probabilidad de que el resultado observado (una diferencia o asociación) se haya producido por azar (y no sea real) es del 6% (MIR).

MIR 04 (7971): Se pretende comparar la frecuencia de complicaciones de dos preparados distintos de un mismo fármaco. Se observó un 5% de complicaciones con un preparado y un 3% con el otro, siendo esta diferencia estadísticamente significativa ($p=0,045$). La interpretación correcta de este resultado es:

- Si ambos preparados tuvieran la misma frecuencia de complicaciones, la probabilidad de encontrar una diferencia igual o mayor a la observada en 0,045.*
- La probabilidad de que ambos preparados tengan la misma frecuencia de complicaciones es de 0,045.
- Los dos preparados tienen distinta frecuencia de complicaciones.
- Los dos preparados no tienen la misma frecuencia de complicaciones.
- La probabilidad de que ambos preparados tengan la misma frecuencia de complicaciones es de 0,955.

MIR 05 (8226): Debemos evaluar los resultados de 3 medicamentos (A, B y C) en el tratamiento de pacientes con déficit cognitivo ligero. Los diseños implementados en los tres casos han sido muy similares y se ha usado placebo como grupo control en los tres casos. Se han observado los siguientes resultados de significación estadística y de porcentaje de reducción absoluta en la progresión a demencia en las comparaciones contra placebo: A versus Placebo: 3%, $p<0,001$; B versus Placebo: 8%, $p=0,041$; C versus Placebo 1% $p=0,021$. ¿Cuál de las siguientes afirmaciones se puede concluir a partir de los presentes resultados?:

- El valor de p demuestra que el medicamento con un efecto de mayor magnitud es del A.
- El valor de significación nos indica que el medicamento con un efecto de menor magnitud es el B.
- El medicamento B es el que muestra un efecto de mayor



magnitud frente a placebo.*

- El medicamento C es mejor que el B ya que es más significativo.
- El valor de significación nos indica que el medicamento con un efecto de menor magnitud es el A.

MIR 07 (8748): Al analizar los resultados de un ensayo clínico, se concluye que existen diferencias entre los parámetros evaluados para cada una de las ramas de tratamiento, con una P de 0,034. Esto significa:

- Que si se repite el estudio, en un 3,4% de ocasiones no se encontrarían diferencias.
- Que en un 3,4% de sujetos de ambas ramas, no se encontraron diferencias.
- Que hay un 3,4% de diferencia entre los tratamientos estudiados.
- Que la magnitud de la diferencia entre tratamientos es mayor que si $P = 0,05$.
- Que hay una probabilidad de 3,4% de que la diferencia obtenida se deba al azar.*

MIR 09 (9235): ¿Cuál es la interpretación de la significación estadística (valor de la "p") de una prueba de contraste de hipótesis?

- La probabilidad de rechazar la hipótesis nula.
- La probabilidad de aceptar la hipótesis nula.
- La probabilidad de rechazar la hipótesis nula cuando es cierta.*
- La probabilidad de rechazar la hipótesis nula cuando es falsa.
- La probabilidad de cometer un error en la decisión.

MIR 12 (9938): En un contraste de hipótesis estadístico, ¿a qué definición corresponde con más exactitud el valor "p"?:

- La probabilidad de observar los resultados del estudio, u otros más alejados de la hipótesis nula, si la hipótesis nula fuera cierta.*
- La probabilidad de que la hipótesis nula sea cierta.
- La probabilidad de observar los resultados del estudio si la hipótesis nula fuera cierta.
- La probabilidad de que los resultados observados sean debidos al azar.
- La probabilidad de observar los resultados del estudio, u otros más alejados de la hipótesis nula, si la hipótesis alternativa fuera cierta.

MIR 13 (10203) (204): En un ensayo clínico aleatorizado de fase III se comparó la eficacia de un nuevo analgésico (experimental) con un tratamiento control (tramadol) en pacientes con dolor crónico. La hipótesis de trabajo era que el tratamiento experimental reduce el dolor más que el tramadol. El efecto de los dos tratamientos se determinó a las 48 horas mediante la reducción de la puntuación marcada por el paciente en una escala analógica - visual de 0 a 100 mm. La reducción media en el grupo tramadol fue de -27 y en el grupo experimental de -31. Se hizo el contraste de hipótesis para las diferencias, con la correspondiente prueba estadística y se obtuvo un valor de $p = 0,03$. Respecto al estudio anterior, ¿cuál de las siguientes conclusiones le parece más correcta?

- El estudio demostró diferencias clínicamente relevantes.
- Las diferencias en el efecto analgésico entre los dos tratamientos estudiados fueron significativas.*
- El beneficio - riesgo del tratamiento experimental fue mejor que el del tramadol.
- El tratamiento experimental fue un 20 % mejor que el tramadol.
- Podemos recomendar el uso generalizado del tratamiento experimental, porque es más eficaz que el tramadol en el tratamiento del dolor crónico.

Una vez calculado p, debemos juzgar si la probabilidad es demasiado alta o nos parece razonablemente baja. ¿Cómo? Viendo si es mayor o menor que el error α o nivel de significación asumidos en el estudio.

B. CONTRASTE DE HIPÓTESIS:

1. METODOLOGÍA:

Para valorar la significación estadística de una diferencia o una asociación se hace lo siguiente:

- Establecemos una hipótesis que se asume como cierta de entrada (H_0 o hipótesis nula). En una asociación a estudio (tabaco y cáncer de pulmón) la H_0 sería "no hay asociación". En una diferencia de proporciones a estudio (diferencia entre % curados con A y con B) sería "no existe diferencia". [A H_0 se le suelen asignar los valores contrarios a lo que pretendemos encontrar: "no hay diferencia", "no hay asociación"]
- Medimos el grado de significación p o probabilidad de que el azar haya generado el efecto observado en el caso de que la hipótesis nula (no hay asociación o diferencia) fuera cierta. Calculamos p con alguno de los diferentes tests de contraste de hipótesis a partir de los datos del estudio.
- Decidimos si el resultado es significativo ($p < \text{error } \alpha$) o no ($p > \text{error } \alpha$) (MIR). Ej: si $p=0,03$ y error $\alpha=0,05$ (establecido a priori), la probabilidad de que el azar en el muestreo haya generado la asociación o diferencia obtenidas es de 0,03 (demasiado baja). Merece la pena pensar que sí existe tal asociación o diferencia. El punto de corte es α , de forma que si $p=0,06$ no habríamos arrojado la suficiente evidencia estadística para decir que existe tal diferencia o asociación (la probabilidad es mayor de lo que estábamos dispuestos a asumir de mano: ¡Y hay que mantener la palabra!).

$p > \alpha$	Probabilidad p demasiado ALTA.	NO estadísticamente significativo.
$p < \alpha$	Probabilidad p suficientemente baja.	Estadísticamente significativo.
p	Probabilidad de que el azar haya generado un efecto observado (diferencia o asociación)..	Calculado con los datos del estudio.
α	Probabilidad de que, posteriormente en el estudio, aceptemos como real una diferencia o asociación falsas, generadas por el azar.	Asumido a priori (no calculado). Suele ser 0.05 ó 5%.



Únicamente cuando p ("estatura de los individuos") sea inferior a la altura del "rótulo $p < \alpha$ ", entrará en el terreno de lo "estadísticamente significativo".

2. DEFINICIONES:

- Hipótesis nula (H_0). Proposición de entrada cierta, cuya veracidad se somete a examen. Generalmente consiste en postular "no hay diferencia" o "no hay asociación", justo lo contrario de lo que queremos demostrar. Ej: Para "no hay diferencia" se postularía:

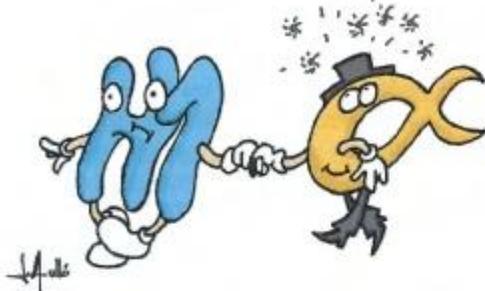
$$H_0 \equiv P(\text{curaciones en A}) - P(\text{curaciones en B}) = 0$$

b) Hipótesis alternativa (H_1). Proposición que se acepta si H_0 resulta rechazada. Generalmente, se asigna a que existe diferencia o asociación. Siguiendo el mismo ejemplo,

$$H_1 = P(\text{curaciones en A}) - P(\text{curaciones en B}) \neq 0$$

c) Error tipo I (error α error de primera especie) (3MIR): Error aleatorio consistente en rechazar H_0 cuando es cierta (2MIR). Es decir, es la probabilidad de que, no existiendo tal diferencia o asociación, interpretemos que existen: resultado falso positivo (MIR). Se trata de una probabilidad condicionada:

$$\text{Riesgo } \alpha: P(\text{rechazar } H_0/H_0 \text{ cierta})$$



© Curso Intensivo MIR Asturias

El error alfa es la probabilidad de aceptar H_1 , siendo H_0 cierta. Representa un falso positivo, por ver una diferencia o asociación que no existen. Es un error aleatorio.

MIR 01 (7180): ¿Qué es el error de tipo I en un estudio clínico?:

1. Un error sistemático debido a la falta de validez de los procedimientos del estudio.
2. Un error aleatorio consistente en rechazar la hipótesis nula cuando es cierta.*
3. Un error aleatorio consistente en no rechazar la hipótesis nula cuando es falsa.
4. Un error sistemático consistente en rechazar la hipótesis nula cuando es cierta.
5. Un error sistemático consistente en no rechazar la hipótesis nula cuando es falsa.

La opción 3 define el error beta.

MIR 08 (8994): Sobre el error de tipo I cuando se estudian las diferencias entre dos tratamientos:

1. Se llama también riesgo Beta.
2. Su opuesto representa la potencia de la prueba.
3. Lleva a concluir que hay una diferencia, cuando en realidad no la hay.*
4. Implica que la hipótesis nula estaba mal planteada.
5. Implica que se debería hacer de nuevo el análisis.

MIR 10 (9466): En un ensayo clínico que evalúa la eficacia de un nuevo agonista dopaminérgico para el tratamiento de la enfermedad de Parkinson los investigadores concluyen que el nuevo fármaco es eficaz cuando en realidad no es mejor que el placebo. Han cometido:

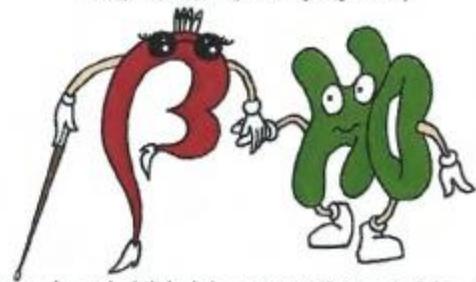
1. Un error tipo I.*
2. Un sesgo del observador.
3. Un sesgo de confusión.
4. Un error tipo II.
5. Un sesgo de clasificación.

Sinónimos (en significación estadística):

Es decir, si el investigador asume un riesgo α del 5% ó 0.05 da por supuesto que si repitiese el estudio infinitas veces se equivocaría en un 5% al interpretar como significativos resultados que no lo fueran. Es razonable: 1 entre 20. [Es un valor establecido por consenso]

d) Error tipo II (error β error de segunda especie) (3MIR): Error aleatorio consistente en no rechazar H_0 cuando es falsa (MIR): resultado falso negativo (MIR). Es decir, la probabilidad de que, existiendo una diferencia o asociación ciertas, interpretemos que no las hay (MIR) (no las vemos significativas).

Riesgo β : $P(\text{aceptar } H_0/H_0 \text{ falsa})$



El error beta es la probabilidad de aceptar H_0 , siendo falsa. Representa un falso negativo, por no ver una diferencia o asociación cuando realmente existe. Es un error aleatorio.

MIR 00 (6925): ¿Qué es el error de tipo II (error beta) en un estudio clínico?:

1. Un error sistemático debido a la falta de validez de los procedimientos del estudio.
2. Un error aleatorio consistente en rechazar la hipótesis nula cuando es cierta.
3. Un error aleatorio consistente en no rechazar la hipótesis nula cuando es falsa.*
4. Un error sistemático consistente en rechazar la hipótesis nula cuando es cierta.
5. Un error sistemático consistente en no rechazar la hipótesis nula cuando es falsa.

La opción 2 define el error tipo I.

MIR 07 (8734): En un contraste de hipótesis estadístico, si la hipótesis nula fuera cierta y se rechazara:

1. Se comete un error de tipo II.
2. Se toma una decisión correcta.
3. La potencia aumenta.
4. Se comete un error de tipo I.*
5. Se toma la decisión más conservadora.

e) Potencia (6MIR) ($1-\beta$: Complementario de β): Probabilidad de detectar como significativa la diferencia o asociación cuando existe (MIR), es decir, la probabilidad de detectar hipótesis alternativas (MIR) o rechazar la hipótesis nula siendo falsa (MIR).



La potencia es la probabilidad de detectar como significativa una diferencia o asociación que realmente existen (hipótesis alternativas)

Es recomendable un valor de potencia >70 u 80% (MIR). Si es insuficiente, cometeremos un falso negativo: deberemos aumentar el tamaño muestral (4MIR). En un estudio donde no se encuentre una diferencia o asociación estadísticamente significativas, debemos fijarnos en su potencia y su tamaño muestral, especialmente si se ha encontrado un efecto clínicamente significativo.

MIR 00 (6926): ¿Cuál es la interpretación correcta de un ensayo clínico en el que no se detectan diferencias estadísticamente significativas entre las intervenciones y en el que no parece haber errores sistemáticos?:

1. Las intervenciones son equivalentes.
2. El estudio ha sido no concluyente, ya que la probabilidad de error de tipo I puede ser muy alta.
3. La interpretación dependerá del tamaño muestral del estudio.*
4. En general, las intervenciones son similares, pero hay que investigar subgrupos de pacientes (formados por edades, sexo, gravedad de la enfermedad, etc.) ya que sigue sien-

do posible que en alguno de ellos un tratamiento sea superior.

- La interpretación dependerá primordialmente del grupo de control elegido.

Nos están hablando de la potencia: Quizá no encontró nada el estudio porque no tenía la suficiente potencia, ¿y de qué depende la potencia? Del tamaño muestral.

MIR 01 (7189): En un ensayo clínico sobre la erradicación de *Helicobacter Pylori* se compararon dos tratamientos administrados cada uno de ellos en unos 50 pacientes. Los porcentajes de erradicación fueron del 73% con un tratamiento y del 72% con el otro. Las diferencias no fueron estadísticamente significativas. Se calculó que el poder o potencia estadística, con un contraste unilateral y para diferencias de hasta el 12%, era del 53%. ¿Cuál es la interpretación correcta de este estudio?:

- Que ambos tratamientos son iguales.
- Que quedan razonablemente descartadas diferencias de porcentajes de erradicación que sean superiores al 6%.
- Que no es improbable que el estudio haya pasado por alto diferencias de hasta el 12% en los porcentajes de erradicación.*
- Que la diferencia de eficacia entre tratamientos es tan pequeña, que la decisión sobre cuál usar deberá basarse sobre otros aspectos, como diferencias de precio, de toxicidad, etc.
- La probabilidad de un error de tipo II (no encontrar diferencias, aunque existan) es del 53%.

MIR 04 (7963): El representante de un laboratorio farmacéutico le informa de la reciente comercialización de un nuevo antibiótico para el tratamiento de la neumonía que permite una administración menos frecuente (una vez al día) que los actualmente existentes y le resalta la importancia de cambiar el antibiótico que actualmente utiliza por el que presenta. Como demostración de su eficacia le presenta un ensayo clínico en el que el riesgo relativo (experimental/referencia) en la tasa de curaciones entre ambos es del 0.97 (intervalo de confianza: 0.60-1.30; no significativo). Basándose en estos resultados los autores concluyen que ambos tratamientos tienen una eficacia similar. Su opinión sobre la eficacia de dicho fármaco y la posibilidad de cambiar de antibiótico sería:

- Los resultados del ensayo y la opinión de los investigadores, que tienen la experiencia con el fármaco, le llevaría a aceptar su conclusión sobre la eficacia del fármaco y a utilizarlo.
- Cree que efectivamente ambos antibióticos pueden considerarse similares a efectos prácticos, ya que la diferencia de eficacia entre ambos (3% de curaciones) es muy pequeña y clínicamente irrelevante.
- La autorización del fármaco por las autoridades es suficiente garantía para aceptar su eficacia y sustituir el antibiótico que estaba utilizando hasta ahora.
- El hecho de que la información se la suministre un representante del laboratorio, le hace descartar la utilización del nuevo producto.
- La posible diferencia de eficacia entre ambos tratamientos incluye cifras demasiado amplias para que puedan considerarse equivalentes terapéuticos.*

MIR 09 (9238): En un contraste de hipótesis estadístico, la probabilidad de rechazar la hipótesis nula cuando ésta es falsa se denomina:

- Error tipo II.
- Error tipo I.
- Potencia.*
- Eficacia.
- Eficiencia.

MIR 11 (9702): En un contraste de hipótesis estadístico si la hipótesis nula fuera cierta y se rechazara ¿Cuál de las respuestas es CORRECTA?:

- Se comete un error de tipo II.
- Se toma una decisión correcta.
- La potencia aumenta.
- Se comete un error de tipo I.*
- Indica un tamaño muestral excesivo.

Sin embargo, una vez alcanzado un resultado concluyente no importa lo probable o improbable que fuese el llegar a él antes de hacer el estudio (mayor o menor potencia) (2MIR).



La probabilidad de que toque la lotería es pequeña. Sin embargo, una vez en posesión del boleto ganador, poco nos preocuparía este hecho irrefutable. De modo análogo, podemos partir de una escasa probabilidad de obtener significación estadística en un estudio (por tamaño muestral supuestamente insuficiente, por ejemplo). En el caso de que aún así alcancemos significación estadística en el mismo, los resultados del estudio SÍ son creíbles

MIR 01 (7192): Suponga que en la fase de diseño se calcula que para tener suficiente poder estadístico en el estudio de cierta intervención terapéutica mediante un ensayo clínico es necesario incluir 100 sujetos. Sin embargo, el ensayo se realiza con sólo 50 pacientes y el resultado es estadísticamente significativo ($P < 0,001$). Suponga que en todo lo demás el estudio parece correcto. En vista del pequeño número de pacientes incluidos, ¿son creíbles los resultados?:

- No, ya que en los ensayos demasiado pequeños los errores sistemáticos son inevitables.
- No, ya que hay muchas posibilidades de que las diferencias encontradas sean un resultado falso positivo.
- Si, ya que una vez alcanzado un resultado concluyente no importa lo probable o improbable que fuese el llegar a él antes de hacer el estudio.*
- No, ya que la probabilidad de encontrar diferencias, si éstas existen, es baja cuando el estudio es pequeño.
- Si, siempre y cuando estos resultados se utilicen exclusivamente para unirlos a los de otros estudios pequeños sobre la misma cuestión y hacer un meta-análisis o una revisión.

MIR 02 (7438): Cuando hay diferencias estadísticamente significativas, ¿es creíble el resultado de un ensayo clínico en el que, siendo todo lo demás correcto, el número de pacientes incluidos es menor que el que parecía necesario antes de realizar el estudio?:

- Si, pues el riesgo de error de tipo I es bajo al ser el resultado estadísticamente significativo.*
- No, pues hay gran riesgo de sesgo.
- No, pues hay gran riesgo de error de tipo I.
- No, pues hay gran riesgo de error de tipo II.
- No, ya que el resultado es no concluyente por no basarse en el número suficiente de pacientes.

		Falsos negativos	
		H_0 verdadera (no diferencia o asociación)	H_1 verdadera (existe diferencia o asociación)
Decisión	H_0 aceptada	1- α	Error tipo II
	H_0 rechazada	Error tipo I	Potencia (1- β)

- Error tipo beta (MIR)
- Potencia insuficiente (MIR)
- Tamaño muestral insuficiente (MIR)
- Sesgos (MIR)

Reflexión: Los riesgos α y β no son probabilidades de sucesos complementarios, es decir, sería erróneo pensar que $\alpha + \beta = 1$. Del mismo modo, tampoco son necesariamente contrapuestos: un aumento del tamaño muestral permite disminuir ambos (MIR), por ejemplo.

No confundir ...	con ...
Hipótesis nula (H_0): Proposición cuya veracidad se somete a examen en un test de contraste de hipótesis. Ej.: la proporción de curaciones de los fármacos A y B es igual.	Hipótesis alternativa (H_1): Proposición que se acepta si H_0 resulta rechazada. Ej.: la proporción de curaciones de los fármacos A y B es diferente.
Error tipo I: Rechazar H_0 siendo cierta. La probabilidad de cometerlo es α .	Error tipo II: Aceptar H_0 , siendo H_1 cierta. La probabilidad de cometerlo es β .
Potencia: capacidad de rechazar H_0 , cuando es falsa. Capacidad de detectar H_1 . Si es insuficiente, podemos no detectar diferencias o asociaciones existentes. Se corrige aumentando el tamaño muestral.	



Un estudio con un error alfa asumido de un 5% (el valor por consenso habitual), supone un riesgo máximo de equivocarse de 5 entre 100, es decir, 1 entre 20, lo que se puede interpretar como una "garantía de un 5% de error"

REPASO

- El grado de significación p es la probabilidad calculada de que una diferencia o una asociación observada se haya generado por azar, si tal diferencia o asociación no existiese (MIR). No es una medida de la magnitud del efecto (no es mayor el tamaño de una diferencia o la fuerza de una asociación por ser menor el valor de p); aunque sí depende de la magnitud de la diferencia o asociación observadas (MIR) y del tamaño muestral (MIR).
- Calculamos p utilizando unos test de contraste de hipótesis (MIR) (siguiente capítulo). En función del tipo de variable y las características del estudio utilizaremos unos u otros.
- El error α , a diferencia de p , se establece a priori, antes de realizar el estudio. Generalmente es 0.05 (MIR) (ó 0.01).
- Si p , probabilidad de que todo sea por el azar, es mayor que α , riesgo que estamos dispuestos a asumir, decimos que no es estadísticamente significativo (MIR). Si es menor ($p < \alpha$), sí se considera estadísticamente significativo.
- La probabilidad ASUMIDA de rechazar H_0 siendo cierta es el error tipo I, mientras p es la probabilidad CALCULADA de que realmente se rechace H_0 siendo cierta.
- es el riesgo asumido de que, existiendo lo que buscamos (diferencia o asociación), interpretemos que no existe. También se establece a priori. Su complementario $(1-\beta)$ se denomina potencia del estudio, y es la probabilidad de ver lo que buscamos (MIR). Si es insuficiente deberemos aumentar el tamaño muestral (MIR).
- α y β no son complementarios ni errores contrapuestos: podemos disminuirlos aumentando el tamaño muestral.

El término ...	Quiere decir que ...
Se encuentra diferencia estadísticamente significativa en el porcentaje de curaciones para los fármacos A y B, para un grado de significación $p < 0,05$ (9MIR).	Si A y B fuesen iguales (H_0), la probabilidad de que las diferencias observadas en la comparación se hayan debido al azar es de $< 0,05$ ó $< 5\%$ (MIR) (tan pequeño que pensamos que debe haber "algún efecto" influenciando en el resultado). Y nos inclinamos por H_1 (A y B son diferentes).

repeMIR

Si H_0 fuese cierta, la probabilidad de que ocurrieran por azar las diferencias encontradas es menor que 0,05. La diferencia observada es significativa con un nivel de significación $p < 0,05$. (9+)

El término ...

No se encuentra diferencia estadísticamente significativa en el porcentaje de curaciones para los fármacos A y B, para un grado de significación $p > 0,05$ (MIR).

Quiere decir que ...

No se dispone de suficiente evidencia estadística a favor de la diferencia (3MIR) y si A y B fuesen iguales (H_0), la probabilidad de que las diferencias observadas en la comparación se hayan debido al azar es grande ($p > 0,05$ ó $p > 5\%$). Aceptamos H_0 (A y B son iguales).

repeMIR

Si aceptamos H_0 , concluimos que no hemos encontrado diferencia significativa entre la eficacia de los fármacos A y B, con un nivel de significación $p > 0,05$ (3+)

El término ...

Se encuentra diferencia estadísticamente significativa en el porcentaje de curaciones para los fármacos A y B, para un nivel de significación $p < 0,05$.

NO implica que ...

- A es mejor que B o viceversa.
- A y B son distintos.
- La probabilidad de que A sea mejor que B (o viceversa) es del 5% (o del 95%).
- A cura un 5% (ó 95%) más de pacientes que B (o viceversa).
- A y B son iguales.
- La diferencia observada es debida al azar.
- La probabilidad de que A sea mejor que B (o viceversa) es del 5% (o del 95%).
- A cura un 5% (ó 95%) más de pacientes que B (o viceversa).

No se encuentra diferencia estadísticamente significativa en el porcentaje de curaciones para los fármacos A y B, para un nivel de significación $p > 0,05$.

En todo caso en estadística generalmente no se demuestra nada (MIR): ni que todo haya sido por azar, ni que seguro se debe a diferencias entre grupos, ni nada que "nos suene" demasiado taxativo o afirmativo. Únicamente calculamos probabilidades. Y es con lo que nos manejamo.

MIR 05 (8227): Acabamos de finalizar un estudio cuyo objetivo era evaluar cual de los dos protocolos de actuación en nuestro centro hospitalario es mejor. Se ha obtenido un resultado no estadísticamente significativo ($p > 0,05$) en nuestro contraste de hipótesis. ¿Cuál de las siguientes respuestas es correcta?:

1. Debemos seguir explorando los datos dividiendo nuestra muestra en diferentes subgrupos para ver si en alguno de los grupos encontramos algún valor de p significativo.
2. Nos hemos de plantear si el tamaño de muestra es suficiente, si pensamos que no lo era continuaremos añadiendo más casos a nuestro estudio hasta obtener un resultado estadísticamente significativo.
3. Si el tamaño de muestra era insuficiente, utilizaremos la información del presente estudio para el diseño de uno nuevo revisando nuestras hipótesis de trabajo y la predefinición del tamaño de la muestra.*
4. Acabaremos concluyendo que los dos protocolos son iguales, al no haberse detectado diferencias significativas.
5. Nos hemos dado cuenta de la existencia de un sesgo que no habíamos tenido en cuenta ni en el diseño ni en la ejecución, por tanto, debemos utilizar un método estadístico más sofisticado que permita la corrección del mismo para seleccionar el protocolo mejor sin la menor duda.

C. TEST DE CONTRASTE DE HIPÓTESIS:

a. CONCEPTO:

Existen diferentes pruebas de contraste de hipótesis. En ellas, utilizando los datos obtenidos en el estudio se calcula el valor de un estadístico determinado (χ^2 , t) y con este valor (...unos cálculos y utilizando unas tablas...) se calcula p . Según nos dé un valor de p u otro valoraremos, junto al error α , si la significación estadística existe o no.

b. SITUACIÓN HIPOTÉTICA:

Imaginémonos que estamos realizando un estudio en el que comparamos el % de curados con el fármaco A y el % de curados con el fármaco B. Nos ha salido 60% para el A y 50% para el B. Repetimos el experimento: 50% para el A y 60% para el B. Más tarde nos enteramos en la bibliografía médica de que ambos fármacos A y B presentaban la misma eficacia. Es decir, hemos visto diferencias donde no las había. Una vez a favor de A, y otra a favor de B.

Conclusión: De nuevo esto se explica por la eterna variabilidad en el muestreo. Hemos elegido de la misma población 2 muestras (hicimos el experimento 2 veces), ambas representativas, y arrojan valores diferentes. ¿Cómo solucionar el problema?

c. FUNCIONES DE PROBABILIDAD:

Estas pruebas de contraste de hipótesis son, en realidad, funciones de probabilidad, como lo era la distribución Normal. Todas ellas se construyen con la distribución de los valores del estadístico correspondiente, bajo el supuesto de que la hipótesis nula (no existe asociación o diferencia) sea cierta. Es decir,

Son las curvas de probabilidad que se construirían a partir de los resultados de repetir el experimento infinitas veces, abarcando así el espacio muestral E de todos los resultados posibles generados por la variabilidad en el muestreo, siendo H_0 cierta (no asociación o diferencia).

Y así, no existiendo diferencia entre dos fármacos, podríamos obtener: a) muchas veces valores cercanos (50 y 55%, 40 y 45%, 60 y 55%) y b) raras veces valores aparentemente diferentes (40 y 80%, 30 y 70%, 85 y 20%), siendo en realidad iguales (pues así lo hemos establecido: H_0 cierta). Los valores más repetidos serán más probables en el contexto H_0 cierta (aceptaremos H_0) y los más raros menos probables (rechazaremos H_0 y aceptaremos H_1). Hablamos pues de probabilidad.

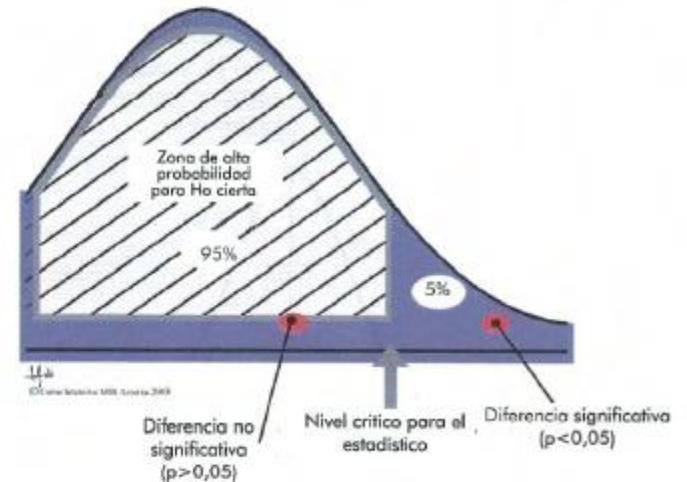
De este modo, el efecto observado en un estudio se puede considerar como el resultado de extraer un elemento concreto de la población o universo de todos los resultados posibles (de la distribución de probabilidad). Al situar sobre la curva de probabilidad el resultado de nuestra observación mediante el valor del estadístico que lo representa, valoraremos la probabilidad de presentación de ese resultado particular bajo el supuesto de hipótesis nula verdadera. Dicho de otro modo, calcularemos p , es decir, la probabilidad de que la diferencia o la asociación observada se deba simplemente al azar, si H_0 es cierta (no hay asociación o diferencia).

d. PUNTO DE CORTE α :

Podemos entonces marcar en la distribución el punto crítico, a partir del cual consideraremos cualquier valor del estadístico (p) con demasiadas probabilidades de ser cierto, es decir, que $p < \alpha$

probabilidad de que el efecto se deba al azar, sea cierto. Este punto de corte no podía ser otro que el error α .

Por ello, si calculamos el estadístico, cuyo valor representa a p , y cae en la zona de alta probabilidad de que realmente no haya asociación o diferencia (H_0), entonces diremos que no es un resultado estadísticamente significativo (MIR), y probablemente la asociación o diferencia detectadas pueden haberse producido por el azar. En caso contrario, existiría suficiente evidencia estadística para decir que es un resultado estadísticamente significativo.



Si el valor del estadístico cae en la zona de alta probabilidad, con H_0 cierta, significa que lo más probable es que los resultados del estudio se deban al azar, aceptando H_0 y rechazando H_1 . El caso contrario sucede si cae en la zona de baja probabilidad para H_0 cierta.

f. ELECCIÓN DEL ESTADÍSTICO ADECUADO (12MIR):

Las distintas pruebas disponibles se diferencian básicamente en: a) el tipo de variables y de contraste de hipótesis a que se pueden aplicar, y b) la distribución de probabilidad que utilizan. De estas propiedades y de las características de las distribuciones de datos resultantes depende la elección de la prueba correcta para un contraste en particular (será tema del siguiente capítulo).

No confundir ...	con ...
Grado de significación estadística: Probabilidad de que, bajo el supuesto de que H_0 es cierta, la diferencia o la asociación observada se deba al azar. Se denomina p . El valor de p se juzga según un nivel crítico establecido (típicamente 0,05 ó 0,01), cuyo valor coincide con el aceptado para α .	Potencia del test estadístico: Probabilidad complementaria de β . Potencia = $1 - \beta$. Es la probabilidad de que el test detecte la hipótesis alternativa siendo ésta correcta, o probabilidad de detectar una diferencia o una asociación si es que existe.

2.2. Significación estadística mediante intervalos de confianza

Hemos visto que una forma de resolver la significación de un efecto observado es mediante la aplicación de un test de contraste de hipótesis, gracias al que se calcula p o probabilidad de que el azar en el muestreo haya generado la diferencia o la asociación observada, si la hipótesis nula fuese verdadera.

Alternativamente, un contraste de hipótesis también se puede resolver utilizando intervalos de confianza para decidir si una diferencia o una asociación observada es significativa.

Al igual que para la media poblacional, recordemos que mediante intervalos de confianza se puede estimar el valor verdadero del tamaño de una diferencia o de una medida de fuerza de asociación (un riesgo relativo o una odds ratio). En consecuencia, se puede decidir si esa diferencia o asociación observada es significativa para un riesgo α determinado.

A. CONTRASTE DE HIPÓTESIS SOBRE UNA DIFERENCIA:

Encontrada una diferencia (entre la proporción de curados por A y por B, 50 y 60%), el paso siguiente es juzgar si esa diferencia es estadísticamente significativa. Utilizando intervalos de

MIR 05 (8231): Tenemos que evaluar los resultados de un ensayo clínico que compara un nuevo antihipertensivo respecto a otro considerado desde el punto de vista clínico como un buen estándar, y donde la reducción de la presión arterial diastólica (TAD) se predefinió como la variable principal. Suponemos que tanto el diseño como la ejecución del estudio son correctos. Los resultados indican que el nuevo tratamiento es más efectivo ya que reduce más la TAD, concretamente en media (intervalo confianza al 95% bilateral) reduce 0,5 (0,2 a 0,7) mmHg más que el grupo control, con $p=0,001$. ¿Cuál de las siguientes afirmaciones es correcta?:

1. El valor de p (0.001) encontrado demuestra con una probabilidad nula de equivocarnos que el nuevo tratamiento es mejor.
2. En base al valor observado de p (0.001) se puede concluir que la magnitud de la reducción de TAD del nuevo medicamento en relación al control es de gran relevancia clínica.
3. Si yo acepto que el tratamiento nuevo es el mejor, me equivocaré sólo con una probabilidad de 0,001.
4. La reducción de TAD es mayor con el nuevo tratamiento, pero la mejoría que en promedio ofrece en relación al tratamiento control no sobrepasaría 0,7 mmHg en mejor de los casos, teniendo en cuenta un error alfa o tipo I del 5% bilateral.*
5. La estimación puntual y los intervalos de confianza no aportan información de la magnitud del efecto de la comparación entre ambos tratamientos.

MIR 05 (8278): En un estudio sobre la prevención secundaria de accidentes cerebrovasculares con ácido acetil salicílico se encontró que la proporción de eventos en el grupo control fue 0,07 y en el grupo que recibió el antiagregante fue 0,09, lo que supuso una reducción de -29% del riesgo relativo. Los límites del intervalo de confianza al 95% de dicha reducción fueron -100% y +43%. ¿Cuál es la interpretación de este resultado?:

1. Es igualmente probable que el fármaco doble el riesgo, que lo aumente en un 29% o que lo reduzca en un 43%, por lo que el estudio no aclara si el fármaco es eficaz.
2. Se puede concluir que los pacientes que recibiesen el fármaco tendrían un riesgo de un 29% mayor.
3. El 95% de los estudios iguales a éste mostraría una reducción del riesgo relativo de -100% a +43%, aunque serían más frecuentes los valores más próximos a -29% que los más próximos a -100% o a +43%. Por tanto, el estudio no ha sido concluyente.*
4. Para poder interpretar estos resultados es imprescindible conocer el número de pacientes que se incluyeron en el estudio.
5. Dada la gran amplitud del intervalo de confianza, lo más probable es que en este estudio se haya cometido un error de tipo I.

MIR 10 (9468): Un artículo de una revista científica, en el que se estudian factores pronósticos del fracaso de una intervención terapéutica para la lumbalgia, informa que uno de estos factores es que la duración del episodio sea superior a un año con un "odds ratio" de 2,0 y un intervalo de confianza al 95% de 1,3 a 2,3. ¿Cuál de las siguientes afirmaciones es FALSA?:

1. La probabilidad de que el verdadero "odds ratio" esté comprendido 1,3 y 2,3 es 0,95.
2. Si se repitiera el estudio muchas veces, en un 95% de ellas se obtendría un "odds ratio" 1,3 y 2,3.
3. El intervalo de confianza está mal calculado, puesto que no está centrado en el "odds ratio".*
4. La probabilidad de que fracase el tratamiento en un individuo cuyo episodio ha durado más de un año, es aproximadamente el doble que si la duración hubiera sido menor.
5. Como el intervalo de confianza no incluye el 1, la asociación entre duración del episodio y fracaso del tratamiento es estadísticamente significativa.

MIR 11 (9704): Se ha realizado un ensayo controlado y aleatorizado en pacientes con hipercolesterolemia para valorar la efectividad de un nuevo fármaco. La variable final valorada es la presencia de infarto de miocardio. El Riesgo Relativo (RR) de infarto de miocardio en relación con el fármaco habitual es de 0,39 con un IC al 95% de 0,21 a 1,18. Ante este resultado usted concluiría que:

1. Debería recomendarse la utilización del nuevo fármaco.
2. El nuevo fármaco reduce el riesgo relativo en más de un 60%.
3. El tamaño muestral es muy elevado.
4. Existen diferencias estadísticamente significativas entre el efecto de los fármacos estudiados.
5. Los resultados revelan una precisión pobre.*

MIR 11 (9711): Se ha realizado un estudio de cohorte retrospectivo para conocer si los pacientes que toman antipsicóticos presentan un mayor riesgo de muerte súbita que la población que no utiliza antipsicóticos. Una vez realizado el ajuste por posibles factores de confusión se ha obtenido un riesgo relativo de 2,39 (intervalo de confianza al 95% de 1,77-3,22). ¿Cuál es la interpretación más correcta del resultado?:

1. El resultado es compatible con un incremento de riesgo asociado al uso de antipsicóticos, pero no es estadísticamente significativo.
2. El resultado sugiere que los antipsicóticos protegen frente al riesgo de muerte súbita.
3. El resultado no es interpretable porque no se ha hecho una asignación aleatoria de los tratamientos.
4. Hay un incremento de riesgo pero es pequeño e irrelevante desde un punto de vista clínico.
5. El resultado apoya la hipótesis de que el uso de antipsicóticos aumenta el riesgo de muerte súbita.*

MIR 12 (9939): En una población, el valor medio del colesterol total es de 216 mg/dL, con una desviación típica de 5 mg/dL. El porcentaje de personas cuyo nivel de colesterol es mayor de 226 mg/dL es, aproximadamente:

1. El 0.025%.
2. El 0.5%.
3. El 2.5%.*
4. El 5%.
5. El 10%.

Se describe una distribución normal, donde nos dan la media (216) y la desviación estándar (5), y nos piden que calculemos cuántos individuos hay por encima de un determinado valor. Como podemos comprobar, ese valor (226) es la media más 2 desviaciones estándar. Por las leyes de la distribución normal sabemos que entre la media MENOS dos desviaciones estándar y la media MÁS dos desviaciones estándar se encuentran el 95% de valores centrales, y por lo tanto por fuera queda el 5%. La mitad del este 5% de los valores de la distribución (es decir 2,5%) se halla por encima de la media más dos desviaciones estándar, lo que nos piden en esta pregunta.

MIR 13 (10204) (205): Se realiza un ensayo clínico en pacientes hipertensos para valorar la efectividad de un nuevo fármaco en la reducción de aparición de insuficiencia cardiaca. El Riesgo Relativo (RR) de insuficiencia cardiaca en relación con el fármaco habitual es de 0,69 con un IC al 95% de 0,31 a 1,17. ¿Que significan estos resultados?:

1. El nuevo fármaco disminuye el riesgo de insuficiencia cardiaca de forma significativa.
2. El nuevo fármaco es muy eficaz y debería comercializarse
3. La reducción del riesgo de insuficiencia cardiaca con el nuevo fármaco es irrelevante clínicamente.
4. No existen diferencias estadísticamente significativas entre el efecto de los fármacos estudiados.*
5. El nuevo fármaco aumenta el riesgo de insuficiencia cardiaca pero de forma no significativa.



RESUMEN DE INFERENCIA ESTADÍSTICA

1. INTRODUCCIÓN

- La Estadística Inferencial intenta extraer conclusiones a partir de los datos obtenidos en una muestra. Dos son las situaciones más frecuentes:

A. ESTIMACIÓN DE PARÁMETROS POBLACIONALES:

- Se pretende estimar un parámetro poblacional, generalmente la media poblacional (ej media de colesterolemia), a partir de los datos de una única muestra representativa, ofreciendo un intervalo de confianza de valores de la variable que incluya con una alta probabilidad el verdadero valor de dicha media poblacional.

B. VALORACIÓN DE SIGNIFICACIÓN ESTADÍSTICA:

- Cuando se valora la significación estadística partimos de un valor calculado (digamos la media poblacional), estimada a partir de los datos muestrales, y lo que se pretende es valorar el significado real de ese valor: si la estadística dice que es significativo o no, es decir, si hay "razones matemáticas" para pensar que no ha sido fruto del azar o el azar puede explicar el valor calculado. Un resultado será significativo cuando la probabilidad de que el azar haya generado dichos resultado sea muy pequeña.

2. ESTIMACIÓN DE PARÁMETROS POBLACIONALES

A. LA NECESIDAD DE ESTIMAR:

- Si de una distribución poblacional, con media poblacional μ , extrájeramos correctamente infinitas muestras de igual tamaño, todas ellas representativas, obtendríamos infinitas medias muestrales cercanas unas a otras pero sin duda ligeramente diferentes. ¿Cuál de todos ellos es el verdadero valor de la media poblacional μ ? Intentaremos dar a esta pregunta una respuesta probabilística: calcularemos un intervalo de confianza (IC), es decir, un intervalo de valores de la variable (colesterolemia) que contenga con una alta probabilidad el verdadero valor de μ .

B. ERROR α Y NIVEL DE CONFIANZA:

- El investigador debe establecer antes de calcular dicho intervalo:
 - El error α que está dispuesto a asumir: la probabilidad que asume de que el intervalo que se construya no incluya el verdadero valor poblacional.
 - El nivel de confianza "1 - α " (complementario de α): la probabilidad de que el intervalo que se construya incluya el verdadero valor poblacional. Suelen utilizarse intervalos de confianza con niveles de confianza del 95%, o lo que es lo mismo, con error α de 0.05 (5%).

C. DISTRIBUCIÓN DE MEDIAS MUESTRALES (UNA DISTRIBUCIÓN NORMAL):

- Si se muestrea infinitamente una distribución poblacional y se calculan sendas medias de las infinitas muestras resultantes, se podría construir una nueva distribución con las medias de todas las muestras; esta nueva distribución se denomina distribución de medias muestrales. Su media, es la media de todas las medias muestrales, y por lo tanto coincide con la media poblacional (si el tamaño muestral es lo suficientemente grande). La desviación estándar de la distribución de medias muestrales se conoce como error estándar de la media, cuyo cálculo depende del tamaño de la muestra obtenida n y de su desviación estándar S , tal como sigue:

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

- Así, a mayor tamaño muestral, menor error estándar de la media (medida éste del error aleatorio y variabilidad en el muestreo), por lo que se cumple que a mayor tamaño muestral, mayor precisión en la estimación de los resultados (2MIR).
- Aunque la distribución poblacional considerada no se ajustara a una distribución Normal, el teorema del límite central de la media, nos garantiza que, si extraemos muestras de tamaño suficientemente grande y calculamos las medias de dichas muestras, la distribución de las medias muestrales SI será una distribución Normal, pudiendo ser aplicadas las propiedades de dicha distribución.

D. CÁLCULO DEL INTERVALO DE CONFIANZA (A PARTIR DE LA MEDIA MUESTRAL):

- Utilizando la media (de una única muestra representativa cualquiera) y el error estándar de la media (de la distribución de medias muestrales), el cual puede ser calculado mediante la fórmula anterior, podemos obtener los siguientes intervalos de confianza para estimar la media poblacional, aplicando las propiedades de la distribución Normal:

$$[\bar{X}_m \pm S_{\bar{x}}] \Rightarrow \text{Intervalo de confianza del 68\%}$$

$$[\bar{X}_m \pm 2S_{\bar{x}}] \Rightarrow \text{Intervalo de confianza del 95\% (6MIR)}$$

$$[\bar{X}_m \pm 3S_{\bar{x}}] \Rightarrow \text{Intervalo de confianza del 99\%}$$

SI EN UNA PREGUNTA MIR...		
Nos hablan de...	Se refieren a....	Y buscaremos la opción del tipo...
Desviación estándar (o típica)	Medida de dispersión de una distribución Normal cualquiera (muestral, poblacional...)	El 95% de los valores centrales de... están entre... y
Error estándar	Medida de dispersión de la distribución de las medias muestrales.	Existe un 95% de confianza de que entre... y... esté el verdadero valor de... .

3. VALORACIÓN DE SIGNIFICACIÓN ESTADÍSTICA

- La valoración de significación estadística se puede realizar de dos maneras: a) calculando el grado de significación p o b) calculando un intervalo de confianza. Dos son las situaciones en las que lo realizaremos: a) comparación de proporciones (viendo diferencias entre porcentajes de curados con dos fármacos) y b) asociaciones entre variables (viendo asociaciones entre factores de riesgo y enfermedades).

3.1. MEDIANTE EL GRADO DE SIGNIFICACIÓN P

- Es la probabilidad calculada de que una diferencia o asociación observada en un estudio sean debidas al azar, sin que existan realmente (MIR). Si $p=0.06$, significa que la probabilidad de que la diferencia o asociación objetivadas en el estudio no existan y tales resultados se expliquen por azar (MIR) (variabilidad).
- El grado de significación p es un valor calculado a partir de los datos del estudio, depende del tamaño muestral y de la magnitud del efecto observado pero no es una medida de tal magnitud (MIR). Para su calculo se utilizan pruebas de contraste de hipótesis.

A. CONTRASTE DE HIPÓTESIS.

- Su metodología sufre tres etapas: 1ª) **Se establece el valor de la hipótesis nula H_0** , que se toma como cierta "de entrada"; 2ª) **Se calcula p** (pruebas de contraste de hipótesis) a partir de los datos del estudio; y 3ª) **Valoramos si existe significación estadística** (p bajo, menor que el error α asumido a priori) o no existe (p alto, mayor que el error α).

B. DEFINICIONES.

- **Hipótesis nula H_0** : proposición cierta de entrada, cuya veracidad se somete a examen. Habitualmente "no hay diferencia" o "no hay asociación".
- **Hipótesis alternativa H_1** : proposición que se acepta si H_0 es rechazada (por ser muy baja la probabilidad de ser cierta).
- **Error tipo α o tipo I**: probabilidad asumida a priori de que, siendo cierta H_0 (no hay asociación o diferencia), la rechazemos (3MIR). Se toma 0.05 (MIR) por defecto.
- **Error tipo β o tipo II**: probabilidad asumida a priori de que, siendo falsa H_0 , la aceptemos y no la rechazemos (MIR). α y β no son complementarios ni errores contrapuestos, y así se pueden disminuir ambos aumentando el tamaño muestral (MIR).
- **Potencia (complementario de β)**: probabilidad de detectar como significativa una diferencia o asociación cuando existen (2MIR) (probabilidad de detectar hipótesis alternativas (MIR)). Si resulta insuficiente (30%), se puede aumentar la potencia aumentando el tamaño muestral (4MIR). Sin embargo, una vez alcanzado un resultado concluyente no importa lo probable o improbable que fuese el llegar a él antes de hacer el estudio (mayor o menor potencia) (2MIR).
- Pueden producirse falsos negativos por: un error beta demasiado alto (MIR), una potencia insuficiente (MIR), generalmente por un tamaño insuficiente (MIR), existencia de sesgos en el estudio (MIR)...

C. INTERPRETACIÓN DE LOS RESULTADOS.

- Si el resultado en valorar las diferencias de eficacia entre el fármaco A y B es estadísticamente significativo significa que, si no hubiera diferencias entre ellos, si la eficacia fuera igual para ambos, la probabilidad de encontrar ese resultado sería del valor de p (menor que 5%, por ser significativo) (9MIR). Y si no fuera estadísticamente significativo significaría que no habríamos encontrado suficiente evidencia estadística a favor de la diferencia (3MIR). En todo caso en estadística generalmente no se demuestra nada (MIR).

3.2. MEDIANTE INTERVALOS DE CONFIANZA.

- Mediante un intervalo de confianza también se puede valorar la significación estadística:
- Si el valor asignado a H_0 cae dentro del intervalo (de alta probabilidad de ser cierto), se considera que el resultado **no es significativo**.
- Si el valor asignado a H_0 cae fuera del intervalo, se considera que el resultado es **significativo**, ya que la probabilidad de ser cierta (no asociación o diferencia) es pequeña.
- El valor asignado a H_0 varía en función de que se trate de una diferencia o una asociación:

A. DIFERENCIA

- Si el intervalo de confianza para la diferencia excluye, con un 95% de probabilidad, el valor establecido para la hipótesis nula (excluye el valor 0) la diferencia será estadísticamente significativa (MIR). Si el intervalo de confianza para la diferencia incluye con un 95% de probabilidad el valor establecido para la hipótesis nula (incluye el valor 0) la diferencia no será estadísticamente significativa (MIR).

B. ASOCIACIÓN

- Al valorar la asociación entre dos variables pueden utilizarse medidas de fuerza de asociación distintas, a saber, riesgo relativo (RR), *odds ratio* (OR) o coeficiente de correlación de Pearson. Como veremos cuando el valor del RR u OR es igual a 1, se considera que las variables estudiadas no están asociadas, son independientes. Si son distintas de 1, estarán asociadas.
- Si el intervalo de confianza para el RR u OR excluye con un 95% de probabilidad el valor establecido para la hipótesis nula (excluye el valor 1), se concluye que existe asociación estadísticamente significativa.
- Si el intervalo de confianza para el RR u OR incluye con un 95% de probabilidad el valor establecido para la hipótesis nula (incluye el valor 1), se concluye que no existe asociación estadísticamente significativa (3MIR). No se habría obtenido suficiente evidencia para descartar que la causalidad sea la responsable de la relación "observada inicialmente" entre las variables (MIR).

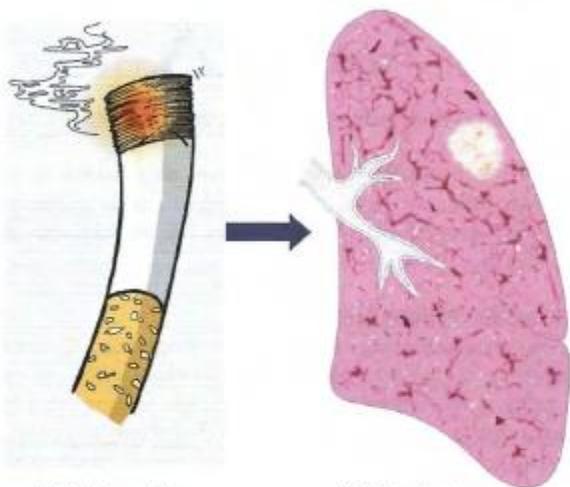
1. Criterios de elección

Sabemos del capítulo previo que, para el hallar el valor de p , se calculan los valores que toma un determinado estadístico. Existen diferentes pruebas de contraste de hipótesis y debemos elegir la correcta en cada caso en función de los siguientes criterios:

A. TIPO DE VARIABLE: cualitativas (nominal, ordinal) o cuantitativa (discreta, continua).

Se distinguen tres grandes grupos de pruebas según se usen en:

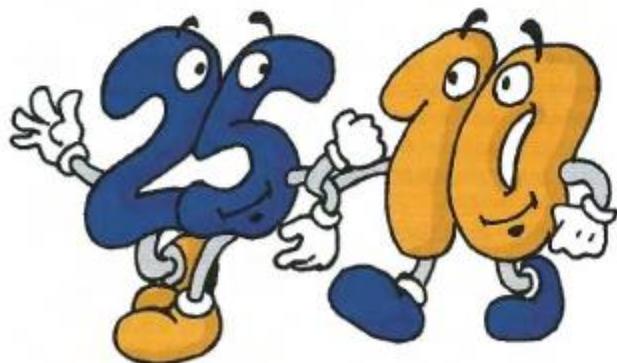
- a) **comparación de proporciones** (curados y no curados en A y en B: cualitativas nominales)
- b) **comparación de medias** valorando si existe diferencia significativa (niveles de colesterolemia en diabéticos y en no diabéticos: cuantitativa continua de desenlace y cualitativa nominal la predictor) a)
- c) **asociaciones entre variables** valorando si existe tal asociación o no (RR, OR y coeficiente de correlación entre factor causal y enfermedad: número de cigarrillos e incidencia de cáncer de pulmón; ambas, factor de riesgo y enfermedad, cuantitativas continuas).



Variable predictor Variable desenlace
En el caso de "el tabaco causa cáncer de pulmón", el tabaco es la variable predictor y el cáncer de pulmón la variable de desenlace.

B. TIPO DE RECOGIDA DE DATOS EN EL ESTUDIO:

- a) **datos apareados:** en un mismo grupo se mide una variable (tensión arterial, por ejemplo) antes y después de una medida terapéutica a estudio (antihipertensivos), y vemos si ha habido cambios significativos. En datos apareados se estudian las mismas personas en dos tiempos diferentes.



Datos apareados: se toman valores antes y después de la medida de intervención.

- b) **datos independientes:** se mide una variable en grupos diferentes, habiéndose administrado diferentes medidas en cada uno: a un grupo, un fármaco, y al otro, placebo u otro medicamento. Comparamos resultados y vemos si hay diferencias significativas. Aquí estudiamos personas diferentes "al mismo tiempo".



© Curso Intensivo MIR Asturias

Datos independientes: se toman medidas de una variable en los diferentes grupos después de una intervención.

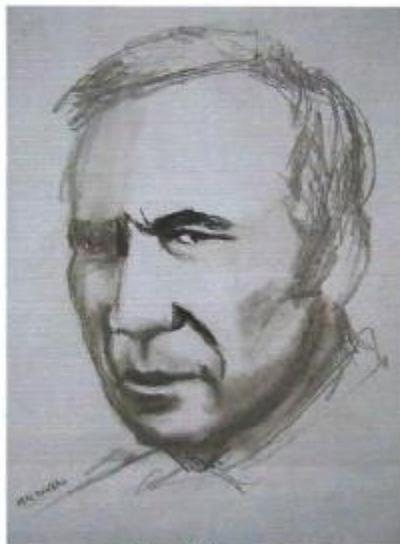
C. NÚMERO DE DISTRIBUCIONES DEL ESTUDIO:

- a) **dos grupos** (colesterolemia en grupos A y B tras un determinado tratamiento)
- b) **más de 2 grupos:** sea el estudio de una única variable (colesterolemia en grupos A, B y C tras tratamiento) o de más de 2 variables a estudio (tabaco, colesterol y sedentarismo en grupos de cardiopatas y sanos)

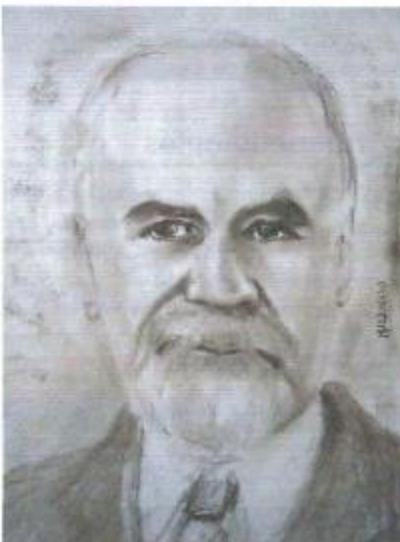
D. ¿DISTRIBUCIÓN PARAMÉTRICA?

Si el grupo de datos que tenemos (por ejemplo, 50 mediciones de colesterol sérico) sigue una distribución normal, podremos usar pruebas paramétricas (que se basan en los parámetros media y desviación estándar, y son más potentes) para los cálculos estadísticos. En caso contrario tendremos que usar pruebas no paramétricas.

Para saber si una muestra de datos sigue una distribución normal, se usa la prueba de Kolmogorov-Smirnov. También se usa el test D'Agostino la prueba de Chi cuadrado o el test de Saphiro-Wilk.



Andrei Kolmogorov



Smirnov (Es cierto, tiene nombre de vodka...)

MIR 08 (8993): Para el empleo de los métodos paramétricos en el análisis estadístico de los datos, la distribución de la variable dependiente debe ser:

1. Emparejada por la edad.
2. Dicotómica.
3. Lineal.
4. Nominal.
5. Normal.*

2. Comparación de proporciones

Ejemplo típico: valoramos las diferencias entre los porcentajes (proporciones o frecuencias) de curados y no curados con dos fármacos distintos o un fármaco y un placebo. Ojo, ya que aunque hablemos de % realmente es una variable nominal de tipo dicotómico: en A habrá curados y no curados (2 opciones), al igual que en B.

Utilizaremos las siguientes pruebas según sean datos apareados o independientes.

2.1. Datos apareados

Para comparar proporciones en dos categorías de datos apareados usaremos el test de McNemar.

Ejemplo: un grupo de pacientes antes y después de un tratamiento en que se mide una variable desenlace nominal con dos categorías (curados-no curados, HTA-no HTA).



McNemar

Pero si queremos comparar un grupo antes y después de un tratamiento (datos apareados), donde la variable desenlace es cualitativa nominal con más de dos categorías (por ejemplo, sufrir efectos secundarios gástricos, neurológicos, cutáneos o visuales), usaremos el test **Q de Cochran**.

2.2. Datos independientes

Situación: dos grupos con una enfermedad en los que se valora la eficacia de dos tratamientos distintos, uno en cada grupo.

A. Z DE COMPARACIÓN DE PROPORCIONES:

Sólo podemos utilizarla en caso de un contraste de hipótesis con dos proporciones (éxitos y fracasos -variable desenlace- en tratados con A y B -variable predictora). Es decir, limitado a tablas de contingencia de 2x2 (véase tabla).

	Éxitos	Fracasos	Suma
A	A	b	a+b
B	C	d	c+d
Suma	a+c	b+d	n

B. PRUEBA DE JI AL CUADRADO (CONOCIDA COMO CHI-CUADRADO):

Es la más frecuentemente utilizada y se puede aplicar para la comparación de dos o más proporciones (2MIR).

MIR 09 (9236): Para estudiar la posible asociación, entre la rubéola materna y las cataratas congénitas, se selecciona una muestra de 20 niños con esta enfermedad y 25 niños con antecedentes y edad semejantes que no la presentan. Una entrevista a la madre de cada niño determina si tuvo o no la rubéola durante el embarazo. ¿Qué test estadístico es el más adecuado para realizar este estudio?:

1. T-student de datos independientes.
2. T-student de datos apareados.
3. Ji-cuadrado.*
4. Análisis de correlación.
5. Análisis de la varianza.

a. TIPO DE VARIABLE PREDICTORA:

Es aplicable al contraste de variables predictoras cualitativas (MIR) (nominales u ordinales), cuantitativas discretas o cuantitativas continuas distribuidas en intervalos de clase.

$$\chi_0 \rightarrow \chi^2$$

La chi-cuadrado (χ^2) se aplica en la comparación de proporciones (%).

Se usa para la comparación de variables distribuidas en clases excluyentes (MIR), es decir, incompatibles: no se puede pertenecer a dos grupos al mismo tiempo (el tratado con A, no tratado con B).

b. APLICACIONES FRECUENTES:

- Test de asociación con tablas de contingencia.** Consiste en comprobar si dos características cualitativas están relacionadas entre sí. Ej anterior: ver si la variable fármaco A o B está relacionada o asociada con la variable curado o no curado.
- Test de homogeneidad de varias muestras cualitativas.** Se trata de comprobar si varias muestras de carácter cualitativo proceden de igual población, con distribución homogénea de sus variables. Ej.: contrastar si tres muestras determinadas de sujetos proceden de poblaciones con igual distribución de grupos sanguíneos.
- Test de bondad de ajuste a distribuciones.** Consiste en valorar el grado de ajuste en que los datos observados siguen una determinada distribución teórica. Ej: la variable estatura sigue una distribución ajustable a la distribución normal. Valorar este grado o bondad de ajuste.

c. CARACTERÍSTICAS DE SU CÁLCULO:

Inicialmente, para el cálculo del estadístico χ^2 , los datos se recogen en forma de tablas de frecuencias (llamadas tablas de contingencia), de modo que, a partir de los valores marginales (porque están en los márgenes de la tabla, sumatorios de las filas y columnas), se determina el número de casos que se esperaría que hubiera en cada casilla, si no existieran diferencias entre los grupos.

Seguidamente, se calcula el valor del estadístico chi-cuadrado, basándose en valoración de las diferencias que existen entre las frecuencias esperadas y las frecuencias realmente observadas (MIR). De hecho, se basa en contrastar los cuadrados de las "frecuencias esperadas" y las "frecuencias observadas". El valor del estadístico representa mayores diferencias cuanto más elevado sea.

El número de grados de libertad, expresión de la capacidad de variación de las categorías, se calcula multiplicando (filas-1) x (columnas-1) (MIR).

Una vez que calculamos los grados de libertad de la tabla y que calculamos χ^2 tenemos que comparar el valor obtenido con el valor que nos da la tabla para esos grados de libertad. Si es mayor, la prueba es significativa (y consideramos que existen diferencias significativas entre los grupos, es decir, rechazamos H_0).



d. CONDICIONANTES DE VALIDEZ:

- a) Ninguna frecuencia esperada debe ser menor que 1 (ó dos según autores)
- b) No más del 20% de ellas inferior o igual a 5 (MIR).

e. SOLUCIONES:

Si no se cumplieran los anteriores requisitos podríamos optar por:

- a) Aumentar el tamaño de la muestra (no siempre posible).
- b) Agrupar celdas, filas y/o columnas, si la estructura de la tabla de contingencia y la pérdida de información que esto ocasiona lo permiten.
- c) Utilizar la corrección por discontinuidad de Yates (MIR). Se trata de cierta transformación que se realiza en el cálculo del valor del estadístico, sobre cuyo uso existe cierta controversia: puede utilizarse como ajuste cuando se incumple la condición de los valores mínimos que pueden alcanzar las frecuencias esperadas, pero algunos autores defienden que debe aplicarse siempre (MIR), y otros aceptan que es conveniente cuando el número de clases es menor o igual que 4.



Frank Yates

- d) Aplicar la PRUEBA EXACTA DE FISHER (sólo tablas de 2x2) cuando no es aplicable χ^2 por manejar en el estudio tamaños de muestra muy pequeños (MIR).



Ronald A. Fisher

Tablas de contingencia de 2x2

Z de comparación de proporciones y de medias: sólo útiles en contrastes de dos proporciones o medias para dos opciones terapéuticas (2x2).

Z de comparación de proporciones y de medias: sólo útiles en contrastes de dos proporciones o medias para dos opciones terapéuticas (2x2).

COMPARACIÓN DE PROPORCIONES

Datos apareados

MCNEMAR

Datos independientes

Z COMPARACIÓN DE PROPORCIONES:

- Limitado a tablas de contingencia de 2x2.

CHI-CUADRADO:

- Variables de desenlace nominales distribuidas en clases excluyentes (2 categorías: curado/ no curado).
- Valora diferencias entre frecuencias esperadas y observadas, siendo los grados de libertad: (filas-1) por (columnas-1).
- Condicionantes de validez: a) ninguna frecuencia esperada debe ser menor de 1, y b) no más de 20% de ellas inferior o igual a 5.
- Soluciones: a) aumentar tamaño muestral o agrupar celdas; b) corrección con discontinuidad de Yates (siempre?); y c) prueba exacta de Fisher (efectivos esperados reducidos; limitado a tablas 2x2).

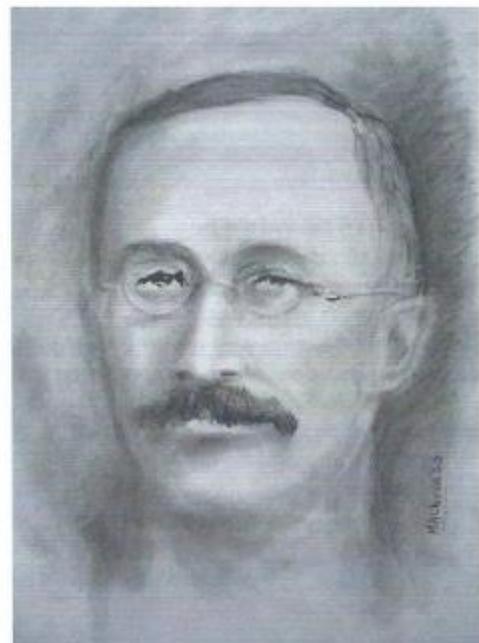
3. Comparación de medias

Valoraremos dos situaciones: comparar medias de dos grupos o comparar medias de más de dos grupos. Y en cada una de estas situaciones veremos qué pruebas utilizamos en datos apareados y en datos independientes. La media siempre la referimos a variables cuantitativas continuas (2MIR).

3.1. Medias de dos grupos

Situación: queremos comparar el efecto de dos fármacos sobre las cifras de presión arterial diastólica (variable desenlace cuantitativa continua) de dos grupos tratados cada uno con un fármaco (variable predictora nominal).

Aunque existe una prueba z de comparación de medias limitada al estudio de tablas de contingencia de 2x2, la prueba más utilizada en este caso es la t de Student-Fisher. Su cálculo y condiciones de validez varían en función de si son datos apareados o independientes.



William Gosset fue el inventor de la "t de student". Firmó sus trabajos como "student" (estudiante), de ahí el nombre de la prueba.

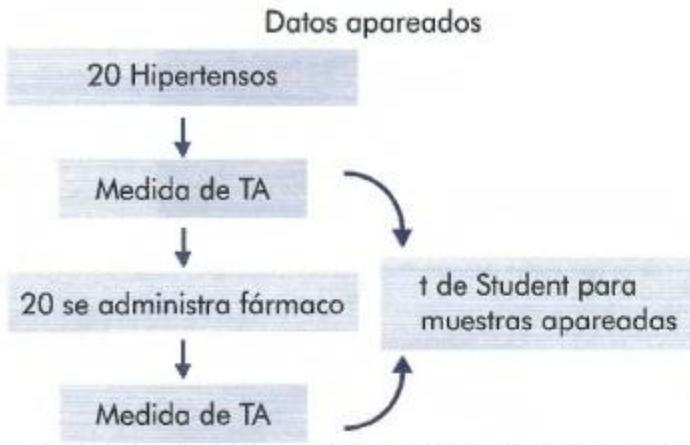
A. PRUEBA T DE STUDENT-FISHER PARA DATOS INDEPENDIENTES (2MIR):

a. TIPO DE VARIABLE:

Únicamente variables de desenlace cuantitativas continuas (colesterolemia, tensión arterial, estatura, peso, uricemia...). Utilizamos la prueba t de Student para datos independientes cuando se realiza una medida de una variable cuantitativa continua en dos grupos a los que se les ha aplicado dos medidas terapéuticas distintas.

Si no podemos...	Utilizaremos...
T de Student por ser varianzas distintas	Test de Welch: si datos independientes
χ^2 por efectivos insuficientes	Prueba exacta de Fisher: si tablas de 2x2

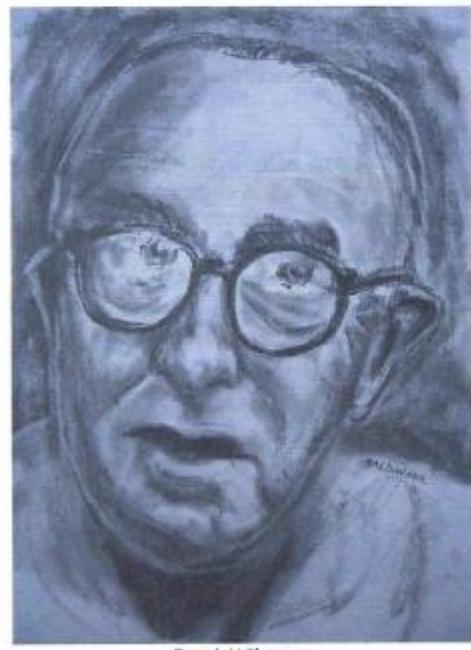
B. PRUEBA T DE STUDENT-FISHER PARA DATOS APAREADOS (2MIR):



La t de Student para datos apareados compara los datos de antes y después de la intervención.

- a. **CONDICIONANTE DE VALIDEZ:**
- a) Las diferencias en este caso deben seguir una distribución normal, si bien esta limitación es menos importante si se estudian más de 20 sujetos.
- b. **SOLUCIONES:**

Cuando la t de Student para datos apareados no es aplicable por incumplir este condicionante se puede recurrir a la **prueba de Wilcoxon o prueba de los rangos con signo** (pruebas paramétricas). También utilizada en variables de desenlace ordinales.



Frank Wilcoxon

MIR 07 (8732): En un estudio clínico, a una serie de pacientes se les trata con un nuevo fármaco para estudiar, si en un periodo de tiempo después de la administración de dicho fármaco, el nivel de bilirrubina ha disminuido. Se acepta que la distribución de la bilirrubina es normal, para este diseño. ¿Cuál es la prueba estadística de elección?:

1. T de Student para datos apareados.*
2. T de Student para datos independientes.
3. Ji-cuadrado.
4. Mann-Whitney.
5. Prueba exacta de Fisher.

3.2. Medias de más de dos grupos

Habíamos adelantado ya que, en los llamados contrastes de hipótesis múltiples, se puede comparar:

- a) medias de una única variable en varios (3 ó más) grupos distintos: utilizaremos el análisis de la varianza o ANOVA.
- b) medias de más de una variable (≥ 2) en dos o más grupos distintos: utilizaremos el llamado análisis multivariante.

En realidad esto es lo más habitual en el curso de un proyecto de investigación. Normalmente se necesita establecer comparaciones múltiples (más correcto sería decir hipótesis múltiples). Esto se refiere a comparar más de dos grupos para una determinada variable (ejemplo, comparar el nivel de uricemia entre tres grupos que han recibido tratamientos diferentes), o bien comparar hipótesis para distintas variables en dos o más grupos (relaciones entre tabaco, sedentarismo, colesterolemia en cardiopatas y sanos).

A. PRUEBAS DE COMPARACIONES MÚLTIPLES O PRUEBAS A POSTERIORI:

Al realizar una comparación simple entre dos hipótesis, el riesgo α asumido de cometer un error tipo I es la probabilidad de identificar como significativa una diferencia generada por el azar en el muestreo. Es decir, inclinarnos por la hipótesis alternativa siendo verdadera la hipótesis nula. El valor de esta probabilidad coincide con el nivel de significación estadística, o "punto crítico de corte" con el que juzgamos como significativo o no el resultado del estudio, según el valor de p obtenido.

Al realizar test de contraste de hipótesis múltiples, el riesgo de identificar como significativa una diferencia o una asociación generada por azar **AUMENTA**, por lo que el estudio no será concluyente (4MIR). De forma que si se realizan 100 comparaciones estableciendo el nivel de significación nominal (en cada una de ellas) de 0,05, 5 de las 100 comparaciones pueden resultar significativas por azar. Por lo tanto, una vez realizado el contraste de hipótesis múltiple debemos valorar adecuadamente sus resultados con el uso de técnicas correctoras: pruebas a posteriori o pruebas de comparaciones múltiples.

MIR 00 (6921): Se trató con un suplemento dietético más dieta a 15 insuficientes renales y solamente con dieta a 16. Se compararon entre ambos grupos unas 100 variables y en 6 de tales comparaciones se encontraron diferencias significativas, al nivel nominal del 5%, a favor del suplemento. ¿Cómo interpretaría estos resultados?:

1. El estudio parece no concluyente, ya que con tantas comparaciones es esperable que unas pocas resulten significativas al nivel nominal del 5%, aunque no haya diferencia entre los tratamientos.*
2. El suplemento dietético es moderadamente eficaz.
3. Este estudio no demuestra nada, ya que no es doble ciego.
4. Aunque el número de pacientes es pequeño, lo inocuo del tratamiento y la consistencia de los resultados aconsejan prescribir el suplemento en la insuficiencia renal.
5. Este estudio no demuestra nada, ya que hay más variables que pacientes.

B. ANÁLISIS DE LA VARIANZA (=ANOVA) (9MIR):

a. **TIPO DE VARIABLE:**
El análisis de la varianza (ANOVA) nos sirve para comparar medias de más de dos grupos (8MIR) para una variable cuantitativa continua. Ejemplo: comparar la uricemia en tres grupos de sujetos que han sido tratados con respectivos fármacos diferentes.

repeMIR

Para comparar las medias en más de dos grupos para una variable cuantitativa continua se usa el análisis de la varianza (ANOVA). (7+)

MIR 04 (7972): En un ensayo clínico se comparan 3 tratamientos (p.e. placebo, tratamiento establecido y un tratamiento nuevo). La variable respuesta es continua (p.e. nivel de glucosa en sangre). Aceptando que la variable tiene una distribución normal, el test correcto para comparar la respuesta es:

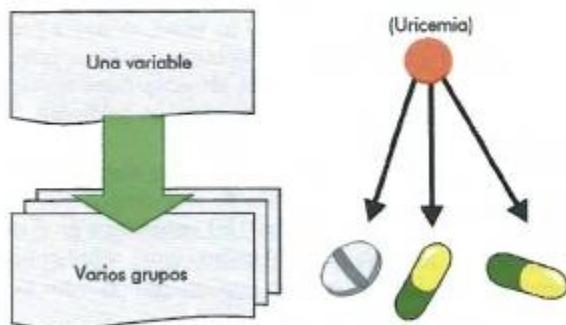
1. La t de Student.
2. El test de Wilcoxon.
3. Análisis de la varianza.*
4. El test de Kruskal-Wallis.
5. El test χ^2 -cuadrado.

MIR 06 (8489): Se ha realizado un ensayo clínico encaminado a valorar el descenso de la presión portal obtenido después de la administración aguda de tres alternativas de tratamiento (carvedilol, propranolol y placebo) en pacientes cirróticos con varices esofágicas. La variable de respuesta es el gradiente de presión venosa hepática que es continua y de distribución normal. ¿Cuál de las siguientes pruebas estadísticas es la correcta para comparar la respuesta?:

1. Una prueba de la t de Student.
2. Una prueba de Wilcoxon.
3. Una prueba de Fisher.
4. Un análisis de la varianza.*
5. Una prueba U de Mann-Whitney.

b. FUNDAMENTO:

El fundamento del análisis de la varianza consiste en estudiar la **variabilidad de las observaciones intragrupalmente** (dispersión entre los datos respecto de la media EN MISMO grupo) e **intergrupalmente** (dispersión de las medias ENTRE DISTINTOS grupos), utilizando la prueba de la F de Snedecor (compara variabilidad).



El ANOVA se utiliza para comparar medias de una variable de tres o más grupos.

c. CONDICIONES DE VALIDEZ:

- Los valores deben seguir o ajustarse a una distribución normal en cada grupo.
- Varianzas iguales u homocedasticidad. NOTA: El que las varianzas deban ser iguales no impide que la variabilidad inter e intragrupalmente puedan ser diferentes, siendo precisamente lo que valora esta prueba utilizando la F de Snedecor.

d. TIPOS:

- Análisis de la varianza de una vía: para datos independientes. [El clásico. Pensad en éste si no se dice nada]
- Análisis de la varianza de dos vías o análisis de la varianza para medidas repetidas: para datos apareados ("antes y después" de una intervención terapéutica, para ver su eficacia).

e. SOLUCIONES:

Cuando la ANOVA no es aplicable por incumplir sus condiciones se puede recurrir a:

- La prueba de Kruskal-Wallis si los datos son independientes
- El test de Friedman si los datos son apareados (recuerda, Morgan Freeman está casado, o sea, apareado...)



El test de Friedman se usa cuando no podemos aplicar ANOVA para datos apareados (Morgan Freeman está casado, es decir, emparejado...)



Milton Friedman, premio Nobel de Economía e inventor del test que lleva su nombre

f. INTERPRETACIÓN:

El análisis de la varianza es una prueba global que determina si existe alguna diferencia entre el conjunto de las medias comparadas. Si se obtiene un resultado estadísticamente significativo a favor de la diferencia, se concluye que no todas las medias son iguales, pero no define cuál o cuáles originan la diferencia significativa (MIR). Se requerirá utilizar posteriormente algún método de comparaciones múltiples o pruebas a posteriori (MIR).



©Curso Intensivo MIR Asturias

El Análisis de la Varianza nos indica de dónde si existe o no diferencias entre las medias de 3 ó más grupos, pero no de qué grupo viene esa diferencia.

MIR 05 (8212): En un ensayo clínico se comparan 3 tratamientos (p.e. placebo, tratamiento establecido y un tratamiento nuevo). La variable respuesta es continua (p.e. nivel de glucosa en sangre). ¿Si la variable no tiene una distribución normal, el test correcto para comparar la respuesta es?:

1. La t de Student.
2. El test de Wilcoxon.
3. Análisis de la varianza.
4. El test de Kruskal-Wallis.*
5. El test ji-cuadrado.

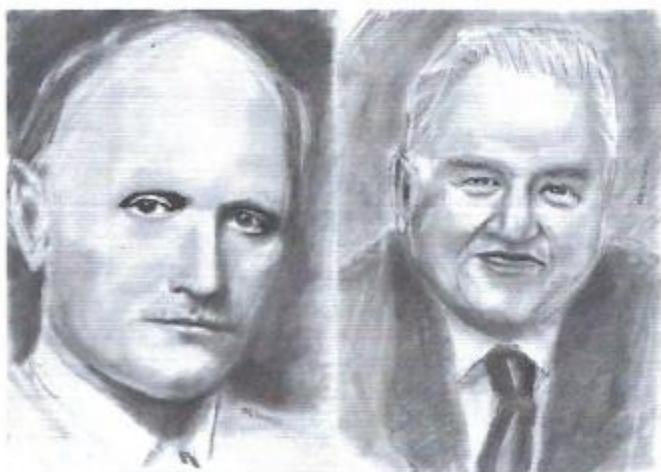
g. MÉTODOS DE COMPARACIONES MÚLTIPLES:

Ejemplos: corrección de Bonferroni, pruebas de Scheffé, de la diferencia significativa franca de Tukey, o de la mínima diferencia significativa de Fisher.

Como dijimos antes, lo que hacen estas métodos de comparaciones múltiples es reducir el riesgo α para el conjunto o suma de las comparaciones que se realicen; en caso contrario, en estudios de muchas comparaciones individuales el sumatorio de todos los "0.05"s sería demasiado alto, y por ello, muy probable encontrar una diferencia aunque realmente no existiera (error tipo I: cierta H_0).

Pruebas a posteriori...

- Disminuyen el error tipo I
- Nos dicen de dónde viene la diferencia



Claudio Bonferroni

John Tukey



Henry Scheffé, autor de uno de los tests de comparaciones múltiples

MIR 00 (6921): Se trató con un suplemento dietético más dieta a 15 insuficientes renales y solamente con dieta a 16. Se compararon entre ambos grupos unas 100 variables y en 6 de tales comparaciones se encontraron diferencias significativas, al nivel nominal del 5%, a favor del suplemento. ¿Cómo interpretaría estos resultados?:

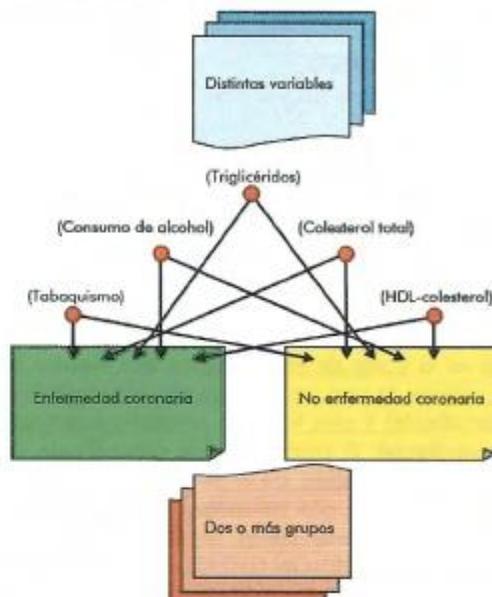
1. El estudio parece no concluyente, ya que con tantas comparaciones es esperable que unas pocas resulten significativas al nivel nominal del 5%, aunque no haya diferencia entre los tratamientos.*

2. El suplemento dietético es moderadamente eficaz.
3. Este estudio no demuestra nada, ya que no es doble ciego.
4. Aunque el número de pacientes es pequeño, lo inocuo del tratamiento y la consistencia de los resultados aconsejan prescribir el suplemento en la insuficiencia renal.
5. Este estudio no demuestra nada, ya que hay más variables que pacientes.

C. ANÁLISIS MULTIVARIANTE:

a. FUNDAMENTO:

Lo que hacemos es contrastar hipótesis para más de una variable sobre dos o más grupos. Es decir, realizar un estudio simultáneo de más de una variable.



El análisis multivariante permite controlar factores de confusión, por poder evaluar las interacciones entre múltiples variables

b. IMPORTANCIA:

El análisis multivariante agrupa bajo su definición un conjunto de sofisticadas técnicas -de la fase de análisis del estudio- que permiten controlar sesgos de confusión (MIR). Esto es debido a que hacemos un estudio simultáneo de todas las variables, no individuales de cada variable por separado, de forma que podremos controlar las posibles relaciones que tengan entre sí (factor de confusión se asocia a enfermedad, a factor de riesgo y no es intermediario de la cadena causal). Ej: tabaco, alta colesterolemia, sedentarismo, diabetes en cardiopatía coronaria.

Para su aplicación se requiere haber recogido previamente la información pertinente para las variables consideradas. No es posible controlar sesgos de otra naturaleza, como los originados en el muestreo (sesgos de selección) o en la recogida de datos (sesgos de información) (2MIR).

Control de sesgos de confusión

diseño	-Muestreo aleatorio –especialmente de tipo estratificado. -Emparejamiento o matching. -Criterios de restricción.
análisis	-Análisis estratificado con técnicas de ajuste o estandarización. -Análisis multivariante.

COMPARACIÓN DE MEDIAS

Dos grupos

Z COMPARACIÓN DE MEDIAS:

- Limitado a tablas 2x2

T DE STUDENT:

- Tipos según recogida de datos: t Student para datos independientes y t Student para datos apareados.
- Medias (cifras) en 2 grupos (colesterolemia en A y B).
- Condicionantes de validez (globalmente): a) ajustado a distribución Normal y b) varianzas iguales (homoscedasticidad). Test de Welch: si datos independientes y varianzas diferentes.

Más de dos grupos

- Requieren pruebas a posteriori o de comparaciones múltiples para: a) controlar el AUMENTO del error tipo I al realizar múltiples comparaciones y b) para indicar de dónde viene la diferencia.
- Siempre que se hacen múltiples comparaciones el error alfa global aumenta, con lo que el estudio puede no ser concluyente al encontrar diferencias estadísticamente significativas.

ANÁLISIS DE LA VARIANZA (ANOVA):

- Medias (cifras) en tres o más grupos (uricemia en A, B y C).
- Tipos: a) ANOVA de una vía (independientes) y b) ANOVA de dos vías o medidas repetidas (apareados).
- Condicionantes de validez: a) ajuste a distribución Normal y b) varianzas iguales.

ANÁLISIS MULTIVARIANTE:

- Contrasta hipótesis para más de una variable sobre dos o más grupos.
- Permiten controlar sesgos de confusión.

4. Asociación entre variables

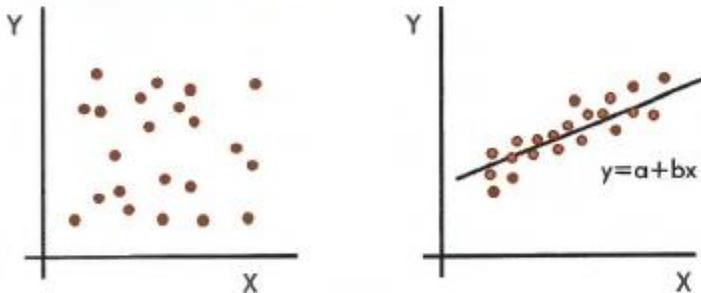
4.1. Regresión y correlación

A. DEFINICIÓN:

Regresión es la parte de la Estadística que estudia la **relación** entre los valores de una variable aleatoria cuantitativa X (variable independiente) y otra también aleatoria cuantitativa Y (variable dependiente). Y, como variable dependiente, irá **adquiriendo valores en función de X**, si existiera **correlación** entre ambas. En el caso de dos variables, como éste, hablamos de regresión simple, si fueran varias, regresión múltiple.

B. CÁLCULO DE REGRESIÓN LINEAL:

El cálculo de regresión (para relacionar dos variables aleatorias X e Y) establece una función $y=f(x)$ que define el comportamiento de Y en función de X. Dicha función puede tener una representación **lineal o curva**, dependiendo de su ecuación.



Una nube de puntos –a la izquierda– que no tiende a ninguna representación, expresa independencia entre variables (no asociadas). Si tiende a una línea –a la derecha–, existe asociación matemática de tipo lineal (correlación lineal).

Para una **correlación lineal** tendríamos una ecuación del siguiente tipo:

$$y = a + bx \Rightarrow \text{Ecuación de la recta de regresión lineal (MIR), donde}$$

$$a \Rightarrow \text{Valor de } y \text{ para } x = 0 \text{ (MIR)}$$

$$b \Rightarrow \text{Pendiente de la recta (MIR)}$$

Ej: $y=3x+2 \Rightarrow$ Conclusiones: la recta corta al eje de ordenadas (el de Y) en 2. La pendiente de la recta es 3 (por cada aumento de X, se triplica el aumento de Y). Para un valor de X dado calculamos el correspondiente a Y ($X=2 \rightarrow Y=8$; $X=1 \rightarrow Y=5$).

MIR 07 (8733): Supongamos un **modelo de regresión lineal simple** que relaciona los niveles medios de Proteína C Reactiva (PCR) medida en mmol/L como variable dependiente con el número de articulaciones dolorosas o hipersensibles de un paciente con artritis idiopática (variable independiente). El modelo ajustado en una muestra de 100 pacientes es el siguiente: $\mu g = 0,47 + 0,08x$ En el modelo. ¿Cómo se interpreta el coeficiente de 0.08?

- Dado que es superior a 0,05, quiere decir que no existe asociación entre las variables dependiente e independiente.
- Es el cambio obtenido en la media de PCR como consecuencia de pasar a tener una articulación adicional inflamada.*
- Es el valor medio de la PCR para los sujetos con una sola articulación afecta.
- Su cuadrado se interpreta como el porcentaje de varianza de la variable dependiente explicado por el modelo lineal.
- Es el valor medio basal de PCR para los individuos sanos (sin articulaciones afectas).

4.2. Coeficiente de correlación de Pearson (R)



Karl Pearson

Entre las variables X e Y **puede o no existir una asociación**, y ésta puede ser más o menos intensa. Este aspecto es medido por el **coeficiente de correlación**.

A. CARACTERÍSTICAS:

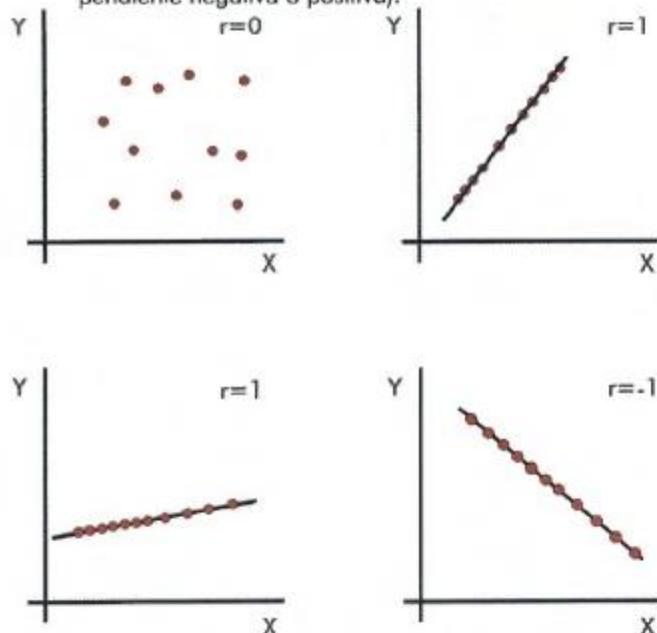
- Expresa la intensidad con que están relacionadas las variables cuantitativas X e Y, es decir, mide la **fuerza de asociación** entre ambas.

Al igual que el riesgo relativo y la odds ratio, el coeficiente de correlación mide la fuerza de asociación entre dos variables (MIR)

MIR 04 (7969): Un investigador está interesado en determinar si existe una asociación entre las cifras de tensión arterial diastólica (medida en mm de Hg) y los niveles de colesterol (medidos en mgr/ml). Para ello, ha realizado estas mediciones a 230 voluntarios. ¿Qué prueba estadística es la más apropiada para examinar esta asociación?:

- Regresión logística.
 - Prueba de la t.
 - Prueba de chi cuadrado.
 - Correlación de Pearson.***
 - Prueba de Fisher.
- Puede ser **muestral (r)** o **poblacional (p)**, dependiendo de que se obtenga a partir de los **datos de una muestra** o sea **inferido para una población**.
 - Sus valores oscilan entre -1 y 1. De forma que,
 - Tiene el **mismo signo** que la **pendiente** de la recta de regresión $y = a + bx$.
 - Su **valor absoluto** expresa la fuerza de la asociación (mayor cuanto más "aplastada" o más junta está la nube de puntos).
 - Si es igual a **cero**, no existe correlación (x e y son entonces variables independientes).
 - Si es igual a **-1 ó +1**, la correlación es **máxima** y la

relación entre x e y sigue exactamente una recta (de pendiente negativa o positiva).



Cuando $R=1$ todos los puntos están sobre la recta (sea negativa o positiva). Si $R=0$, la nube de puntos está dispersa y no existe relación entre las variables estudiadas

d) Ej: coeficiente de correlación R de $0.34 \rightarrow$ relación positiva entre X e Y , pero escasa fuerza de asociación (5MIR); si fuera 0.70 ó 0.80 sería importante ya, siendo máxima la fuerza de asociación si fuera igual a 1.00 . En el caso -0.20 la correlación hablaría a favor de una escasa fuerza de asociación negativa (MIR), es decir, a mayor valor de X , menor sería el resultante de Y .

1	Correlación positiva perfecta: nube de puntos traza exactamente una recta de pendiente positiva.
De 0 a 1	Correlación positiva variable: si cercana a 1 (0.8), buena; si cercana a 0 (0.3), débil.
0	Variables independientes: la variabilidad de una no depende de la variabilidad de la otra. Nube de puntos no tiende a una representación lineal.
De 0 a -1	Correlación negativa variable: si cercana a -1 (-0.8), buena; si cercana a 0 (-0.3), débil.
-1	Correlación negativa perfecta: nube de puntos traza exactamente una recta, con una pendiente negativa.

NOTA: El grado de correlación NO ES LA PENDIENTE de la recta. Una correlación determinada puede tener distintas pendientes. Con el coeficiente sólo sabemos (de forma directa) si es positiva o negativa, pero no el valor de la pendiente.

El coeficiente de correlación mide la fuerza de asociación entre dos variables, mostrando su signo si es una correlación positiva o negativa. (6+)

MIR 13 (10191) (192): El coeficiente de correlación de Pearson indica que existe asociación estadística entre dos variables cuando:

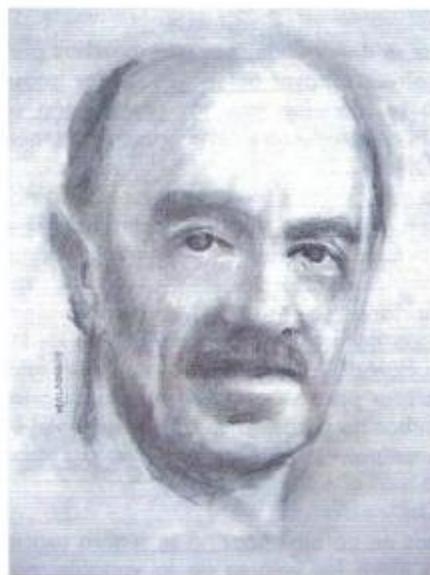
1. Su valor es positivo
2. Su valor está entre -1 (menos uno) y 1 (uno)
3. Su valor se aproxima a cero.
4. Su valor es igual o muy similar al tamaño muestral utilizado para su cálculo.
5. Su valor se acerca a sus valores extremos posibles -1 (menos uno) o 1 (uno). *

B. CONDICIONES DE VALIDEZ:

La prueba de correlación de Pearson tiene como condicionante de aplicación la distribución normal de la variable X respecto de Y y de Y respecto de X .

En caso de no poder aplicarse por incumplimiento de este

criterio, la prueba no paramétrica adecuada es la correlación de Spearman (también aplicable al estudio de correlación entre variables ordinales, como veremos a continuación).



Charles Spearman

MIR 10 (9469): La técnica estadística que debe aplicarse para analizar la concordancia de dos observadores en la medición del volumen espiratorio forzado en el primer segundo (FEV1) es:

1. El coeficiente de correlación de Spearman.
2. El coeficiente de correlación intraclass.*
3. La prueba de Kappa.3
4. La prueba de Kappa ponderada.
5. El test de Cohen.

Para el estudio de concordancia de variables cualitativas se utiliza la prueba de Kappa propuesto por Cohen en 1960 (por ejemplo, la clasificación como norma/patológico de una serie de radiografías de tórax). Para la concordancia de variables cuantitativas (como el FEV1), se usa el coeficiente de correlación intraclass. No es correcto usar pruebas de correlación, como el coeficiente de Pearson o el de Spearman, debido a que, si las diferencias entre los dos observadores fuesen sistemáticamente iguales, hallaríamos un coeficiente de Pearson igual a 1, a pesar de que no habría concordancia entre los observadores. La prueba de Kappa ponderada es una forma especial de Kappa que da un valor a las discrepancias entre observadores en función de lo diferentes que sean entre sí esas diferencias (es decir, no es lo mismo una diferencia entre observadores sano/enfermo que dudoso/enfermo).

MIR 11 (9705): El objetivo de una investigación es determinar la probabilidad de sufrir cirrosis hepática en función de la presencia o no de cinco variables: sexo, edad, consumo de alcohol, consumo de drogas y nivel de actividad física. La técnica estadística más adecuada para evaluar el objetivo propuesto es:

1. El cálculo de la matriz de correlación entre los factores.
2. El cálculo de la regresión lineal.
3. El análisis de componentes principales.
4. El cálculo de la regresión logística.*
5. La prueba de "ji" cuadrado.

MIR 12 (9940): La prueba kappa o test de Cohén:

1. Mide la validez interna de una prueba diagnóstica.
2. Mide la validez externa de una prueba diagnóstica.
3. Se mantiene estable para una misma prueba diagnóstica al modificar la frecuencia de la enfermedad.
4. Se mantiene estable para una misma prueba diagnóstica tanto si se utiliza como cribado o prueba clínica.
5. Tiende a disminuir su valor al aumentar el número de categorías de la prueba diagnóstica.*

5. Pruebas no paramétricas

A. INTRODUCCIÓN:

Las principales pruebas estadísticas para variables cuantitativas que hasta el momento hemos visto, es decir, t de Student para datos independientes y apareados, análisis de la varianza y correlación de Pearson, emplean en sus cálculos tanto paráme-

fros (medias, desviaciones típicas o varianzas) como propiedades de la distribución Normal. Por este motivo, para su aplicación exigen ciertas condiciones de validez referentes a la distribución de las variables en los grupos o poblaciones que se comparan.

Estas pruebas se denominan, por ello, pruebas paramétricas y son las que ofrecen mayor potencia y mayor "robustez", sobre todo cuando se manejan muestras de tamaño elevado. [La prueba de Chi-cuadrado no es una prueba paramétrica]

B. DEFINICIÓN:

Sin embargo, en ciertas ocasiones las condiciones de validez no se cumplen, o bien no se puede determinar si se cumplen, especialmente cuando se manejan tamaños muestrales pequeños.

En estas situaciones se recurre al empleo de ciertas pruebas estadísticas que no presuponen nada acerca de la naturaleza de la distribución de los datos y no requieren estos condicionantes de validez para su aplicabilidad. Estas pruebas se denominan pruebas no paramétricas. Y presentan el inconveniente de ser menos potentes que las paramétricas y, habitualmente, de cálculo más complejo.

C. CÁLCULO:

En las pruebas no paramétricas no se realiza exactamente una comparación entre los valores de la variable, no se utilizan medias ni varianzas, sino una comparación de rangos, con la consiguiente pérdida de información. Por esta misma razón, estas pruebas se pueden aplicar también al estudio de variables cualitativas ordinales.

D. APLICACIONES:

- Con variables cualitativas ordinales (MIR) (o cuantitativa continua en intervalos de clase), donde las pruebas paramétricas no se pueden utilizar.
- Cuando no se cumplen las condiciones de validez de las pruebas paramétricas, en cuanto a ajuste a distribución Normal o heterogeneidad en varianzas (no homocedasticidad).

Prueba paramétrica	Prueba no paramétrica
<ul style="list-style-type: none"> Utiliza "parámetros" y necesita el cálculo de medias y varianzas Necesita varianzas iguales y ajustes a distribución Normal Útil para variables (desenlace) cuantitativas continuas Ofrece mayor robustez y fácil cálculo matemático 	<ul style="list-style-type: none"> Utiliza escalas de rangos sin necesidad de calcular medias ni varianzas No necesita varianzas iguales ni ajuste a distribución Normal Útil en variables ordinales y cuantitativas en intervalos de clase Menor robustez y cálculo complejo
t de Student datos independientes	U de Mann-Whitney Kruskal-Wallis para dos grupos Prueba Wilcoxon (MIR) (suma de rangos)
t de Student datos apareados	Prueba de Wilcoxon (rangos con signo)
ANOVA de una vía	Prueba de Kruskal-Wallis
ANOVA para medidas repetidas	Test de Friedman
Correlación de Pearson	Correlación de Spearman

COMPARACIÓN DE PROPORCIONES

Apareados	McNemar		
	Z de comparación de proporciones	- Tablas 2 x 2	
Independientes	Chi-cuadrado (variables predictoras cualitativas, cuantitativas discretas y continuas en intervalos de clase)	- Ninguna frecuencia esperada < 1 - No más del 20% ≤ 5	- Aumentar tamaño muestral o agrupar celdas - Corrección por discontinuidad de Yates - Prueba exacta de Fisher (2x2)

COMPARACIÓN DE MEDIAS

(variables desenlace cuantitativas continuas)

	Z de comparación de medias	Tablas 2x2	
Dos grupos	T de Student para datos independientes	- Distribución Normal de la variable (menos, si > 30) - Homocedasticidad (varianzas iguales)	- Prueba no paramétrica - Test de Welch (si varianzas distintas)
	T de Student para datos apareados	- Distribución Normal de la variable (menos, si > 20) - Distribución Normal de la variable	- Prueba no paramétrica
Más de dos grupos	ANOVA o análisis de la varianza (una variable en ≥ 3 grupos)	- Homocedasticidad (varianzas iguales) - Requiere pruebas a posteriori - Recogida correcta de información (controla factores de confusión) - Requiere pruebas a posteriori.	- Prueba no paramétrica
	Análisis multivariante (> 1 variable en ≥ 2 grupos)		



Fin de la estadística.



REPASO

No hay que confundir ...

Comparar dos distribuciones de una variable: Ej.: estatura en dos poblaciones A y B. Test: t de Student para datos independientes.

Comparar n distribuciones de una variable: Ej.: uricemia en tres grupos de sujetos que recibieron respectivos tratamientos A, B y C. Test: análisis de la varianza.

Comparar tres o más distribuciones de una variable: Ej.: uricemia en tres grupos de sujetos que recibieron respectivos tratamientos A, B y C. Test: análisis de la varianza.

Significación de una prueba de contraste de hipótesis: Estudio sobre niveles plasmáticos de un fármaco en un grupo de sujetos y su triglicéridemia. Es la probabilidad de que una correlación positiva observada se haya debido al azar, bajo el supuesto de que no existe asociación. Ej Para $\alpha=5\%$, si $p<0,05$ la correlación es significativa; si $p>0,05$ la correlación no es significativa.

con ...

Comparar dos variables: Ej.: estatura y niveles séricos de hormona del crecimiento en un grupo de niños de cierta edad. Test: correlación de Pearson.

Comparar la variabilidad relativa de dos o más distribuciones de una o distintas variables: Ej.: comparar la dispersión relativa de valores de uricemia de un grupo de sujetos, con la de su edad y estatura. Medida: coeficiente de variación.

Estudiar el comportamiento conjunto de diversas variables para controlar factores de confusión: Ej.: En un estudio sobre consumo de etanol e incidencia de cardiopatía coronaria, analizar los datos en relación a la distribución de la edad, colesterolemia, consumo de tabaco y presión arterial en los sujetos a estudio. Técnica: análisis multivariante.

Intensidad o fuerza de una asociación:

Para el mismo estudio sobre niveles plasmáticos de un fármaco y triglicéridemia. Si r (coeficiente de correlación de Pearson) es igual a 0,2 la correlación positiva observado es débil. Si $r = 0,8$ la correlación es intensa.

No hay que confundir ...

Estudiar la tendencia de una variable dicotómica según los valores de una variable cuantitativa continua: Ej.: Estudiar, según sus niveles plasmáticos de calcitonina, la aparición o no de osteoporosis en los individuos de un grupo,. Técnica: regresión logística.

Comparar dos distribuciones de una variable cualitativa ordinal, para detectar una diferencia: Ej.: Distribución de niveles de dolor según leve/moderado/intenso en un grupo de sujetos tratados, frente a sujetos no tratados. Test: U de Mann-Whitney (suma de rangos de Wilcoxon).

con ...

Comparar dos distribuciones de una variable cuantitativa continua: Ej.: Comparar la distribución de valores de calcitonina sérica de un grupo de sujetos con osteoporosis, con la de un grupo de sujetos sin osteoporosis. Test: t de Student para datos independientes.

Realizar un test de asociación entre las categorías de dos variables cualitativas, mediante sus proporciones en una tabla de contingencia: Ej.: homogeneidad de los porcentajes en que se presenta en sujetos tratados y no tratados cada grado de dolor (leve/moderado/intenso). Test: Chi-cuadrado.

6. Principales pruebas de contraste de hipótesis y sus aplicaciones

<div style="display: flex; align-items: center;"> <div style="writing-mode: vertical-rl; transform: rotate(180deg);">VARIABLE PREDICTORA</div> <div style="margin: 0 10px;">→</div> <div style="writing-mode: vertical-rl; transform: rotate(180deg);">VARIABLE DE DESENLACE</div> </div>		Cualitativa nominal, dos categorías <i>Ejemplo:</i> Curación/No curación	Cualitativa nominal, más de dos categorías <i>Ejemplo:</i> Grupo sanguíneo A, B, AB, O.	Cualitativa ordinal o cuantitativa en intervalos <i>Ejemplo:</i> intensidad de un síntoma, leve/moderado/intenso	Cuantitativa continua <i>Ejemplo:</i> Nivel sérico de colesterol
Cualitativo: dos grupos independientes <i>Ejemplo:</i> Tratados/No tratados, poblaciones A y B	- z de comparación de proporciones - Chi-cuadrado - Prueba exacta de Fisher	- Chi-cuadrado	- U de Mann-Whitney - Wilcoxon (suma de rangos) - Kruskal-Wallis para dos grupos	- t de Student para datos independientes	
Cualitativo: dos grupos apareados <i>Ejemplo:</i> Una muestra de sujetos antes y después de un tratamiento	- Prueba de McNemar - Prueba exacta de Fisher	- Q de Cochran	- Wilcoxon (rangos con signo)	- t de Student para datos apareados	
Cualitativo: más de dos grupos independientes <i>Ejemplo:</i> Poblaciones A, B y C.	- Chi-cuadrado	- Chi-cuadrado	- Kruskal-Wallis	- Análisis de la varianza	
Ordinal <i>Ejemplo:</i> Nivel socioeconómico bajo/medio/alto	- Chi-cuadrado	- Chi-cuadrado	- Correlación de Spearman	- Correlación de Spearman	
Cuantitativo discreto o en intervalos de clase <i>Ejemplo:</i> Número de hijos	- Chi-cuadrado	- Chi-cuadrado	- Correlación de Spearman	- Correlación de Spearman	
Cuantitativo continuo <i>Ejemplo:</i> Renta anual, Nivel plasmático de un fármaco.	- Regresión logística		- Correlación de Spearman	- Correlación de Pearson	

Pruebas de contraste de hipótesis apropiadas según características de las variables medidas y agrupamiento en el estudio.



RESUMEN DE PRUEBAS DE CONTRASTE DE HIPÓTESIS

1. INTRODUCCIÓN

- Para calcular el valor del grado de significación p es necesario aplicar a los datos del estudio una determinada prueba estadística de contraste de hipótesis. Debe ser elegida la correcta, y para ello utilizamos los siguientes criterios: a) tipo de variable (predictora y desenlace); b) tipo de recogida de datos; y c) número de distribuciones (variables o grupos a estudio).

1.1. TIPO DE VARIABLE

- Se pueden distinguir **tres grandes grupos**: a) **comparación de proporciones** (variables predictora y de desenlace nominales: proporción de curados versus no curados en grupos A y B); b) **comparación de medias** (variable predictora nominal y de desenlace cuantitativa continua: medias de colesterolemia en grupos A y B); y c) **asociación entre variables** (variables predictora y de desenlace cuantitativas continuas: asociación entre número de cigarrillos e incidencia de cáncer de pulmón).

1.2. TIPO DE RECOGIDA DE DATOS

- Se distinguen **dos tipos**: a) **datos apareados**: se estudian las **mismas personas** en dos momentos, **antes y después** de administrar **a todo el grupo** la misma medida terapéutica (se mide tensión arterial antes y después de administrar a todos el mismo antihipertensivo); y b) **independientes**: se estudian personas diferentes "al mismo tiempo", tras haber administrado **a cada grupo una medida terapéutica distinta** (fármaco antihipertensivo A y B o un antihipertensivo versus placebo y se comparan diferencias en tensión arterial de unos y otros).

1.3. NÚMERO DE DISTRIBUCIONES

- Se puede diferenciar entre: a) una única variable estudiada en dos grupos (uricemia en A y B tras tratamiento); b) una única variable estudiada en más de dos grupos (uricemia en A, B y C tras tratamiento); y c) más de una variable en dos o más grupos (tabaquismo, colesterolemia, tensión arterial... en distintos grupos).

2. COMPARACIÓN DE PROPORCIONES

- Pruebas de contraste de hipótesis utilizadas: prueba de McNemar, Z de comparación de proporciones, **Chi-cuadrado** (con o sin prueba de discontinuidad de Yates) y prueba exacta de Fisher.

2.1. PARA DATOS APAREADOS

- **PRUEBA DE MCNEMAR.**

2.2. PARA DATOS INDEPENDIENTES

A. Z DE COMPARACIÓN DE PROPORCIONES.

- Puede utilizarse para contraste de hipótesis de dos proporciones (éxitos y fracasos en tratados con A y B). Se limita a tablas de contingencia del tipo 2×2 .

B. CHI-CUADRADO.

- **La más utilizada** para variables de desenlace nominales y datos independientes, puede utilizarse para variables predictoras cualitativas (MIR), cuantitativas discretas y continuas distribuidas en intervalos de clase. Compara variables distribuidas en clases excluyentes (MIR).
- El **cálculo de la Chi-cuadrado** se basa en la valoración de las diferencias que existen entre las frecuencias esperadas y las frecuencias realmente observadas (MIR). El número de grados de libertad se calcula multiplicando (filas - 1) por (columnas - 1) (MIR).
- **Aplicaciones frecuentes**: a) test de asociación con tablas de contingencia; b) test de homogeneidad de varias muestras cualitativas; y c) test de bondad de ajuste a distribuciones.
- **Condiciones de validez**: a) ninguna frecuencia esperada debe ser menor de 1 y b) no más de 20% de ellas inferior o igual a 5 (MIR). Si no se cumplen estos condicionantes puede: a) aumentar el tamaño muestral o agrupar celdas (filas y/o columnas); b) utilizar la corrección por discontinuidad de Yates (MIR) (algunos autores creen que debe aplicarse siempre (MIR)); y c) aplicar la prueba exacta de Fisher, siempre que no sea aplicable la Chi-cuadrado por manejar efectivos esperados muy reducidos y se trate de una tabla de contingencia de 2×2 (MIR).

3. COMPARACIÓN DE MEDIAS

- Pruebas de contraste de hipótesis utilizadas: z de comparación de medias, t de Student-Fisher (datos apareados e independientes) y test de Welch.

3.1. PARA MEDIAS DE DOS GRUPOS

- **Z DE COMPARACIÓN DE MEDIAS.** Puede utilizarse para contraste de hipótesis de dos medias (colesterolemia en tratados con A y B). Se limita a tablas de contingencia del tipo 2×2 .
- **T DE STUDENT.** Según sea el tipo de recogida de datos utilizaremos las pruebas t de Student para datos apareados (2MIR) o t de Student para datos independientes (2MIR). En ambos casos la variable de desenlace es cuantitativa continua (2MIR) (colesterolemia, uricemia, estatura, peso...).
- **Condiciones de validez**: a) **datos independientes**: las varianzas de los dos grupos deben ser iguales (=homoscedasticidad) y la variable a estudio ha de seguir una distribución Normal; y b) **datos apareados**: las diferencias de la variable entre "antes y después" de la medida terapéutica deben seguir una distribución Normal.
- Si no se cumplen estos criterios podemos utilizar **pruebas no paramétricas** distintas según el caso (v. más adelante). El **test de Welch** es una modificación de la t de Student para datos independientes en casos de varianzas distintas.

3.2. PARA MEDIAS DE MÁS DE DOS GRUPOS

- Distinguimos ahora entre: a) una variable de desenlace cuantitativa continua estudiada en 3 ó más grupos (ANOVA) y b) más de una variable estudiada en dos o más grupos (análisis multivariante).
- En ambas situaciones se debe tener en cuenta que: a) valoran diferencias globales, dicen si existe o no una diferencia entre los múltiples grupos estudiados; y b) al realizar múltiples comparaciones a nivel nominal de un 0.05 ó 5% en cada uno, se produce un **AUMENTO del error tipo I global**, lo que hace que el estudio no sea concluyente (4MIR), dado que ese porcentaje es acumulativo con cada grupo o variable. Ej: de 100 comparaciones a 5% cada una sería esperable, simplemente por azar, que 5 saliera estadísticamente significativa ($5\% \times 100 = 5$).
- Por todo esto es necesario realizar después de obtener resultados "a favor de la diferencia" pruebas a posteriori o algún método de comparaciones múltiples (MIR), las cuales nos ayudarán a saber de dónde viene la diferencia (qué grupo) y nos ayudarán a **controlar el aumento del error tipo I**.



A. ANÁLISIS DE LA VARIANZA (ANOVA)

- Compara medias (variables cuantitativas continuas) de tres o más grupos (7MIR) "Comparar uricemia entre tres o más grupos A, B, C...". Valora la variabilidad inter e intragrupalmente con la F de Snedecor.
- **Tipos:** a) ANOVA de una vía (el clásico): para datos independientes y b) ANOVA de dos vías o para medidas repetidas: para datos apareados.
- **Condicionantes de validez:** los valores deben ajustarse a la distribución Normal y tener los diferentes grupos varianzas iguales (homoscedasticidad).

B. ANÁLISIS MULTIVARIANTE.

- **Compara medias (cifras) de dos o más variables en dos o más grupos.** "Comparar el efecto de uricemia, tensión arterial, tabaquismo... en cardíopatas, y no cardíopatas". Es un grupo de diferentes técnicas que permiten controlar factores de confusión (MIR) pero no otro tipo de sesgos (selección, información) (2MIR).

4. ASOCIACIÓN ENTRE VARIABLES

4.1. INTRODUCCIÓN

- La **Regresión** es la parte de la Estadística que se ocupa del estudio de la asociación entre variables, pudiendo ésta ser de tipo lineal o curva (logarítmica, exponencial). Cuando una **variable Y o variable dependiente** va tomando valores en función de otra **variable X o variable independiente** (MIR), se considera que ambas están asociadas, siguen una **correlación**.

4.2. CORRELACIÓN LINEAL

- Si dos variables X e Y siguen una correlación lineal, los valores de Y pueden predecirse en función de los valores de X, según la fórmula $y=ax+b$ (MIR), donde "a" es el valor de la pendiente de la recta (positiva o negativa) (MIR) y "b" es el valor que toma Y cuando X vale 0 (MIR). Ej.: Tensión arterial sistólica = $81.5 + 1.2 \times \text{edad}$ → por cada año más de edad se incrementa la TAS en 1.2 mmHg (MIR).

4.3. COEFICIENTE DE CORRELACIÓN DE PEARSON (R)

- Mide la fuerza de asociación entre dos variables cuantitativas continuas (edad y tensión arterial) (MIR).
- El valor absoluto (MIR) del coeficiente R indica la fuerza de asociación. Sus valores **oscilan entre -1 y 1**, donde: a) 0, indica que no están asociadas (son independientes); b) -1 y 1, indican correlación perfecta entre variables, negativa y positiva respectivamente; y c) **valores intermedios entre 0 y 1 ó 0 y -1**, indican, según sea el valor absoluto, una fuerza de asociación más fuerte (0.8) o más débil (0.3), y positiva o negativa según el signo de R (6MIR) (el signo de R indica el **signo de la pendiente de la recta**, y NO el valor de la pendiente de la recta).
- Ej: 0.3, débil correlación positiva (5MIR). -0.3, débil correlación negativa (MIR). 0.8, fuerte correlación positiva. -0.8, fuerte correlación negativa.
- **Condicionantes de validez:** distribución ajustable a la Normal de Y respecto a X y de X respecto a Y.

5. PRUEBAS NO PARAMÉTRICAS

5.1. DEFINICIÓN

- Las **pruebas paramétricas** (t de Student, ANOVA y correlación de Pearson) son pruebas de contraste de hipótesis de **mayor robustez estadística y cálculo más sencillo**, sin embargo **no pueden utilizarse cuando se incumplen sus condicionantes de validez:** necesitan un ajuste a la distribución Normal y varianzas iguales en la mayoría de los casos. Se denominan paramétricas por necesitar calcular parámetros del tipo **medias y varianzas**.
- Las **pruebas no paramétricas** ofrecen menor robustez y son de cálculo más complejo, aunque pueden utilizarse cuando se incumplen los condicionantes de validez de las paramétrica, además de poder utilizarse en variables de desenlace ordinales (MIR). Cada tipo de prueba paramétrica tiene su prueba de elección cuando se incumplen sus condicionantes:

Pruebas paramétricas	Pruebas NO paramétricas
<ul style="list-style-type: none"> • t de Student datos independientes • Test de Welch (para varianzas distintas) 	<ul style="list-style-type: none"> • U de Mann-Whitney • Kruskal-Wallis para dos grupos • Prueba Wilcoxon (MIR) (suma de rangos)
<ul style="list-style-type: none"> • t de Student datos apareados 	<ul style="list-style-type: none"> • Prueba de Wilcoxon (rangos con signo)
<ul style="list-style-type: none"> • ANOVA de una vía 	<ul style="list-style-type: none"> • Prueba de Kruskal-Wallis
<ul style="list-style-type: none"> • ANOVA para medidas repetidas 	<ul style="list-style-type: none"> • Test de Friedman
<ul style="list-style-type: none"> • Correlación de Pearson 	<ul style="list-style-type: none"> • Correlación de Spearman

