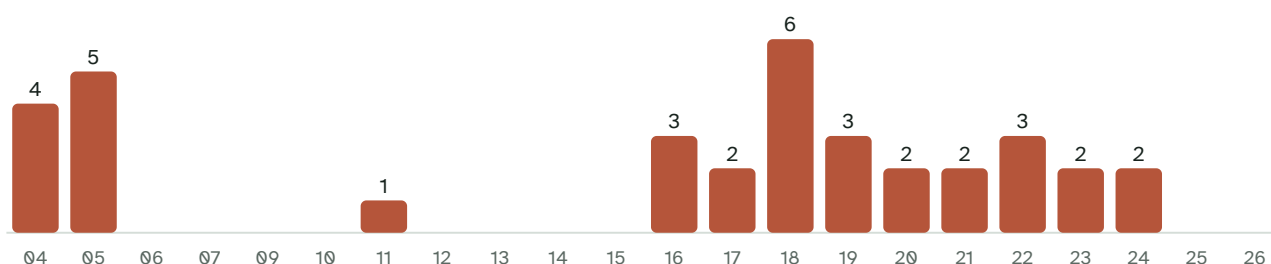


Estadística descriptiva e inferencial

La estadística descriptiva resume lo que se ha medido en una muestra; la inferencial generaliza esos resultados a la población de la que procede, con una probabilidad de error conocida. En el banco hay 39 preguntas de este tema entre 2004 y 2026 y casi todas piden un dato concreto: qué tipo de variable es, qué medida de dispersión acompaña a la mediana, qué representa la línea de un diagrama de cajas, cómo se interpreta un intervalo de confianza, qué es el error tipo I y, sobre todo, qué prueba estadística corresponde a cada situación (t de Student, ANOVA, U de Mann-Whitney, Wilcoxon, Kruskal-Wallis, Friedman, ji cuadrado).

Apuntes para el EIR · versión 2026-10-01 · **sin revisión enfermera**. Todas las preguntas son reales, de exámenes EIR, con su año y su número; ninguna está escrita por IA. Las respuestas están en el solucionario, al final.

Cómo cae en el EIR



35 preguntas de este tema en 22 exámenes (2004-2026, ordinarias, clasificación segura).

Lo más repetido

- Interpretación del intervalo de confianza del 95 % (2 exámenes)
- Test de Kruskal-Wallis: no paramétrico para más de 2 grupos independientes (2 exámenes)
- t de Student: variable cuantitativa normal frente a cualitativa dicotómica (2 exámenes)
- Error tipo I (alfa): rechazar H_0 siendo cierta (2 exámenes)

Índice

Resumen para repasar	3
Puntos clave	6
1. Población, muestra y tipos de variables	7
2. Muestreo y tamaño de la muestra	10
3. Estadística descriptiva: centralización, posición y dispersión	13
4. Representación gráfica	16
5. Probabilidad	20
6. Inferencia: intervalo de confianza y contraste de hipótesis	23
7. Elección de la prueba estadística	27
8. Correlación y regresión	33
Todas las preguntas del tema (38)	36
Preguntas que también podrían ser de este tema (1)	46
Solucionario	47
Fuentes	49

Resumen para repasar

1. Población, muestra y tipos de variables

- Descriptiva: **presenta y sintetiza** los datos; inferencial: **generaliza de la muestra a la población** ^{F13}.
- Población diana: **grupo a estudiar, sobre el que se quiere información y estimaciones** ^{F4}; accesible: **subgrupo alcanzable por razones prácticas**; elegible: la que dejan los **criterios de selección** ^{F5}.
- Cuantitativa **continua** (edad, peso) o **discreta** (**número de hijos**) ^{F1}.
- Cualitativa **dicotómica**, **nominal** (sin orden: color de ojos, sexo, profesión) u **ordinal** (con orden: estadiaje); la ordinal sigue siendo **cualitativa** ^{F1,F36}.
- Escalas: nominal, ordinal, **intervalo** (sin cero verdadero) y **razón** (con cero verdadero) ^{F28}.

2. Muestreo y tamaño de la muestra

- Probabilístico o aleatorio frente a no probabilístico (**cuotas, conveniencia, bola de nieve**), que no permite generalizar ^{F5}.
- **Aleatorio simple**: todos con la **misma probabilidad** de ser elegidos; **sistemático**: a **intervalos iguales** de una lista ^{F4}.
- **Estratificado**: muestra de cada estrato; **conglomerados**: se eligen grupos y se estudian **todos** sus elementos; **polietápico**: selección **por etapas** dentro de las unidades de la etapa anterior ^{F4,F5}.
- Estimar una proporción o una media: **confianza ($Z\alpha$)**, **precisión y variabilidad** (p esperada, o 0,5 si se desconoce; o varianza) ^{F6}.
- Comparar grupos: además **diferencia a detectar**, **α** , **potencia ($1-\beta$)** y hipótesis uni o bilateral ^{F6}.

3. Estadística descriptiva: centralización, posición y dispersión

- Categóricas: **frecuencias y porcentajes** ^{F1}.
- Centralización: **media** (sensible a extremos), **mediana** (resistente; elección en datos asimétricos) y **moda** ^{F1,F29}.
- **Mediana = P50 = Q2 = D5**; cuartiles y percentiles son medidas de **posición** ^{F1,F7,F36,F38}.
- Dispersión: **rango**, **RIC ($Q3 - Q1$)**, **varianza**, **desviación típica y CV**; la **mediana no es de dispersión** ^{F1,F7}.
- **RIC con la mediana; desviación típica con la media** ^{F7}.
- Varianza: **nunca negativa** (se elevan al cuadrado las diferencias) ^{F7}, 0 si todos los datos son iguales ^{F1}; restar una constante **baja la media** en esa cantidad y **no cambia la varianza** ^{F36}.

4. Representación gráfica

- **Barras y sectores** → **cualitativas** (sectores: pocas categorías); **histograma** → **cuantitativa continua**, rectángulos contiguos ^{F2,F3}.
- **Pictograma**: frecuencias representadas **con figuras o símbolos** (por área o por número de símbolos) ^{F3}.
- **Dispersión** → **relación entre dos cuantitativas**; **líneas** → **tendencias temporales** ^{F2,F3}.
- Diagrama de cajas: caja = **50 % central (Q1 a Q3)**, línea interior = **mediana**, puntos aislados = **valores extremos** ^{F2}.

5. Probabilidad

- Probabilidad entre **0 y 1**; complementario: **$1 - P(A)$** ^{F13}.
- **Unión**: $P(A) + P(B) - P(A \text{ y } B)$ (sin restar si son incompatibles); **independientes**: $P(A \text{ y } B) = P(A) \cdot P(B)$ ^{F13}.
- Dos portadores de un gen **autosómico recesivo**: **25 %** de hijo afectado en cada embarazo ^{F14}.

6. Inferencia: intervalo de confianza y contraste de hipótesis

- El IC describe la **incertidumbre y la precisión** de la estimación ^{F15,F16}.
- IC 95 % de una media = **media $\pm$ 1,96 x error estándar** ($EE = DE/\sqrt{n}$); también para proporciones, RR u OR; habitual 95 %, también 90 % o 99 % ^{F7,F15}.
- Interpretación estricta: si se repitiera el estudio **muchas veces, el 95 % de los intervalos contendría el valor poblacional**; aproximada: **95 % de seguridad** de que el valor de la **población** está dentro ^{F15}.
- **Error tipo I (α)**: rechazar H_0 **siendo cierta** (falso positivo); **tipo II (β)**: aceptar H_0 siendo falsa; **potencia = $1 - \beta$** ^{F16,F17}.
- p: según Fistera, **probabilidad de rechazar H_0 cuando es verdadera**; técnicamente, probabilidad del efecto observado o mayor si H_0 fuera cierta ^{F15,F16}.
- IC 95 % que **excluye el valor nulo** $\Leftrightarrow p < 0,05$ ^{F15}.

7. Elección de la prueba estadística

- Paramétricas solo con **distribución normal** ^{F18}; normalidad: **Shapiro-Wilk** ($n < 50$), **Kolmogorov-Smirnov** ($n \geq 50$) ^{F21}.
- Igualdad de varianzas: **F de Snedecor** (cociente de varianzas) y **Levene** ^{F20,F22,F23}.
- Normal: **t de Student** (2 grupos, también dicotómica frente a cuantitativa), **ANOVA** (3 o más) ^{F18,F34}.
- No normal: **Mann-Whitney** (2 independientes), **Wilcoxon** (2 relacionadas), **Kruskal-Wallis** (3 o más independientes), **Friedman** (3 o más relacionadas) ^{F18}.
- Cualitativas: **ji cuadrado** (independientes), **McNemar** (relacionadas), **Fisher** si menos del 80 % de celdas tienen esperados > 5 ^{F18,F24,F25}.
- Homogeneidad: la ji cuadrado comprueba si **dos o más muestras independientes difieren en su distribución** de una misma variable ^{F39}.

8. Correlación y regresión

- **Covarianza** puede ser **positiva o negativa**, al contrario que la **varianza** ^{F37}.
- **Pearson** de -1 a $+1$ ^{F27,F37}; **Spearman** si no hay normalidad u ordinales ^{F18}; correlación **no implica causalidad** ^{F27}.
- $r = 0$: cada recta de regresión es una **constante** ($Y^* = \bar{y}$, $X^* = x$); $r = \pm 1$: **coinciden** ^{F37}.
- Respuesta binaria y **odds ratio** → **regresión logística**; razón de tasas → Poisson; hazard ratio → Cox ^{F26,F33}.

Puntos clave

- 1 Número de hijos: cuantitativa **discreta** ^{F1}; las cualitativas son **nominales** (sin orden) u **ordinales**, y una variable ordinal sigue siendo **cualitativa** ^{F1,F36}.
- 2 Población **diana**: el grupo a estudiar, delimitado por sus características, sobre el que se quiere información y estimaciones; la **accesible** es el subgrupo al que se llega por razones prácticas ^{F4,F5}.
- 3 Muestreo **aleatorio simple**: todos los miembros tienen la misma probabilidad de ser elegidos; **polietápico**: selección por etapas dentro de las unidades elegidas en la etapa anterior ^{F4}.
- 4 Para **estimar** una media o una proporción: confianza, precisión y variabilidad; la **potencia** solo interviene al comparar grupos ^{F6}.
- 5 **Mediana = P50 = Q2 = D5**; el **RIC** ($Q3 - Q1$) acompaña a la mediana y la desviación típica a la media ^{F7,F36,F38}.
- 6 Restar una constante: la **media** baja esa constante y la **varianza no cambia** ^{F36}.
- 7 En el diagrama de cajas, la línea interior es la **mediana**; los bordes de la caja, $Q1$ y $Q3$; los puntos aislados, valores extremos ^{F2}.
- 8 IC 95 %: si se repitiera el estudio muchas veces, el **95 % de los intervalos contendría el parámetro poblacional**; mide la **precisión** de la estimación ^{F15,F16}.
- 9 **Error tipo I (α)**: rechazar H_0 siendo cierta; **tipo II (β)**: aceptar H_0 siendo falsa; **potencia** = $1 - \beta$ ^{F16,F17}.
- 10 No paramétricas: **Mann-Whitney** (2 independientes), **Wilcoxon** (2 relacionadas), **Kruskal-Wallis** (3 o más independientes), **Friedman** (3 o más relacionadas) ^{F18}.
- 11 Normalidad: **Shapiro-Wilk** si $n < 50$, **Kolmogorov-Smirnov** si $n \geq 50$; varianzas: **F de Snedecor** y **Levene** ^{F21,F22,F23}.
- 12 Ji cuadrado: independencia, bondad de ajuste y **homogeneidad** (si dos o más muestras independientes difieren en su distribución) ^{F39}; **Fisher** si las frecuencias esperadas son pequeñas ^{F24,F25}.
- 13 $r = 0$: cada recta de regresión es una **constante**; $r = \pm 1$: **coinciden** ^{F37}; odds ratio $\rightarrow$ **regresión logística** ^{F26}.

1. Población, muestra y tipos de variables

La estadística se clasifica en **descriptiva**, cuando se usa para presentar y sintetizar la información recogida en un estudio, e **inferencial**, cuyo objetivo es **generalizar la información obtenida en una muestra a resultados válidos para la población de la que procede** ^{F13}. La base matemática de la inferencia es el cálculo de probabilidades ^{F13}.

La **población** es el conjunto grande de individuos que se desea estudiar y generalmente es inaccesible; la **muestra** es el conjunto menor de individuos, subconjunto accesible y limitado, sobre el que se hacen las mediciones con la idea de obtener conclusiones generalizables a la población ^{F1}. La muestra debe ser **representativa**: cualquier individuo de la población en estudio debe haber tenido la misma probabilidad de ser elegido ^{F1}. Lo que se estudia en cada individuo son las **variables**, y los **datos** son los valores que toma la variable en cada caso ^{F1}.

Nivel	Qué es ^{F5}
Población diana o blanco	La delimitación del grupo a estudiar según características clínicas, demográficas, sociales, estilos de vida, etc.; cuanto más específica, más probable es la generalización de los hallazgos . La OCDE la define como el conjunto de elementos sobre los que se quiere información y se requieren estimaciones ^{F4} : es la población a la que se refieren las conclusiones del estudio.
Población accesible	Subgrupo de la población diana que se determina por consideraciones prácticas , según las posibilidades o recursos de los investigadores (por ejemplo, solo los pacientes que acuden a un hospital).
Población elegible	La que queda delimitada por los criterios de selección : inclusión, exclusión y eliminación.
Muestra	La fracción de la población que realmente se estudia; debe elegirse de forma que sea representativa.

EIR 2021 · 45

P39 · respuesta en el solucionario

Población diana: definición

La Población definida por la pregunta de investigación a la que se generalizan los resultados obtenidos en un estudio, se denomina:

1. Población accesible.
2. Muestra del estudio.
3. Población diana.
4. Sujetos que cumplen los criterios de selección.

La naturaleza de la variable decide el método estadístico. Las **cuantitativas** pueden medirse o expresarse numéricamente; las **cualitativas** representan una cualidad o atributo que clasifica cada caso en una categoría ^{F1}. Además de las dicotómicas, hay cualitativas con más de dos categorías, como el color de los ojos, el **grupo sanguíneo** o la profesión ^{F1}.

Tipo	Definición	Ejemplos
Cuantitativa continua	Admite cualquier valor dentro de un rango numérico ^{F1}	Edad, peso, talla ^{F1}
Cuantitativa discreta	No admite todos los valores intermedios; suele tomar solo valores enteros ^{F1}	Número de hijos , de partos, de hermanos ^{F1}
Cualitativa dicotómica o binaria	Solo dos categorías ^{F1}	Hombre/mujer, enfermo/sano, fumador/no fumador ^{F1}
Cualitativa nominal	Categorías sin relación de orden entre sí ^{F1} ; ni siquiera admiten una ordenación en sus valores ^{F36}	Color de ojos, sexo, profesión ^{F1} ; estado civil (datos cualitativos que no siguen ningún orden) ^{F3}
Cualitativa ordinal	Hay un cierto orden o jerarquía entre las categorías ^{F1}	Grados de disnea, estadiaje de un tumor ^{F1}

Ordinal sigue siendo cualitativa

Una variable ordinal es **cualitativa**: las cualitativas se dividen en **nominales** y **ordinales**, y las ordinales, a pesar de tomar valores no numéricos, admiten una ordenación natural; por ejemplo, la nota de una asignatura (suspense, aprobado, notable, sobresaliente, matrícula de honor) es cualitativa ordinal ^{F36}. Y «discreta» o «continua» solo se aplica a las variables **cuantitativas** ^{F1,F36}.

El CDC usa además la clasificación por **escalas de medida**: nominal (categorías sin orden), ordinal (se pueden ordenar, pero no están igualmente espaciadas), de **intervalo** (unidades iguales sin cero verdadero, como la fecha de nacimiento) y de **razón** (intervalo con cero verdadero, como la talla en centímetros); nominales y ordinales son categóricas o cualitativas, y las de intervalo y razón, continuas o cuantitativas ^{F28}. Una misma variable puede medirse en más de una escala: la prueba de la tuberculina puede registrarse como positiva o negativa (nominal) o en milímetros de induración (razón) ^{F28}.

EIR 2020 · 135

P14 · respuesta en el solucionario

Número de hijos: variable cuantitativa discreta

El número de hijos de una familia es una variable:

1. Cualitativa dicotómica.
2. Cuantitativa continua.
3. Cuantitativa discreta.
4. Cuantitativa nominal.

Grupo sanguíneo: variable cualitativa nominal

La variable “grupo sanguíneo”, ¿qué tipo de variable es?:

1. Cuantitativa discreta.
2. Cuantitativa continua.
3. Cualitativa nominal.
4. Cualitativa ordinal.

Variable nivel socioeconómico: cualitativa (ordinal)

En una investigación la variable “nivel socioeconómico”, especificada en los posibles valores como bajo, medio-bajo, medio-medio, medioalto y alto, es un tipo de variable:

1. Cuantitativa.
2. Discreta.
3. Continua.
4. Diferencial.
5. Cualitativa.

2. Muestreo y tamaño de la muestra

La mejor forma de que la muestra permita hacer inferencias es elegirla de manera **aleatoria**, para que todos los elementos de la población tengan la **misma probabilidad** de ser incluidos ^{F5}. Los procedimientos de muestreo se dividen en **probabilísticos o aleatorios** y **no probabilísticos**; los no probabilísticos no sirven para hacer generalizaciones, porque no hay certeza de que la muestra sea representativa ^{F5}.

Muestreo	Cómo se hace
Aleatorio simple	Cada miembro de la población tiene la misma probabilidad de ser elegido ^{F4} . Se numera a cada individuo y se eligen tantos como haga falta con tablas de números aleatorios o números generados por ordenador; tiene poca utilidad práctica si la población es muy grande ^{F5} .
Sistemático	Se obtiene por un método sistemático, por ejemplo tomando individuos de una lista a intervalos iguales (intervalo de muestreo) ^{F4} .
Estratificado	Se divide la población en estratos homogéneos respecto a alguna característica (sexo, profesión, municipio...) y se toma parte de la muestra de cada estrato, para asegurar que todos están representados; exige conocer bien la población ^{F4,F5} .
Por conglomerados	La unidad muestral es un grupo de elementos (hospitales, departamentos, áreas geográficas); se seleccionan al azar conglomerados y se estudian todos sus elementos ^{F4,F5} .
Polietápico	La muestra se selecciona por etapas : las unidades de cada etapa se submuestrean dentro de las unidades (más grandes) elegidas en la etapa anterior; hay unidades de primera etapa, de segunda etapa, etc. ^{F4}
Por cuotas (no probabilístico)	Se fijan cuotas de individuos con ciertas condiciones y se eligen los primeros que las cumplen; se parece al estratificado, pero sin aleatoriedad ^{F5} .
Intencional o de conveniencia (no probabilístico)	Selección no aleatoria; lo más frecuente es usar a los individuos de fácil acceso ^{F5} .
Bola de nieve (no probabilístico)	Unos individuos conducen a otros hasta completar la muestra; se usa con poblaciones difíciles de localizar ^{F5} .



Reconocer el polietápico

Si se eligen al azar unidades grandes (por ejemplo, distritos) y **dentro de ellas** otras más pequeñas (secciones censales), hay **varias etapas de selección**: es un muestreo **polietápico** ^{F4}. Que en la última etapa se incluya a todas las personas de la unidad elegida no cambia que la selección se haya hecho por etapas. Para llamarlo **estratificado** hace falta dividir antes la población en estratos y muestrear dentro de cada uno ^{F4,F5}.

EIR 2017 · 21

P25 · respuesta en el solucionario

Muestreo aleatorio simple: definición

¿Qué es el muestreo aleatorio simple?:

1. Es la técnica que decide el nº de individuos que tienen que participar en el estudio.
2. Es la técnica de selección de sujetos donde no interviene el azar.
3. Es la técnica aleatoria de selección de variables de estudio.
4. Es aquella técnica en la que cada unidad del marco muestral tiene la misma probabilidad de ser escogida.

EIR 2018 · 147

P23 · respuesta en el solucionario

Muestreo polietápico

Para el diseño muestral de una investigación se decide seleccionar aleatoriamente un número de distritos de una ciudad. Posteriormente, sobre estos distritos se eligen un número de secciones censales al azar, de las cuales se incluyen para el estudio todas las personas que viven en dichas secciones. Se trata de:

1. Un muestreo polietápico.
2. Un muestreo sistemático.
3. Un muestreo estratificado polietápico.
4. Un muestreo semiprobabilístico.

El **tamaño muestral** se calcula en la fase de diseño. Si no se hace, el estudio puede quedarse sin pacientes suficientes para estimar con precisión y para encontrar diferencias que sí existen, o bien estudiar pacientes de más, con pérdida de tiempo y recursos ^{F6}. Hay que distinguir dos situaciones: estudios para **estimar un parámetro** (una proporción o una media poblacional) y estudios para **contrastar hipótesis** (comparar medias o proporciones) ^{F6}.

Objetivo	Qué hay que conocer ^{F6}
Estimar una proporción	Nivel de confianza o seguridad ($1-\alpha$), que da el coeficiente $Z\alpha$ (1,96 para el 95 %; 2,58 para el 99 %); la precisión deseada (d); y una idea de la proporción esperada (p). Si no se conoce, se usa $p = 0,5$, que maximiza el tamaño muestral. Fórmula: $n = Z\alpha^2 \cdot p \cdot q / d^2$, con $q = 1 - p$.
Estimar una media	Nivel de confianza ($Z\alpha$); precisión ($2 \cdot d$ es la amplitud del intervalo de confianza); y una idea de la varianza (S^2) de la variable en la población. Fórmula: $n = Z\alpha^2 \cdot S^2 / d^2$.
Contrastar hipótesis (comparar dos medias o dos proporciones)	Magnitud de la diferencia clínicamente relevante a detectar; una idea de los parámetros de la variable; seguridad (riesgo α); poder estadístico ($1-\beta$); y si la hipótesis es unilateral o bilateral (la bilateral es más conservadora).

★ La potencia solo entra al comparar

Para **estimar** una media o una proporción intervienen la confianza, la precisión y la variabilidad (varianza o proporción esperada). La **potencia estadística** (**$1-\beta$**) y la diferencia a detectar solo se añaden cuando el estudio **compara** grupos (contraste de hipótesis) ^{F6}. En todos los casos se aumenta la muestra por las pérdidas previstas: $n \text{ ajustada} = n \cdot (1 / (1 - R))$, siendo R la proporción esperada de pérdidas ^{F6}.

EIR 2022 · 52

P8 · respuesta en el solucionario

Tamaño muestral para estimar media o proporción: la potencia no interviene

Respecto al cálculo del número de sujetos necesarios para la realización de un estudio cuyo objetivo es la estimación de una media o una proporción, señale qué elemento NO interviene:

1. Variabilidad del parámetro que se desea estimar.
2. Precisión con la que se desea estimar el parámetro.
3. Potencia estadística o capacidad del estudio para detectar una determinada diferencia.
4. Nivel de confianza deseado.

3. Estadística descriptiva: centralización, posición y dispersión

Las variables **categorías** se describen con el número de casos de cada categoría y el **porcentaje** que representa del total, en una tabla de frecuencias ^{F1}. Las **numéricas** se describen respondiendo a dos preguntas: alrededor de qué valor se agrupan los datos (**medidas de centralización**) y cómo de concentrados o dispersos están (**medidas de dispersión**) ^{F1}.

Medida de centralización	Definición	Sensibilidad a valores extremos
Media aritmética	Suma de todos los valores dividida entre el número total de datos ^{F1}	Le afecta cualquier valor extremo; es la medida de elección con datos de distribución normal ^{F29}
Mediana	La observación equidistante de los extremos: deja la mitad de los datos por encima y la otra mitad por debajo ^{F1}	No suele afectarle uno o dos valores extremos; es la medida de elección con datos asimétricos o no normales ^{F29}
Moda	El valor de la variable con mayor frecuencia ^{F1}	No suele afectarle uno o dos valores extremos ^{F29}

Cuando la media y la mediana son iguales o están próximas, la distribución se considera **simétrica**; la media es muy sensible a la variación de las puntuaciones y la mediana, menos ^{F1}. Los **cuartiles y percentiles** no son medidas de tendencia central sino **medidas de posición** ^{F1}. El percentil P es el valor que tiene el P % de las observaciones en él o por debajo; los **cuartiles** dividen los datos en cuatro partes iguales: el primer corte es el percentil 25, el segundo el **percentil 50, que es la mediana**, y el tercero el percentil 75 ^{F7}. Los **deciles** son los percentiles múltiplos de 10: el primer decil es el percentil 10, el **quinto decil, el percentil 50**, y el noveno, el percentil 90 ^{F38}; la mediana coincide con el quinto decil ^{F36}. El cuartil central se llama mediana ^{F4}.

Cuatro nombres, un valor

Mediana = percentil 50 = segundo cuartil = quinto decil ^{F7,F36,F38}.

EIR 2005 · 17

P33 · respuesta en el solucionario

Percentil 50 = mediana = cuartil 2 = decil 5

El percentil 50:

1. Son valores referentes a las desviaciones típicas.
2. Es un tipo de desviación típica para distribuciones continuas.
3. Coincide con la moda en cualquier distribución.
4. Tiene el mismo valor que la mediana, cuartil segundo y decil quinto.
5. Significa que una distribución está sesgada en su punto central.

Medida de dispersión	Definición	Con qué se usa
Rango, recorrido o amplitud	Diferencia entre el valor mayor y el menor F1,F4	Señala los valores extremos; con la mediana en distribuciones asimétricas F1
Rango o recorrido intercuartílico (RIC)	Distancia entre el cuartil superior y el inferior ($Q3 - Q1$, del percentil 25 al 75); contiene la mitad central de las observaciones F4,F7	Se usa con la mediana , sobre todo en distribuciones asimétricas F7
Varianza (S^2)	Media de los cuadrados de las diferencias entre cada valor y la media aritmética; se expresa en las unidades de la variable al cuadrado F1	Describe la variabilidad alrededor de la media; su raíz cuadrada es la desviación típica F1
Desviación típica o estándar (S)	Raíz cuadrada de la varianza; se expresa en las mismas unidades que la variable F1	Se usa con la media F7 ; es la medida de dispersión más utilizada F1
Coeficiente de variación (CV)	Desviación típica dividida por la media y multiplicada por 100 F1	Compara la dispersión de variables con distintas unidades F1

Con datos simétricos se describen la **media y la desviación típica**; con datos asimétricos son más adecuadas la **mediana** y la amplitud, junto con cuartiles y percentiles **F1**. En la misma línea, las variables cuantitativas discontinuas o continuas sin distribución normal se expresan como **mediana y rango intercuartílico** (también es correcto dar el mínimo y el máximo) **F18**.

Ni la mediana dispersa ni el RIC va con la media

La **mediana** es una medida de **centralización** (y de posición), no de dispersión: rango, varianza, desviación típica y RIC sí son de dispersión **F1,F7**. El **rango intercuartílico** mide la amplitud del 50 % central y se acompaña de la **mediana**, no de la media; la desviación típica es la que acompaña a la media **F7**.

EIR 2025 · 209

P2 · respuesta en el solucionario

Medidas de dispersión: la mediana no lo es

En estadística descriptiva, ¿cuál de las siguientes medidas descriptivas NO es una medida de dispersión?:

1. Varianza.
2. Desviación típica.
3. Rango.
4. Mediana.

Rango intercuartílico: se acompaña de la mediana

Señale la opción INCORRECTA en relación al rango intercuartílico en estadística descriptiva:

1. Se define como la diferencia entre el tercer y el primer cuartil (percentiles 75 y 25, respectivamente).
2. Indica la amplitud del 50% central de la muestra.
3. Se usa como medida de dispersión cuando la variable presenta valores extremos.
4. Suele ir acompañado de la media como medida de tendencia central.

Propiedades de la varianza que caen: las diferencias respecto a la media se elevan al cuadrado para eliminar los números negativos ^{F7}, de modo que la varianza **nunca es negativa** ^{F36}, y vale **0** cuando todos los datos son iguales ^{F1,F36}. En la práctica, como interesa hacer inferencias a la población, en el denominador de la desviación típica se usa $n - 1$ en lugar de n ^{F1}. Si se **suma una constante** a a todos los valores ($X + a$), la nueva **media** es la media más a , mientras que la **varianza no cambia**, $v(X + a) = v(X)$, porque la transformación traslada los datos pero mantiene las diferencias; si se **multiplican** todos por una constante (aX), la varianza sí cambia, con la constante al cuadrado: $v(aX) = a^2 \cdot v(X)$ ^{F36}.



Sumar o restar una constante

Si a todos los valores se les resta 4, la **media disminuye 4** y la **varianza no cambia** (tampoco la desviación típica: $d(X + a) = d(X)$): los datos se trasladan, pero se mantienen las diferencias ^{F36}.

Efecto de restar una constante sobre media y varianza

¿De qué manera se ven afectadas la media y la varianza de una distribución, si a todos los valores de la muestra se les resta 4?:

1. La media disminuye 4 y la varianza no varía.
2. La media disminuye 4 y la varianza 16.
3. La media y varianza no varían.
4. La media no varía y la varianza sí.

4. Representación gráfica

El gráfico adecuado depende del **tipo de variable**. Para variables **categóricas** (sexo, estadio TNM, profesión) se representa la frecuencia y el porcentaje de cada categoría con **diagramas de barras o de sectores** ^{F2}; el diagrama de sectores puede calcularse para cualquier tipo de variable, aunque suele aplicarse a variables **cualitativas** ^{F36}. Recuerda que las cualitativas son **nominales** u ordinales, y que «discreta» y «continua» son tipos de variable **cuantitativa** ^{F36}. Para variables numéricas continuas el más usado es el **histograma** ^{F2}.

Gráfico	Para qué variable o dato	Cómo es
Barras	Frecuencias de una variable cualitativa o discreta ^{F3} ; también variables numéricas discretas con pocos valores ^{F2}	Una barra por categoría, con altura proporcional a su frecuencia o porcentaje ^{F2}
Sectores («tarta»)	Frecuencias relativas de una variable cualitativa o discreta ^{F3} ; variables categóricas ^{F2}	Círculo dividido en tantas porciones como categorías, cada una con un arco proporcional a su frecuencia ^{F2} ; útil con pocas categorías ^{F3}
Pictograma	Frecuencias de una variable cualitativa o discreta ^{F3}	Representa las frecuencias mediante figuras o símbolos : símbolos cuyas áreas representan las frecuencias, o barras formadas por símbolos ^{F3} ; presentación visual con dibujos del tema estudiado ^{F4}
Histograma	Frecuencias de una variable cuantitativa continua ^{F3}	Rectángulos contiguos (sin separación) sobre los intervalos de clase, con área proporcional a la frecuencia ^{F3,F4}
Polígono de frecuencias	Variable cuantitativa	Línea que une los puntos medios de los extremos superiores de las barras del histograma ^{F2,F3}
Líneas	Tendencias a lo largo del tiempo ^{F2,F3}	Puntos conectados por rectas; el tiempo va en el eje horizontal ^{F3}
Diagrama de dispersión (nube de puntos)	Relación entre dos variables cuantitativas ^{F2,F4}	Cada individuo es un punto con coordenadas (x, y); la forma de la nube muestra la relación ^{F4,F3}
Diagrama de cajas (box plot)	Variable numérica, sobre todo si se aleja de la normal; comparación de grupos ^{F2}	Caja, mediana, bigotes y valores extremos (ver abajo) ^{F2,F8}
Barras de error	Comparar la media de una variable continua entre grupos ^{F2}	Media de cada grupo con su intervalo de confianza del 95 % ^{F2}

Gráfico	Para qué variable o dato	Cómo es
Pirámide de población	Estructura por sexo y edad ^{F3}	Histograma bidireccional ^{F3}

★ Qué gráfico cae

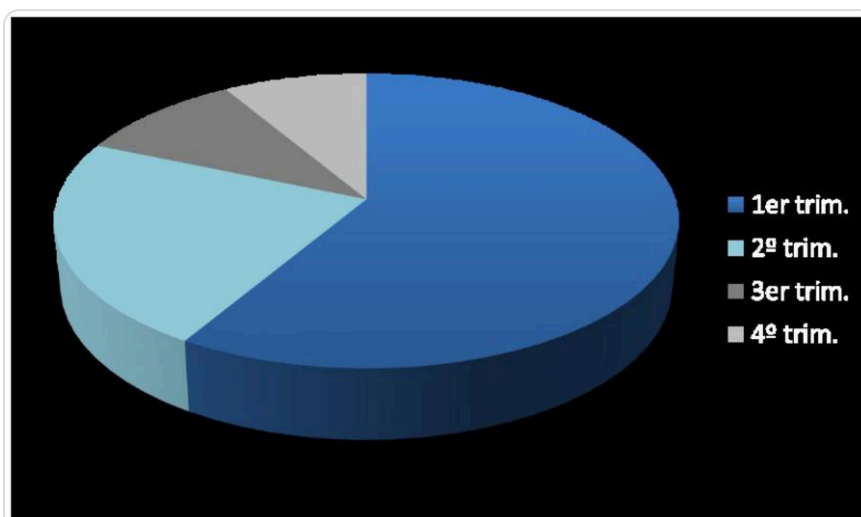
Sectores y barras → **cualitativas** (y también cuantitativas discretas); **histograma** → **cuantitativa continua**; **dispersión** → **relación entre dos cuantitativas**; **líneas** → **evolución temporal**; **pictograma** → **dibujos o símbolos** cuya área o número representa la frecuencia ^{F2,F3}.

EIR 2021 · 1

P12 · respuesta en el solucionario

Gráfico para variables cualitativas nominales (imagen)

Pregunta asociada a la imagen 1. Identifique con qué tipo de variables se utiliza este modelo de gráfico:



1. Cuantitativa continua.
2. Cuantitativa ordinal.
3. Cualitativa nominal.
4. Cualitativa discreta.

EIR 2005 · 16

P34 · respuesta en el solucionario

Pictogramas

Los pictogramas:

1. Son gráficos que reflejan una evolución temporal.
2. Es un tipo de diagrama para representar las variabilidades.
3. Se emplean para mostrar la información mediante un dibujo alusivo al colectivo que se estudia, con el área (o su altura) proporcional a la frecuencia que representa.
4. Algunos pictogramas se utilizan en el cálculo de la covarianza.
5. Son cambios de variable.

Relación entre dos variables cuantitativas: diagrama de dispersión

En un estudio se quiere determinar si existe relación entre el colesterol total en sangre y la edad, ¿qué gráfico se utilizará para mostrar la relación entre ambas variables? :

1. Histograma.
2. Diagrama de barras.
3. Diagrama de dispersión.
4. Diagrama de sectores.

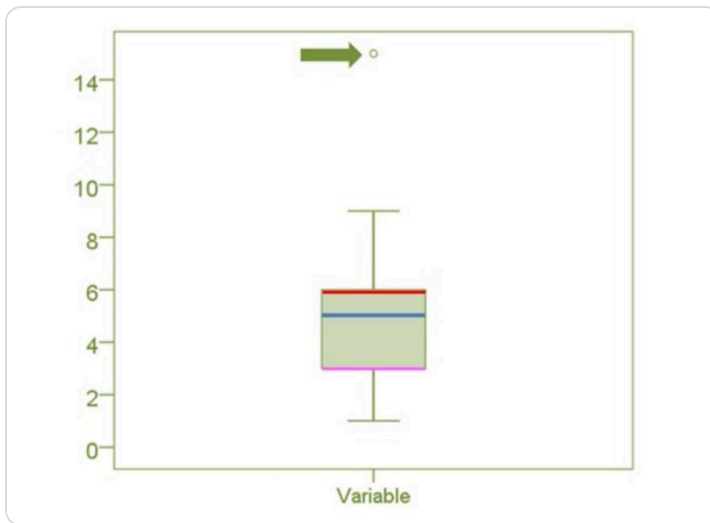
El **diagrama de cajas** resume una variable numérica mediante percentiles. La **caja central** indica el rango en el que se concentra el **50 % central** de los datos; sus extremos son el **primer y el tercer cuartil** ^{F2}. La **línea central de la caja es la mediana**: si la variable es simétrica, queda en el centro de la caja ^{F2}. Los **bigotes** llegan a los valores que delimitan el 95 % central de los datos, aunque a veces coinciden con los valores extremos ^{F2}; en la versión del CDC llegan al mínimo y al máximo ^{F8}. Las observaciones que quedan fuera se dibujan aparte como **valores atípicos o extremos** (outliers) ^{F2}. En la variante con **vallas** del manual del NIST, el bigote llega hasta el valor más alejado que no supera $Q1 - 1,5 \times RIC$ o $Q3 + 1,5 \times RIC$, y los puntos más allá se dibujan como **círculos**: más allá de 1,5 RIC son atípicos leves y más allá de 3 RIC, atípicos extremos ^{F31,F32}. Por eso un punto aislado por encima del bigote superior es un valor extremo de la variable.

La línea de la caja no es la media

En un diagrama de cajas, la línea horizontal **dentro** de la caja es la **mediana**, no la media ^{F2,F8}. En una caja vertical, de abajo arriba: bigote inferior, **borde inferior de la caja = $Q1$** , **línea interior = mediana**, **borde superior de la caja = $Q3$** , bigote superior y, aislados, los **valores extremos** ^{F2,F31}.

Interpretación de gráfico con la media (imagen)

Pregunta asociada a la imagen 5. Identifique la opción INCORRECTA:



1. La línea roja se corresponde con el Q3.
2. La línea azul se corresponde con la media.
3. La línea rosa se corresponde con el Q1.
4. El símbolo señalado con la flecha se corresponde con el valor extremo de la variable a estudio.

5. Probabilidad

La **probabilidad** de un suceso es la proporción de veces que ocurriría si se repitiese la observación un gran número de veces en condiciones similares (definición frecuentista); se mide entre **0** (suceso que no ocurre nunca) y **1** (suceso que ocurre siempre) ^{F13}.

Regla	Fórmula	Cuándo se usa ^{F13}
Suceso complementario	$P(\text{no } A) = 1 - P(A)$	Probabilidad de que no ocurra A
Suma para sucesos incompatibles	$P(A \text{ o } B) = P(A) + P(B)$	Sucesos mutuamente excluyentes , que no pueden darse a la vez (cara o cruz en una moneda)
Suma general (unión)	$P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B)$	Sucesos que pueden darse a la vez: se resta la intersección para no contarla dos veces
Producto (intersección)	$P(A \text{ y } B) = P(A) \cdot P(B/A)$	Probabilidad de que ocurran los dos
Producto para sucesos independientes	$P(A \text{ y } B) = P(A) \cdot P(B)$	El resultado de uno no afecta al otro

«O» suma, «Y» multiplica

«A o B» se resuelve **sumando** (y restando lo común si pueden coincidir); «A y B» se resuelve **multiplicando** si son independientes ^{F13}.

Con las reglas de la tabla ^{F13} se resuelven los cálculos típicos (la aritmética es propia): si se conocen $P(A)$, $P(B)$ y la probabilidad de tener **al menos uno** (la unión), la de tener **los dos** se despeja de la suma general: $P(A \text{ y } B) = P(A) + P(B) - P(A \text{ o } B)$. Para que dos sucesos **independientes** ocurran a la vez se multiplican sus probabilidades; por eso, para que **dos personas** independientes tengan ambas un rasgo A, la probabilidad es $P(A) \cdot P(A) = P(A)^2$. La probabilidad de que ocurra **solo uno** de dos sucesos independientes es la suma de los dos casos incompatibles: $P(A) \cdot [1 - P(B)] + [1 - P(A)] \cdot P(B)$. Así lo resuelve el manual *Introductory Statistics 2e* de OpenStax: al sacar dos bolas **con reemplazamiento** (las extracciones se consideran entonces independientes) de una urna con 3 rojas y 8 azules, la probabilidad de obtener **una roja y una azul**, en cualquier orden, es $P(RB \text{ o } BR) = (3/11)(8/11) + (8/11)(3/11) = 48/121$ ^{F35}. Si se pide que se cumpla «ambas personas tienen A» o «ambas tienen B», se suman $P(A)^2$ y $P(B)^2$ y se resta la probabilidad de que se cumplan las dos cosas, que, suponiendo que los dos rasgos sean independientes entre sí, es $P(A)^2 \cdot P(B)^2$.

Ejemplo con monedas (cálculo con las reglas anteriores): al lanzar dos monedas independientes, la probabilidad de dos caras es $1/2 \cdot 1/2 = 1/4$ y la de dos cruces también $1/4$; como «dos caras» y «dos cruces» son incompatibles, la de obtener una u otra es $1/4 + 1/4 = 1/2$ ^{F13}. En genética, dos personas no afectadas que son **portadoras** de una copia del gen alterado de un trastorno **autosómico recesivo** tienen un **25 %** de probabilidad, en cada embarazo, de tener un hijo afectado ^{F14}.

Probabilidad de la intersección de sucesos

Si la probabilidad de tener la enfermedad A es del 5%, la de tener la enfermedad B es del 10% y la de tener al menos una de las dos es del 13%, ¿Cuál es la probabilidad de tener las dos?:

1. 0%.
2. 1%.
3. 2%.
4. 8%.

Probabilidad de hijo enfermo con padres portadores de gen recesivo

Los dos miembros de una pareja son ambos portadores de un gen recesivo para una determinada enfermedad. ¿Cuál es la probabilidad de tener un hijo enfermo?

1. 1.
2. 0,75.
3. 0,5.
4. 0,25.

Probabilidad con dos monedas

Se tiran dos monedas regulares al mismo tiempo, determinar la probabilidad de obtener 2 caras ó 2 cruces:

1. 1/4.
2. 1/6.
3. 1/3.
4. 1/2.
5. 2/3.

Probabilidad de sucesos independientes (sensores de puertas)

Un pequeño hospital tiene dos puertas de emergencia, que se abren automáticamente mediante sensores independientes. La probabilidad de funcionar el primer sensor bajo una avería es del 99% y del 95% para el segundo. Si se produjera una avería, ¿cuál es la probabilidad de que solo uno actuase?:

1. 9,5%.
2. 95%
3. 45%
4. 5,1%
5. 5,9%

Probabilidad de la unión de sucesos (rubio u ojos negros)

La probabilidad de ser rubio, es 0,4 y la probabilidad de ojos negros, es 0,3. Determinar la probabilidad de que dos personas sean rubias o tengan los ojos negros:

1. 0,2256
2. 0,4023
3. 0,5156
4. 0,2356
5. 0,2346

6. Inferencia: intervalo de confianza y contraste de hipótesis

La inferencia tiene dos herramientas: la **estimación** de parámetros (con intervalos de confianza) y el **contraste de hipótesis** ^{F16}. El papel del azar debe contemplarse siempre, realizando test de hipótesis o construyendo **intervalos de confianza para conocer la precisión** de la estimación ^{F16}. La **estimación puntual** es el mejor valor único (por ejemplo, una media o una odds ratio); el **intervalo de confianza (IC)** describe la **incertidumbre** de esa estimación: un rango de valores en el que se puede estar razonablemente seguro de que está el verdadero efecto ^{F15}.

El IC de una media se construye con el **error estándar de la media** (desviación típica dividida por la raíz de n), que mide la variabilidad de las medias de muestras repetidas de la misma población: **IC 95 % = media $\pm$ 1,96 $\times$ error estándar** ^{F7}. No hay que confundirlo con la **desviación típica**, que describe la variabilidad de los datos: en una distribución normal, la media $\pm$ 1,96 desviaciones típicas incluye el 95 % de las **observaciones** ^{F7}. Los IC se calculan también para una **proporción**, un riesgo relativo o una odds ratio, no solo para medias ^{F7}. Se pueden dar con cualquier nivel de confianza, aunque lo más habitual es el **95 %** y a veces el **90 %** o el **99 %** ^{F15,F4}. Su **amplitud** depende sobre todo del **tamaño de la muestra** (más muestra, IC más estrecho) y de la variabilidad de la medida ^{F15}.

Interpretación. Un IC del 95 % se interpreta a menudo como un rango en el que se puede estar **95 % seguro de que se encuentra el verdadero valor** (de la población); es una interpretación aproximada, útil como guía ^{F15}. La interpretación **estrictamente correcta** es hipotética: **si el estudio se repitiera infinitas veces** y cada vez se calculara un IC del 95 %, **el 95 % de esos intervalos contendría el verdadero valor** ^{F15}. El CDC da la misma definición estadística y añade que en un estudio concreto se interpreta como el rango de valores compatibles con los datos ^{F7}. En los dos casos el IC se refiere al **parámetro de la población** de la que se extrajo la muestra, no a la media de la muestra (que es conocida) ni al porcentaje de sujetos ^{F7,F15}.

Dos formas de decirlo en el examen

Si entre las opciones está la **interpretación estricta** («si se repite el estudio muchas veces, el 95 % de los intervalos incluirá el valor poblacional»), esa es la correcta; «la media poblacional tiene un 95 % de probabilidad de estar en este intervalo» es solo la versión aproximada ^{F15}. Si no aparece, vale la formulación aproximada **referida a la población**: «con una seguridad del 95 %, la media de la **población** de la que procede la muestra está entre...» ^{F15}. Nunca es correcto referirlo a la media de la **muestra** ni al 95 % de los **sujetos**, que es lo que abarca la media $\pm$ 1,96 desviaciones típicas ^{F7,F15}.

Interpretación del intervalo de confianza del 95 %

En un estudio se calcula la media de glucemia capilar en una muestra de 200 pacientes de una consulta. La media es de 130 mg/dl, desviación estándar de 20 y un Intervalo de Confianza (IC) al 95% (125-136). ¿Cómo se interpretaría el IC?:

1. Con una seguridad del 95% la media de glucemia capilar en la muestra de la población, se encuentra entre 125 y 136 mg/dl.
2. Con una seguridad del 95% existen diferencias estadísticamente significativas en la media de glucemia capilar de la muestra.
3. Con una seguridad del 95% no existen diferencias estadísticamente significativas en la media de glucemia capilar de la muestra.
4. Con una seguridad del 95% la media de glucemia capilar en la población en la que se extrajo la muestra, se encuentra entre 125 y 136 mg/dl.

Interpretación del intervalo de confianza del 95 %

El intervalo de confianza (IC) del 95 % calculado para la media aritmética, indica:

1. Que la media de la población tiene un 95 % de probabilidad de encontrarse dentro de ese rango.
2. Que el 95 % de los sujetos se sitúa dentro del intervalo calculado.
3. Que, si se repite el estudio muchas veces, en el 95 % de los casos el intervalo incluirá la media poblacional.
4. Que existe un 95 % de probabilidad de que la hipótesis nula sea verdadera.

Intervalo de confianza: evalúa la precisión de la estimación

Señala la respuesta correcta respecto al intervalo de confianza (IC):

1. Permite evaluar la precisión con que se ha estimado un parámetro poblacional.
2. Se calcula siempre con una confianza del 95%.
3. Nos indica la estimación puntual de una variable en una muestra.
4. Es aplicable únicamente en estimaciones de variables cuantitativas.

El **contraste de hipótesis** compara dos hipótesis: la **nula (H_0)**, que afirma que no hay asociación entre las variables, y la **alternativa (H_a)**, que afirma que hay algún grado de asociación ^{F16}. Al decidir se pueden cometer dos errores ^{F16,F17}:

	H_0 verdadera (no hay diferencia)	H_0 falsa (sí hay diferencia)
Se rechaza H_0	Error tipo I o α (falso positivo): rechazar H_0 cuando es verdadera ^{F16}	Acierto. Su probabilidad es la potencia ($1 - \beta$) : rechazar H_0 cuando es realmente falsa ^{F17}

	H0 verdadera (no hay diferencia)	H0 falsa (sí hay diferencia)
No se rechaza H0	Acierto	Error tipo II o β (falso negativo): aceptar H0 cuando es falsa ^{F16}

El **nivel de significación** es el riesgo que el investigador asume voluntariamente de equivocarse al rechazar H0 cuando es cierta; se fija normalmente en **0,05 o 0,01** ^{F16}. Si se disminuye α , disminuye también la potencia: reducir el error tipo I aumenta el tipo II ^{F17}. La **potencia** es la capacidad del estudio para encontrar diferencias si realmente las hay; aumenta con el tamaño del efecto a detectar y con el tamaño de la muestra, y disminuye cuanto mayor es la variabilidad de la respuesta ^{F17}.

El **valor de p**: Pita Fernández y Pértega Díaz (Fisterra) lo definen como la probabilidad de aceptar la hipótesis alternativa como cierta cuando la cierta podría ser la nula; es decir, **la probabilidad de rechazar la hipótesis nula cuando en realidad es verdadera** (error de tipo I) ^{F16}. Si $p < 0,05$ se considera significativo y se rechaza H0; si $p > 0,05$ no se rechaza ^{F16}. La p **no mide la fuerza de la asociación ni su importancia**: depende de la magnitud de la diferencia y del tamaño de la muestra, y con muestra suficiente cualquier diferencia puede resultar significativa ^{F16}. La significación estadística tampoco equivale a relevancia clínica ^{F16}. El IC del 95 % excluye el valor nulo (por ejemplo, una odds ratio de 1) **si y solo si** la p es menor de 0,05 ^{F15}.

La definición técnica de p

El Manual Cochrane define la p con más rigor: es la probabilidad de obtener el efecto observado (**o uno mayor**) si la hipótesis nula fuera cierta ^{F15}. En el EIR se ha preguntado con la definición que da Fisterra: la p indica **la probabilidad de rechazar la hipótesis nula cuando es verdadera**, y se relaciona con el error tipo I ^{F16}. No es la probabilidad de aceptar H0 siendo cierta ni la potencia ($1 - \beta$) ^{F16,F17}.

EIR 2019 · 51

P17 · respuesta en el solucionario

Valor p: significado

En una prueba estadística el valor de p indica:

1. La probabilidad de aceptar la hipótesis nula cuando es verdadera.
2. La probabilidad de rechazarla hipótesis alternativa cuando es verdadera.
3. La potencia estadística.
4. La probabilidad de rechazar la hipótesis nula cuando es verdadera.

Error tipo I (alfa): rechazar H_0 siendo cierta

El error de Tipo I es:

1. Equivalente al error típico.
2. Un error secuencial y sistemático de la varianza.
3. El que se comete al aceptar la hipótesis nula.
4. Es el que se comete al rechazar la hipótesis nula siendo verdadera.
5. El error que se comete al considerar la tendencia, en la regresión.

Error tipo I (alfa): rechazar H_0 siendo cierta

En términos probabilísticos, ¿qué significa obtener un error tipo I o alfa?:

1. Aceptar H_0 siendo cierta.
2. Rechazar H_0 siendo cierta.
3. Rechazar H_0 siendo falsa.
4. No existe este tipo de error.

7. Elección de la prueba estadística

Para elegir la prueba hay que considerar tres aspectos: el **diseño** (cuántos grupos y si son independientes o relacionados), el **número de mediciones** y la **escala de medición** de las variables ^{F18}. Las pruebas se dividen en **paramétricas** y **no paramétricas**: las paramétricas corresponden a variables cuantitativas continuas y **solo se pueden usar si los datos siguen una distribución normal** ^{F18} (con un número suficiente de observaciones, digamos más de 30, se justifica igualmente la t de Student ^{F20}); las no paramétricas se usan con variables cualitativas, cuantitativas discontinuas o cuantitativas sin distribución normal ^{F18}.

Muestras independientes o relacionadas (apareadas). Son independientes cuando los grupos están formados por sujetos distintos (por ejemplo, dos tratamientos diferentes) ^{F18}. Las series **dependientes o apareadas** surgen normalmente cuando se evalúa un mismo dato **más de una vez en cada sujeto** (antes y después); también en estudios de casos y controles en los que cada caso se **empareja individualmente** con un control ^{F20}.

Comprobar las condiciones de las pruebas paramétricas. La **normalidad** se comprueba con métodos gráficos (histograma, diagrama de cajas, gráficos de normalidad) o con pruebas como **Kolmogorov-Smirnov** y **Shapiro-Wilk** ^{F18,F20}. La prueba de **Shapiro-Wilk** es la más apropiada para **muestras pequeñas (< 50)**, y la de **Kolmogorov-Smirnov** se usa para $n \geq 50$ ^{F21}. La t de Student clásica para muestras independientes exige además **igualdad de varianzas (homocedasticidad)**; si las varianzas son distintas se usa una modificación, el **test de Welch** ^{F20}. Para comparar dos varianzas se usa la **prueba F** o de la razón de varianzas, cuyo estadístico (el cociente de las dos varianzas) sigue una **distribución F de Snedecor**; es una prueba paramétrica, que supone que las dos poblaciones siguen una distribución normal ^{F20,F22}. La **prueba de Levene** contrasta la igualdad de varianzas entre grupos (su H_0 es que todas son iguales) y sirve para verificar ese supuesto en pruebas como el ANOVA; es menos sensible que la de Bartlett a la falta de normalidad ^{F23}. Como en cualquier contraste, si su p es mayor de 0,05 no se rechaza H_0 y se asume que las varianzas son iguales; si es menor, se rechaza ^{F16}.

EIR 2022 · 51

P9 · respuesta en el solucionario

Comparación de varianzas: F de Fisher-Snedecor

¿Qué prueba paramétrica se utilizará para determinar la varianza?:

1. Test de Wilcoxon.
2. U de Mann-Whitney.
3. F de Fisher-Snedecor.
4. Test de Kruskal-Wallis.

Objetivo (adaptado del cuadro 1 de ^{F18})	Variables y distribución	Muestras relacionadas	Muestras independientes
Comparar 2 promedios	Cuantitativa con distribución normal	t de Student (apareada)	t de Student

Objetivo (adaptado del cuadro 1 de ^{F18})	Variables y distribución	Muestras relacionadas	Muestras independientes
Comparar 2 promedios	Cuantitativa discontinua o continua sin distribución normal	Wilcoxon	U de Mann-Whitney
Comparar 3 o más grupos	Cuantitativa con distribución normal	ANOVA de 2 vías	ANOVA de una vía
Comparar 3 o más grupos	Cuantitativa discontinua o continua sin distribución normal	Friedman	Kruskal-Wallis
Comparar 2 grupos	Cualitativa nominal u ordinal	McNemar	Ji cuadrado (nota del cuadro: prueba exacta de Fisher «si algún grupo tiene valor < 5»; el criterio detallado, con frecuencias esperadas, está más abajo)
Comparar 3 o más grupos	Cualitativa nominal u ordinal	Q de Cochran	Ji cuadrado
Correlación de 2 variables	Cuantitativas con distribución normal / cuantitativas sin distribución normal	—	Pearson / Spearman

La tabla adapta el cuadro de selección de pruebas de Flores-Ruiz y cols. ^{F18}. Comparar **2 grupos** en una variable **cuantitativa continua con distribución normal** es, dicho de otra forma, **comparar promedios entre 2 grupos**, y se elige una prueba **t de Student**, para muestras independientes o relacionadas ^{F18}. Para valorar la asociación entre una variable **cuantitativa** y una **categorica**, que clasifica a los individuos en tantos grupos como categorías tiene, se comparan las medias de la cuantitativa en esos grupos: si la categorica es **dicotómica** (solo dos categorías), la comparación de medias entre dos grupos independientes se hace con la **t de Student**; si tiene tres o más, con el **ANOVA**; y ambas exigen, entre otros requisitos, que la variable cuantitativa tenga **distribución normal** en los grupos y homogeneidad de varianzas ^{F34}. Para comparar **3 o más medias** se usa el **ANOVA**: de una vía con 3 o más grupos independientes y de 2 vías con muestras relacionadas medidas 3 o más veces ^{F18}. Cuando los datos no siguen una distribución normal se comparan las medianas con pruebas no paramétricas: **Wilcoxon** para muestras relacionadas (antes y después), **U de Mann-Whitney** para 2 grupos independientes, **Kruskal-Wallis** para 3 o más grupos independientes (equivalente no paramétrico del ANOVA de una vía) y **Friedman** para 3 o más muestras relacionadas ^{F18}.

Los dos Wilcoxon

Hay dos pruebas con el nombre de Wilcoxon. La **suma de rangos de Wilcoxon** es para **dos muestras independientes** y es esencialmente equivalente a la **U de Mann-Whitney**; la **suma de rangos con signo de Wilcoxon** es para **dos muestras relacionadas** ^{F19}. Cuando el cuadro de selección dice solo «Wilcoxon», se refiere a muestras **relacionadas** ^{F18}. Kruskal-Wallis es la extensión de Mann-Whitney a más de dos grupos, igual que el ANOVA extiende la t de Student ^{F19}.

Paramétrica → no paramétrica

t independiente → **Mann-Whitney**; **t apareada** → **Wilcoxon**; **ANOVA de una vía** → **Kruskal-Wallis**; **ANOVA de 2 vías (muestras relacionadas)** → **Friedman**; **ji cuadrado** → **McNemar** si las muestras están relacionadas ^{F18,F19}.

EIR 2019 · 55

P15 · respuesta en el solucionario

t de Student: variable cuantitativa normal frente a cualitativa dicotómica

Siquisiéramos medir la efectividad del seguimiento de pacientes en los programas de cuidados paliativos especializados domiciliarios (variable independiente “seguimiento en cuidados paliativos”: si/no) en términos de la media de permanencia en casa (variable dependiente “media de días sin ingreso hospitalario”). ¿Qué test estadístico utilizaríamos?:

1. T de Student.
2. Ji cuadrado.
3. Wilcoxon.
4. McNemar.

EIR 2021 · 44

P11 · respuesta en el solucionario

t de Student: variable cuantitativa normal frente a cualitativa dicotómica

En la comparación de una variable cualitativa de dos categorías y otra de tipo cuantitativa en la que se conoce una distribución normal, la prueba a utilizar será:

1. Chi cuadrado.
2. ANOVA.
3. t de Student.
4. Correlación de Pearson.

ANOVA: comparación de medias en tres grupos independientes

En los estudios de investigación cuantitativa, ¿qué test se utilizaría para la comparación de medias entre tres grupos independientes, asumiendo normalidad?:

1. Test de la t de Student.
2. Test de ANOVA.
3. Test de Kruskal-Wallis.
4. Test de Chi Cuadrado.

Test de Kruskal-Wallis: no paramétrico para más de 2 grupos independientes

Cuando se quiere estudiar si existe relación entre una variable cuantitativa en tres grupos diferentes de individuos, ¿qué test no paramétrico se debería utilizar:

1. Test de Kruskal-Wallis.
2. Test de Anova.
3. Test de la suma de rangos de Wilcoxon.
4. Test de Student.

Test de Kruskal-Wallis: no paramétrico para más de 2 grupos independientes

¿Cuál sería la prueba no paramétrica en el caso de muestras independientes >2 grupos?:

1. Kruskal-Wallis.
2. U de Mann-Whitney.
3. Wilcoxon.
4. Friedman.

Test de Friedman: más de dos grupos apareados no normales

¿Qué test se utilizaría para la comparación de medias entre tres o más grupos apareados, en los que la distribución de los datos cuantitativos no es normal?:

1. Test de Friedman.
2. Test de ANOVA.
3. Test de Kruskal-Wallis.
4. Test de Wilcoxon.

Prueba no paramétrica cuantitativa frente a dicotómica: Wilcoxon/U de Mann-Whitney

Identifique la prueba correcta para el análisis estadístico teniendo en cuenta que se enfrenta una variable cuantitativa con una cualitativa dicotómica, que no hay un buen ajuste con la distribución normal y que las variables están apareadas:

1. U de Mann Whitney.
2. Wilcoxon.
3. t de Student.
4. Friedman.

Variables cualitativas. La **ji cuadrado** (χ^2) determina si dos variables cualitativas están asociadas, comparando las frecuencias **observadas** con las **esperadas** si fuesen independientes; es la prueba estándar para grupos independientes ^{F24}. Condición: al menos el **80 % de las celdas** deben tener valores esperados **mayores de 5**; en una tabla 2×2 deberían cumplirlo todas, aunque en la práctica se tolera que una quede ligeramente por debajo ^{F24,F25}. Si no se cumple, se usa la **prueba exacta de Fisher**, que analiza si dos variables dicotómicas están asociadas cuando la muestra es demasiado pequeña ^{F25}. En tablas 2×2 con muestras reducidas se puede aplicar la **corrección de Yates**, sobre la que no hay consenso en la literatura ^{F24}; Flores-Ruiz y cols. la aplican cuando el número de datos es menor de 30 ^{F18}. Si las muestras están relacionadas, se usa **McNemar** ^{F18,F25}.

La ji cuadrado se aplica para comprobar la **independencia** o asociación entre dos variables y la **bondad de ajuste** (si la distribución observada se ajusta a la esperada) ^{F39}. La prueba de **homogeneidad** solo se diferencia de la de independencia en la hipótesis nula, que afirma la homogeneidad o igualdad: se usa para determinar si **dos o más muestras independientes** difieren en su distribución de una misma variable ^{F39}. Se dice que varias poblaciones son **homogéneas** si son idénticas y, por extensión, que los datos de sus muestras también lo son ^{F4}.

Shapiro-Wilk para muestras pequeñas

Error típico de examen: invertir las pruebas de normalidad. **Shapiro-Wilk** → **muestras pequeñas (< 50)**; **Kolmogorov-Smirnov** → **n ≥ 50** ^{F21}. En la pregunta EIR2023-025 (¿cuál es INCORRECTA?) la respuesta oficial es la afirmación que dice lo contrario. Del resto de afirmaciones de esa pregunta, la de Levene coincide con lo expuesto arriba ^{F23} y la de Fisher es compatible en lo esencial, aunque Fistera describe Fisher para tablas 2×2 y con el umbral del 20 % de celdas ^{F24,F25}. **Discrepancia:** la clave da por correcta la afirmación de que las muestras son apareadas **únicamente** si las mediciones se hacen en los mismos individuos, mientras que la fuente considera también apareados los **casos y controles emparejados** individualmente ^{F20}. Para el examen, la respuesta oficial sigue siendo la de las pruebas de normalidad invertidas.

Bondad de ajuste a la normal en variables cuantitativas

A la hora de realizar el análisis estadístico hay que tener en cuenta varios conceptos. Identifique la respuesta INCORRECTA:

1. Dos variables son apareadas o emparejadas únicamente si se relacionan de manera directa (mismos individuos en distintos medios).
2. En una tabla de contingencia de $2 \times n$, con frecuencias esperadas inferiores a 5 en más de una celda (25%), se debe recurrir a la prueba exacta de Fisher.
3. Con variables cuantitativas es necesario evaluar la bondad de ajuste con la distribución normal mediante la prueba Shapiro-Wilk para muestras grandes y la de Kolmogorov-Smirnov para muestras más pequeñas.
4. La t de Student para muestras independientes se debe acompañar de la prueba de Levene para evaluar si hay homocedasticidad ($p > 0,05$, varianzas iguales) o heterocedasticidad ($p < 0,05$, varianzas diferentes).

Contraste de homogeneidad: aplicación de la ji-cuadrado

Un contraste de homogeneidad es:

1. Una función monótona creciente.
2. Una aplicación de la distribución Ji-cuadrado para ver si varias muestras proceden de la misma población.
3. Una aplicación de la Ji-cuadrado para el contraste de la varianza.
4. Una aplicación de la distribución T-student para el contraste de la varianza.
5. Una aplicación de la independencia estocástica.

8. Correlación y regresión

La **covarianza** mide la relación entre las dos componentes de una variable bidimensional, y solo tiene sentido si ambas son cuantitativas ^{F37}. Si los valores grandes de una variable tienden a ir con valores grandes de la otra (y los pequeños con pequeños), la relación es **directa** y la covarianza es **positiva**; si los grandes van con los pequeños, la relación es **inversa** y la covarianza es **negativa** ^{F37}. **Al contrario que la varianza, la covarianza puede ser negativa** ^{F37}; la varianza es la media de los cuadrados de las diferencias con la media ^{F1}.

¿La varianza «siempre positiva»?

En la pregunta EIR2005-020, la respuesta (reconstruida por un tercero) da por correcta la afirmación «la varianza siempre es positiva y la covarianza no». Según la fuente, la varianza **nunca puede ser negativa**, pues es una suma de cuadrados, pero **vale 0** si y solo si todos los valores de la muestra son iguales ^{F36}; la covarianza, en cambio, **sí puede ser negativa** ^{F37}. En sentido estricto, la varianza no es «siempre positiva», sino nunca negativa. Para el examen vale la respuesta de la clave: lo que contrasta la afirmación es que la covarianza puede ser negativa y la varianza no.

EIR 2005 · 20

P30 · respuesta en el solucionario

Varianza siempre positiva, covarianza no

La varianza siempre es positiva y la covarianza no:

1. Son valores complementarios.
2. La afirmación es correcta.
3. La covarianza también es positiva en las distribuciones discretas.
4. Coinciden sus valores en las distribuciones simétricas.
5. Una varianza negativa significaría que no hay correlación entre las variables.

El **coeficiente de correlación lineal de Pearson (r)** es el cociente entre la covarianza y el producto de las desviaciones típicas de las dos variables ^{F37}: la covarianza tiene dimensión (en el ejemplo de Fisterra, kg por cm) y se divide por las desviaciones típicas para que no la tenga y sea un coeficiente; su valor es independiente de cualquier unidad usada para medir las variables ^{F27} y **oscila entre -1 y $+1$** ^{F27,F37}. Una correlación **próxima a cero** indica que no hay relación **lineal** ^{F27}, pero una covarianza nula no significa que las variables no estén relacionadas: puede haber una relación no lineal ^{F37}. El signo de r es el de la covarianza: positivo, relación directa; negativo, inversa ^{F37}. El **coeficiente de determinación (r^2)** indica qué porcentaje de la variabilidad de una variable se explica por la otra ^{F27}. Si alguna de las variables no sigue una distribución normal, o si son ordinales, se usa el coeficiente de **Spearman** ^{F18}. La correlación **no implica causalidad** ^{F27}.

La **regresión** permite predecir una variable a partir de otra. La recta de regresión de **Y sobre X** es $Y^* - \bar{y} = [\text{cov}(X, Y) / v(X)] \cdot (X - \bar{x})$; la de **X sobre Y** se obtiene intercambiando X e Y: $X^* - \bar{x} = [\text{cov}(X, Y) / v(Y)] \cdot (Y - \bar{y})$. Las dos rectas **no coinciden en general** ^{F37}. Posición relativa de las dos rectas ^{F37}:

Valor de r	Rectas de regresión Y/X y X/Y ^{F37}
$r = 0$ ($r^2 = 0$)	Para que r^2 valga 0 la covarianza tiene que anularse, y entonces la recta de regresión es una constante ^{F37} . Sustituyendo $\text{cov}(X, Y) = 0$ en las dos fórmulas de arriba ^{F37} (cálculo propio, no escrito en la fuente), la de Y sobre X queda $Y^* = \bar{y}$ y la de X sobre Y, $X^* = \bar{x}$.
$r = \pm 1$ ($r^2 = 1$)	Las dos rectas de regresión coinciden ^{F37} .
$0 < r < 1$	Cuanto más se acerca $ r $ a 1, mejor es la relación lineal (por debajo de $-0,7$ o por encima de $0,7$, buena; cerca de 0, muy mala) ^{F37} .



Correlación cero

Si $r = 0$, la covarianza es nula y la recta de regresión es una constante ^{F37}; sustituyendo $\text{cov}(X, Y) = 0$ en sus fórmulas (cálculo propio), $Y^* = \bar{y}$ (Y sobre X) y $X^* = \bar{x}$ (X sobre Y). Si $r = \pm 1$, las dos rectas **coinciden** ^{F37}.

EIR 2005 · 19

P31 · respuesta en el solucionario

Correlación cero: rectas de regresión perpendiculares

Si el coeficiente de correlación es de valor cero:

1. Las rectas de regresión de y/x y de x/y se cortan perpendicularmente.
2. La afirmación no es correcta pues nunca toma este valor.
3. Las rectas de regresión de y/x y de x/y , no se cruzan.
4. Coinciden ambas rectas en las distribuciones simétricas.
5. La probabilidad del suceso complementario, es también nula.

En la **regresión lineal** la variable respuesta o dependiente (Y) es cuantitativa y se relaciona con una variable explicativa, predictora o independiente (X) mediante una recta $Y = a + bX$, donde b es la pendiente ^{F30}. La **regresión logística** funciona de forma parecida, pero con una variable respuesta **binaria** (dicotómica), y se usa para obtener **odds ratios** en presencia de más de una variable explicativa: el resultado es el efecto de cada variable sobre la odds ratio del suceso de interés ^{F26}. Se emplea también cuando el protocolo incluye variables de confusión ^{F18}. Otros modelos dan otras medidas de efecto ajustadas: la **regresión logística** da **odds ratios**, la **regresión de Poisson** da **razones de tasas** y el **modelo de riesgos proporcionales de Cox**, en datos de supervivencia, estima el **hazard ratio** ^{F33}.

Odds ratio: regresión logística

¿Qué modelo de regresión se debería utilizar si se tiene como medida de asociación una Odds Ratio?:

1. Regresión lineal.
2. Regresión logística.
3. Regresión de Cox.
4. Regresión de Poisson.

Todas las preguntas del tema (38)

De la más reciente a la más antigua. Las respuestas, en el solucionario.

EIR 2026 · 202

P1 · respuesta en el solucionario

El intervalo de confianza (IC) del 95 % calculado para la media aritmética, indica:

1. Que la media de la población tiene un 95 % de probabilidad de encontrarse dentro de ese rango.
2. Que el 95 % de los sujetos se sitúa dentro del intervalo calculado.
3. Que, si se repite el estudio muchas veces, en el 95 % de los casos el intervalo incluirá la media poblacional.
4. Que existe un 95 % de probabilidad de que la hipótesis nula sea verdadera.

EIR 2025 · 209

P2 · respuesta en el solucionario

En estadística descriptiva, ¿cuál de las siguientes medidas descriptivas NO es una medida de dispersión?:

1. Varianza.
2. Desviación típica.
3. Rango.
4. Mediana.

EIR 2024 · 80

P3 · respuesta en el solucionario

¿Qué test se utilizaría para la comparación de medias entre tres o más grupos apareados, en los que la distribución de los datos cuantitativos no es normal?:

1. Test de Friedman.
2. Test de ANOVA.
3. Test de Kruskal-Wallis.
4. Test de Wilcoxon.

EIR 2024 · 76

P4 · respuesta en el solucionario

En los estudios de investigación cuantitativa, ¿qué test se utilizaría para la comparación de medias entre tres grupos independientes, asumiendo normalidad?:

1. Test de la t de Student.
2. Test de ANOVA.
3. Test de Kruskal-Wallis.
4. Test de Chi Cuadrado.

A la hora de realizar el análisis estadístico hay que tener en cuenta varios conceptos. Identifique la respuesta INCORRECTA:

1. Dos variables son apareadas o emparejadas únicamente si se relacionan de manera directa (mismos individuos en distintos medios).
2. En una tabla de contingencia de $2 \times n$, con frecuencias esperadas inferiores a 5 en más de una celda (25%), se debe recurrir a la prueba exacta de Fisher.
3. Con variables cuantitativas es necesario evaluar la bondad de ajuste con la distribución normal mediante la prueba Shapiro-Wilk para muestras grandes y la de Kolmogorov-Smirnov para muestras más pequeñas.
4. La t de Student para muestras independientes se debe acompañar de la prueba de Levene para evaluar si hay homocedasticidad ($p > 0,05$, varianzas iguales) o heterocedasticidad ($p < 0,05$, varianzas diferentes).

Identifique la prueba correcta para el análisis estadístico teniendo en cuenta que se enfrenta una variable cuantitativa con una cualitativa dicotómica, que no hay un buen ajuste con la distribución normal y que las variables están apareadas:

1. U de Mann Whitney.
2. Wilcoxon.
3. t de Student.
4. Friedman.

¿Cuál sería la prueba no paramétrica en el caso de muestras independientes > 2 grupos?:

1. Kruskal-Wallis.
2. U de Mann-Whitney.
3. Wilcoxon.
4. Friedman.

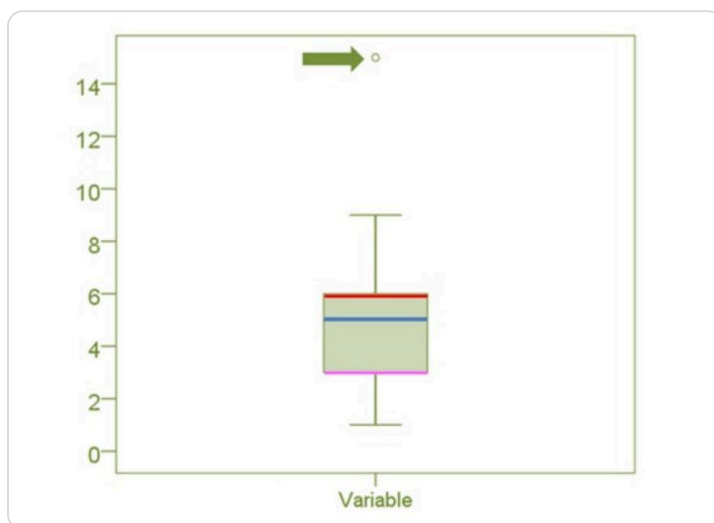
Respecto al cálculo del número de sujetos necesarios para la realización de un estudio cuyo objetivo es la estimación de una media o una proporción, señale qué elemento NO interviene:

1. Variabilidad del parámetro que se desea estimar.
2. Precisión con la que se desea estimar el parámetro.
3. Potencia estadística o capacidad del estudio para detectar una determinada diferencia.
4. Nivel de confianza deseado.

¿Qué prueba paramétrica se utilizará para determinar la varianza?:

1. Test de Wilcoxon.
2. U de Mann-Whitney.
3. F de Fisher-Snedecor.
4. Test de Kruskal-Wallis.

Pregunta asociada a la imagen 5. Identifique la opción INCORRECTA:

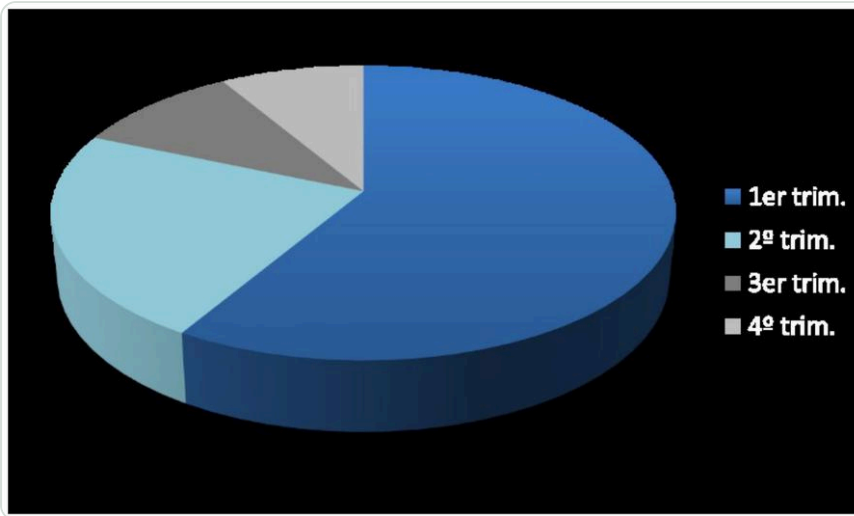


1. La línea roja se corresponde con el Q3.
2. La línea azul se corresponde con la media.
3. La línea rosa se corresponde con el Q1.
4. El símbolo señalado con la flecha se corresponde con el valor extremo de la variable a estudio.

En la comparación de una variable cualitativa de dos categorías y otra de tipo cuantitativa en la que se conoce una distribución normal, la prueba a utilizar será:

1. Chi cuadrado.
2. ANOVA.
3. t de Student.
4. Correlación de Pearson.

Pregunta asociada a la imagen 1. Identifique con qué tipo de variables se utiliza este modelo de gráfico:



1. Cuantitativa continua.
2. Cuantitativa ordinal.
3. Cualitativa nominal.
4. Cualitativa discreta.

Señale la opción INCORRECTA en relación al rango intercuartílico en estadística descriptiva:

1. Se define como la diferencia entre el tercer y el primer cuartil (percentiles 75 y 25, respectivamente).
2. Indica la amplitud del 50% central de la muestra.
3. Se usa como medida de dispersión cuando la variable presenta valores extremos.
4. Suele ir acompañado de la media como medida de tendencia central.

El número de hijos de una familia es una variable:

1. Cualitativa dicotómica.
2. Cuantitativa continua.
3. Cuantitativa discreta.
4. Cuantitativa nominal.

Si quisiéramos medir la efectividad del seguimiento de pacientes en los programas de cuidados paliativos especializados domiciliarios (variable independiente “seguimiento en cuidados paliativos”: sí/no) en términos de la media de permanencia en casa (variable dependiente “media de días sin ingreso hospitalario”). ¿Qué test estadístico utilizaríamos?:

1. T de Student.
2. Ji cuadrado.
3. Wilcoxon.
4. McNemar.

Señala la respuesta correcta respecto al intervalo de confianza (IC):

1. Permite evaluar la precisión con que se ha estimado un parámetro poblacional.
2. Se calcula siempre con una confianza del 95%.
3. Nos indica la estimación puntual de una variable en una muestra.
4. Es aplicable únicamente en estimaciones de variables cuantitativas.

En una prueba estadística el valor de p indica:

1. La probabilidad de aceptar la hipótesis nula cuando es verdadera.
2. La probabilidad de rechazarla hipótesis alternativa cuando es verdadera.
3. La potencia estadística.
4. La probabilidad de rechazar la hipótesis nula cuando es verdadera.

Cuando se quiere estudiar si existe relación entre una variable cuantitativa en tres grupos diferentes de individuos, ¿qué test no paramétrico se debería utilizar:

1. Test de Kruskal-Wallis.
2. Test de Anova.
3. Test de la suma de rangos de Wilcoxon.
4. Test de Student.

En un estudio se quiere determinar si existe relación entre el colesterol total en sangre y la edad, ¿qué gráfico se utilizará para mostrar la relación entre ambas variables? :

1. Histograma.
2. Diagrama de barras.
3. Diagrama de dispersión.
4. Diagrama de sectores.

En un estudio se calcula la media de glucemia capilar en una muestra de 200 pacientes de una consulta. La media es de 130 mg/dl, desviación estándar de 20 y un Intervalo de Confianza (IC) al 95% (125-136). ¿Cómo se interpretaría el IC?:

1. Con una seguridad del 95% la media de glucemia capilar en la muestra de la población, se encuentra entre 125 y 136 mg/dl.
2. Con una seguridad del 95% existen diferencias estadísticamente significativas en la media de glucemia capilar de la muestra.
3. Con una seguridad del 95% no existen diferencias estadísticamente significativas en la media de glucemia capilar de la muestra.
4. Con una seguridad del 95% la media de glucemia capilar en la población en la que se extrajo la muestra, se encuentra entre 125 y 136 mg/dl.

La variable “grupo sanguíneo”, ¿qué tipo de variable es?:

1. Cuantitativa discreta.
2. Cuantitativa continua.
3. Cualitativa nominal.
4. Cualitativa ordinal.

¿Qué modelo de regresión se debería utilizar si se tiene como medida de asociación una Odds Ratio?:

1. Regresión lineal.
2. Regresión logística.
3. Regresión de Cox.
4. Regresión de Poisson.

Para el diseño muestral de una investigación se decide seleccionar aleatoriamente un número de distritos de una ciudad. Posteriormente, sobre estos distritos se eligen un número de secciones censales al azar, de las cuales se incluyen para el estudio todas las personas que viven en dichas secciones. Se trata de:

1. Un muestreo polietápico.
2. Un muestreo sistemático.
3. Un muestreo estratificado polietápico.
4. Un muestreo semiprobabilístico.

Si la probabilidad de tener la enfermedad A es del 5%, la de tener la enfermedad B es del 10% y la de tener al menos una de las dos es del 13%, ¿Cuál es la probabilidad de tener las dos?:

1. 0%.
2. 1%.
3. 2%.
4. 8%.

¿Qué es el muestreo aleatorio simple?:

1. Es la técnica que decide el nº de individuos que tienen que participar en el estudio.
2. Es la técnica de selección de sujetos donde no interviene el azar.
3. Es la técnica aleatoria de selección de variables de estudio.
4. Es aquella técnica en la que cada unidad del marco muestral tiene la misma probabilidad de ser escogida.

¿De qué manera se ven afectadas la media y la varianza de una distribución, si a todos los valores de la muestra se les resta 4?:

1. La media disminuye 4 y la varianza no varía.
2. La media disminuye 4 y la varianza 16.
3. La media y varianza no varían.
4. La media no varía y la varianza sí.

Los dos miembros de una pareja son ambos portadores de un gen recesivo para una determinada enfermedad. ¿Cuál es la probabilidad de tener un hijo enfermo?

1. 1.
2. 0,75.
3. 0,5.
4. 0,25.

En términos probabilísticos, ¿qué significa obtener un error tipo I o alfa?:

1. Aceptar H_0 siendo cierta.
2. Rechazar H_0 siendo cierta.
3. Rechazar H_0 siendo falsa.
4. No existe este tipo de error.

En una investigación la variable “nivel socioeconómico”, especificada en los posibles valores como bajo, medio-bajo, medio-medio, medioalto y alto, es un tipo de variable:

1. Cuantitativa.
2. Discreta.
3. Continua.
4. Diferencial.
5. Cualitativa.

La varianza siempre es positiva y la covarianza no:

1. Son valores complementarios.
2. La afirmación es correcta.
3. La covarianza también es positiva en las distribuciones discretas.
4. Coinciden sus valores en las distribuciones simétricas.
5. Una varianza negativa significaría que no hay correlación entre las variables.

Si el coeficiente de correlación es de valor cero:

1. Las rectas de regresión de y/x y de x/y se cortan perpendicularmente.
2. La afirmación no es correcta pues nunca toma este valor.
3. Las rectas de regresión de y/x y de x/y , no se cruzan.
4. Coinciden ambas rectas en las distribuciones simétricas.
5. La probabilidad del suceso complementario, es también nula.

Se tiran dos monedas regulares al mismo tiempo, determinar la probabilidad de obtener 2 caras ó 2 cruces:

1. $1/4$.
2. $1/6$.
3. $1/3$.
4. $1/2$.
5. $2/3$.

El percentil 50:

1. Son valores referentes a las desviaciones típicas.
2. Es un tipo de desviación típica para distribuciones continuas.
3. Coincide con la moda en cualquier distribución.
4. Tiene el mismo valor que la mediana, cuartil segundo y decil quinto.
5. Significa que una distribución está sesgada en su punto central.

Los pictogramas:

1. Son gráficos que reflejan una evolución temporal.
2. Es un tipo de diagrama para representar las variabilidades.
3. Se emplean para mostrar la información mediante un dibujo alusivo al colectivo que se estudia, con el área (o su altura) proporcional a la frecuencia que representa.
4. Algunos pictogramas se utilizan en el cálculo de la covarianza.
5. Son cambios de variable.

La probabilidad de ser rubio, es 0,4 y la probabilidad de ojos negros, es 0,3. Determinar la probabilidad de que dos personas sean rubias o tengan los ojos negros:

1. 0,2256
2. 0,4023
3. 0,5156
4. 0,2356
5. 0,2346

Un pequeño hospital tiene dos puertas de emergencia, que se abren automáticamente mediante sensores independientes. La probabilidad de funcionar el primer sensor bajo una avería es del 99% y del 95% para el segundo. Si se produjera una avería, ¿cuál es la probabilidad de que solo uno actuase?:

1. 9,5%.
2. 95%
3. 45%
4. 5,1%
5. 5,9%

Un contraste de homogeneidad es:

1. Una función monótona creciente.
2. Una aplicación de la distribución Ji-cuadrado para ver si varias muestras proceden de la misma población.
3. Una aplicación de la Ji-cuadrado para el contraste de la varianza.
4. Una aplicación de la distribución T-student para el contraste de la varianza.
5. Una aplicación de la independencia estocástica.

El error de Tipo I es:

1. Equivalente al error típico.
2. Un error secuencial y sistemático de la varianza.
3. El que se comete al aceptar la hipótesis nula.
4. Es el que se comete al rechazar la hipótesis nula siendo verdadera.
5. El error que se comete al considerar la tendencia, en la regresión.

Preguntas que también podrían ser de este tema (1)

La clasificación automática dudó entre este tema y otro.

EIR 2021 · 45

P39 · respuesta en el solucionario

La Población definida por la pregunta de investigación a la que se generalizan los resultados obtenidos en un estudio, se denomina:

1. Población accesible.
2. Muestra del estudio.
3. Población diana.
4. Sujetos que cumplen los criterios de selección.

Solucionario

Nº	Pregunta	Respuesta oficial	Notas
P1	EIR 2026 · 202	3	
P2	EIR 2025 · 209	4	
P3	EIR 2024 · 80	1	
P4	EIR 2024 · 76	2	
P5	EIR 2023 · 25	3	hoja del archivo oficial
P6	EIR 2023 · 24	2	hoja del archivo oficial
P7	EIR 2022 · 202	1	
P8	EIR 2022 · 52	3	
P9	EIR 2022 · 51	3	
P10	EIR 2022 · 5	2	
P11	EIR 2021 · 44	3	
P12	EIR 2021 · 1	3	
P13	EIR 2020 · 137	4	
P14	EIR 2020 · 135	3	
P15	EIR 2019 · 55	1	hoja del archivo oficial
P16	EIR 2019 · 52	1	hoja del archivo oficial
P17	EIR 2019 · 51	4	hoja del archivo oficial
P18	EIR 2018 · 163	1	hoja del archivo oficial
P19	EIR 2018 · 162	3	hoja del archivo oficial
P20	EIR 2018 · 161	4	hoja del archivo oficial
P21	EIR 2018 · 159	3	hoja del archivo oficial
P22	EIR 2018 · 158	2	hoja del archivo oficial
P23	EIR 2018 · 147	1	hoja del archivo oficial
P24	EIR 2017 · 192	3	
P25	EIR 2017 · 21	4	

N°	Pregunta	Respuesta oficial	Notas
P26	EIR 2016 · 209	1	hoja del archivo oficial
P27	EIR 2016 · 208	4	hoja del archivo oficial
P28	EIR 2016 · 206	2	hoja del archivo oficial
P29	EIR 2011 · 77	5	
P30	EIR 2005 · 20	2	reconstruida por un tercero
P31	EIR 2005 · 19	1	reconstruida por un tercero
P32	EIR 2005 · 18	4	reconstruida por un tercero
P33	EIR 2005 · 17	4	reconstruida por un tercero
P34	EIR 2005 · 16	3	reconstruida por un tercero
P35	EIR 2004 · 75	4	reconstruida por un tercero
P36	EIR 2004 · 74	5	reconstruida por un tercero
P37	EIR 2004 · 72	2	reconstruida por un tercero
P38	EIR 2004 · 71	4	reconstruida por un tercero
P39	EIR 2021 · 45	3	

Fuentes

- **F1** Pita Fernández S, Pérttega Díaz S. Estadística descriptiva de los datos. Fistera (Metodología de la investigación); publicado en: Manual de metodología de trabajo en atención primaria. Jarpyo; 1997. Actualizado el 06/03/2001. · <https://www.fistera.com/formacion/metodologia-investigacion/estadistica-descriptiva-datos/> · consultada 2026-09-25
- **F2** Pérttega Díaz S, Pita Fernández S. Representación gráfica en el análisis de datos. Fistera (Metodología de la investigación); remitido a Cad Aten Primaria (pendiente de publicación según la propia página). Fecha en Fistera: 02/04/2001. · <https://www.fistera.com/formacion/metodologia-investigacion/representacion-grafica-analisis-datos/> · consultada 2026-09-25
- **F3** Instituto Nacional de Estadística (INE). INE Explica: Tipos de gráficos (barras, pirámide, líneas, Pareto, sectores, pictograma, dispersión, cartograma). Versión disponible en 2026. · https://www.ine.es/explica/docs/pasos_tipos_graficos.pdf · consultada 2026-09-25
- **F4** OCDE. OECD Glossary of Statistical Terms. París: OECD Publishing; 2008 (ISBN 978-92-64-02556-1). Entradas: Target population, Simple random sampling, Systematic sample, Stratified sample, Cluster sampling, Multi-stage sampling, Quartile, Range, Interquartile range, Homogeneity, Histogram, Scatter diagram, Pictogram, Confidence interval. · https://www.oecd.org/content/dam/oecd/en/publications/reports/2008/09/oecd-glossary-of-statistical-terms_g1gh9ad7/9789264055087-en.pdf · consultada 2026-10-01
- **F5** Arias-Gómez J, Villasís-Keever MÁ, Miranda-Navales MG. El protocolo de investigación III: la población de estudio. Rev Alerg Mex. 2016;63(2):201-206. doi:10.29262/ram.v63i2.181. · <https://doi.org/10.29262/ram.v63i2.181> · consultada 2026-09-25
- **F6** Pita Fernández S. Determinación del tamaño muestral. Fistera (Metodología de la investigación); publicado en Cad Aten Primaria 1996; 3: 138-141. Fecha en Fistera: 01/12/2010. · <https://www.fistera.com/formacion/metodologia-investigacion/determinacion-tamano-muestral/> · consultada 2026-09-25
- **F7** Centers for Disease Control and Prevention (CDC). Principles of Epidemiology in Public Health Practice, 3.^a ed. Lesson 2: Summarizing Data, Section 7: Measures of Spread (incluye error estándar e intervalo de confianza de la media). Last reviewed: 18 de mayo de 2012 (archivo oficial del CDC); página archivada, sin actualizar desde 2012. · https://archive.cdc.gov/www_cdc.gov/csels/dsepd/ss1978/lesson2/section7.html · consultada 2026-10-01
- **F8** Centers for Disease Control and Prevention (CDC). Principles of Epidemiology in Public Health Practice, 3.^a ed. Lesson 4: Displaying Public Health Data, Section 4: Other Data Displays (dot plots and box plots). Last reviewed: 18 de mayo de 2012 (archivo oficial del CDC); página archivada, sin actualizar desde 2012. · https://archive.cdc.gov/www_cdc.gov/csels/dsepd/ss1978/lesson4/section4.html · consultada 2026-09-25
- **F13** Pérttega Díaz S, Pita Fernández S. Cálculo de probabilidades: nociones básicas. Fistera (Metodología de la investigación). Fecha en Fistera: 28/06/2004. · <https://www.fistera.com/formacion/metodologia-investigacion/calculo-probabilidades-nociones-basicas/> · consultada 2026-09-25
- **F14** MedlinePlus Genetics (National Library of Medicine, EE. UU.). If a genetic disorder runs in my family, what are the chances that my children will have the condition? Última actualización: 12 de mayo de 2021. · <https://medlineplus.gov/genetics/understanding/inheritance/riskassessment/> · consultada 2026-09-25
- **F15** Schünemann HJ, Vist GE, Higgins JPT, et al. Chapter 15: Interpreting results and drawing conclusions (última actualización: agosto de 2023). En: Higgins JPT, Thomas J, et al. (eds.). Cochrane Handbook for Systematic Reviews of Interventions, versión 6.5. Cochrane; 2024. Apartados 15.3.1 y 15.3.2. ·

<https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-15> · consultada 2026-09-25

- **F16** Pita Fernández S, Pértega Díaz S. Significancia estadística y relevancia clínica. Fistera (Metodología de la investigación). Fecha en Fistera: 01/12/2010. · <https://www.fistera.com/formacion/metodologia-investigacion/significancia-estadistica-relevancia-clinica/> · consultada 2026-09-25
- **F17** Pértega Díaz S, Pita Fernández S. Cálculo del poder estadístico de un estudio. Fistera (Metodología de la investigación). Fecha en Fistera: 20/01/2003. · <https://www.fistera.com/formacion/metodologia-investigacion/calculo-poder-estadistico-estudio/> · consultada 2026-09-25
- **F18** Flores-Ruiz E, Miranda-Novales MG, Villasís-Keever MÁ. El protocolo de investigación VI: cómo elegir la prueba estadística adecuada. Estadística inferencial. Rev Alerg Mex. 2017;64(3):364-370 (cuadro 1). doi:10.29262/ram.v64i3.304. · <https://doi.org/10.29262/ram.v64i3.304> · consultada 2026-09-25
- **F19** Pértega Díaz S, Pita Fernández S. Métodos no paramétricos para la comparación de dos muestras. Fistera (Metodología de la investigación). Fecha en Fistera: 17/08/2007. · <https://www.fistera.com/formacion/metodologia-investigacion/metodos-no-parametricos-para-comparacion-dos-muestras/> · consultada 2026-09-25
- **F20** Pértega Díaz S, Pita Fernández S. Métodos paramétricos para la comparación de dos medias. t de Student. Fistera (Metodología de la investigación); Cad Aten Primaria 2001; 8: 37-41. Fecha en Fistera: 23/03/2001. · <https://www.fistera.com/formacion/metodologia-investigacion/metodos-parametricos-para-comparacion-dos-medias-t-student/> · consultada 2026-09-25
- **F21** Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. Ann Card Anaesth. 2019;22(1):67-72. doi:10.4103/aca.ACA_157_18. Texto completo abierto en PubMed Central (PMC6350423). · https://doi.org/10.4103/aca.ACA_157_18 · consultada 2026-09-25
- **F22** NIST/SEMATECH e-Handbook of Statistical Methods. 1.3.5.9. F-Test for Equality of Two Variances. National Institute of Standards and Technology (EE. UU.); versión web vigente en 2026. · <https://www.itl.nist.gov/div898/handbook/eda/section3/eda359.htm> · consultada 2026-09-25
- **F23** NIST/SEMATECH e-Handbook of Statistical Methods. 1.3.5.10. Levene Test for Equality of Variances. National Institute of Standards and Technology (EE. UU.); versión web vigente en 2026. · <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm> · consultada 2026-09-25
- **F24** Pita Fernández S, Pértega Díaz S. Asociación de variables cualitativas: test de chi-cuadrado. Fistera (Metodología de la investigación). Fecha en Fistera: 03/11/2004. · <https://www.fistera.com/formacion/metodologia-investigacion/asociacion-variables-cualitativas-test-chi-cuadrado/> · consultada 2026-09-25
- **F25** Pértega Díaz S, Pita Fernández S. Asociación de variables cualitativas: el test exacto de Fisher y el test de McNemar. Fistera (Metodología de la investigación). Fecha en Fistera: 14/11/2004. · <https://www.fistera.com/formacion/metodologia-investigacion/asociacion-variables-cualitativas-test-exacto-fisher-test-mcnemar/> · consultada 2026-09-25
- **F26** Sperandei S. Understanding logistic regression analysis. Biochem Med (Zagreb). 2014;24(1):12-18. doi:10.11613/BM.2014.003. · <https://doi.org/10.11613/BM.2014.003> · consultada 2026-09-25
- **F27** Pita Fernández S, Pértega Díaz S. Relación entre variables cuantitativas. Fistera (Metodología de la investigación). Fecha en Fistera: 30/03/2001. · <https://www.fistera.com/formacion/metodologia-investigacion/relacion-entre-variables-cuantitativas/> · consultada 2026-09-25
- **F28** Centers for Disease Control and Prevention (CDC). Principles of Epidemiology in Public Health Practice, 3.ª ed. Lesson 2: Summarizing Data, Section 2: Types of Variables. Archivo oficial del CDC (revisado en 2012);

página archivada, sin actualizar desde 2012. ·

https://archive.cdc.gov/www_cdc.gov/csels/dsepd/ss1978/lesson2/section2.html · consultada 2026-09-25

- **F29** Centers for Disease Control and Prevention (CDC). Principles of Epidemiology in Public Health Practice, 3.ª ed. Lesson 2: Summarizing Data, Section 6: Measures of Central Location. Archivo oficial del CDC (revisado en 2012); página archivada, sin actualizar desde 2012. ·
https://archive.cdc.gov/www_cdc.gov/csels/dsepd/ss1978/lesson2/section6.html · consultada 2026-09-25
- **F30** Pérttega Díaz S, Pita Fernández S. Técnicas de regresión: regresión lineal simple. Fistera (Metodología de la investigación). Fecha en Fistera: 20/08/2001. · <https://www.fistera.com/formacion/metodologia-investigacion/tecnicas-regresion-regresion-lineal-simple/> · consultada 2026-09-25
- **F31** NIST/SEMATECH e-Handbook of Statistical Methods. 1.3.3.7. Box Plot (incluye «Box plots with fences»). National Institute of Standards and Technology (EE. UU.); versión web vigente en 2026. ·
<https://www.itl.nist.gov/div898/handbook/eda/section3/boxplot.htm> · consultada 2026-09-25
- **F32** NIST/SEMATECH e-Handbook of Statistical Methods. 7.1.6. What are outliers in the data? National Institute of Standards and Technology (EE. UU.); versión web vigente en 2026. ·
<https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm> · consultada 2026-09-25
- **F33** Higgins JPT, Li T, Deeks JJ (eds.). Chapter 6: Choosing effect measures and computing estimates of effect (última actualización: agosto de 2023). En: Cochrane Handbook for Systematic Reviews of Interventions, versión 6.5. Cochrane; 2024. · <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-06> · consultada 2026-09-25
- **F34** Rubio Hurtado MJ, Berlanga Silvente V. Cómo aplicar las pruebas paramétricas bivariadas t de Student y ANOVA en SPSS. Caso práctico. REIRE, Revista d'Innovació i Recerca en Educació (Universitat de Barcelona). 2012;5(2):83-100. doi:10.1344/reire2012.5.2527. Apartado 3: Condiciones de parametricidad. ·
<https://doi.org/10.1344/reire2012.5.2527> · consultada 2026-09-28
- **F35** Illowsky B, Dean S. Introductory Statistics 2e. Houston (Texas): OpenStax, Rice University; 13 de diciembre de 2023. Capítulo 3: apartado 3.2 (Independent and Mutually Exclusive Events: el muestreo con reemplazamiento da sucesos independientes) y apartado 3.5, ejemplo 3.24 c (P(RB OR BR) con reemplazamiento). · <https://openstax.org/books/introductory-statistics-2e/pages/3-5-tree-and-venn-diagrams> · consultada 2026-09-28
- **F36** Miranda Menéndez P. Estadística Básica. Madrid: Ediciones Complutense; 2025. Capítulo 3: Estadística descriptiva unidimensional, pp. 57-101 (clasificación de variables, diagrama de sectores, cuartiles y deciles, propiedades de la media, la varianza y la desviación típica ante $X + a$ y aX). doi:10.5209/docm.003.03. Libro de acceso abierto en Docta, repositorio oficial de la Universidad Complutense. ·
<https://doi.org/10.5209/docm.003.03> · consultada 2026-10-01
- **F37** Miranda Menéndez P. Estadística Básica. Madrid: Ediciones Complutense; 2025. Capítulo 4: Estadística descriptiva bidimensional. Regresión lineal, pp. 103-133 (covarianza, rectas de regresión de Y sobre X y de X sobre Y, coeficientes de determinación y de correlación). doi:10.5209/docm.003.04. Libro de acceso abierto en Docta, repositorio oficial de la Universidad Complutense. · <https://doi.org/10.5209/docm.003.04> · consultada 2026-10-01
- **F38** National Institute of Standards and Technology (NIST). DATAPLOT Reference Manual, vol. 2 (LET Subcommands), cap. 2: DECILE, pp. 2-21 y 2-22. 18 de marzo de 1997. ·
<https://www.itl.nist.gov/div898/software/dataplot/refman2/ch2/decile.pdf> · consultada 2026-10-01
- **F39** Egbuchulem KI. The Karl Pearson's chi-square: a medical research libero, and a versatile test statistic: an editorial. Ann Ib Postgrad Med. 2024;22(2):5-8. PMID 40007702. Texto completo abierto en PubMed Central

Pendiente de verificar

- EIR2004-075 (respuesta reconstruida por un tercero): el enunciado «que dos personas sean rubias o tengan los ojos negros» es ambiguo; la cifra 0,2356 solo sale interpretando «ambas rubias o ambas con ojos negros» con independencia entre los dos rasgos ($0,16 + 0,09 - 0,0144$), y ninguna fuente del tema dice que el color del pelo y el de los ojos sean independientes: es un supuesto de la pregunta. Refuerzo del 28/09/2026: se buscó el cuadernillo y la plantilla oficial de 2004 (web del Ministerio de Sanidad, que solo publica las convocatorias recientes; `fuentes/oficial/` empieza en 2022) sin resultado; las copias de academias y colegios no valen. Sigue sin confirmar.
- Pasada «material de cursos fuera» (01/10/2026): se retiran como fuente los apuntes de Bioestadística de la Universidad de Granada (antes F9-F12), que son material de un curso. Sus datos se cotejaron en Miranda 2025 ^{F36,F37}, el glosario de la OCDE ^{F4}, el CDC ^{F7}, el manual DATAPLOT del NIST ^{F38} y el editorial de Egbuchulem 2024 ^{F39}. Quedan sin fuente válida y salen de la teoría: (a) que en la prueba de homogeneidad el investigador fija de antemano el tamaño de cada muestra (buscado en Fistera, glosario de la OCDE, manual del NIST, Miranda 2025 caps. 3, 4 y 8 y Egbuchulem 2024: no consta); (b) la palabra «perpendiculares» para las dos rectas de regresión con $r = 0$: solo la usan apuntes de cursos y webs de academias; el tema da, según Miranda 2025, que con covarianza nula la recta es una constante ($Y^* = \bar{y}$; $X^* = \bar{x}$); (c) el grupo sanguíneo llamado literalmente «nominal»: Fistera lo cita como cualitativa de más de dos categorías, junto al color de los ojos y la profesión, pero ninguna fuente válida leída lo llama nominal (la guía de resultados del IdISBa que lo hacía da 404 y no tiene copia archivada); (d) «nacionalidad» como ejemplo de nominal y la frase de que la moda no coincide con la mediana «salvo en casos particulares».