Estadística para la investigación

Pensar con datos, estimar con incertidumbre y comunicar con honestidad

Pablo Sánchez-Núñez
Universidad Carlos III de Madrid

26 de noviembre de 2026

Índice

Presentación 1
Usar R en este libro 3
Capítulo 1. Pensar con datos 5
Capítulo 2. Describir y visualizar 11
Capítulo 3. Estimar con incertidumbre 16
Capítulo 4. Contrastes y valores *p* 21
Capítulo 5. Tamaño del efecto y potencia 27
Capítulo 6. Regresión 32
Capítulo 7. Introducción a la inferencia bayesiana 37
Chuleta de R 42
Laboratorio interactivo 44
Guía para el profesorado 46
Glosario 47
Referencias 49
Bibliografía 49

Presentación

ℹ Ficha del libro

- **Autor:** Pablo Sánchez-Núñez
- **ORCID:** https://orcid.org/0000-0001-7845-9506
- **Web del autor:** https://pablosancheznunez.com
- **Afiliación:** Universidad Carlos III de Madrid
- **DOI:** https://doi.org/10.5281/zenodo.24534415
- **Licencia:** CC BY 4.0
- **Versión web interactiva:** https://pablosancheznunez.com/estadistica-investigacion/
- **Cómo citarlo:** Sánchez-Núñez, P. (2026). *Estadística para la investigación*. Universidad Carlos III de Madrid. Zenodo. https://doi.org/10.5281/zenodo.24534415

ℹ Sobre esta versión en PDF

Esta es la versión para leer e imprimir. El código de R aparece sin ejecutar, las preguntas de autoevaluación se muestran sin corrección automática y las actividades interactivas aparecen en papel, con su solución; los laboratorios solo funcionan en la web. Para ejecutar el código, hacer los ejercicios con corrección y responder a las preguntas, usa la versión web: https://pablosancheznunez.com/estadistica-investigacion/.

Casi todas las tesis acaban necesitando estadística, y casi todos los errores estadísticos publicados se deben a lo mismo: confundir lo que dicen los datos con lo que queremos que digan. Este libro no es un recetario de pruebas. Enseña a **pensar con datos**: formular bien la pregunta, describir antes de contrastar, **estimar con su incertidumbre**, interpretar un valor p sin mitos, planificar la muestra con la potencia y comunicar los resultados con honestidad.

Todo se practica en R, en tu navegador, con un conjunto de datos simulado de 120 doctorandos ficticios (horas de estudio, sueño, ansiedad y nota en un examen de métodos) y con laboratorios que simulan miles de estudios para que veas cómo se comportan los métodos.

ℹ Cómo citar este libro

Sánchez-Núñez, P. (2026). *Estadística para la investigación: pensar con datos, estimar con incertidumbre y comunicar con honestidad*. Universidad Carlos III de Madrid. Zenodo. https://doi.org/10.5281/zenodo.24534415

Autor: Pablo Sánchez-Núñez, Universidad Carlos III de Madrid
ORCID: 0000-0001-7845-9506
Web: pablosancheznunez.com

DOI: 10.5281/zenodo.24534415.
 Texto bajo licencia CC BY 4.0.

Mapa del libro

Capítulo	Pregunta
1. Pensar con datos	¿Qué pregunta hago y qué diseño puede responderla?
2. Describir y visualizar	¿Cómo son mis datos antes de cualquier prueba?
3. Estimar con incertidumbre	¿Qué rango de valores es compatible con mis datos?
4. Contrastes y valores p	¿Qué dice (y qué no) un valor p?
5. Tamaño del efecto y potencia	¿Cuánto importa el efecto y cuántos datos necesito?
6. Regresión	¿Cómo se relacionan varias variables a la vez?
7. Introducción a la inferencia bayesiana	¿Cómo actualizar lo que sabía con lo que observo?

Si nunca has usado R, empieza por Usar R en este libro; el apéndice Chuleta de R reúne todas las funciones del libro.

Tu progreso se guarda solo en tu navegador. El libro funciona sin conexión una vez visitado.

Usar R en este libro

Todos los capítulos tienen ejercicios de R que se ejecutan **en tu navegador**, sin instalar nada. Este capítulo te explica lo justo en quince minutos; si ya usas R, puedes saltarlo.

> **Objetivos**
>
> - Ejecutar y modificar código de R en el navegador.
> - Guardar valores y textos en objetos y trabajar con vectores.

Cómo funcionan las celdas

Cada bloque de código tiene un botón **Run Code**. Puedes cambiar el código y volver a ejecutarlo; **Start Over** recupera la versión original. La primera vez que abres una página, el navegador descarga el motor de R (unos segundos).

```
notas <- c(7.5, 6, 8.2, 4.9, 9.1)  # cinco notas
sum(notas)                          # total
mean(notas)                         # media
median(notas)                       # mediana
```

Con `<-` se guarda un valor en un objeto. Un **vector** reúne varios valores con `c()`. Los textos van entre comillas:

```
d <- read.csv("datos/estudio.csv")   # el conjunto de datos del libro
head(d)                              # primeras filas
d$nota[1:5]                          # una columna con $
nrow(d)                              # número de filas
```

Ejercicio 1

De 480 participantes, 24 abandonan el estudio. Calcula la proporción de abandonos.

> **Ejercicio: completa los huecos _____**
>
> ```
> pertinentes <- 24
> resultados <- _______
> pertinentes / resultados
> ```

> **Pista**
>
> Sustituye el hueco por el número total de participantes.

> **Solución**
>
> ```
> pertinentes <- 24
> resultados <- 480
> pertinentes / resultados
> ```
>
> 0,05: un 5 % de abandonos.

Ejercicio 2

¿Cuántas filas tiene el conjunto de datos del libro? Usa `nrow()`.

Ejercicio: completa los huecos _____

```
_____(d)
```

Pista

La función que cuenta filas es `nrow()`.

Solución

```
nrow(d)
```

120 doctorandos ficticios. `ncol(d)` daría el número de columnas.

Autoevaluación

Responde en el papel y comprueba después con las explicaciones del recuadro. En la versión web estas preguntas se corrigen automáticamente. No cuentan para la nota.

1. En R, los textos se escriben entre comillas. *[V / F]*
2. Para crear un vector con varios valores se usa la función *[length() / c() / toupper()]*
3. `length(c("a", "b", "c"))` devuelve _____
4. Para hacer los ejercicios del libro hay que instalar R en el ordenador. *[V / F]*

Explicaciones

- Sin comillas, R buscaría un objeto con ese nombre.
- `c()` combina valores en un vector.
- Tres elementos.
- Se ejecutan en el navegador con webR.

Capítulo 1. Pensar con datos

Objetivos

- Formular la pregunta de investigación según su tipo: descriptiva, exploratoria, inferencial, predictiva, causal o mecanicista.
- Clasificar variables y elegir el diseño que puede responder a la pregunta.
- Distinguir correlación de causalidad y reconocer las variables de confusión.

Glosario del capítulo

Población y muestra · Variable · Estudio observacional · Experimento aleatorizado · Variable de confusión · Validez interna y externa

Descargas

Datos del libro (CSV) Plantilla de entrega del reto (.qmd) Tema de estilo (.scss)

Lectura obligatoria

Leek, J. T.; Peng, R. D. (2015). What is the question? *Science*, 347(6228), 1314-1315.

1. Primero, la pregunta

Leek y Peng (2015) sostienen que el error estadístico más frecuente no es de cálculo, sino de **pregunta**: usar un análisis que responde a una pregunta distinta de la que se quería responder. Proponen seis tipos:

Tipo	Pregunta	Ejemplo
Descriptiva	¿Cómo son estos datos?	¿Cuántas horas estudian, de media, los doctorandos de la muestra?
Exploratoria	¿Qué patrones aparecen?	¿Parece que quienes duermen más sacan mejor nota?
Inferencial	¿Se mantiene el patrón en la población?	¿Existe esa relación en todos los doctorandos de la universidad?
Predictiva	¿Puedo predecir un valor para alguien nuevo?	¿Qué nota sacará una persona que estudia 20 horas?
Causal	¿Qué pasa si cambio una variable?	¿Mejora la nota una tutoría semanal?
Mecanicista	¿Por qué ocurre, paso a paso?	¿Por qué vía actúa la tutoría: más horas, menos ansiedad…?

Cada tipo exige más que el anterior. Una correlación responde a una pregunta exploratoria o inferencial, no a una causal.

2. Variables y escalas

Tipo	Valores	Ejemplos en nuestros datos	Resúmenes típicos
Categórica nominal	Categorías sin orden	`grupo` (tutoría, control)	Frecuencias, proporciones
Categórica ordinal	Categorías con orden	Nivel de acuerdo de 1 a 5	Mediana, percentiles
Cuantitativa discreta	Recuentos	Número de artículos publicados	Media, mediana
Cuantitativa continua	Cualquier valor en un intervalo	`horas`, `sueno`, `nota`	Media, desviación típica, cuantiles

```r
d <- read.csv("datos/estudio.csv")
str(d)
table(d$grupo)
```

3. Diseños: quién decide quién recibe qué

- En un **estudio observacional**, el investigador mide lo que ocurre sin intervenir. Puede describir y encontrar asociaciones, pero las **variables de confusión** (factores que influyen a la vez en la supuesta causa y en el efecto) amenazan cualquier conclusión causal.
- En un **experimento aleatorizado**, el azar decide quién recibe el tratamiento. Si la muestra es suficiente, la aleatorización equilibra en promedio las variables de confusión, **medidas y no medidas**: por eso permite inferir causas.
- Entre ambos están los **cuasiexperimentos** (diferencias en diferencias, regresión discontinua, variables instrumentales), que aprovechan situaciones naturales para aproximarse a una comparación justa.

> ! En nuestros datos
>
> `grupo` se asignó **al azar** (es un experimento simulado), así que la diferencia entre grupos estima el efecto causal de la tutoría. En cambio, `horas`, `sueno` y `ansiedad` se **observaron**: su relación con la nota es una asociación, que puede deberse a confusores (por ejemplo, la motivación).

4. Validez: interna y externa

La **validez interna** pregunta si la conclusión es correcta para la muestra estudiada (¿hay confusión, sesgo de medida, abandonos?). La **validez externa** pregunta si se puede generalizar a otras personas, contextos o momentos. Una muestra de doctorandos de una sola universidad puede tener buena validez interna y una validez externa limitada.

Ejercicio 1.1

¿Qué proporción del grupo de tutoría tiene una nota de 5 o más? Completa el código.

> i Ejercicio: completa los huecos `_____`
>
> ```r
> tut <- d[d$grupo == "_____", ]
> mean(tut$nota >= 5)
> ```

Pista

El grupo se llama `tutoria` (sin tilde).

Solución

```
tut <- d[d$grupo == "tutoria", ]
mean(tut$nota >= 5)
```

La media de un vector lógico es la proporción de `TRUE`. Es una pregunta **descriptiva** sobre la muestra.

Caso práctico resuelto

De la pregunta al análisis: un plan antes de mirar los datos

Pregunta: ¿duermen menos los doctorandos con más ansiedad?

1. **Tipo de pregunta:** asociación (no causal), en una muestra observacional.
2. **Variables:** ansiedad (escala continua), horas de sueño (continua); covariables posibles: horas de trabajo, grupo.
3. **Análisis previsto:** correlación y regresión lineal del sueño sobre la ansiedad, con intervalo de confianza.
4. **Decisiones fijadas antes:** exclusiones (cuestionarios incompletos), tratamiento de valores extremos, umbral de significación o, mejor, interpretación por el tamaño del efecto.
5. **Interpretación prevista:** «se asocia con», nunca «causa».

Escribir el plan antes de ver los datos (y, si es posible, preregistrarlo) protege frente al «jardín de los senderos que se bifurcan».

Ejercicios graduados

Básico. Clasifica la pregunta «¿Mejora la nota un taller de técnicas de estudio asignado al azar?»: descriptiva, de asociación o causal.

Explicaciones

Causal: hay una intervención asignada al azar, lo que permite atribuir las diferencias al taller.

Avanzado. Escribe un plan de análisis de cinco puntos para una pregunta de tu tesis.

Explicaciones

Debe incluir tipo de pregunta, variables y unidades, análisis principal, decisiones fijadas de antemano y cómo se interpretará un resultado nulo.

Actividades

Actividad: ¿Qué tipo de pregunta es?

Categorías: Descriptiva · Exploratoria · Inferencial · Predictiva · Causal · Mecanicista

- ¿Cuál es la nota media de los 120 participantes?
- ¿Qué variables parecen relacionadas con la nota?
- ¿La relación entre horas y nota se da en la población de doctorandos?
- ¿Qué nota sacará un nuevo doctorando según sus horas y su sueño?
- ¿La tutoría hace que suba la nota?
- ¿La tutoría mejora la nota porque reduce la ansiedad?

Solución

- ¿Cuál es la nota media de los 120 participantes? ⟶ **Descriptiva**
- ¿Qué variables parecen relacionadas con la nota? ⟶ **Exploratoria**
- ¿La relación entre horas y nota se da en la población de doctorandos? ⟶ **Inferencial**
- ¿Qué nota sacará un nuevo doctorando según sus horas y su sueño? ⟶ **Predictiva**
- ¿La tutoría hace que suba la nota? ⟶ **Causal**
- ¿La tutoría mejora la nota porque reduce la ansiedad? ⟶ **Mecanicista**

Actividad: ¿Qué tipo de variable es?

Categorías: Nominal · Ordinal · Discreta · Continua

- Universidad de procedencia
- Satisfacción con la dirección de tesis (de 1 a 5)
- Número de artículos publicados
- Horas de sueño
- Rama de conocimiento
- Tiempo hasta defender la tesis (en meses, con decimales)

Solución

- Universidad de procedencia ⟶ **Nominal**
- Satisfacción con la dirección de tesis (de 1 a 5) ⟶ **Ordinal**
- Número de artículos publicados ⟶ **Discreta**
- Horas de sueño ⟶ **Continua**
- Rama de conocimiento ⟶ **Nominal**
- Tiempo hasta defender la tesis (en meses, con decimales) ⟶ **Continua**

Actividad: ¿Permite concluir causalidad?

Categorías: Sí, en principio · No, solo asociación

- Se sortea qué doctorandos reciben tutoría y se comparan las notas
- Se comparan las notas de quienes eligieron ir a tutoría con quienes no
- Se encuesta a 2000 doctorandos sobre su sueño y su productividad
- Se asigna al azar el orden de presentación de dos cuestionarios

Solución

- Se sortea qué doctorandos reciben tutoría y se comparan las notas → **Sí, en principio**
- Se comparan las notas de quienes eligieron ir a tutoría con quienes no → **No, solo asociación**. Autoselección: quienes eligen ir pueden estar más motivados.
- Se encuesta a 2000 doctorandos sobre su sueño y su productividad → **No, solo asociación**
- Se asigna al azar el orden de presentación de dos cuestionarios → **Sí, en principio**

Reto evaluable

Reto del capítulo 1 · La pregunta de tu tesis

1. Escribe la pregunta principal de un estudio de tu tesis y clasifícala con la tipología de Leek y Peng.
2. Haz una **tabla de variables**: nombre, tipo, cómo se mide y su papel (resultado, exposición, covariable).
3. Dibuja un **diagrama causal** sencillo (con flechas) con al menos dos posibles confusores.
4. Justifica qué diseño puede responder a tu pregunta y qué limitaciones de validez interna y externa tendrá.

Entrega: con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. Una correlación entre horas de sueño y nota demuestra que dormir más mejora la nota. *[V / F]*
2. Una variable que influye a la vez en la supuesta causa y en el efecto se llama variable de _____
3. La aleatorización equilibra, en promedio, los confusores *[solo medidos / medidos y no medidos / solo no medidos]*
4. El nivel de acuerdo de 1 a 5 es una variable *[nominal / ordinal / continua]*
5. Generalizar los resultados a otras universidades es una cuestión de validez *[interna / externa]*
6. Según Leek y Peng, el error más frecuente es responder con el análisis a una pregunta distinta de la planteada. *[V / F]*
7. En `mean(x >= 5)`, el resultado es *[la media de x / la proporción de valores mayores o iguales que 5 / el número de valores mayores que 5]*
8. ¿Qué tipo de pregunta exige más del diseño?

(a) Descriptiva (b) Exploratoria (c) Causal o mecanicista

Explicaciones

- Puede haber confusores (motivación, salud…).
- *Confounder*.
- Esa es su gran ventaja.
- Categorías ordenadas.
- Externa: generalización.
- *What is the question?*
- Media de un lógico = proporción.
- Hacen falta diseños que controlen la confusión.

Capítulo 2. Describir y visualizar

Objetivos

- Resumir variables con medidas de centro y de dispersión adecuadas a su forma.
- Elegir gráficos que muestren los datos, no solo sus medias.
- Detectar valores atípicos, asimetrías y errores antes de cualquier contraste.

Glosario del capítulo

Media y mediana · Desviación típica · Rango intercuartílico · Valor atípico · Correlación

Descargas

Script R del capítulo Datos del libro (CSV) Plantilla de entrega del reto (.qmd)

Lectura obligatoria

Weissgerber, T. L.; Milic, N. M.; Winham, S. J.; Garovic, V. D. (2015). Beyond bar and line graphs: Time for a new data presentation paradigm. *PLOS Biology*, 13(4), e1002128.

1. Centro y dispersión

Medida	Qué resume	Cuándo usarla
Media	Valor «de equilibrio»	Distribuciones aproximadamente simétricas
Mediana	Valor central	Distribuciones asimétricas o con atípicos
Desviación típica	Distancia típica a la media	Junto a la media
Rango intercuartílico (IQR)	Amplitud del 50 % central	Junto a la mediana

```
d <- read.csv("datos/estudio.csv")
summary(d$nota)
c(media = mean(d$nota), desv_tipica = sd(d$nota), IQR = IQR(d$nota))
tapply(d$nota, d$grupo, median)   # mediana por grupo
```

2. Mostrar los datos, no solo las medias

Weissgerber et al. (2015) revisaron artículos de fisiología y encontraron que la mayoría presentaba datos continuos con **gráficos de barras** de medias. El problema: distribuciones muy distintas (con atípicos, bimodales, con pocos datos) producen la misma barra. Recomiendan **gráficos que muestren cada observación** (diagramas de puntos o de dispersión) o, con muchos datos, **diagramas de caja** o de violín.

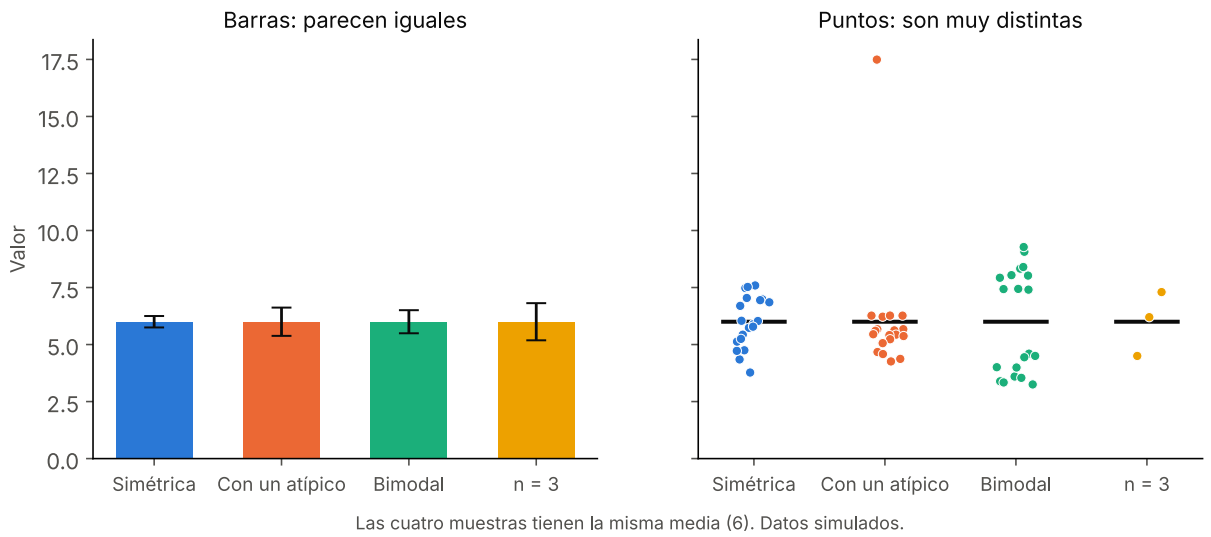

Figura 1: La misma media puede esconder datos muy distintos. Las cuatro muestras tienen la misma media y casi la misma barra, pero sus puntos cuentan historias diferentes.

```
d <- read.csv("datos/estudio.csv")
stripchart(nota ~ grupo, data = d, vertical = TRUE, method = "jitter", pch = 19,
           col = c("#eb6834", "#2a78d6"), ylab = "Nota", main = "Cada punto, una
persona")
boxplot(nota ~ grupo, data = d, add = TRUE, outline = FALSE, col = NA, border =
"grey40")
```

3. Relaciones entre dos variables

La **correlación de Pearson** (r) mide la fuerza de una relación **lineal** entre −1 y 1. No mide relaciones curvas, es sensible a atípicos y no implica causalidad. Antes de calcularla, **dibuja** el diagrama de dispersión.

```
d <- read.csv("datos/estudio.csv")
plot(d$horas, d$nota, pch = 19, col = "#2a78d6", xlab = "Horas de estudio a la semana",
ylab = "Nota")
cor(d$horas, d$nota)
```

Laboratorio: el mismo resumen, datos distintos

ℹ Laboratorio interactivo

El **cuarteto de Anscombe** (1973) son cuatro conjuntos de datos con la misma media, la misma varianza y la misma correlación ($r \approx 0{,}82$). Elige uno y mira el gráfico.

Este laboratorio es interactivo en la versión web: pablosancheznunez.com/estadistica-investigacion

Ejercicio 2.1

Calcula la **mediana** de las horas de estudio de cada grupo.

ℹ Ejercicio: completa los huecos _____

```
tapply(d$horas, d$grupo, ______)
```

Pista

`tapply(variable, grupo, función)` aplica la función a cada grupo; la función es `median`.

Solución

```
tapply(d$horas, d$grupo, median)
```

Ejercicio 2.2

Calcula la correlación entre **ansiedad** y **sueño**. ¿Es positiva o negativa?

Ejercicio: completa los huecos _____

```
cor(d$ansiedad, d$_____)
```

Pista

La columna del sueño se llama `sueno` (sin eñe).

Solución

```
cor(d$ansiedad, d$sueno)
```

Negativa: en estos datos simulados, quienes duermen más tienden a tener menos ansiedad. Es una asociación, no una prueba de causalidad.

Caso práctico resuelto

Describir una variable asimétrica

Las citas de un grupo de 20 artículos: 0, 0, 1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 6, 6, 7, 8, 9, 10, 12, 140.

- **Media:** 11,4. **Mediana:** 4,5. La media se dispara por un solo artículo.
- **Desviación típica** muy alta; el **rango intercuartílico** (de 2 a 7,5) describe mejor la dispersión típica.
- **Gráfico:** un histograma o un diagrama de caja muestra la asimetría y el valor atípico; un gráfico de barras con la media lo ocultaría.

Regla práctica: con distribuciones asimétricas, informa de la **mediana y el rango intercuartílico**, y muestra los datos.

Ejercicios graduados

Básico. ¿Qué medida de tendencia central usarías para los ingresos de una población? ¿Por qué?

Explicaciones

La mediana: los ingresos son muy asimétricos y unos pocos valores muy altos elevan la media.

Avanzado. Con el conjunto `estudio.csv` del libro, describe la variable `horas` con medidas y un gráfico adecuados y justifica la elección.

Explicaciones

Comprueba primero la forma de la distribución; si es aproximadamente simétrica, media y desviación típica son adecuadas; si no, mediana y rango intercuartílico. El gráfico debe mostrar la distribución, no solo un resumen.

Actividades

Actividad: ¿Media o mediana?

Categorías: Media · Mediana

- Ingresos de los hogares de un barrio
- Altura de un grupo de personas adultas
- Número de citas de los artículos de un departamento
- Temperatura media diaria de un mes
- Tiempo hasta terminar la tesis, con algunos casos muy largos

Solución

- Ingresos de los hogares de un barrio ⟶ **Mediana**. Muy asimétricos.
- Altura de un grupo de personas adultas ⟶ **Media**
- Número de citas de los artículos de un departamento ⟶ **Mediana**
- Temperatura media diaria de un mes ⟶ **Media**
- Tiempo hasta terminar la tesis, con algunos casos muy largos ⟶ **Mediana**

Actividad: ¿Qué gráfico usarías?

Categorías: Diagrama de puntos o de caja · Diagrama de dispersión · Histograma · Barras de frecuencias

- Comparar la nota de dos grupos de 15 personas
- Ver la relación entre horas de estudio y nota
- Ver la forma de la distribución de 5000 tiempos de respuesta
- Mostrar cuántos doctorandos hay en cada rama

Solución

- Comparar la nota de dos grupos de 15 personas → **Diagrama de puntos o de caja**
- Ver la relación entre horas de estudio y nota → **Diagrama de dispersión**
- Ver la forma de la distribución de 5000 tiempos de respuesta → **Histograma**
- Mostrar cuántos doctorandos hay en cada rama → **Barras de frecuencias**

Reto evaluable

! Reto del capítulo 2 · Rehacer una figura

1. Busca en tu campo un artículo que presente datos continuos con un **gráfico de barras** de medias.
2. Explica qué información oculta (tamaño de la muestra, distribución, atípicos).
3. Con los datos del libro (o con datos tuyos), crea en R una figura que muestre **cada observación** y escribe un pie de figura completo y autosuficiente.
4. Añade una tabla descriptiva con las medidas de centro y dispersión adecuadas.

Entrega: con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. Con una distribución muy asimétrica, la mediana describe mejor el valor típico que la media. *[V / F]*
2. La medida de dispersión que acompaña a la mediana es el rango _____
3. Una correlación de Pearson igual a 0 significa que no hay ninguna relación entre las variables. *[V / F]*
4. Según Weissgerber et al. (2015), para datos continuos con muestras pequeñas conviene usar *[gráficos de barras / gráficos que muestren cada observación / gráficos de tarta]*
5. Los cuatro conjuntos del cuarteto de Anscombe tienen *[gráficos idénticos / resúmenes estadísticos casi idénticos / correlaciones muy distintas]*
6. `tapply(d$nota, d$grupo, mean)` calcula *[la media global / la media de la nota en cada grupo / la nota de cada persona]*
7. La correlación de Pearson mide la fuerza de una relación _____
8. ¿Qué debes hacer antes de calcular una correlación?

(a) Un contraste de hipótesis (b) Dibujar el diagrama de dispersión (c) Eliminar los valores atípicos

Explicaciones

- La media se deja arrastrar por los extremos.
- IQR: percentil 75 menos percentil 25.
- Puede haber una relación curva fuerte con $r \approx 0$.
- Puntos individuales, cajas o violines.
- Esa es su lección.
- Una media por grupo.
- Solo lineal.
- Y no eliminar atípicos sin justificarlo.

Capítulo 3. Estimar con incertidumbre

Objetivos

- Entender la variabilidad de muestreo y el error típico.
- Calcular e interpretar correctamente un intervalo de confianza.
- Pasar de «¿hay efecto?» a «¿cuánto efecto y con qué precisión?».

Glosario del capítulo

Distribución muestral · Error típico · Intervalo de confianza · Estimación · Remuestreo (*bootstrap*)

Descargas

Script R del capítulo Plantilla de entrega del reto (.qmd)

Lectura obligatoria

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7-29.

1. Una muestra es una de muchas posibles

Si repitieras el estudio con otros 120 doctorandos, la media sería otra. La **distribución muestral** describe cómo varía un estadístico (la media, una diferencia, una pendiente) de muestra en muestra. Su desviación típica es el **error típico**: para una media, aproximadamente `sd / sqrt(n)`. Cuanto mayor es la muestra, más precisa es la estimación, pero la precisión crece con la **raíz** de *n*: para reducir el error a la mitad hace falta cuadruplicar la muestra.

2. El intervalo de confianza

Un **intervalo de confianza del 95 %** es un procedimiento que, si se repitiera el estudio muchas veces, contendría el valor verdadero en el 95 % de las repeticiones. Para un estudio concreto, una forma práctica de leerlo (Cumming, 2014) es: el intervalo reúne los **valores del parámetro razonablemente compatibles con los datos**, y los valores del centro son más plausibles que los de los extremos.

Interpretaciones erróneas frecuentes

- «Hay un 95 % de probabilidad de que el valor verdadero esté en *este* intervalo». En la estadística frecuentista, el valor verdadero es fijo: o está o no está; el 95 % se refiere al procedimiento.
- «El 95 % de los datos está dentro del intervalo». El intervalo es para el parámetro (la media), no para las observaciones individuales.
- «Si dos intervalos se solapan, no hay diferencia». Pueden solaparse algo y la diferencia ser significativa: hay que calcular el intervalo de la **diferencia**.

```
d <- read.csv("datos/estudio.csv")
t.test(d$nota)$conf.int                       # IC 95 % de la nota media
t.test(nota ~ grupo, data = d)$conf.int       # IC 95 % de la diferencia (control −
tutoría)
```

Laboratorio: el baile de los intervalos

Laboratorio interactivo

Simulamos 50 estudios de la misma población (media verdadera = 50, desviación típica = 10). Cada línea es el intervalo de confianza de un estudio; en naranja, los que **no** contienen el valor verdadero. Cambia el tamaño de la muestra y el nivel de confianza, y pulsa **Repetir**.

Este laboratorio es interactivo en la versión web: pablosancheznunez.com/estadistica-investigacion

3. Remuestreo: el *bootstrap*

Cuando no hay una fórmula sencilla (por ejemplo, para una mediana), el **bootstrap** aproxima la distribución muestral **remuestreando con reemplazo** los propios datos muchas veces.

```
d <- read.csv("datos/estudio.csv")
set.seed(1)
medianas <- replicate(2000, median(sample(d$nota, replace = TRUE)))
quantile(medianas, c(0.025, 0.975))    # IC 95 % bootstrap de la mediana
```

Ejercicio 3.1

Calcula el **error típico** de la nota media: la desviación típica entre la raíz del tamaño de la muestra.

Ejercicio: completa los huecos _____

```
sd(d$nota) / _____(nrow(d))
```

Pista

La raíz cuadrada en R es `sqrt()`.

Solución

```
sd(d$nota) / sqrt(nrow(d))
```

Unos 0,14 puntos. El IC 95 % es, aproximadamente, la media ± 2 errores típicos.

Ejercicio 3.2

Calcula el intervalo de confianza del **99 %** de la nota media.

Ejercicio: completa los huecos _____

```
t.test(d$nota, conf.level = _____)$conf.int
```

Pista

El 99 % se escribe 0.99.

Solución

```
t.test(d$nota, conf.level = 0.99)$conf.int
```

Más ancho que el del 95 %: más confianza exige un intervalo mayor.

Caso práctico resuelto

Interpretar un intervalo de confianza

En una muestra de 120 doctorandos, la media de horas de sueño es 6,8 con un intervalo de confianza del 95 % de [6,6; 7,0].

- **Interpretación correcta:** el procedimiento que genera estos intervalos captura el valor poblacional en el 95 % de las muestras; los valores entre 6,6 y 7,0 son compatibles con los datos.
- **Interpretación incorrecta:** «hay un 95 % de probabilidad de que la media verdadera esté entre 6,6 y 7,0» (en el enfoque frecuentista, el parámetro es fijo; la probabilidad se refiere al procedimiento).
- **Precisión:** un intervalo estrecho indica una estimación precisa; para reducirlo a la mitad hace falta aproximadamente cuadruplicar la muestra.

Ejercicios graduados

Básico. Si con n = 100 el margen de error es ±0,4, ¿qué margen aproximado tendrás con n = 400?

Explicaciones

Aproximadamente ±0,2: el error típico depende de $1/\sqrt{n}$, y $\sqrt{400} = 2 \times \sqrt{100}$.

Avanzado. Simula en R 100 intervalos de confianza de muestras de una población conocida y cuenta cuántos contienen el valor verdadero.

Explicaciones

Deberían ser alrededor de 95. Este ejercicio muestra que la confianza es una propiedad del procedimiento, no de un intervalo concreto.

Actividades

Actividad: ¿Interpretación correcta del IC 95 % [5,99; 6,53]?

Categorías: Correcta · Incorrecta

- Los valores entre 5,99 y 6,53 son los razonablemente compatibles con los datos para la nota media
- El 95 % de los doctorandos tiene una nota entre 5,99 y 6,53
- Si repitiéramos el estudio muchas veces, el 95 % de los intervalos así calculados contendría la media verdadera
- Hay un 95 % de probabilidad de que la media verdadera esté en este intervalo concreto

Solución

- Los valores entre 5,99 y 6,53 son los razonablemente compatibles con los datos para la nota media → **Correcta**
- El 95 % de los doctorandos tiene una nota entre 5,99 y 6,53 → **Incorrecta**. El intervalo es para la media, no para las personas.
- Si repitiéramos el estudio muchas veces, el 95 % de los intervalos así calculados contendría la media verdadera → **Correcta**
- Hay un 95 % de probabilidad de que la media verdadera esté en este intervalo concreto → **Incorrecta**. Esa es una lectura bayesiana; en frecuentista, la probabilidad se refiere al procedimiento.

Actividad: ¿El intervalo se estrecha o se ensancha?

Categorías: Se estrecha · Se ensancha

- Aumentar el tamaño de la muestra
- Pasar del 95 % al 99 % de confianza
- Medir con un instrumento más ruidoso
- Reducir la variabilidad entre participantes con un diseño intrasujeto

Solución

- Aumentar el tamaño de la muestra → **Se estrecha**
- Pasar del 95 % al 99 % de confianza → **Se ensancha**
- Medir con un instrumento más ruidoso → **Se ensancha**
- Reducir la variabilidad entre participantes con un diseño intrasujeto → **Se estrecha**

Reto evaluable

! Reto del capítulo 3 · Reescribe unos resultados en clave de estimación

1. Toma un artículo de tu campo que informe solo de valores p («la diferencia fue significativa, $p < 0{,}05$»).
2. Con los datos que dé (medias, desviaciones, n), calcula los **intervalos de confianza** de los efectos principales.
3. Reescribe el párrafo de resultados centrado en **cuánto** efecto y con **qué precisión**, al estilo de Cumming (2014).
4. Discute si las conclusiones del artículo cambian con esta lectura.

Entrega: con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. Para reducir a la mitad el error típico de una media hay que multiplicar la muestra por *[2 / 4 / 10]*
2. Un intervalo del 99 % es más estrecho que uno del 95 % con los mismos datos. *[V / F]*
3. El IC 95 % contiene el 95 % de las observaciones individuales. *[V / F]*
4. La desviación típica de la distribución muestral de un estadístico se llama error _____
5. Remuestrear con reemplazo los propios datos para estimar la incertidumbre es el _____
6. Si dos intervalos se solapan un poco, la diferencia nunca puede ser significativa. *[V / F]*
7. En el laboratorio, con intervalos del 95 %, ¿cuántos de cada 50 esperas que no contengan la media verdadera? *[0 / 2 o 3 / 25]*
8. ¿Qué propone Cumming con la «nueva estadística»?

(a) Abandonar los datos cuantitativos (b) Centrarse en estimar efectos con intervalos y acumular evidencia con metaanálisis (c) Usar solo valores p más pequeños

Explicaciones

- Precisión ∝ raíz de n.
- Más confianza, más anchura.
- Es para el parámetro.
- *Standard error*.
- Efron, 1979.
- Hay que mirar el intervalo de la diferencia.
- El 5 % de 50 es 2,5.
- Estimación y metaanálisis.

Capítulo 4. Contrastes y valores *p*

Objetivos

- Entender la lógica del contraste de hipótesis y lo que significa un valor *p*.
- Reconocer y evitar las interpretaciones erróneas más frecuentes.
- Elegir y aplicar los contrastes básicos en R, comprobando sus supuestos.

Glosario del capítulo

Hipótesis nula · Valor *p* · Nivel de significación · Error de tipo I y II · Prueba t · Comparaciones múltiples

Descargas

Script R del capítulo Plantilla de entrega del reto (.qmd)

Lectura obligatoria

Wasserstein, R. L.; Lazar, N. A. (2016). The ASA statement on *p*-values: Context, process, and purpose. *The American Statistician*, 70(2), 129-133.

Lecturas complementarias

- Greenland et al. (2016)
- Halsey et al. (2015)

1. La lógica del contraste

1. Se plantea una **hipótesis nula** (H_0), normalmente «no hay efecto» o «no hay diferencia».
2. Se calcula un estadístico que mide cuánto se alejan los datos de lo que predice H_0.
3. El **valor *p*** es la probabilidad de obtener un estadístico **tan extremo o más** que el observado **si H_0 y todos los supuestos del modelo fueran ciertos**.
4. Si *p* es menor que un umbral α fijado de antemano (por convención, 0,05), se rechaza H_0.

Este procedimiento controla a largo plazo dos errores: el de **tipo I** (rechazar H_0 cuando es cierta, con probabilidad α) y el de **tipo II** (no rechazarla cuando es falsa, con probabilidad β; la potencia es $1 - \beta$, capítulo 5).

2. Seis principios de la ASA

En 2016, la American Statistical Association publicó por primera vez una declaración sobre un procedimiento concreto (Wasserstein y Lazar, 2016):

1. Los valores *p* pueden indicar cuán incompatibles son los datos con un modelo estadístico concreto.
2. **No miden** la probabilidad de que la hipótesis sea cierta, ni la de que los datos se deban solo al azar.
3. Las conclusiones científicas y las decisiones no deberían basarse solo en si *p* supera un umbral.
4. Una inferencia correcta exige **informar de todo** con transparencia.
5. Un valor *p* **no mide el tamaño del efecto** ni la importancia del resultado.
6. Por sí solo, un valor *p* no es una buena medida de la evidencia sobre un modelo o una hipótesis.

3. Laboratorio: ¿cómo se distribuyen los valores *p*?

> **i** Laboratorio interactivo
>
> Simulamos 2000 estudios con dos grupos. Si **no hay efecto**, los valores *p* se reparten de manera **uniforme** entre 0 y 1: un 5 % cae por debajo de 0,05 por puro azar. Si **hay efecto**, se amontonan cerca de 0, tanto más cuanto mayor es la potencia.
>
> Este laboratorio es interactivo en la versión web: pablosancheznunez.com/estadistica-investigacion

4. Contrastes básicos en R

Pregunta	Datos	Contraste	En R
¿Difieren dos medias?	Dos grupos independientes	*t* de Welch	`t.test(y ~ grupo)`
¿Cambia una medida antes y después?	Medidas emparejadas	*t* emparejada	`t.test(antes, despues, paired = TRUE)`
¿Difieren dos distribuciones sin suponer normalidad?	Dos grupos	Wilcoxon-Mann-Whitney	`wilcox.test(y ~ grupo)`
¿Difieren varias medias?	Tres o más grupos	ANOVA	`aov(y ~ grupo)`
¿Están asociadas dos variables categóricas?	Tabla de frecuencias	Ji cuadrado	`chisq.test(tabla)`
¿Hay relación lineal?	Dos cuantitativas	Correlación	`cor.test(x, y)`

```r
d <- read.csv("datos/estudio.csv")
t.test(nota ~ grupo, data = d)
```

La salida da el estadístico *t*, los grados de libertad, el valor *p*, el **intervalo de confianza de la diferencia** y las medias. Lee primero el intervalo: la tutoría podría mejorar la nota en casi un punto o empeorarla ligeramente; los datos son compatibles con un efecto nulo, pero también con uno moderado. «No significativo» **no** es «no hay efecto».

5. Muchas pruebas, muchos falsos positivos

Con 20 contrastes independientes y $\alpha = 0{,}05$, la probabilidad de al menos un falso positivo es $1 - 0{,}95^{20} \approx 64$ %. Las correcciones (Bonferroni, Holm, tasa de falsos descubrimientos) ajustan los valores *p* cuando se hacen muchas pruebas:

```r
p <- c(0.001, 0.012, 0.03, 0.04, 0.2)
p.adjust(p, method = "holm")
p.adjust(p, method = "BH")   # tasa de falsos descubrimientos
```

Ejercicio 4.1

¿Está asociada la ansiedad con la nota? Calcula `cor.test()` y extrae el valor *p*.

Ejercicio: completa los huecos _____

```
cor.test(d$ansiedad, d$nota)$_____
```

Pista

El elemento con el valor *p* se llama `p.value`.

Solución

```
cor.test(d$ansiedad, d$nota)$p.value
```

Mira también `$conf.int`: el intervalo de la correlación dice más que el valor *p*.

Ejercicio 4.2

Calcula la probabilidad de al menos un falso positivo al hacer **10** contrastes independientes con $\alpha = 0{,}05$.

Ejercicio: completa los huecos _____

```
1 - (1 - 0.05)^_____
```

Pista

El exponente es el número de contrastes.

Solución

```
1 - (1 - 0.05)^10
```

Alrededor del 40 %.

Caso práctico resuelto

Leer bien un valor p

Un estudio informa: diferencia de 3 puntos en la nota entre quienes hicieron un taller y quienes no; $t(118) = 2{,}1$; $p = 0{,}038$.

Lo que significa: si no hubiera ninguna diferencia en la población, sería poco frecuente (en torno al 4 % de las veces) observar una diferencia como esta o mayor.

Lo que **no** significa:

- No es la probabilidad de que la hipótesis nula sea cierta.
- No dice que el efecto sea grande o importante: hay que mirar el tamaño del efecto y su intervalo.
- $p = 0{,}038$ y $p = 0{,}052$ no son resultados opuestos: la evidencia es muy parecida.

Informe completo: diferencia de 3 puntos, IC 95 % [0,2; 5,8], d de Cohen ≈ 0,38, $p = 0{,}038$.

Ejercicios graduados

Básico. Un estudio obtiene $p = 0{,}30$. ¿Demuestra que no hay efecto?

Explicaciones

No: un resultado no significativo no demuestra la ausencia de efecto. Puede deberse a poca potencia; hay que mirar el intervalo de confianza para ver qué efectos son compatibles con los datos.

Avanzado. Busca en un artículo de tu campo un resultado con valor p y reescribe su descripción con tamaño del efecto e intervalo de confianza.

Explicaciones

La reescritura debe sustituir «significativo» por la magnitud del efecto con su incertidumbre y valorar su relevancia práctica.

Actividades

Actividad: Con $p = 0{,}03$, ¿la afirmación es correcta?

Categorías: Correcta · Incorrecta

- Hay un 3 % de probabilidad de que la hipótesis nula sea cierta
- Si H_0 y el modelo fueran ciertos, datos tan extremos o más aparecerían un 3 % de las veces
- El efecto es grande e importante
- Hay un 97 % de probabilidad de que el resultado se replique
- Los datos son poco compatibles con la hipótesis nula, según este modelo

Solución

- Hay un 3 % de probabilidad de que la hipótesis nula sea cierta → **Incorrecta**. El valor p supone que H_0 es cierta; no calcula su probabilidad.
- Si H_0 y el modelo fueran ciertos, datos tan extremos o más aparecerían un 3 % de las veces → **Correcta**
- El efecto es grande e importante → **Incorrecta**. El valor p no mide el tamaño del efecto.
- Hay un 97 % de probabilidad de que el resultado se replique → **Incorrecta**
- Los datos son poco compatibles con la hipótesis nula, según este modelo → **Correcta**

Actividad: ¿Qué contraste usarías?

Categorías: t de Welch · t emparejada · Ji cuadrado · ANOVA · Correlación

- Comparar la nota media de tutoría y control
- Comparar la ansiedad de las mismas personas antes y después de un curso
- Ver si el abandono (sí/no) depende de la rama (cinco ramas)
- Comparar la nota media de cuatro métodos de enseñanza
- Ver si las horas de estudio se relacionan con la nota

Solución

- Comparar la nota media de tutoría y control → **t de Welch**
- Comparar la ansiedad de las mismas personas antes y después de un curso → **t emparejada**
- Ver si el abandono (sí/no) depende de la rama (cinco ramas) → **Ji cuadrado**
- Comparar la nota media de cuatro métodos de enseñanza → **ANOVA**
- Ver si las horas de estudio se relacionan con la nota → **Correlación**

Reto evaluable

Reto del capítulo 4 · Auditoría de los valores *p*

1. Elige dos artículos de tu campo y localiza todas las afirmaciones basadas en valores *p*.
2. Clasifícalas según los seis principios de la ASA: ¿cuáles son correctas y cuáles caen en alguno de los errores de Greenland et al. (2016)?
3. Comprueba si informan de tamaños del efecto e intervalos y si corrigen las comparaciones múltiples.
4. Reescribe dos conclusiones con un lenguaje correcto.

Entrega: con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. El valor *p* es la probabilidad de que la hipótesis nula sea cierta. *[V / F]*
2. Si no hay efecto, los valores *p* se distribuyen de forma *[concentrada cerca de 0 / uniforme entre 0 y 1 / concentrada cerca de 1]*

3. Rechazar la hipótesis nula cuando es cierta es un error de tipo *[I / II]*
4. «No significativo» equivale a «no hay efecto». *[V / F]*
5. Para medidas antes y después en las mismas personas se usa la prueba *t* _____
6. Con 20 contrastes independientes y $\alpha = 0{,}05$, la probabilidad de al menos un falso positivo es aproximadamente *[5 % / 20 % / 64 % / 100 %]*
7. Un valor *p* pequeño indica que el efecto es grande. *[V / F]*
8. ¿Qué debe leerse primero en la salida de `t.test()`?

(a) Los grados de libertad (b) El intervalo de confianza de la diferencia, que dice cuánto efecto es compatible con los datos (c) El nombre del contraste

Explicaciones

- Principio 2 de la ASA.
- Uniforme bajo H_0.
- Tipo I: falso positivo.
- Ausencia de evidencia no es evidencia de ausencia.
- *Paired.*
- $1 - 0{,}95^{20} \approx 0{,}64$.
- Con muestras enormes, efectos diminutos dan *p* minúsculos.
- Estimación antes que dicotomía.

Capítulo 5. Tamaño del efecto y potencia

Objetivos

- Calcular, informar e interpretar tamaños del efecto estandarizados y no estandarizados.
- Entender la potencia estadística y sus determinantes.
- Justificar el tamaño de la muestra de un estudio.

Glosario del capítulo

Tamaño del efecto · d de Cohen · Potencia · Efecto mínimo de interés · Justificación de la muestra

Descargas

Script R del capítulo Plantilla de entrega del reto (.qmd)

Lectura obligatoria

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, 8(1), 33267.

Lecturas complementarias

- Lakens (2013)

1. Cuánto, no solo si

Un **tamaño del efecto** cuantifica la magnitud de un fenómeno. Puede expresarse en las **unidades originales** (la tutoría sube la nota 0,4 puntos) o de forma **estandarizada**, sin unidades, para comparar estudios:

Medida	Para	Fórmula o función
Diferencia de medias	Dos grupos, unidades comprensibles	`mean(a) - mean(b)`
d de Cohen	Dos grupos, estandarizado	diferencia / desviación típica combinada
r	Relación entre dos variables	`cor(x, y)`
R^2	Variación explicada por un modelo	`summary(lm(...))$r.squared`
Razón de odds, riesgo relativo	Resultados binarios	`glm(..., family = binomial)`

Las etiquetas de Cohen (d = 0,2 pequeño, 0,5 mediano, 0,8 grande) eran orientativas y él mismo pedía usarlas solo cuando no hubiera nada mejor. Lo que importa es si el efecto es **relevante en su contexto**: medio punto de nota puede ser mucho o poco según para qué.

```
d <- read.csv("datos/estudio.csv")
a <- d$nota[d$grupo == "tutoria"]; b <- d$nota[d$grupo == "control"]
dif <- mean(a) - mean(b)
d_cohen <- dif / sqrt((var(a) + var(b)) / 2)
c(diferencia = dif, d = d_cohen)
```

2. Potencia

La **potencia** es la probabilidad de detectar un efecto de un tamaño dado si existe. Depende de cuatro cosas, y conocidas tres se calcula la cuarta:

1. El **tamaño del efecto** que se quiere detectar.
2. La **variabilidad** de los datos.
3. El **tamaño de la muestra**.
4. El umbral α.

Una potencia del 80 % es la convención mínima; muchos estudios publicados tienen bastante menos (véase el capítulo 3 de *Ciencia de la ciencia y metaciencia*). Con poca potencia, los resultados significativos tienden a exagerar el efecto y a no replicarse.

```
# ¿Qué potencia tenía nuestro estudio (60 por grupo) para detectar 0,4 puntos con sd =
1,3?
power.t.test(n = 60, delta = 0.4, sd = 1.3)$power
# ¿Cuántas personas por grupo harían falta para una potencia del 80 %?
power.t.test(delta = 0.4, sd = 1.3, power = 0.8)$n
```

Nuestro estudio solo tenía un 39 % de potencia para ese efecto: no es extraño que la diferencia no fuera significativa. Para tener un 80 %, harían falta unas **167 personas por grupo**.

Laboratorio: potencia

> **i** Laboratorio interactivo
>
> Ajusta el efecto (d) y la muestra, y mira cómo se solapan las distribuciones del estadístico bajo la hipótesis nula (gris) y bajo el efecto real (azul). La potencia es el área azul más allá del valor crítico.
>
> Este laboratorio es interactivo en la versión web: pablosancheznunez.com/estadistica-investigacion

3. Justificar el tamaño de la muestra

Lakens (2022) distingue seis formas legítimas de justificar una muestra: medir a **toda la población**, limitaciones de **recursos**, análisis de **potencia** a priori, búsqueda de una **precisión** dada (anchura del intervalo), **heurísticos** y **ninguna justificación** (que hay que reconocer). Para el análisis de potencia, la clave es el **efecto mínimo de interés**: el más pequeño que tendría importancia teórica o práctica, no el que esperamos ni el de un estudio piloto pequeño (que suele estar inflado).

Ejercicio 5.1

¿Cuántas personas por grupo harían falta para detectar una diferencia de **0,5** puntos (sd = 1,3) con una potencia del **90 %**?

> **i** Ejercicio: completa los huecos _____
>
> ```
> power.t.test(delta = 0.5, sd = 1.3, power = _____)$n
> ```

> **i** Pista
>
> El 90 % se escribe 0.9.

Solución

```
power.t.test(delta = 0.5, sd = 1.3, power = 0.9)$n
```

Unas 143 por grupo (redondea siempre hacia arriba).

Ejercicio 5.2

Calcula la ***d* de Cohen** de la diferencia en horas de estudio entre los grupos.

Ejercicio: completa los huecos _____

```
(mean(a) - mean(b)) / sqrt((var(a) + ______) / 2)
```

Pista

Falta la varianza del segundo grupo: `var(b)`.

Solución

```
(mean(a) - mean(b)) / sqrt((var(a) + var(b)) / 2)
```

Un efecto muy pequeño: la aleatorización produjo grupos parecidos en horas de estudio.

Caso práctico resuelto

Calcular el tamaño muestral antes de empezar

Quieres comparar la ansiedad de dos grupos. Estudios previos sugieren un efecto pequeño-moderado, d = 0,3. Con α = 0,05 bilateral y potencia del 80 %:

- En R: `power.t.test(delta = 0.3, sd = 1, sig.level = 0.05, power = 0.8)` devuelve unas **176 personas por grupo**.
- Si solo puedes reclutar 50 por grupo, la potencia para d = 0,3 es de un 32 %: lo más probable es no detectar el efecto aunque exista.

Opciones: colaborar con otros grupos para aumentar la muestra, usar un diseño intrasujeto (más eficiente), medir con menos error o plantear el estudio como piloto y estimar el efecto en lugar de contrastarlo.

Ejercicios graduados

Básico. ¿Qué pasa con el tamaño muestral necesario si el efecto esperado se reduce a la mitad?

Explicaciones

Se multiplica aproximadamente por cuatro, porque el tamaño muestral depende del inverso del cuadrado del efecto.

Avanzado. Calcula con `power.t.test` el tamaño muestral para d = 0,2, 0,5 y 0,8 y comenta qué implicaciones tiene para tu investigación.

Explicaciones

Aproximadamente 394, 64 y 26 por grupo. Los efectos pequeños, frecuentes en las ciencias sociales, exigen muestras grandes.

Actividades

Actividad: ¿La potencia sube o baja?

Categorías: Sube · Baja

- Aumentar el tamaño de la muestra
- Querer detectar un efecto más pequeño
- Usar α = 0,005 en lugar de 0,05
- Medir con menos ruido (más fiabilidad)
- Usar un diseño intrasujeto en lugar de entre grupos

Solución

- Aumentar el tamaño de la muestra ⟶ **Sube**
- Querer detectar un efecto más pequeño ⟶ **Baja**
- Usar α = 0,005 en lugar de 0,05 ⟶ **Baja**
- Medir con menos ruido (más fiabilidad) ⟶ **Sube**
- Usar un diseño intrasujeto en lugar de entre grupos ⟶ **Sube**. Si las medidas están correlacionadas.

Actividad: ¿Qué justificación de la muestra es [@lakens2022]?

Categorías: Toda la población · Recursos · Análisis de potencia · Precisión

- Estudiamos los 47 programas de doctorado de la universidad
- Solo hay presupuesto para 80 participantes
- Calculamos la muestra para un 90 % de potencia ante el efecto mínimo de interés
- Queremos que el intervalo de la media tenga una anchura de ±0,2 puntos

Solución

- Estudiamos los 47 programas de doctorado de la universidad ⟶ **Toda la población**
- Solo hay presupuesto para 80 participantes ⟶ **Recursos**. Hay que decir qué se podrá concluir con ellos.
- Calculamos la muestra para un 90 % de potencia ante el efecto mínimo de interés ⟶ **Análisis de potencia**
- Queremos que el intervalo de la media tenga una anchura de ±0,2 puntos ⟶ **Precisión**

Reto evaluable

! Reto del capítulo 5 · Justifica tu muestra

1. Para tu próximo estudio, define el **efecto mínimo de interés** y justifícalo (teoría, relevancia práctica, coste).
2. Haz un **análisis de potencia** con R para varios escenarios y presenta una curva de potencia.
3. Si tus recursos limitan la muestra, calcula qué efectos podrás detectar con ella (análisis de sensibilidad).
4. Redacta el párrafo de justificación de la muestra para tu preregistro o tu artículo, siguiendo a Lakens (2022).

Entrega: con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. La *d* de Cohen es la diferencia de medias dividida entre la desviación típica _____
2. Las etiquetas de Cohen (0,2; 0,5; 0,8) deben usarse siempre para decidir si un efecto es importante. *[V / F]*
3. La potencia es la probabilidad de *[cometer un falso positivo / detectar un efecto si existe / que la hipótesis sea cierta]*
4. Para un análisis de potencia, lo mejor es usar el efecto de un estudio piloto pequeño. *[V / F]*
5. Si conoces el efecto, la variabilidad y α, puedes calcular la *[media / muestra necesaria para una potencia dada / correlación]*
6. Nuestro estudio (60 por grupo) tenía una potencia de alrededor del *[10 % / 39 % / 80 % / 99 %]* para detectar 0,4 puntos
7. El efecto más pequeño que tendría importancia teórica o práctica se llama efecto mínimo de _____
8. ¿Por qué un resultado no significativo en un estudio con poca potencia es poco informativo?

(a) Porque los estudios pequeños tienen más errores de cálculo (b) Porque aunque exista un efecto relevante, el estudio tenía pocas probabilidades de detectarlo (c) Porque el valor p es siempre mayor que 0,5

Explicaciones

- *Pooled standard deviation.*
- Son orientativas; importa el contexto.
- $1 - \beta$.
- Los pilotos pequeños dan efectos muy inciertos y a menudo inflados.
- Es el análisis a priori.
- `power.t.test(n = 60, delta = 0.4, sd = 1.3)`.
- *Smallest effect size of interest* (SESOI).
- La ausencia de significación no distingue «no hay efecto» de «no tenía forma de verlo».

Capítulo 6. Regresión

Objetivos

- Ajustar e interpretar modelos de regresión lineal simple y múltiple.
- Comprobar los supuestos con los residuos.
- Saber qué puede y qué no puede decir un coeficiente sobre causas.

Glosario del capítulo

Regresión lineal · Coeficiente · Residuo · R^2 · Variable centrada · Sobreajuste

Descargas

Script R del capítulo Plantilla de entrega del reto (.qmd)

Lectura obligatoria

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, 1(2), 103-113.

1. La recta que mejor se ajusta

La **regresión lineal** modela una variable resultado como una combinación lineal de predictores más un error:

`nota = b0 + b1 · horas + error`

`b1` es el cambio **esperado** en la nota por cada hora más de estudio; `b0`, la nota esperada con 0 horas (a menudo, sin sentido práctico: por eso conviene **centrar** los predictores, como recomienda Schielzeth, 2010).

```
d <- read.csv("datos/estudio.csv")
m1 <- lm(nota ~ horas, data = d)
summary(m1)$coefficients
confint(m1)
plot(d$horas, d$nota, pch = 19, col = "#2a78d6", xlab = "Horas de estudio", ylab =
"Nota")
abline(m1, col = "#eb6834", lwd = 2)
```

Cada hora más de estudio se asocia con unos 0,14 puntos más de nota (IC 95 %: 0,09 a 0,19). El **R^2** indica qué proporción de la variabilidad de la nota explica el modelo (aquí, en torno al 22 %).

2. Varios predictores a la vez

En la regresión **múltiple**, cada coeficiente es la asociación con el resultado **manteniendo constantes** los demás predictores del modelo.

```
d <- read.csv("datos/estudio.csv")
d$horas_c <- d$horas - mean(d$horas)      # centrar
d$sueno_c <- d$sueno - mean(d$sueno)
d$ansiedad_c <- d$ansiedad - mean(d$ansiedad)
m2 <- lm(nota ~ grupo + horas_c + sueno_c + ansiedad_c, data = d)
round(summary(m2)$coefficients, 3)
summary(m2)$r.squared
```

Con los predictores centrados, la constante es la nota esperada en el grupo de referencia (control) para alguien con horas, sueño y ansiedad medios. Fíjate en que el coeficiente de `grupotutoria` (el efecto de la tutoría

ajustado por las demás variables) se estima con algo más de precisión que en la simple comparación de medias: ajustar por covariables que predicen el resultado **aumenta la precisión** en un experimento.

3. Comprobar los supuestos

La regresión lineal supone, sobre todo: relación **lineal**, **independencia** de los errores, **varianza constante** y, para los intervalos con muestras pequeñas, errores aproximadamente **normales**. Se comprueba mirando los **residuos**:

```
d <- read.csv("datos/estudio.csv")
m1 <- lm(nota ~ horas, data = d)
par(mfrow = c(1, 2))
plot(fitted(m1), resid(m1), pch = 19, col = "#2a78d6", xlab = "Valores ajustados", ylab
= "Residuos"); abline(h = 0, lty = 2)
qqnorm(resid(m1), pch = 19, col = "#2a78d6"); qqline(resid(m1))
```

Buscas que los residuos no formen patrones (curvas, embudos) y que los puntos del gráfico Q-Q sigan la recta.

Laboratorio: un punto puede cambiarlo todo

i Laboratorio interactivo

Diez puntos sin relación aparente y un undécimo que puedes mover. Un punto alejado en el eje x (con mucha **palanca**) puede crear o destruir una pendiente.

Este laboratorio es interactivo en la versión web: pablosancheznunez.com/estadistica-investigacion

Ejercicio 6.1

Usa el modelo `m1` para **predecir** la nota de alguien que estudia 20 horas a la semana.

i Ejercicio: completa los huecos _____

```
predict(m1, newdata = data.frame(horas = ______))
```

i Pista

Sustituye el hueco por 20.

Solución

```
predict(m1, newdata = data.frame(horas = 20))
```

Unos 7,2 puntos. Con `interval = "prediction"` verás el intervalo para una persona concreta, mucho más ancho que el de la media.

Ejercicio 6.2

Calcula el $\mathbf{R^2}$ del modelo con horas y sueño.

Ejercicio: completa los huecos _____

```
m <- lm(nota ~ horas + ______, data = d)
summary(m)$r.squared
```

Pista

La variable del sueño se llama `sueno`.

Solución

```
m <- lm(nota ~ horas + sueno, data = d)
summary(m)$r.squared
```

Algo mayor que con solo las horas. El R^2 siempre sube al añadir predictores, aunque no aporten nada: por eso existe el R^2 ajustado.

Caso práctico resuelto

Interpretar los coeficientes de una regresión

Modelo: `nota ~ horas + sueno` con los datos del libro. Supongamos que el resultado es: intersección 4,1; horas 0,05; sueño 0,30.

- **horas = 0,05:** por cada hora semanal más de estudio, la nota aumenta en promedio 0,05 puntos, **manteniendo constante** el sueño.
- **sueno = 0,30:** cada hora más de sueño se asocia con 0,3 puntos más, a igualdad de horas de estudio.
- **Intersección:** nota prevista con 0 horas de estudio y 0 de sueño: no tiene sentido sustantivo; centrar las variables la hace interpretable (nota prevista para valores medios).
- **R^2** indica la proporción de la variación explicada, no si el modelo es causal ni si está bien especificado: hay que revisar los residuos.

Ejercicios graduados

Básico. ¿Qué significa un coeficiente de 0,30 para «sueño» en la regresión anterior?

Explicaciones

Que, a igualdad de horas de estudio, cada hora adicional de sueño se asocia con 0,3 puntos más de nota media en la muestra. No implica causalidad.

Avanzado. Ajusta el modelo con las variables centradas e interpreta de nuevo la intersección.

Explicaciones

Con `scale(x, scale = FALSE)` o restando la media, la intersección pasa a ser la nota prevista para un doctorando con horas y sueño medios; los demás coeficientes no cambian.

Actividades

i Actividad: El coeficiente de horas es 0,14 (IC 95 %: 0,09 a 0,19). ¿La interpretación es correcta?

Categorías: Correcta · Incorrecta

- Cada hora más de estudio se asocia con 0,14 puntos más de nota, en promedio
- Si un doctorando estudia una hora más, su nota subirá 0,14 puntos
- Los datos son compatibles con asociaciones entre 0,09 y 0,19 puntos por hora
- Las horas explican el 14 % de la nota

Solución

- Cada hora más de estudio se asocia con 0,14 puntos más de nota, en promedio → **Correcta**
- Si un doctorando estudia una hora más, su nota subirá 0,14 puntos → **Incorrecta**. Es un estudio observacional: no autoriza la lectura causal individual.
- Los datos son compatibles con asociaciones entre 0,09 y 0,19 puntos por hora → **Correcta**
- Las horas explican el 14 % de la nota → **Incorrecta**. Eso sería el R^2, no el coeficiente.

i Actividad: ¿Qué problema muestra el gráfico de residuos?

Categorías: Relación no lineal · Varianza no constante · Ninguno aparente

- Los residuos forman una U
- Los residuos se abren en abanico al aumentar los valores ajustados
- Una nube sin forma alrededor de cero

Solución

- Los residuos forman una U → **Relación no lineal**
- Los residuos se abren en abanico al aumentar los valores ajustados → **Varianza no constante**
- Una nube sin forma alrededor de cero → **Ninguno aparente**

Reto evaluable

! Reto del capítulo 6 · Un modelo bien contado

1. Con los datos del libro (o con los tuyos), plantea una pregunta y ajusta un modelo de regresión múltiple justificado (no un modelo con «todo lo que hay»).
2. Centra los predictores continuos e interpreta cada coeficiente con su intervalo, en las unidades originales.
3. Comprueba los supuestos con los residuos y revisa los puntos influyentes.
4. Escribe el párrafo de resultados y un párrafo de limitaciones sobre causalidad.

Entrega: con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. En una regresión múltiple, cada coeficiente se interpreta manteniendo constantes los demás predictores del modelo. *[V / F]*
2. Restar la media a un predictor se llama _____
3. El R^2 siempre aumenta al añadir predictores. *[V / F]*
4. Un residuo es *[el valor predicho / la diferencia entre el valor observado y el predicho / el coeficiente de una variable]*
5. Un coeficiente de regresión en un estudio observacional prueba una relación causal. *[V / F]*
6. Un punto con un valor de x muy alejado del resto tiene mucha _____
7. En un experimento, ajustar por covariables que predicen el resultado *[introduce sesgo siempre / suele mejorar la precisión del efecto / no cambia nada]*
8. ¿Qué intervalo es más ancho?

(a) El de la nota media esperada para quienes estudian 20 horas (b) El de la nota de una persona concreta que estudia 20 horas (c) Son iguales

Explicaciones

- «Ceteris paribus», dentro del modelo.
- Hace interpretable la constante.
- Por eso existe el R^2 ajustado.
- Observado menos predicho.
- Asociación ajustada, no causa.
- *Leverage*.
- Reduce el error típico del efecto.
- El de predicción incluye la variabilidad individual.

Capítulo 7. Introducción a la inferencia bayesiana

Objetivos

- Entender la lógica bayesiana: distribución previa, verosimilitud y distribución posterior.
- Calcular e interpretar una posterior y un intervalo de credibilidad en un caso sencillo.
- Conocer el factor de Bayes y cuándo conviene el enfoque bayesiano.

Glosario del capítulo

Distribución previa · Verosimilitud · Distribución posterior · Intervalo de credibilidad · Factor de Bayes

Descargas

Script R del capítulo Plantilla de entrega del reto (.qmd)

Lectura obligatoria

Kruschke, J. K.; Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25(1), 178-206.

Lecturas complementarias

- Wagenmakers et al. (2018)

1. Probabilidad como grado de creencia

En la estadística frecuentista, la probabilidad describe frecuencias a largo plazo y los parámetros son fijos. En la **bayesiana**, la probabilidad expresa **incertidumbre sobre los parámetros**, y se actualiza con los datos mediante el teorema de Bayes:

$$\text{posterior} \propto \text{verosimilitud} \times \text{previa}$$

- La **distribución previa** recoge lo que se sabía (o se supone) antes de los datos.
- La **verosimilitud** dice qué valores del parámetro hacen más probables los datos observados.
- La **distribución posterior** combina ambas: es lo que se sabe después.

Su gran ventaja es responder directamente a la pregunta que solemos querer: ¿qué valores del parámetro son **creíbles** dados los datos? Su gran exigencia es explicitar y justificar la previa.

2. Un caso que se calcula a mano: una proporción

Queremos estimar la proporción de doctorandos que aprueban un examen. Con una previa **Beta(a, b)** y k aprobados de n, la posterior es **Beta(a + k, b + n − k)**. La previa Beta(1, 1) es plana: todos los valores entre 0 y 1 igual de creíbles.

```
d <- read.csv("datos/estudio.csv")
k <- sum(d$nota >= 5); n <- nrow(d)
a <- 1; b <- 1                                  # previa plana
qbeta(c(0.025, 0.975), a + k, b + n - k)        # intervalo de credibilidad del 95 %
curve(dbeta(x, a + k, b + n - k), 0.5, 1, col = "#2a78d6", lwd = 2, xlab = "Proporción
de aprobados", ylab = "Densidad")
curve(dbeta(x, a, b), add = TRUE, col = "grey50", lty = 2)
```

El **intervalo de credibilidad** del 95 % sí admite la lectura intuitiva: dada la previa y los datos, hay un 95 % de probabilidad de que la proporción esté entre esos valores.

Laboratorio: actualizar con los datos

> **i** Laboratorio interactivo
>
> Elige una previa (su media y cuánta información aporta, en «observaciones equivalentes») y unos datos. Mira cómo la posterior se desplaza desde la previa hacia los datos y cómo los datos dominan cuando son abundantes.
>
> Este laboratorio es interactivo en la versión web: pablosancheznunez.com/estadistica-investigacion

3. Contrastar hipótesis: el factor de Bayes

El **factor de Bayes** (BF_{10}) compara cuánto mejor predicen los datos una hipótesis (H_1) que otra (H_0). Un BF_{10} = 6 significa que los datos son seis veces más probables bajo H_1 que bajo H_0. A diferencia del valor *p*, puede aportar evidencia **a favor** de la hipótesis nula. Depende, eso sí, de cómo se defina la previa de H_1.

4. ¿Cuándo usar el enfoque bayesiano?

- Cuando hay **información previa** legítima que conviene incorporar.
- Con **muestras pequeñas** o modelos complejos (multinivel), donde las previas estabilizan las estimaciones.
- Cuando se quiere cuantificar la evidencia **a favor de la nula**.
- Cuando se quiere hablar en términos de **probabilidad de los parámetros**.

En R, paquetes como **brms** o **rstanarm** ajustan modelos bayesianos con una sintaxis parecida a `lm()`, y **BayesFactor** calcula factores de Bayes. Fuera del navegador, son el siguiente paso natural tras este capítulo.

Ejercicio 7.1

Con una previa Beta(1, 1) y 15 aprobados de 20, calcula la **media de la posterior**: (a + k) / (a + b + n).

> **i** Ejercicio: completa los huecos _____

```
a <- 1; b <- 1; k <- 15; n <- 20
(a + k) / (a + b + _____)
```

> **i** Pista
>
> El denominador suma a, b y el número de observaciones n.

> Solución

```
a <- 1; b <- 1; k <- 15; n <- 20
(a + k) / (a + b + n)
```

0,727: un poco por debajo de la proporción observada (0,75), atraída hacia la media de la previa (0,5).

Ejercicio 7.2

Calcula el intervalo de credibilidad del 95 % de esa posterior, Beta(16, 6).

Ejercicio: completa los huecos _____

```
qbeta(c(0.025, _______), 16, 6)
```

Pista

El límite superior es el cuantil 0,975.

Solución

```
qbeta(c(0.025, 0.975), 16, 6)
```

Caso práctico resuelto

Actualizar una creencia con datos

Quieres estimar la proporción de doctorandos que publican durante el primer año. A priori, por estudios previos, crees que ronda el 20 % (distribución Beta(2, 8)). Observas 15 que publican de 40.

- **Posterior:** Beta(2 + 15, 8 + 25) = Beta(17, 33).
- **Media posterior:** 17 / 50 = 0,34. Los datos (15/40 = 0,375) acercan la estimación hacia ellos, moderados por la información previa.
- **Intervalo de credibilidad del 95 %:** aproximadamente [0,22; 0,47]; aquí sí puede decirse que hay un 95 % de probabilidad (según el modelo) de que la proporción esté en ese intervalo.

Con más datos, la información previa pesa cada vez menos.

Ejercicios graduados

Básico. Con una previa Beta(1, 1) y 15 éxitos en 20 intentos, ¿cuál es la media posterior?

Explicaciones

(1 + 15) / (1 + 1 + 20) = 16 / 22 ≈ 0,73.

Avanzado. Compara la posterior con tres previas distintas (poco informativa, optimista y pesimista) para los mismos datos y comenta la sensibilidad del resultado.

Explicaciones

Con pocos datos, la previa influye bastante; con muchos, las tres posteriores convergen. Informar del análisis de sensibilidad es una buena práctica.

Actividades

Actividad: ¿De qué concepto se trata?

Categorías: Previa · Verosimilitud · Posterior · Factor de Bayes

- Lo que se sabía de la tasa de aprobados antes de este curso
- Cómo de probables son los datos observados para cada valor del parámetro
- La distribución de la tasa de aprobados una vez vistos los datos
- Los datos son 8 veces más probables si la tutoría tiene efecto que si no

Solución

- Lo que se sabía de la tasa de aprobados antes de este curso → **Previa**
- Cómo de probables son los datos observados para cada valor del parámetro → **Verosimilitud**
- La distribución de la tasa de aprobados una vez vistos los datos → **Posterior**
- Los datos son 8 veces más probables si la tutoría tiene efecto que si no → **Factor de Bayes**

Actividad: ¿Frecuentista o bayesiana?

Categorías: Frecuentista · Bayesiana

- «El 95 % de los intervalos calculados así contendría el valor verdadero»
- «Hay un 95 % de probabilidad de que el parámetro esté entre 0,6 y 0,8»
- «Si la hipótesis nula fuera cierta, datos así aparecerían un 3 % de las veces»
- «Los datos apoyan la hipótesis nula frente a la alternativa con un factor de 5»

Solución

- «El 95 % de los intervalos calculados así contendría el valor verdadero» → **Frecuentista**
- «Hay un 95 % de probabilidad de que el parámetro esté entre 0,6 y 0,8» → **Bayesiana**
- «Si la hipótesis nula fuera cierta, datos así aparecerían un 3 % de las veces» → **Frecuentista**
- «Los datos apoyan la hipótesis nula frente a la alternativa con un factor de 5» → **Bayesiana**

Reto evaluable

Reto del capítulo 7 · Dos lecturas de los mismos datos

1. Elige un resultado de tu tesis o de un artículo con una proporción o una diferencia de medias.
2. Analízalo de forma **frecuentista** (intervalo y valor p) y **bayesiana** (posterior e intervalo de credibilidad) con al menos dos previas distintas: una poco informativa y otra informativa justificada.
3. Compara las conclusiones y discute la sensibilidad a la previa.
4. Explica, para un público no especialista, qué significa cada intervalo.

Entrega: con la plantilla de entrega. Este reto **no tiene solución en el libro**.

Autoevaluación

Responde a todas las preguntas y pulsa **Mostrar respuestas**. No cuenta para la nota.

1. En la inferencia bayesiana, la posterior es proporcional a la verosimilitud por la _____
2. Con previa Beta(1, 1) y 15 éxitos de 20, la posterior es *[Beta(15, 5) / Beta(16, 6) / Beta(1, 1)]*
3. El intervalo de credibilidad admite la lectura «hay un 95 % de probabilidad de que el parámetro esté aquí». *[V / F]*
4. Un factor de Bayes puede aportar evidencia a favor de la hipótesis nula. *[V / F]*
5. Con muchos datos, la posterior depende sobre todo de *[la previa / los datos / el investigador]*
6. Un paquete de R para ajustar modelos bayesianos con sintaxis parecida a `lm()` es *[ggplot2 / brms / dplyr]*
7. El enfoque bayesiano no necesita justificar ninguna decisión. *[V / F]*
8. ¿Qué hace la previa con una muestra pequeña?

(a) No tiene ningún efecto (b) Atrae la estimación hacia sus valores y la estabiliza (c) Hace que el intervalo sea infinito

Explicaciones

- Teorema de Bayes.
- a + k = 16; b + n − k = 6.
- Esa es su ventaja interpretativa.
- A diferencia del valor *p*.
- Los datos dominan cuando abundan.
- *Bayesian regression models using Stan*.
- La previa debe justificarse y analizarse su sensibilidad.
- Efecto de contracción (*shrinkage*).

Chuleta de R

Todas las funciones que usa el libro, agrupadas por tarea.

Datos

Tarea	Código
Leer un CSV	`d <- read.csv("datos/estudio.csv")`
Ver la estructura	`str(d)`, `head(d)`, `summary(d)`
Una columna	`d$nota`
Filtrar filas	`d[d$grupo == "tutoria", ]`
Crear una columna	`d$horas_c <- d$horas - mean(d$horas)`
Frecuencias	`table(d$grupo)`

Describir

Tarea	Código
Media, mediana	`mean(x)`, `median(x)`
Desviación típica, IQR	`sd(x)`, `IQR(x)`
Cuantiles	`quantile(x, c(0.25, 0.5, 0.75))`
Resumen por grupo	`tapply(d$nota, d$grupo, mean)`
Proporción que cumple algo	`mean(x >= 5)`
Correlación	`cor(x, y)`

Gráficos

Tarea	Código
Puntos por grupo	`stripchart(y ~ g, data = d, vertical = TRUE, method = "jitter")`
Caja	`boxplot(y ~ g, data = d)`
Dispersión	`plot(x, y)`
Histograma	`hist(x)`
Recta de regresión	`abline(lm(y ~ x))`

Inferencia

Tarea	Código
IC de una media	`t.test(x)$conf.int`
Dos grupos	`t.test(y ~ g, data = d)`
Emparejado	`t.test(a, b, paired = TRUE)`
No paramétrico	`wilcox.test(y ~ g, data = d)`
Correlación con IC	`cor.test(x, y)`
Tabla de contingencia	`chisq.test(table(a, b))`
Corregir comparaciones múltiples	`p.adjust(p, method = "holm")`

Tarea	Código
Bootstrap	`replicate(2000, median(sample(x, replace = TRUE)))`

Efecto y potencia

Tarea	Código
Potencia de un estudio	`power.t.test(n = 60, delta = 0.4, sd = 1.3)`
Muestra necesaria	`power.t.test(delta = 0.4, sd = 1.3, power = 0.8)`
d de Cohen	`(mean(a) - mean(b)) / sqrt((var(a) + var(b)) / 2)`

Modelos

Tarea	Código
Regresión	`m <- lm(y ~ x1 + x2, data = d)`
Coeficientes e IC	`summary(m)$coefficients`, `confint(m)`
R^2	`summary(m)$r.squared`
Predicción	`predict(m, newdata = data.frame(x1 = 20, x2 = 7))`
Residuos	`resid(m)`, `fitted(m)`, `qqnorm(resid(m))`
Puntos influyentes	`cooks.distance(m)`

Bayes (proporción)

Tarea	Código
Posterior Beta	`Beta(a + k, b + n - k)`
Intervalo de credibilidad	`qbeta(c(0.025, 0.975), a + k, b + n - k)`
Densidad	`curve(dbeta(x, a + k, b + n - k), 0, 1)`

Laboratorio interactivo

Este apéndice reúne laboratorios y actividades para practicar las ideas del libro con simulaciones que se ejecutan en tu navegador.

Laboratorio: el teorema central del límite

Laboratorio interactivo

Elige la forma de la población y el tamaño de cada muestra. El laboratorio extrae 3000 muestras, calcula sus medias y dibuja cómo se distribuyen. Aunque la población sea asimétrica o tenga dos picos, la distribución de las medias se acerca a una normal al crecer n, y su dispersión es $\sigma/\sqrt{n}$: el **error típico**.

Este laboratorio es interactivo en la versión web: pablosancheznunez.com/estadistica-investigacion

Actividad: ¿qué prueba usarías?

Actividad: Elige la prueba para cada pregunta de investigación

Categorías: t para dos grupos · ANOVA de un factor · Chi cuadrado · Correlación de Pearson · Regresión logística

- ¿Difiere la nota media entre el grupo con tutoría y el grupo sin tutoría?
- ¿Difiere el tiempo de lectura entre tres formatos de texto?
- ¿Está asociada la facultad con publicar en acceso abierto (sí/no)?
- ¿Se relacionan las horas de estudio y la nota final?
- ¿Predicen la edad y la experiencia previa si alguien abandona el curso (sí/no)?
- ¿Tienen más citas los artículos con datos abiertos que los que no los tienen?

Solución

- ¿Difiere la nota media entre el grupo con tutoría y el grupo sin tutoría? → **t para dos grupos**. Una variable cuantitativa comparada entre dos grupos.
- ¿Difiere el tiempo de lectura entre tres formatos de texto? → **ANOVA de un factor**. Una cuantitativa comparada entre tres o más grupos.
- ¿Está asociada la facultad con publicar en acceso abierto (sí/no)? → **Chi cuadrado**. Dos variables categóricas.
- ¿Se relacionan las horas de estudio y la nota final? → **Correlación de Pearson**. Dos variables cuantitativas.
- ¿Predicen la edad y la experiencia previa si alguien abandona el curso (sí/no)? → **Regresión logística**. Resultado binario con varios predictores.
- ¿Tienen más citas los artículos con datos abiertos que los que no los tienen? → **t para dos grupos**. Dos grupos; con citas muy asimétricas, considera transformarlas o usar Mann-Whitney.

Laboratorio: conversor de tamaños del efecto

> ℹ Laboratorio interactivo
>
> Los estudios informan del efecto en métricas distintas. Este conversor pasa entre la *d* de Cohen, la correlación *r* y la razón de odds (OR), y traduce el efecto a probabilidades que se entienden sin estadística. Las conversiones suponen dos grupos de igual tamaño y distribuciones normales.
>
> Fíjate en que incluso un efecto «grande» (d = 0,8) deja un solapamiento de casi el 70 %: muchas personas del grupo control puntúan más que muchas del grupo tratado.
>
> Este laboratorio es interactivo en la versión web: pablosancheznunez.com/estadistica-investigacion

Para ver y seguir aprendiendo

Son enlaces, no vídeos incrustados: el vídeo se abre en la web de quien lo publica, así esta página no carga rastreadores de terceros. La mayoría está en inglés; activa los subtítulos si los necesitas.

- **Dance of the p values** — Geoff Cumming, La Trobe University. Muestra con simulaciones lo mucho que cambia el valor p al repetir el mismo experimento.

Guía para el profesorado

El libro sirve como manual de un curso de doctorado de estadística aplicada de unas 30-40 horas (siete sesiones de 4-5 horas) o como material autónomo. No exige conocimientos previos de R.

Propuesta de evaluación

Actividad	Peso orientativo	Apoyo en el libro
Retos evaluables (elegir 5 de 7)	60 %	Un reto por capítulo, con rúbrica en la plantilla de entrega
Análisis final con datos propios	30 %	Integra los retos 1, 3, 5 y 6
Autoevaluaciones y actividades	10 %	En cada capítulo y en Aula Global

La rúbrica común valora la pregunta y el diseño (20 %), el análisis correcto (30 %), la interpretación de la incertidumbre (25 %), la reproducibilidad (15 %) y la comunicación (10 %).

Datos

`datos/estudio.csv` contiene 120 doctorandos **ficticios** simulados con semilla fija: grupo asignado al azar (tutoría o control), horas de estudio semanales, horas de sueño, ansiedad (10-60) y nota (0-10). El efecto simulado de la tutoría es pequeño y el estudio tiene poca potencia para detectarlo, a propósito: es el hilo que conecta los capítulos 3 a 6.

Actividades interactivas y laboratorios

Laboratorios que se usan sin programar: cuarteto de Anscombe, baile de los intervalos de confianza, distribución de los valores *p*, potencia, puntos con palanca en regresión y actualización bayesiana Beta-Binomial. En el PDF y el EPUB, las actividades aparecen en papel con su solución.

Autoevaluación en Aula Global

Las preguntas están exportadas a **Moodle XML** en la carpeta `moodle/` del repositorio.

Descarga el banco de preguntas: todo el libro (Moodle XML) · por capítulo: 00 · 01 · 02 · 03 · 04 · 05 · 06 · 07

Diapositivas

diapositivas/ (F, pantalla completa; S, notas).

Mantenimiento

```
python3 herramientas/preparar.py        # sw.js y flujo de integración continua
python3 herramientas/extraer_r.py       # scripts .R de cada capítulo
python3 herramientas/moodle.py          # banco de preguntas
python3 herramientas/figuras.py         # figuras
python3 herramientas/pdf.py             # PDF y EPUB
quarto render
python3 herramientas/diapositivas.py    # diapositivas de clase
```

Glosario

Términos del libro ordenados por capítulo. Cada entrada enlaza con el capítulo donde se explica.

Pensar con datos

Población y muestra
La población es el conjunto sobre el que se quiere concluir; la muestra, la parte que se observa. → Capítulo 1

Variable
Característica que toma valores distintos en las unidades estudiadas: nominal, ordinal, discreta o continua. → Capítulo 1

Estudio observacional
Estudio en el que se mide sin intervenir; permite describir y encontrar asociaciones. → Capítulo 1

Experimento aleatorizado
Estudio en el que el azar decide quién recibe cada condición; permite inferir causas. → Capítulo 1

Variable de confusión
Factor que influye a la vez en la supuesta causa y en el efecto, y crea asociaciones engañosas. → Capítulo 1

Validez interna y externa
Interna: si la conclusión es correcta para la muestra. Externa: si se puede generalizar. → Capítulo 1

Describir y visualizar

Media y mediana
La media es el valor de equilibrio; la mediana, el valor central. La mediana resiste mejor los extremos. → Capítulo 2

Desviación típica
Distancia típica de los datos a su media. → Capítulo 2

Rango intercuartílico
Diferencia entre los percentiles 75 y 25: amplitud del 50 % central de los datos. → Capítulo 2

Valor atípico
Observación muy alejada del resto; hay que investigarla, no eliminarla sin justificación. → Capítulo 2

Correlación
Medida de la fuerza de una relación lineal entre dos variables, entre −1 y 1. → Capítulo 2

Estimar con incertidumbre

Distribución muestral
Distribución de un estadístico si el estudio se repitiera muchas veces. → Capítulo 3

Error típico
Desviación típica de la distribución muestral; para una media, aproximadamente $sd/\sqrt{n}$. → Capítulo 3

Intervalo de confianza
Intervalo calculado con un procedimiento que, repetido, contiene el valor verdadero en un porcentaje fijado de las veces. → Capítulo 3

Estimación
Cuantificación de un parámetro con su incertidumbre, en lugar de una decisión de sí o no. → Capítulo 3

Remuestreo (*bootstrap*)
Aproximación a la distribución muestral remuestreando con reemplazo los propios datos. → Capítulo 3

Contrastes y valores p

Hipótesis nula
Hipótesis de referencia, normalmente de ausencia de efecto, frente a la que se evalúan los datos. → Capítulo 4

Valor *p*
Probabilidad de obtener datos tan extremos o más que los observados si la hipótesis nula y el modelo fueran ciertos. → Capítulo 4

Nivel de significación
Umbral α, fijado de antemano, por debajo del cual se rechaza la hipótesis nula. → Capítulo 4

Error de tipo I y II
Tipo I: rechazar una nula cierta (falso positivo). Tipo II: no rechazar una nula falsa (falso negativo). → Capítulo 4

Prueba t
Contraste para comparar medias; la versión de Welch no supone varianzas iguales. → Capítulo 4

Comparaciones múltiples
Problema de hacer muchos contrastes: aumenta la probabilidad de algún falso positivo. → Capítulo 4

Tamaño del efecto y potencia

Tamaño del efecto
Magnitud de un fenómeno, en unidades originales o estandarizada. → Capítulo 5

d de Cohen
Diferencia de medias dividida entre la desviación típica combinada. → Capítulo 5

Potencia
Probabilidad de detectar un efecto de un tamaño dado si existe. → Capítulo 5

Efecto mínimo de interés
Efecto más pequeño que tendría importancia teórica o práctica. → Capítulo 5

Justificación de la muestra
Argumento explícito de por qué se usa un tamaño de muestra concreto. → Capítulo 5

Regresión

Regresión lineal
Modelo que expresa un resultado como combinación lineal de predictores más un error. → Capítulo 6

Coeficiente
Cambio esperado en el resultado por unidad de un predictor, con los demás constantes. → Capítulo 6

Residuo
Diferencia entre el valor observado y el predicho por el modelo. → Capítulo 6

R^2
Proporción de la variabilidad del resultado explicada por el modelo. → Capítulo 6

Variable centrada
Variable a la que se resta su media para que el cero sea interpretable. → Capítulo 6

Sobreajuste
Modelo que se ajusta al ruido de la muestra y predice mal datos nuevos. → Capítulo 6

Inferencia bayesiana

Distribución previa
Lo que se sabe o se supone de un parámetro antes de ver los datos. → Capítulo 7

Verosimilitud
Cuán probables son los datos observados para cada valor del parámetro. → Capítulo 7

Distribución posterior
Distribución del parámetro una vez combinadas la previa y los datos. → Capítulo 7

Intervalo de credibilidad
Intervalo que contiene el parámetro con una probabilidad dada, según la posterior. → Capítulo 7

Factor de Bayes
Cociente que compara cuánto mejor predicen los datos una hipótesis que otra. → Capítulo 7

Referencias

Bibliografía

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7-29. https://doi.org/10.1177/0956797613504966

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., y Altman, D. G. (2016). Statistical tests, \emph{P} values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337-350. https://doi.org/10.1007/s10654-016-0149-3

Halsey, L. G., Curran-Everett, D., Vowler, S. L., y Drummond, G. B. (2015). The fickle \emph{P} value generates irreproducible results. *Nature Methods*, *12*(3), 179-185. https://doi.org/10.1038/nmeth.3288

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for \emph{t}-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), 33267. https://doi.org/10.1525/collabra.33267

Leek, J. T., y Peng, R. D. (2015). What is the question?. *Science*, *347*(6228), 1314-1315. https://doi.org/10.1126/science.aaa6146

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., y others. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35-57. https://doi.org/10.3758/s13423-017-1343-3

Wasserstein, R. L., y Lazar, N. A. (2016). The ASA statement on \emph{p}-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129-133. https://doi.org/10.1080/00031305.2016.1154108

Weissgerber, T. L., Milic, N. M., Winham, S. J., y Garovic, V. D. (2015). Beyond bar and line graphs: Time for a new data presentation paradigm. *PLOS Biology*, *13*(4), e1002128. https://doi.org/10.1371/journal.pbio.1002128