UNIVERSIDAD DE GRANADA

Técnicas Cuantitativas II

Temario de la asignatura

Ismael Sallami Moreno

Recursos Ingeniería Informática y Ade

Licencia

Este trabajo está bajo una Licencia Creative Commons BY-NC-ND 4.0.

Permisos: Se permite compartir, copiar y redistribuir el material en cualquier medio o formato.

Condiciones: Es necesario dar crédito adecuado, proporcionar un enlace a la licencia e indicar si se han realizado cambios. No se permite usar el material con fines comerciales ni distribuir material modificado.

Técnicas Cuantitativas II

Ismael Sallami Moreno

https://elblogdeismael.github.io

Índice general

Capítulo 1

Introducción: modelos continuos, muestra y estadísticos

Tema 1 del programa. Los modelos continuos de variable aleatoria —uniforme, exponencial, gamma, beta, normal y las asociadas a la normal—, los conceptos de muestra y estadístico, y las primeras propiedades de la media y la varianza muestrales.

1.1 Modelos continuos

1.1.1 Uniforme

$$X \sim U(a,b): \qquad f(x) = \frac{1}{b-a} \text{ en } [a,b]$$

$$E[X] = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

Modela la ignorancia total dentro de un rango, y es la base de la simulación: cualquier distribución se genera transformando una uniforme.

1.1.2 Exponencial

$$X \sim \mathrm{Exp}(\lambda): \qquad f(x) = \lambda e^{-\lambda x} \text{ en } [0,\infty)$$

$$E[X] = \frac{1}{\lambda}, \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}, \qquad F(x) = 1 - e^{-\lambda x}$$

Modela **tiempos de espera** entre sucesos que ocurren según un proceso de Poisson: si las llamadas llegan a razón de λ por hora, el tiempo entre dos llamadas es exponencial de parámetro λ.

Proposición 1.1 (Falta de memoria).

$$P(X > s+t \mid X > s) = P(X > t)$$

La exponencial es la única distribución continua con esta propiedad.

Es la versión continua de la geométrica del curso anterior, con la misma lectura: **un componente sin desgaste no envejece**. Que sea la única con esa propiedad es lo que la hace tan usada, y también lo que limita su realismo: los equipos reales sí se desgastan, y por eso la fiabilidad usa la Weibull.

1.1.3 Gamma

$$X \sim \Gamma(\alpha, \lambda): \qquad f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0$$

$$E[X] = \frac{\alpha}{\lambda}, \qquad \mathrm{Var}(X) = \frac{\alpha}{\lambda^2}$$

Generaliza la exponencial, que es el caso $\alpha = 1$. La suma de α exponenciales independientes del mismo parámetro es una gamma, así que modela el tiempo hasta el α-ésimo suceso.

1.1.4 Beta

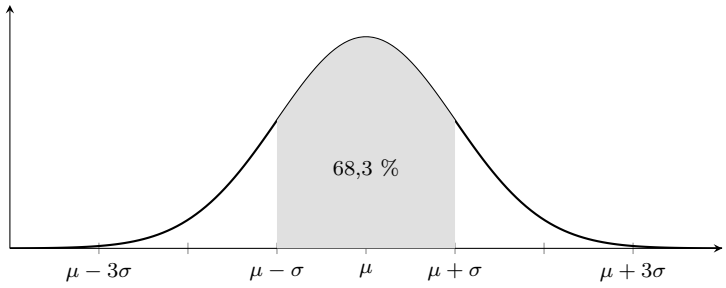
$$X \sim \mathrm{Beta}(p, q): \qquad f(x) = \frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}, \quad x \in (0,1)$$

$$E[X] = \frac{p}{p+q}, \qquad \mathrm{Var}(X) = \frac{pq}{(p+q)^2(p+q+1)}$$

Su soporte es el intervalo $(0, 1)$, así que modela proporciones y porcentajes: cuotas de mercado, tasas de aceptación, fracciones defectuosas. Su forma es muy flexible: con $p = q = 1$ es la uniforme, con $p = q > 1$ es acampanada y simétrica, y con parámetros distintos es asimétrica.

1.1.5 Normal

$$X \sim N(\mu, \sigma): \qquad f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Propiedad	Enunciado
Simetría	respecto de μ; media, mediana y moda coinciden
Tipificación	$Z = (X - \mu)/\sigma \sim N(0, 1)$
Reproductividad	la suma de normales independientes es normal
Combinación lineal	$aX + b \sim N(a\mu + b,\ \lvert a\rvert\sigma)$
Regla empírica	el 68,3 %, 95,4 % y 99,7 % de la masa en $\mu \pm \sigma$, $\pm 2\sigma$ y $\pm 3\sigma$

La tipificación es lo que permite usar una sola tabla para todas las normales, y es el cálculo más repetido de la asignatura.

Teorema 1.1 (Central del límite). Si $X_1, \ldots, X_n$ son independientes e idénticamente distribuidas

con media μ y varianza σ^2 finitas, entonces

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow[n\to\infty]{} N(0,1)$$

sea cual sea la distribución de partida.

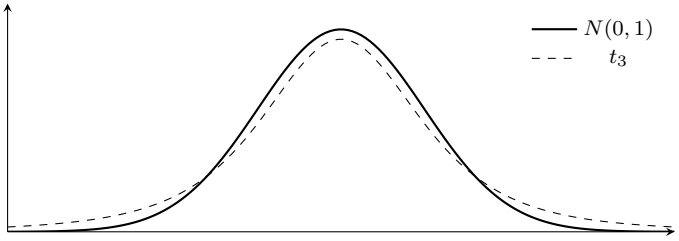
Es el resultado que sostiene toda la inferencia del curso: **la media muestral es aproximadamente normal aunque los datos no lo sean**, y por eso los intervalos y los contrastes de los temas siguientes funcionan con datos reales. En la práctica se aplica a partir de $n \geq 30$, y con menos si la población ya era simétrica.

1.1.6 Distribuciones asociadas a la normal

Distribución	Definición	Se usa para
χ^2_n	suma de n normales tipificadas al cuadrado	varianzas
t_n	$\dfrac{Z}{\sqrt{\chi^2_n/n}}$ con Z y χ^2 independientes	medias con σ desconocida
$F_{n,m}$	$\dfrac{\chi^2_n/n}{\chi^2_m/m}$	cocientes de varianzas

Distribución	Media	Forma
χ^2_n	n	asimétrica a la derecha; tiende a la normal con n grande
t_n	0 si $n > 1$	simétrica, con colas más pesadas que la normal
$F_{n,m}$	$m/(m-2)$ si $m > 2$	asimétrica a la derecha

La t tiende a la normal al crecer n, y a partir de unos 30 grados de libertad la diferencia es despreciable. Sus colas más pesadas son la penalización por no conocer σ y tener que estimarla: los intervalos salen algo más anchos.



1.2 Muestra y estadístico

Definición 1.1 (Muestra aleatoria simple). Conjunto de n variables aleatorias $X_1, \ldots, X_n$ independientes y con la misma distribución que la población.

Definición 1.2 (Estadístico). Cualquier función de la muestra que no dependa de parámetros desconocidos.

Un estadístico es a su vez una variable aleatoria, porque cambia de una muestra a otra. Su distribución se llama distribución en el muestreo, y es el objeto de estudio del tema 3.

Estadístico	Definición
Media muestral	$\bar{X} = \frac{1}{n}\sum X_i$
Varianza muestral	$S^2 = \frac{1}{n}\sum(X_i - \bar{X})^2$
Cuasivarianza muestral	$\hat{S}^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$
Proporción muestral	$\hat{p} = \frac{X}{n}$ con X el número de éxitos

Nota 1.1 . Varianza y cuasivarianza se diferencian solo en el denominador y **no son intercambiables**. La segunda es la que aparece en los intervalos y los contrastes, porque es la insesgada. Confundirlas produce resultados sistemáticamente sesgados, tanto más cuanto menor es n: con $n = 5$ la diferencia es del 25 %.

1.3 Valor esperado y varianza de los estadísticos

Proposición 1.2 (Media muestral).

$$E[\bar{X}] = \mu, \qquad \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

Las dos igualdades dicen cosas distintas y las dos importan:

- $E[\bar{X}] = \mu$: la media muestral **acierta en promedio**, sin sesgo.
- $\mathrm{Var}(\bar{X}) = \sigma^2/n$: su precisión mejora con el tamaño de muestra, pero **como** $1/\sqrt{n}$. Para dividir el error por dos hay que cuadruplicar la muestra, y esa es la ley de rendimientos decrecientes de toda la estadística.

Proposición 1.3 (Varianza y cuasivarianza).

$$E[S^2] = \frac{n-1}{n}\sigma^2, \qquad E[\hat{S}^2] = \sigma^2$$

La varianza muestral **subestima** sistemáticamente σ^2, y la cuasivarianza no. La razón intuitiva: las desviaciones se miden respecto de $\bar{X}$, que está calculada con los propios datos y por tanto queda más cerca de ellos de lo que estaría la μ verdadera. Dividir por $n-1$ compensa exactamente ese efecto, y el $n-1$ se llama grados de libertad porque una de las n desviaciones queda determinada por las otras.

Ejemplo 1.1 . Una población tiene $\mu = 50$ y $\sigma = 10$. Con muestras de tamaño 25:

$E[\bar{X}] = 50$ y $\mathrm{Var}(\bar{X}) = 100/25 = 4$, así que el error típico de la media es $\sigma/\sqrt{n} = 2$.

Por el teorema central del límite, $\bar{X} \approx N(50,\ 2)$, y por tanto en torno al 95 % de las muestras dan

una media entre 46 y 54. Con $n = 100$ el error típico baja a 1 y el intervalo se estrecha a 48-52: cuadruplicar la muestra ha reducido el error a la mitad.

1.4 Ejercicios

Ejercicio 1.4.1. El tiempo entre llegadas a una ventanilla es exponencial con media 4 minutos. ¿Cuál es la probabilidad de esperar más de 6 minutos? ¿Y de esperar más de 6 sabiendo que ya se han esperado 3?

Solución 1.4.1. $\lambda = 1/4$, así que $P(X > 6) = e^{-6/4} = e^{-1,5} = 0{,}2231$.

Por la falta de memoria, $P(X > 9 \mid X > 3) = P(X > 6) = 0{,}2231$: haber esperado tres minutos no cambia nada. Y $P(X > 6 \mid X > 3) = P(X > 3) = e^{-0,75} = 0{,}4724$.

Ejercicio 1.4.2. Las ventas diarias de una tienda tienen media 500 y desviación típica 120, con distribución desconocida. ¿Cuál es la probabilidad de que la media de 36 días supere 540?

Solución 1.4.2. Por el teorema central del límite, $\bar{X} \approx N(500,\ 120/6) = N(500, 20)$. Tipificando, $z = (540 - 500)/20 = 2$, y $P(Z > 2) = 0{,}0228$.

Nótese que no ha hecho falta conocer la distribución de las ventas diarias: con $n = 36$ el teorema ya se aplica, y esa es toda su utilidad.

Ejercicio 1.4.3. De una muestra de 5 datos se obtiene $\sum(x_i - \bar{x})^2 = 40$. Calcular la varianza y la cuasivarianza muestrales, y decir cuál estima mejor σ^2.

Solución 1.4.3. $S^2 = 40/5 = 8$ y $\hat{S}^2 = 40/4 = 10$. La cuasivarianza es la insesgada, así que es la que estima bien σ^2; la varianza muestral subestima en un factor $(n-1)/n = 0{,}8$, un 20 % en este caso. Con $n = 100$ el sesgo sería del 1 %, y por eso la distinción importa sobre todo con muestras pequeñas.

Los modelos continuos y los conceptos de muestreo están desarrollados en Canavos, 1987 y Herrerías, Palacios y Callejón, 2012, con problemas resueltos en Herrerías, Palacios, Pérez et al., 2012 y Casas et al., 2006.

Capítulo 2

Estimación puntual de parámetros

Tema 2 del programa. El concepto de estimador, los métodos de máxima verosimilitud y de los momentos, y las propiedades deseables: insesgadez, consistencia, eficiencia y suficiencia.

2.1 Concepto

Definición 2.1 (Estimador). Estadístico $\hat{\theta} = T(X_1, \ldots, X_n)$ que se usa para aproximar un parámetro desconocido θ de la población.

Conviene separar dos cosas que la notación confunde:

Término	Qué es
Estimador	la fórmula, que es una variable aleatoria
Estimación	el número que sale con una muestra concreta

Preguntar por la probabilidad de que la estimación acierte no tiene sentido: el número ya está fijado. Lo que tiene distribución es el estimador, y es de él de lo que se habla al decir que un método es bueno.

2.2 Método de máxima verosimilitud

La idea: **elegir el valor del parámetro que hace más probable lo observado**.

Definición 2.2 (Función de verosimilitud).

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

El estimador de máxima verosimilitud es el valor de θ que la maximiza.

En la práctica se maximiza el logaritmo, que tiene el mismo máximo y convierte el producto en suma:

$$\ell(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i; \theta), \qquad \frac{d\ell}{d\theta} = 0$$

Ejemplo 2.1 (Bernoulli). Con n observaciones de las que k son éxitos,

$$L(p) = p^k(1-p)^{n-k}, \qquad \ell(p) = k\ln p + (n-k)\ln(1-p)$$

$$\frac{d\ell}{dp} = \frac{k}{p} - \frac{n-k}{1-p} = 0 \implies \hat{p} = \frac{k}{n}$$

La proporción muestral, que es lo que la intuición dictaba. Que el método lo confirme es lo que da confianza en aplicarlo donde la intuición no llega.

Ejemplo 2.2 (Normal).

$$\hat{\mu} = \bar{X}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum(X_i - \bar{X})^2 = S^2$$

La media muestral y **la varianza muestral, no la cuasivarianza**. El estimador de máxima verosimilitud de σ^2 es sesgado, lo que muestra que el método no garantiza insesgadez.

Ventaja	Inconveniente
Es consistente y asintóticamente eficiente	puede ser sesgado en muestras pequeñas
Invariante: si $\hat{\theta}$ estima θ, $g(\hat{\theta})$ estima $g(\theta)$	a veces no tiene solución cerrada
Usa toda la información de la muestra	exige conocer la forma de la distribución

La invarianza es muy cómoda: si $\hat{\sigma}^2$ es el estimador máximo verosímil de la varianza, $\sqrt{\hat{\sigma}^2}$ lo es de la desviación típica, sin volver a derivar nada. La insesgadez, en cambio, no se conserva al transformar.

2.3 Método de los momentos

Igualar los momentos muestrales a los poblacionales y despejar:

$$a_r = \frac{1}{n}\sum X_i^r \;=\; E[X^r] \qquad r = 1, 2, \ldots$$

Se plantean tantas ecuaciones como parámetros haya.

Ejemplo 2.3 (Uniforme). Para $U(0,\theta)$, $E[X] = \theta/2$, así que igualando a $\bar{X}$ sale $\hat{\theta} = 2\bar{X}$.

Es un estimador insesgado y **puede dar valores imposibles**: si la muestra es $\{1, 2, 9\}$, entonces $\hat{\theta} = 8$, menor que el 9 observado. El de máxima verosimilitud, $\hat{\theta} = \max X_i = 9$, nunca produce esa incoherencia.

Método	Ventaja	Inconveniente
Momentos	sencillo, no exige conocer la densidad completa	puede dar estimaciones absurdas
Máxima verosimilitud	mejores propiedades asintóticas	más laborioso

2.4 Propiedades deseables

2.4.1 Insesgadez

Definición 2.3 . $\hat{\theta}$ es insesgado si $E[\hat{\theta}] = \theta$ para todo valor de θ. En otro caso, su sesgo es $\text{Sesgo}(\hat{\theta}) = E[\hat{\theta}] - \theta$.

Estimador	¿Insesgado?
$\bar{X}$ para μ	sí
$\hat{p}$ para p	sí
S^2 para σ^2	**no**: $E[S^2] = \frac{n-1}{n}\sigma^2$
$\hat{S}^2$ para σ^2	sí
$\hat{S}$ para σ	**no**, aunque $\hat{S}^2$ sí lo sea

La última fila recoge la advertencia importante: **la insesgadez no se conserva al transformar**, porque $E[g(X)] \neq g(E[X])$ salvo que g sea lineal.

2.4.2 Eficiencia

Entre dos estimadores insesgados, es preferible el de menor varianza. El criterio que compara todos, sesgados incluidos, es el **error cuadrático medio**:

$$\text{ECM}(\hat{\theta}) = E\big[(\hat{\theta} - \theta)^2\big] = \text{Var}(\hat{\theta}) + \text{Sesgo}(\hat{\theta})^2$$

Nota 2.1 . La descomposición del error cuadrático medio explica por qué un estimador sesgado puede ser preferible: si a cambio del sesgo reduce mucho la varianza, su error total es menor. Es el compromiso entre sesgo y varianza, y reaparece idéntico en aprendizaje automático.

Teorema 2.1 (Cota de Cramér-Rao). Bajo condiciones de regularidad, todo estimador insesgado cumple

$$\text{Var}(\hat{\theta}) \geq \frac{1}{n\, I(\theta)}$$

con $I(\theta)$ la información de Fisher. Un estimador que alcanza la cota se llama eficiente.

La cota dice que **hay un límite a lo que se puede afinar con n datos**, y no depende del ingenio del estadístico sino de la propia distribución.

2.4.3 Consistencia

Definición 2.4 . $\hat{\theta}_n$ es consistente si converge en probabilidad a θ:

$$\lim_{n\to\infty} P\big(|\hat{\theta}_n - \theta| > \varepsilon\big) = 0 \quad \text{para todo } \varepsilon > 0$$

Proposición 2.1 . Si $E[\hat{\theta}_n] \to \theta$ y $\text{Var}(\hat{\theta}_n) \to 0$, entonces $\hat{\theta}_n$ es consistente.

La condición suficiente es la que se usa en la práctica, y con $\bar{X}$ se comprueba de inmediato: es

insesgado y su varianza σ^2/n tiende a cero.

La consistencia es un requisito mínimo. Un estimador que no mejora al crecer la muestra no sirve para nada, por muchas otras propiedades que tenga.

2.4.4 Suficiencia

Definición 2.5 . T es suficiente para θ si la distribución de la muestra condicionada a T no depende de θ.

Teorema 2.2 (Factorización de Fisher-Neyman). T es suficiente si y solo si la verosimilitud se factoriza como

$$L(\theta) = g\big(T(x), \theta\big) \cdot h(x)$$

con h sin depender de θ.

Un estadístico suficiente **resume la muestra sin perder información** sobre el parámetro. En la Bernoulli, el número de éxitos es suficiente para p: saber además en qué orden salieron no aporta nada.

El criterio de factorización convierte la comprobación en un ejercicio algebraico: basta escribir la verosimilitud y ver si el parámetro aparece solo a través de una función de los datos.

2.5 Resumen de estimadores habituales

Parámetro	Estimador	Insesgado	Consistente
μ	$\bar{X}$	sí	sí
σ^2	$\hat{S}^2$	sí	sí
σ^2	S^2	no	sí
p	$\hat{p} = X/n$	sí	sí
$\mu_1 - \mu_2$	$\bar{X}_1 - \bar{X}_2$	sí	sí

2.6 Ejercicios

Ejercicio 2.6.1. Obtener el estimador de máxima verosimilitud del parámetro λ de una distribución de Poisson.

Solución 2.6.1.

$$L(\lambda) = \prod \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}, \qquad \ell(\lambda) = -n\lambda + \Big(\sum x_i\Big)\ln\lambda - \sum \ln(x_i!)$$

Derivando, $-n + \sum x_i/\lambda = 0$, de donde $\hat{\lambda} = \bar{X}$. Coincide con el estimador por el método de los momentos, porque en la Poisson $E[X] = \lambda$.

Ejercicio 2.6.2. Dados dos estimadores insesgados de μ con $\mathrm{Var}(T_1) = 4$ y $\mathrm{Var}(T_2) = 9$, ¿cuál se prefiere? ¿Y si T_2 tuviera sesgo 1 y varianza 2?

Solución 2.6.2. En el primer caso, T_1: entre insesgados manda la varianza.

En el segundo, $\text{ECM}(T_1) = 4$ y $\text{ECM}(T_2) = 2 + 1^2 = 3$, así que se prefiere T_2 pese a ser sesgado. Es el compromiso entre sesgo y varianza: lo que se minimiza es el error total, no el sesgo.

Ejercicio 2.6.3. Demostrar que $\bar{X}$ es consistente para μ.

Solución 2.6.3. $E[\bar{X}] = \mu$ para todo n, así que la primera condición se cumple sin necesidad de límite. Y $\text{Var}(\bar{X}) = \sigma^2/n \to 0$. Por la condición suficiente, $\bar{X}$ es consistente. Alternativamente, es consecuencia directa de la ley débil de los grandes números.

Los métodos de estimación y sus propiedades están desarrollados en Gómez Villegas, 2007 y Canavos, 1987, con problemas resueltos en Casas et al., 2006 y Herrerías, Palacios, Pérez et al., 2012.

Capítulo 3

Distribuciones de los estadísticos muestrales

Temas 3 y 4 del programa. Las distribuciones en el muestreo de la media, la varianza y la proporción en poblaciones normales, y las de la diferencia de medias, el cociente de varianzas y la diferencia de proporciones para dos poblaciones.

Todo lo que sigue es la maquinaria que los temas 5 y 6 usan sin volver a deducirla: **cada intervalo de confianza y cada contraste sale de una de estas distribuciones**.

3.1 Una población normal

3.1.1 Media con varianza conocida

$$\bar{X} \sim N\left(\mu,\ \frac{\sigma}{\sqrt{n}}\right) \qquad \Longrightarrow \qquad Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

La cantidad $\sigma/\sqrt{n}$ se llama **error típico de la media**, y no debe confundirse con σ: la primera mide cuánto varía la media muestral y la segunda cuánto varían los datos.

Por el teorema central del límite, el resultado vale aproximadamente aunque la población no sea normal, siempre que n sea suficientemente grande.

3.1.2 Varianza y cuasivarianza

Teorema 3.1 . Si la población es $N(\mu,\sigma)$,

$$\frac{(n-1)\hat{S}^2}{\sigma^2} = \frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$$

y además $\bar{X}$ y $\hat{S}^2$ son independientes.

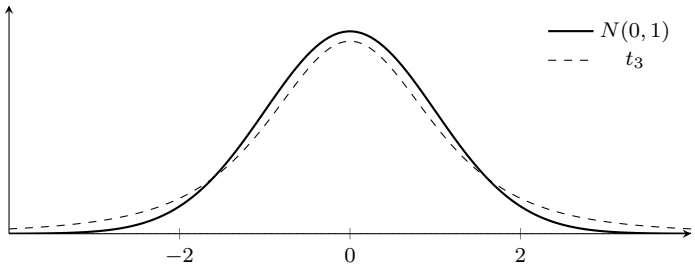
La independencia entre media y cuasivarianza es exclusiva de la normal, y es lo que permite construir la t de Student del apartado siguiente. En cualquier otra distribución no se cumple, y toda la inferencia clásica se apoya en ella.

Los grados de libertad son $n-1$ y no n porque una desviación queda determinada por las otras: $\sum(X_i - \bar{X}) = 0$.

3.1.3 Media con varianza desconocida

Sustituyendo σ por $\hat{S}$, el cociente deja de ser normal:

$$T = \frac{\bar{X} - \mu}{\hat{S}/\sqrt{n}} \sim t_{n-1}$$



La t es más achatada y con colas más pesadas: **estimar σ añade incertidumbre**, y eso se traduce en valores críticos mayores y por tanto en intervalos más anchos. Al crecer n converge a la normal, y a partir de 30 grados de libertad la diferencia deja de importar en la práctica.

3.1.4 Proporción

Para n grande, con $\hat{p} = X/n$ y $X \sim \mathrm{B}(n, p)$:

$$\hat{p} \approx N\left(p,\ \sqrt{\frac{p(1-p)}{n}}\right)$$

La aproximación se admite cuando $np \geq 5$ y $n(1-p) \geq 5$. Con proporciones muy pequeñas o muy grandes no vale, y hay que trabajar con la binomial exacta.

3.2 Resumen para una población

Se estudia	Estadístico	Distribución	Condición
μ, σ conocida	$\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	$N(0,1)$	normal, o n grande
μ, σ desconocida	$\dfrac{\bar{X} - \mu}{\hat{S}/\sqrt{n}}$	t_{n-1}	población normal
σ^2	$\dfrac{(n-1)\hat{S}^2}{\sigma^2}$	χ^2_{n-1}	población normal
p	$\dfrac{\hat{p} - p}{\sqrt{p(1-p)/n}}$	$N(0,1)$ aprox.	$np \geq 5$, $n(1-p) \geq 5$

3.3 Dos poblaciones normales independientes

3.3.1 Diferencia de medias

Caso	Estadístico	Distribución
Varianzas **conocidas**	$\dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$	$N(0,1)$
Desconocidas e **iguales**	el mismo con S_p	$t_{n_1+n_2-2}$
Desconocidas y **distintas**	aproximación de Welch	t_ν con ν aproximado
Desconocidas, n grandes	con $\hat{S}_1^2$ y $\hat{S}_2^2$	$N(0,1)$ aprox.

Con varianzas iguales se usa la **cuasivarianza combinada**, que promedia las dos ponderando por grados de libertad:

$$S_p^2 = \frac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2}$$

Y con varianzas distintas, la **aproximación de Welch** ajusta los grados de libertad:

$$\nu \approx \frac{\left(\dfrac{\hat{S}_1^2}{n_1} + \dfrac{\hat{S}_2^2}{n_2}\right)^2}{\dfrac{(\hat{S}_1^2/n_1)^2}{n_1 - 1} + \dfrac{(\hat{S}_2^2/n_2)^2}{n_2 - 1}}$$

que en general no es entero y se redondea hacia abajo.

Nota 3.1 . Decidir entre los casos segundo y tercero exige contrastar antes la igualdad de varianzas con la F, y ese contraste previo es sensible a la falta de normalidad. Por eso la práctica moderna recomienda usar **Welch siempre**: cuando las varianzas son iguales pierde muy poca potencia, y cuando no lo son evita un error grave.

3.3.2 Cociente de varianzas

$$F = \frac{\hat{S}_1^2/\sigma_1^2}{\hat{S}_2^2/\sigma_2^2} \sim F_{n_1-1,\ n_2-1}$$

Bajo la hipótesis de igualdad de varianzas, el cociente se reduce a $\hat{S}_1^2/\hat{S}_2^2$, que es el estadístico del contraste.

La F tiene una propiedad que ahorra tablas:

$$F_{n,m;\,\alpha} = \frac{1}{F_{m,n;\,1-\alpha}}$$

y por eso las tablas solo recogen la cola derecha.

3.3.3 Diferencia de proporciones

$$\hat{p}_1 - \hat{p}_2 \approx N\left(p_1 - p_2,\ \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

Con las mismas condiciones de aproximación en cada muestra.

3.4 Resumen para dos poblaciones

Se estudia	Estadístico	Distribución
$\mu_1 - \mu_2$, varianzas conocidas	tipificado	$N(0,1)$
$\mu_1 - \mu_2$, desconocidas iguales	con S_p	$t_{n_1+n_2-2}$
$\mu_1 - \mu_2$, desconocidas distintas	Welch	t_ν
σ_1^2/σ_2^2	$\hat{S}_1^2/\hat{S}_2^2$	F_{n_1-1,n_2-1}
$p_1 - p_2$	tipificado	$N(0,1)$ aprox.

Nota 3.2 . Todo lo anterior supone **muestras independientes**. Si los datos están emparejados —las mismas personas antes y después, o mediciones de dos aparatos sobre las mismas piezas— hay que trabajar con las diferencias individuales y aplicar el caso de una sola población. Tratar datos emparejados como independientes desperdicia la información del emparejamiento y suele impedir detectar diferencias que existen.

3.5 Ejercicios

Ejercicio 3.5.1. De una población normal con $\sigma = 12$ se toma una muestra de 16 elementos. ¿Cuál es la probabilidad de que la media muestral se aleje de μ más de 5 unidades?

Solución 3.5.1. El error típico es $12/\sqrt{16} = 3$, así que $\bar{X} \sim N(\mu, 3)$. Tipificando, $z = 5/3 = 1{,}667$, y

$$P(|\bar{X} - \mu| > 5) = 2\,P(Z > 1{,}667) = 2 \cdot 0{,}0478 = 0{,}0956$$

Un 9,6 %. Con $n = 64$ el error típico bajaría a 1,5 y la probabilidad a 0,0009.

Ejercicio 3.5.2. En una muestra de 10 elementos de una población normal se obtiene $\hat{S}^2 = 25$. ¿Qué distribución sigue $9\hat{S}^2/\sigma^2$?

Solución 3.5.2. Una χ^2 con $n - 1 = 9$ grados de libertad. Su valor observado es $225/\sigma^2$, y esa relación es la que permite construir el intervalo de confianza para σ^2 del tema siguiente sin más que buscar dos cuantiles de la tabla.

Ejercicio 3.5.3. Dos muestras de tamaños 10 y 15 dan cuasivarianzas 40 y 25. ¿Qué grados de libertad tiene el estadístico F para contrastar la igualdad de varianzas, y cuánto vale?

Solución 3.5.3. $F = 40/25 = 1{,}6$ con 9 y 14 grados de libertad. Conviene poner en el numerador la cuasivarianza mayor, porque así el valor supera 1 y basta consultar la cola derecha de la tabla, que es la única tabulada.

Las distribuciones en el muestreo están desarrolladas en Canavos, 1987 y Herrerías, Palacios y Callejón, 2012, con problemas resueltos en Herrerías, Palacios, Pérez et al., 2012 y Espejo Miranda, 2016.

Capítulo 4

Estimación por intervalos de confianza

Temas 5 y 6 del programa. El concepto de intervalo de confianza, y los intervalos para la media, la varianza y la proporción de una población, y para la diferencia de medias, la razón de varianzas y la diferencia de proporciones de dos.

4.1 Concepto

Definición 4.1 (Intervalo de confianza). Un intervalo aleatorio (L_1, L_2), construido a partir de la muestra, tal que

$$P\big(L_1 < \theta < L_2\big) = 1 - \alpha$$

El valor $1 - \alpha$ es el nivel de confianza.

Nota 4.1 . **Lo aleatorio es el intervalo, no el parámetro.** Una vez calculado con datos concretos, θ está dentro o no lo está, y no hay ninguna probabilidad de por medio. La lectura correcta es frecuentista: si se repitiera el muestreo muchas veces, el 95 % de los intervalos construidos así contendrían el verdadero valor. Decir «hay un 95 % de probabilidad de que μ esté entre 48 y 52» es incorrecto.

Todos los intervalos del tema tienen la misma forma:

$$\text{estimador} \ \pm \ (\text{valor crítico}) \times (\text{error típico})$$

y el semiancho se llama **error máximo admisible**.

Factor	Efecto sobre la amplitud
Mayor nivel de confianza	más ancho
Mayor tamaño de muestra	más estrecho, como $1/\sqrt{n}$
Mayor variabilidad de la población	más ancho

Confianza y precisión compiten. Un intervalo del 99 % es más ancho que uno del 95 % con los mismos datos, y la única forma de ganar en las dos cosas a la vez es aumentar la muestra.

4.2 Una población

4.2.1 Media, con σ conocida

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$1-\alpha$	$z_{\alpha/2}$
90 %	1,645
95 %	1,960
99 %	2,576

4.2.2 Media, con σ desconocida

$$\bar{x} \pm t_{n-1;\,\alpha/2} \frac{\hat{s}}{\sqrt{n}}$$

Con n grande la t se aproxima por la normal, pero conviene usar la t siempre que σ sea desconocida: no cuesta más y es exacta con poblaciones normales.

4.2.3 Tamaño de muestra necesario

Imponiendo que el error máximo no supere E:

$$n \geq \left(\frac{z_{\alpha/2}\,\sigma}{E}\right)^2$$

Ejemplo 4.1 . Para estimar una media con error máximo de 2 unidades al 95 % de confianza, sabiendo que $\sigma \approx 10$:

$$n \geq \left(\frac{1{,}96 \cdot 10}{2}\right)^2 = 96{,}04 \implies n = 97$$

Reducir el error a 1 exigiría $n = 385$: **dividir el error por dos cuadruplica la muestra**, que es la ley de rendimientos decrecientes vista otra vez.

4.2.4 Varianza

$$\left(\frac{(n-1)\hat{s}^2}{\chi^2_{n-1;\,\alpha/2}},\ \frac{(n-1)\hat{s}^2}{\chi^2_{n-1;\,1-\alpha/2}}\right)$$

No es simétrico respecto de $\hat{s}^2$, porque la χ^2 no lo es. Escribirlo como «estimación más menos algo» es un error.

4.2.5 Proporción

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Y el tamaño de muestra, tomando el caso más desfavorable $\hat{p} = 0{,}5$:

$$n \geq \frac{z^2_{\alpha/2}}{4E^2}$$

Ejemplo 4.2 . La fórmula anterior explica un número que aparece en todas las encuestas: con $E = 0{,}03$ y confianza del 95 %,

$$n \geq \frac{1{,}96^2}{4 \cdot 0{,}0009} = 1067$$

De ahí que las encuestas electorales usen sistemáticamente muestras en torno a las 1000-1100 personas, con un margen de error de unos tres puntos.

Nota 4.2 . El tamaño necesario **no depende del tamaño de la población**, salvo que la muestra sea una fracción apreciable de ella. Para estimar la opinión de un país de cuarenta millones y la de uno de cuatro hacen falta las mismas mil personas, y eso es lo que más sorprende de este resultado.

4.3 Dos poblaciones

4.3.1 Diferencia de medias

Caso	Intervalo
Varianzas conocidas	$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
Desconocidas iguales	$(\bar{x}_1 - \bar{x}_2) \pm t_{n_1+n_2-2;\,\alpha/2}\, s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
Desconocidas distintas	ídem con las cuasivarianzas por separado y t_ν de Welch

La lectura clave: si el intervalo contiene el cero, los datos son compatibles con que las dos medias sean iguales. Si no lo contiene, hay evidencia de diferencia, y el signo del intervalo dice en qué sentido.

4.3.2 Razón de varianzas

$$\left(\frac{\hat{s}_1^2}{\hat{s}_2^2} \cdot \frac{1}{F_{n_1-1,n_2-1;\,\alpha/2}},\quad \frac{\hat{s}_1^2}{\hat{s}_2^2} \cdot F_{n_2-1,n_1-1;\,\alpha/2}\right)$$

Aquí lo que se mira es si el intervalo contiene el **uno**, porque el parámetro es un cociente y no una diferencia.

4.3.3 Diferencia de proporciones

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

4.4 Tabla de decisión

Parámetro	Condiciones	Distribución
μ	σ conocida	$N(0,1)$
μ	σ desconocida, población normal	t_{n-1}
σ^2	población normal	χ^2_{n-1}
p	n grande	$N(0,1)$
$\mu_1-\mu_2$	varianzas conocidas	$N(0,1)$
$\mu_1-\mu_2$	desconocidas iguales	$t_{n_1+n_2-2}$
$\mu_1-\mu_2$	desconocidas distintas	t_ν
σ_1^2/σ_2^2	poblaciones normales	F
p_1-p_2	muestras grandes	$N(0,1)$

4.5 Ejercicios

Ejercicio 4.5.1. Una muestra de 25 piezas da una longitud media de 10,2 cm con cuasidesviación típica 0,5 cm. Construir un intervalo al 95 % para la longitud media, suponiendo normalidad.

Solución 4.5.1. σ es desconocida, así que se usa la t con 24 grados de libertad: $t_{24;\,0{,}025} = 2{,}064$.

$$10{,}2 \pm 2{,}064 \cdot \frac{0{,}5}{5} = 10{,}2 \pm 0{,}2064 = (9{,}994,\ 10{,}406)$$

Con la normal saldría $\pm 0{,}196$: algo más estrecho, y sería incorrecto porque σ se ha estimado.

Ejercicio 4.5.2. En una encuesta a 400 personas, 240 se declaran a favor. Intervalo al 95 % para la proporción poblacional.

Solución 4.5.2. $\hat{p} = 0{,}6$ y el error típico es $\sqrt{0{,}6 \cdot 0{,}4/400} = 0{,}0245$.

$$0{,}6 \pm 1{,}96 \cdot 0{,}0245 = 0{,}6 \pm 0{,}048 = (0{,}552,\ 0{,}648)$$

Se cumplen las condiciones: $np = 240$ y $n(1-p) = 160$, los dos muy por encima de 5.

Ejercicio 4.5.3. Dos métodos de producción dan medias 52 y 48 con cuasidesviaciones típicas 6 y 5, sobre muestras de 20 y 25 piezas. Suponiendo varianzas iguales, ¿hay diferencia entre los dos al 95 %?

Solución 4.5.3.

$$s_p^2 = \frac{19 \cdot 36 + 24 \cdot 25}{43} = \frac{684 + 600}{43} = 29{,}86, \qquad s_p = 5{,}46$$

$$4 \pm 2{,}017 \cdot 5{,}46\sqrt{\tfrac{1}{20} + \tfrac{1}{25}} = 4 \pm 2{,}017 \cdot 5{,}46 \cdot 0{,}300 = 4 \pm 3{,}30$$

El intervalo es $(0{,}70,\ 7{,}30)$, que **no contiene el cero**: hay evidencia de que el primer método da una media mayor, con una diferencia de entre 0,7 y 7,3 unidades.

Los intervalos de confianza están desarrollados en Herrerías, Palacios y Callejón, 2012 y Lind et al., 2012, con problemas resueltos en Herrerías, Palacios, Pérez et al., 2012, Casas et al., 2006 y Fernández-Sánchez y Hernández-Bastida, 2023.

Capítulo 5

Contraste de hipótesis sobre parámetros

Tema 7 del programa. Introducción al contraste, y los contrastes para una muestra, para dos y para más de dos.

5.1 Planteamiento

Definición 5.1 . Un contraste enfrenta dos hipótesis complementarias sobre un parámetro:

- H_0, la *hipótesis nula*: la que se somete a prueba y se mantiene salvo que los datos la contradigan;
- H_1, la *alternativa*: lo que se sostiene si H_0 se rechaza.

Tipo	H_0	H_1	Región crítica
Bilateral	$\theta = \theta_0$	$\theta \neq \theta_0$	las dos colas
Unilateral derecho	$\theta \leq \theta_0$	$\theta > \theta_0$	cola derecha
Unilateral izquierdo	$\theta \geq \theta_0$	$\theta < \theta_0$	cola izquierda

La igualdad va siempre en H_0, porque es la hipótesis bajo la cual se calcula la distribución del estadístico. Y la alternativa se fija **antes de mirar los datos**: elegirla después según lo que se ha observado invalida el contraste.

5.2 Los dos errores

	H_0 cierta	H_0 falsa
Se rechaza H_0	error de tipo I (α)	decisión correcta ($1-\beta$)
No se rechaza	decisión correcta	error de tipo II (β)

Concepto	Definición
Nivel de significación α	probabilidad de rechazar H_0 siendo cierta
Potencia $1-\beta$	probabilidad de rechazar H_0 siendo falsa

Nota 5.1 . Los dos errores **no se pueden reducir a la vez** con la misma muestra: bajar α sube β. La única forma de mejorar los dos es aumentar n. Por eso se fija α de antemano —normalmente 0,05 o 0,01— según cuál de los dos errores sea más grave en el contexto.

La asimetría del planteamiento es deliberada: el contraste **protege** H_0. Por eso la hipótesis nula suele ser «no hay efecto», «el proceso funciona» o «los dos grupos son iguales», y se necesita evidencia fuerte para abandonarla.

Nota 5.2 . **No rechazar no es aceptar.** Que los datos sean compatibles con H_0 puede deberse a que sea cierta o a que la muestra sea demasiado pequeña para detectar la diferencia. La conclusión correcta es «no hay evidencia suficiente para rechazar», y nunca «queda demostrado que H_0 es cierta».

5.3 Procedimiento

1. Formular H_0 y H_1.
2. Fijar α.
3. Elegir el estadístico de contraste y su distribución bajo H_0.
4. Determinar la región crítica, o calcular el p-valor.
5. Calcular el estadístico con la muestra.
6. Decidir y **redactar la conclusión en términos del problema**.

5.3.1 El p-valor

Definición 5.2 . Probabilidad de obtener un resultado tan extremo como el observado, o más, siendo H_0 cierta.

La regla de decisión: se rechaza H_0 si $p < \alpha$.

Nota 5.3 . El p-valor **no es la probabilidad de que H_0 sea cierta**, ni la probabilidad de haberse equivocado. Es una probabilidad calculada *suponiendo* H_0 cierta. Y un p pequeño indica que los datos son raros bajo H_0, no que el efecto sea grande: con muestras enormes, diferencias sin ninguna importancia práctica salen significativas. Significación estadística y relevancia no son lo mismo.

5.4 Contrastes para una muestra

Parámetro	Condiciones	Estadístico	Distribución
μ	σ conocida	$\dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$N(0,1)$
μ	σ desconocida	$\dfrac{\bar{x}-\mu_0}{\hat{s}/\sqrt{n}}$	t_{n-1}
σ^2	población normal	$\dfrac{(n-1)\hat{s}^2}{\sigma_0^2}$	χ^2_{n-1}
p	n grande	$\dfrac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$	$N(0,1)$

En el contraste de proporciones, el error típico se calcula con p_0 y no con $\hat{p}$, **porque bajo H_0 el valor del parámetro es** p_0. Es la diferencia con el intervalo de confianza, donde p_0 no existe.

Ejemplo 5.1 . Un fabricante afirma que sus piezas pesan 500 g. Una muestra de 36 da media 495 g y cuasidesviación 12 g. ¿Hay evidencia contra la afirmación al 5 %?

$H_0 : \mu = 500$ frente a $H_1 : \mu \neq 500$. El estadístico es

$$t = \frac{495-500}{12/6} = -2{,}5$$

con 35 grados de libertad. El valor crítico es $t_{35;\,0{,}025} = 2{,}030$, y $|-2{,}5| > 2{,}030$: **se rechaza**. El p-valor es 0,017.

Conclusión en términos del problema: los datos aportan evidencia de que el peso medio difiere de 500 g, y apuntan a que es menor.

5.5 Contrastes para dos muestras

Se compara	Condiciones	Distribución
μ_1 y μ_2	varianzas conocidas	$N(0,1)$
μ_1 y μ_2	desconocidas iguales	$t_{n_1+n_2-2}$
μ_1 y μ_2	desconocidas distintas	t_ν de Welch
μ_1 y μ_2	**datos emparejados**	t_{n-1} sobre las diferencias
σ_1^2 y σ_2^2	poblaciones normales	F_{n_1-1,n_2-1}
p_1 y p_2	muestras grandes	$N(0,1)$ con proporción combinada

En el contraste de proporciones, bajo $H_0 : p_1 = p_2$ se usa la **proporción combinada**:

$$\hat{p} = \frac{x_1+x_2}{n_1+n_2}, \qquad z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

Los datos emparejados merecen atención. Cuando cada observación de una muestra tiene su pareja en la otra —el mismo sujeto antes y después—, se trabaja con las diferencias individuales y el problema se reduce a una sola muestra. Tratarlos como independientes desperdicia la información del emparejamiento y suele impedir detectar diferencias reales.

5.6 Contrastes para más de dos muestras

Comparar k medias no se puede hacer con contrastes dos a dos: con $k = 5$ hay 10 comparaciones, y **la probabilidad de al menos un falso positivo se dispara**. Con $\alpha = 0{,}05$ en cada una, la global es $1 - 0{,}95^{10} = 0{,}40$.

La solución es el **análisis de la varianza**, que contrasta de una vez

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

descomponiendo la variabilidad total:

$$SC_{total} = SC_{entre} + SC_{dentro}$$

$$F = \frac{SC_{entre}/(k-1)}{SC_{dentro}/(N-k)} \sim F_{k-1,\ N-k}$$

Fuente	Suma de cuadrados	Grados de libertad
Entre grupos	$\sum n_i(\bar{x}_i - \bar{x})^2$	$k-1$
Dentro de los grupos	$\sum\sum(x_{ij} - \bar{x}_i)^2$	$N-k$
Total	$\sum\sum(x_{ij} - \bar{x})^2$	$N-1$

La idea: si las medias fuesen iguales, la variabilidad entre grupos sería del mismo orden que la de dentro, y el cociente rondaría 1. Un F grande indica que los grupos difieren más de lo que la variabilidad interna explica.

Hipótesis del ANOVA	Qué exige
Normalidad	dentro de cada grupo
Homocedasticidad	misma varianza en todos
Independencia	entre y dentro de los grupos

Rechazar H_0 solo dice que no todas las medias son iguales, no cuáles difieren. Para eso hacen falta comparaciones múltiples posteriores, que ajustan el nivel para controlar el error global.

5.7 Ejercicios

Ejercicio 5.7.1. Un proceso produce históricamente un 4 % de defectuosas. En una muestra de 500 piezas aparecen 28. ¿Ha empeorado el proceso, al 5 %?

Solución 5.7.1. $H_0 : p \leq 0{,}04$ frente a $H_1 : p > 0{,}04$, unilateral derecho. $\hat{p} = 28/500 = 0{,}056$ y el error típico bajo H_0 es $\sqrt{0{,}04 \cdot 0{,}96/500} = 0{,}00876$.

$$z = \frac{0{,}056 - 0{,}04}{0{,}00876} = 1{,}83$$

El valor crítico es 1,645, y $1{,}83 > 1{,}645$: se rechaza. El p-valor es 0,034. Hay evidencia de empeoramiento al 5 %, aunque no al 1 %.

Ejercicio 5.7.2. Un contraste da un p-valor de 0,08. ¿Qué se concluye a los niveles del 5 % y del 10 %?

Solución 5.7.2. Al 5 % no se rechaza H_0, porque $0{,}08 > 0{,}05$; al 10 % sí, porque $0{,}08 < 0{,}10$. Eso muestra por qué α debe fijarse **antes** de ver los datos: elegirlo después según convenga al resultado convierte el contraste en un ejercicio de autoengaño. Lo correcto es informar del p-valor y dejar que el lector juzgue.

Ejercicio 5.7.3. Tres máquinas producen piezas con medias muestrales 20, 22 y 21, sobre 10 piezas cada una. El ANOVA da $F = 4{,}2$ con 2 y 27 grados de libertad, y el valor crítico al 5 % es 3,35. ¿Qué se concluye?

Solución 5.7.3. $4{,}2 > 3{,}35$, así que se rechaza H_0: hay evidencia de que **no todas** las medias son iguales. El contraste no dice cuáles difieren; para eso hacen falta comparaciones múltiples, que ajustan el nivel para que la probabilidad global de falso positivo siga siendo el 5 %.

Los contrastes de hipótesis están desarrollados en Herrerías, Palacios y Callejón, 2012 y Lind et al., 2012, con problemas resueltos en Herrerías, Palacios, Pérez et al., 2012, Casas et al., 2006 y Espejo Miranda, 2016.

Capítulo 6

Tests no paramétricos

Tema 8 del programa. Los contrastes que no suponen una forma concreta para la distribución de la población.

6.1 Por qué hacen falta

Todos los contrastes del tema anterior suponen normalidad, o un tamaño de muestra suficiente para invocar el teorema central del límite. Cuando ninguna de las dos cosas se cumple, sus conclusiones no son fiables.

Situación	Problema
Muestra pequeña y población no normal	el teorema central del límite no se aplica
Datos ordinales	la media no tiene sentido
Valores atípicos extremos	la media y la varianza se distorsionan
Distribución muy asimétrica	los intervalos simétricos son inadecuados

	Paramétricos	No paramétricos
Suponen	forma de la distribución	poco o nada
Usan	los valores	rangos, signos o frecuencias
Potencia	mayor, si los supuestos se cumplen	menor, pero robusta
Si los supuestos fallan	conclusiones no válidas	siguen valiendo

El precio de la generalidad es la potencia: con datos realmente normales, un test no paramétrico necesita más muestra para detectar el mismo efecto. La regla práctica es usar el paramétrico cuando sus supuestos se sostienen y el no paramétrico cuando no.

6.2 Bondad de ajuste

Contrastan si una muestra procede de una distribución dada.

6.2.1 Chi-cuadrado de Pearson

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{k-1-r}$$

con O_i las frecuencias observadas, E_i las esperadas bajo H_0 y r el número de parámetros estimados a partir de los datos.

Requisito	Detalle
Frecuencias esperadas	$E_i \geq 5$ en todas las clases
Si alguna es menor	se agrupan clases contiguas
Grados de libertad	se pierde uno por cada parámetro estimado
Cola	siempre la derecha

Ejemplo 6.1 . Se lanza un dado 120 veces y salen las frecuencias 15, 25, 18, 22, 20, 20. ¿Es equilibrado?

Bajo H_0, $E_i = 20$ para las seis caras.

$$\chi^2 = \frac{25+25+4+4+0+0}{20} = \frac{58}{20} = 2{,}9$$

con 5 grados de libertad. El valor crítico al 5 % es 11,07, muy por encima: no se rechaza. Los datos son compatibles con un dado equilibrado.

6.2.2 Kolmogorov-Smirnov

Compara la función de distribución empírica con la teórica:

$$D_n = \sup_x |F_n(x) - F_0(x)|$$

Ventaja frente a la χ^2	Limitación
No exige agrupar en clases	solo para distribuciones continuas
Más potente con muestras pequeñas	los parámetros deben conocerse, no estimarse
Usa la información de todos los datos	menos flexible

Cuando los parámetros se estiman de la muestra, los valores críticos cambian, y para el caso normal se usa la corrección de Lilliefors.

6.3 Contrastes de posición

6.3.1 Test de los signos

El más elemental. Para contrastar que la mediana vale m_0, se cuentan cuántas observaciones la superan y cuántas no, y ese recuento es binomial de parámetro $p = 0{,}5$ bajo H_0.

Usa solo el signo de las diferencias, así que **descarta la magnitud** y es poco potente. A cambio, no supone absolutamente nada sobre la distribución.

6.3.2 Test de Wilcoxon de los rangos con signo

Para una muestra o para datos emparejados. Ordena las diferencias por valor absoluto, les asigna rangos y suma los rangos de las positivas.

Aprovecha la magnitud además del signo, así que **es más potente que el de los signos**. Su único supuesto es que la distribución de las diferencias sea simétrica.

Test	Alternativa paramétrica
Signos	t para una muestra
Wilcoxon de rangos con signo	t para una muestra o emparejada
Mann-Whitney	t para dos muestras independientes
Kruskal-Wallis	ANOVA de un factor

6.3.3 Mann-Whitney

Para dos muestras independientes. Se juntan las observaciones, se ordenan, se asignan rangos y se suman los de cada grupo. Si las dos poblaciones fuesen iguales, las sumas serían proporcionales a los tamaños.

Contrasta si una distribución está desplazada respecto de la otra, y con muestras grandes su estadístico se aproxima por la normal.

6.3.4 Kruskal-Wallis

La extensión a k muestras, y el equivalente no paramétrico del ANOVA:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1) \ \sim \ \chi^2_{k-1}$$

con R_i la suma de rangos del grupo i.

Nota 6.1 . Al asignar rangos, los **empates** reciben el rango medio de las posiciones que ocupan, y el estadístico lleva un factor de corrección. Con muchos empates —típico de datos ordinales con pocas categorías— la corrección deja de ser menor y hay que aplicarla.

6.4 Independencia y homogeneidad

6.4.1 Chi-cuadrado de independencia

Sobre una tabla de contingencia $r \times c$:

$$E_{ij} = \frac{n_{i\cdot}\, n_{\cdot j}}{N}, \qquad \chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)(c-1)}$$

Es el contraste que formaliza la independencia estadística del curso anterior: las frecuencias esperadas son justamente el producto de las marginales dividido por el total.

El mismo estadístico sirve para dos preguntas distintas:

Contraste	H_0	Diseño
Independencia	los dos caracteres son independientes	una muestra, dos variables
Homogeneidad	las poblaciones tienen la misma distribución	varias muestras, una variable

La aritmética es idéntica y la interpretación no, así que conviene decir cuál de los dos se está haciendo.

Ejemplo 6.2 . Una tabla 2×2 con observados 30, 20, 20 y 30, y $N = 100$. Las marginales son 50 en las cuatro, así que todas las esperadas valen $50 \cdot 50/100 = 25$.

$$\chi^2 = 4 \cdot \frac{25}{25} = 4$$

con $(2-1)(2-1) = 1$ grado de libertad. El crítico al 5 % es 3,84, y $4 > 3{,}84$: se rechaza la independencia, por muy poco. El p-valor es 0,046.

Nota 6.2 . Rechazar la independencia **no mide la intensidad** de la asociación. Con muestras grandes, asociaciones ínfimas salen significativas. Para cuantificar hace falta una medida de asociación como el coeficiente de contingencia o la V de Cramér, y conviene informar de las dos cosas.

6.5 Aleatoriedad

El **test de rachas** contrasta si una secuencia de observaciones es aleatoria. Una racha es un bloque máximo de valores consecutivos del mismo tipo.

Número de rachas	Indica
Demasiado pocas	agrupamiento, tendencia o persistencia
Demasiadas	alternancia sistemática
Intermedio	compatible con la aleatoriedad

Es un contraste bilateral: **los dos extremos delatan falta de aleatoriedad**. Se usa sobre los residuos de una regresión para comprobar que no queda estructura sin modelar.

6.6 Cómo se elige

Objetivo	Paramétrico	No paramétrico
Una media o mediana	t	signos, Wilcoxon
Dos muestras independientes	t	Mann-Whitney
Dos muestras emparejadas	t emparejada	Wilcoxon
Más de dos muestras	ANOVA	Kruskal-Wallis
Ajuste a una distribución	—	χ^2, Kolmogorov-Smirnov
Asociación entre dos caracteres	correlación de Pearson	χ^2, correlación de Spearman
Aleatoriedad	—	rachas

6.7 Ejercicios

Ejercicio 6.7.1. Se observan 40 llegadas por hora durante 5 horas: 8, 12, 6, 9 y 5. Contrastar al 5 % que las llegadas se reparten uniformemente.

Solución 6.7.1. Bajo H_0, $E_i = 40/5 = 8$.

$$\chi^2 = \frac{0 + 16 + 4 + 1 + 9}{8} = \frac{30}{8} = 3{,}75$$

con 4 grados de libertad. El crítico al 5 % es 9,49, y $3{,}75 < 9{,}49$: no se rechaza. Se cumple además el requisito de que todas las esperadas superen 5.

Ejercicio 6.7.2. ¿Por qué no se usa siempre un test no paramétrico, si exige menos supuestos?

Solución 6.7.2. Porque pierde potencia. Al trabajar con rangos en vez de valores descarta información, y necesita más muestra para detectar el mismo efecto. Con datos realmente normales, la eficiencia de Mann-Whitney frente a la t es de aproximadamente el 95 %, así que la pérdida es modesta; pero con otras alternativas es mayor. La regla es usar el paramétrico cuando sus supuestos se sostienen.

Ejercicio 6.7.3. En un contraste χ^2 de bondad de ajuste a una Poisson se estima λ con la propia muestra, y quedan 6 clases. ¿Cuántos grados de libertad tiene el estadístico?

Solución 6.7.3. $k - 1 - r = 6 - 1 - 1 = 4$. Se pierde uno por la restricción de que las frecuencias sumen n y otro por haber estimado λ. Olvidar el segundo descuento da un valor crítico mayor del que corresponde y hace el contraste demasiado conservador.

Los contrastes no paramétricos están desarrollados en Canavos, 1987 y Lind et al., 2012, con problemas resueltos en Herrerías, Palacios, Pérez et al., 2012 y Espejo Miranda, 2016.

Capítulo 7

Temario práctico

La guía indica que el temario práctico acompaña al teórico: se resuelven ejercicios a mano y con hoja de cálculo o software estadístico, aplicando la inferencia a datos reales.

7.1 Herramientas

Herramienta	Para qué
Hoja de cálculo	funciones de distribución, contrastes básicos, gráficos
R o Python	todo lo anterior, más los no paramétricos y el ANOVA
Tablas estadísticas	comprobar que el resultado del programa es razonable

Las tablas no sobran aunque haya programa. Buscar el valor crítico a mano una vez por cada tipo de contraste fija qué distribución se está usando y cuántos grados de libertad tiene, que es donde se cometen los errores.

Cálculo	Hoja de cálculo	R
Cuantil normal	`INV.NORM.ESTAND(p)`	`qnorm(p)`
Cuantil t	`INV.T.2C(alfa; gl)`	`qt(p, gl)`
Cuantil χ^2	`INV.CHICUAD(p; gl)`	`qchisq(p, gl)`
Cuantil F	`INV.F(p; gl1; gl2)`	`qf(p, gl1, gl2)`
Contraste t	`PRUEBA.T(...)`	`t.test(x, y)`

7.2 Práctica 1. Distribuciones continuas

Actividad	Qué se obtiene
Representar densidades	forma de cada modelo al variar sus parámetros
Calcular probabilidades	áreas bajo la densidad
Tipificar	pasar de $N(\mu, \sigma)$ a $N(0, 1)$ y al revés
Comparar t con normal	el efecto de los grados de libertad
Simular el teorema central del límite	medias de muestras de una población no normal

El experimento del teorema central del límite es el que más convence: se genera una población claramente no normal —exponencial, o uniforme discreta—, se extraen miles de muestras de tamaño n, se representa el histograma de las medias y se repite con $n = 2, 5, 10, 30$. La campana aparece sola, y se ve **a partir de qué** n, que es la pregunta que las tablas no responden.

7.3 Práctica 2. Estimación puntual

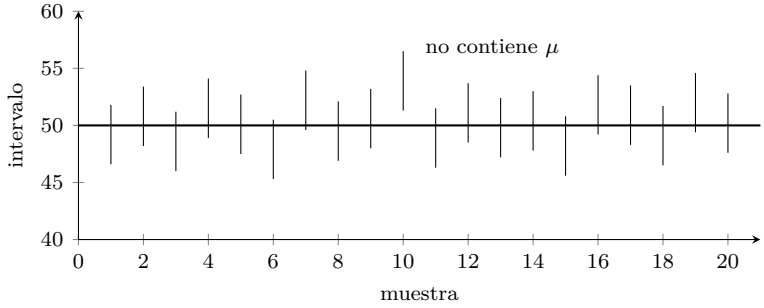
Actividad	Qué se comprueba
Simular muchas muestras y calcular $\bar{X}$	su media coincide con μ: insesgadez
Ídem con S^2 y $\hat{S}^2$	la primera subestima, la segunda no
Ver cómo cambia $\mathrm{Var}(\bar{X})$ con n	la ley $1/\sqrt{n}$, medida
Comparar dos estimadores por su ECM	el compromiso entre sesgo y varianza

La comprobación del sesgo de S^2 es especialmente instructiva: con $n = 5$ y muchas repeticiones, la media de las varianzas muestrales se queda en torno al 80 % de σ^2, exactamente el factor $(n-1)/n$ que predice la teoría.

7.4 Práctica 3. Intervalos de confianza

Actividad	Qué se obtiene
Construir intervalos para μ, σ^2 y p	con datos propios
Variar el nivel de confianza	ver cómo cambia la amplitud
Variar n	ídem
Calcular el tamaño de muestra necesario	para un error dado

El experimento que aclara la interpretación: simular 100 muestras de una población conocida, construir el intervalo del 95 % con cada una y contar cuántos contienen la verdadera μ. Salen unos 95, y los cinco que fallan se ven en el gráfico. Eso es lo que significa «95 % de confianza», y verlo evita la lectura incorrecta que el tema 4 advierte.



7.5 Práctica 4. Contrastes de hipótesis

Actividad	Qué se obtiene
Contrastes para una y dos muestras	con datos reales
Comparar la decisión por región crítica y por p-valor	coinciden siempre
Datos emparejados frente a independientes	el mismo conjunto tratado de las dos formas
Curva de potencia	$1-\beta$ frente al valor real del parámetro
ANOVA de un factor	tabla completa e interpretación

La comparación entre emparejado e independiente es el ejercicio que más enseña: con los mismos datos, el contraste emparejado detecta la diferencia y el independiente no, porque el segundo trata como ruido la variabilidad entre sujetos que el primero elimina.

Y la **curva de potencia** hace visible el error de tipo II, que en el papel es una letra griega: se representa la probabilidad de rechazar frente al valor verdadero del parámetro, y se ve que cerca de H_0 la potencia es casi nula por muy grande que sea la muestra.

7.6 Práctica 5. Tests no paramétricos

Actividad	Qué se obtiene
χ^2 de bondad de ajuste	comprobar si unos datos siguen un modelo
χ^2 de independencia	sobre una tabla de contingencia real
Mann-Whitney y Wilcoxon	comparados con la t sobre los mismos datos
Kruskal-Wallis	comparado con el ANOVA
Test de rachas	sobre residuos de una regresión

El experimento de cierre: aplicar el test paramétrico y el no paramétrico **a los mismos datos**, primero con datos normales y después contaminados con un valor atípico extremo. Con datos limpios las conclusiones coinciden; con el atípico, la t cambia de decisión y Mann-Whitney no. Eso es la robustez, medida en vez de enunciada.

7.7 Sobre la memoria

Lo que se entrega:

1. Origen de los datos y su descripción.
2. **Comprobación de los supuestos** antes de aplicar cada técnica.
3. Cálculos, con la distribución y los grados de libertad indicados.
4. Decisión, con el p-valor.
5. **Conclusión redactada en términos del problema**, no en jerga estadística.
6. Limitaciones.

El punto 2 es el que separa aplicar una fórmula de hacer inferencia: antes de una t hay que mirar si la normalidad es defendible, y antes de un ANOVA, si las varianzas son comparables.

Y el punto 5 se evalúa por sí solo. «Se rechaza H_0 con $p = 0{,}03$» no es una conclusión; «los datos aportan evidencia de que el nuevo proceso reduce el tiempo medio de montaje, con una diferencia estimada de entre 1,2 y 4,8 minutos» sí lo es.

Nota 7.1 . Y la advertencia que cierra la asignatura: **el programa siempre devuelve un número**. Aplica una t a datos que no lo permiten, calcula un χ^2 con frecuencias esperadas de 0,5 y da un p-valor con cuatro decimales en los tres casos. Comprobar los supuestos es responsabilidad de quien analiza, y ningún software lo hace por su cuenta.

El material de las prácticas sigue Fernández-Sánchez y Hernández-Bastida, 2023 y Herrerías, Palacios, Pérez et al., 2012, con el enfoque de hoja de cálculo de Black y Eldredge, 2001 y los conjuntos de datos de Lind et al., 2012.

Bibliografía

Black, K., & Eldredge, D. (2001). *Business and Economic Statistics Using Microsoft Excel*. South-Western.

Canavos, G. C. (1987). *Probabilidad y estadística: aplicaciones y métodos*. McGraw-Hill.

Casas, J. M., García, C., Rivera, L. F., & Zamora, A. I. (2006). *Ejercicios de inferencia estadística y muestreo para Economía y Administración de Empresas*. Pirámide.

Espejo Miranda, I. (2016). *Inferencia estadística: teoría y problemas*. Universidad de Cádiz.

Fernández-Sánchez, M. P., & Hernández-Bastida, A. (2023). *Cuaderno de Estadística*. Técnica Avicam.

Gómez Villegas, M. A. (2007). *Inferencia estadística*. Díaz de Santos.

Herrerías, R., Palacios, F., & Callejón, J. (2012). *Técnicas cuantitativas para la inferencia*. Delta.

Herrerías, R., Palacios, F., Pérez, E., Chica, J., Callejón, J., & Cano, R. (2012). *Ejercicios resueltos de técnicas cuantitativas para la inferencia*. Delta.

Lind, D. A., Marchal, W. G., & Wathen, S. A. (2012). *Estadística aplicada a los negocios y la economía*. McGraw-Hill Interamericana.