

ESTADÍSTICA I

Tema 5: Contraste de hipótesis

- ▶ Planteamiento del problema. Conceptos básicos: hipótesis nula y alternativa, tipos de errores, nivel de significación, región crítica o de rechazo, tamaño del test, potencia
- ▶ Contrastes paramétricos habituales en una población
- ▶ p-valor
- ▶ Contrastes para dos muestras. Distribución F de Fisher-Snedecor
- ▶ Construcción de tests: test de razón de verosimilitudes
- ▶ Contrastes uniformemente más potentes y lema de Neyman-Pearson

Planteamiento del problema. Conceptos básicos

El objetivo de la teoría de contraste de hipótesis es **elegir entre dos posibilidades mutuamente excluyentes (hipótesis nula e hipótesis alternativa)** relativas al valor de un parámetro poblacional, a partir de la información proporcionada por la muestra.

Sea X_1, \dots, X_n una muestra aleatoria de una v.a. X con función de distribución F_θ , donde $\theta \in \Theta$.

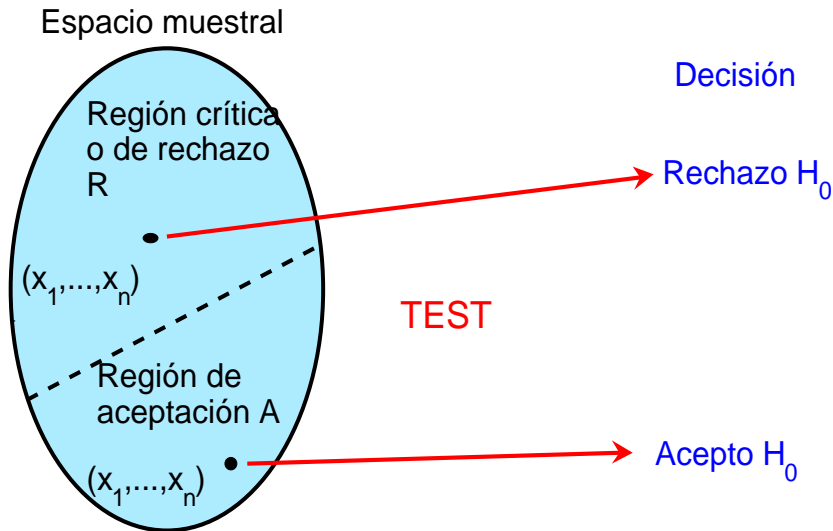
Objetivo: Dada una partición del espacio paramétrico $\Theta = \Theta_0 \cup \Theta_1$, deseamos decidir, en base a la muestra obtenida, si $\theta \in \Theta_0$ o si $\theta \in \Theta_1$. Queremos contrastar

$$H_0 : \theta \in \Theta_0 \quad (\text{hipótesis nula})$$

$$H_1 : \theta \in \Theta_1 \quad (\text{hipótesis alternativa})$$

Un test para contrastar estas dos hipótesis consiste en proporcionar una regla de decisión que, a cada posible observación de la muestra (x_1, \dots, x_n) , le asigne una decisión: aceptar o rechazar H_0 .

Los contrastes habituales (no aleatorizados) se definen mediante una **región crítica** o **región de rechazo** $R \subset \mathbb{R}^n$, de tal manera que, cuando $(x_1, \dots, x_n) \in R$, se rechaza la hipótesis nula.



Es importante destacar que la metodología de contraste de hipótesis **no “demuestra” la validez de la hipótesis** que se acepta en cada caso (en el sentido en el que se demuestra algo mediante un método deductivo, por ejemplo).

La manera correcta de interpretar los resultados es decir que “los datos disponibles proporcionan (o no proporcionan) evidencia estadística suficiente en contra de la hipótesis nula”. En todo caso, la conclusión depende de información incompleta y aleatoria, procedente de una o varias muestras, y siempre existe la posibilidad de cometer un error aceptando una hipótesis equivocada.

Los procedimientos que se utilizan habitualmente se suelen denominar “contrastes o tests de hipótesis”.

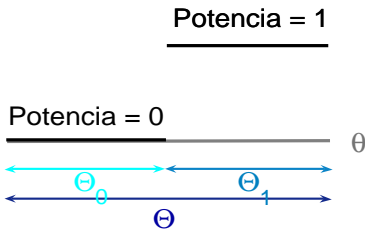
Posibles errores de un test:

- ▶ **Error de tipo I:** Rechazar H_0 cuando H_0 es cierta.
- ▶ **Error de tipo II:** Aceptar H_0 cuando H_0 es falsa.

La *función de potencia* de un test con región de rechazo R para contrastar $H_0 : \theta \in \Theta_0$ frente a $H_1 : \theta \in \Theta_1$ es la función

$$\begin{aligned} \beta_n : \Theta &\longrightarrow [0, 1] \\ \theta &\longmapsto \beta_n(\theta) = \mathbb{P}_\theta\{(X_1, \dots, X_n) \in R\} \end{aligned}$$

Lo que nos gustaría:



Lo que en realidad se suele hacer (teoría de Neyman-Pearson):

1. Acotar la máxima probabilidad de error de tipo I.

- Se fija un *nivel de significación* $\alpha \in (0, 1)$. Típicamente $\alpha = 0.05$.
- Se define el *tamaño* de un test como la máxima probabilidad de error de tipo I: $\max_{\theta \in \Theta_0} \mathbb{P}_{\theta}(R) = \max_{\theta \in \Theta_0} \beta_n(\theta)$.
- Se busca una región de rechazo R tal que $\max_{\theta \in \Theta_0} \mathbb{P}_{\theta}(R) \leq \alpha$.

2. Minimizar la probabilidad de error de tipo II. Se intenta buscar una región de rechazo R que maximice la función de potencia cuando $\theta \in \Theta_1$.

Las hipótesis H_0 y H_1 no son simétricas.

Como vemos, los test de hipótesis están diseñados para controlar la probabilidad máxima de rechazar H_0 cuando es cierta. En consecuencia, suelen ser “conservadores” con la hipótesis nula: hace falta mucha evidencia muestral para rechazar H_0 .

Observemos que es posible que, con los mismos datos, H_0 se rechace para un nivel de significación $\alpha = 0.05$ y se acepte para $\alpha = 0.01$.

En una primera aproximación, los problemas de contraste de hipótesis pueden clasificarse en **problemas de una muestra** (cuando hay una sola población de interés) y **problemas de dos muestras** (cuando se quiere comparar dos poblaciones y se dispone de una muestra de cada una de ellas). Presentaremos las ideas básicas en el caso de los problemas de una muestra pero pueden extenderse de modo análogo a los de dos muestras.

Un ejemplo ilustrativo y sus consecuencias

Ejemplo: Se analiza un envío de botellas de aceite envasado con un mecanismo del que se afirma que, en media, rellena las botellas con 100 cl. de aceite. Examinada una muestra de 5 botellas se obtiene que el promedio es 95 cl. y la varianza (insesgada) es 1.1. Suponemos que la v.a. $X =$ “contenido de aceite (en cl.) en una botella” sigue una distribución $N(\mu, \sigma)$. ¿Hay suficiente evidencia empírica para afirmar que el contenido medio de las botellas no es 100 cl.?

Queremos contrastar

$$H_0 : \mu = 100, \text{ frente a } H_1 : \mu \neq 100.$$

Otra posibilidad sería preguntarse si existe evidencia empírica suficiente para afirmar que el consumidor recibe, en promedio, menos cantidad de la que indica la etiqueta. En ese caso, el planteamiento correcto sería contrastar

$$H_0 : \mu \geq 100, \text{ frente a } H_1 : \mu < 100.$$

Ejemplo (cont.): Tenemos, por tanto, un problema del tipo:

contrastar $H_0 : \mu = \mu_0$ frente a $H_1 : \mu \neq \mu_0$

(siendo μ_0 un valor prefijado) a partir de una muestra X_1, \dots, X_n extraída de $N(\mu, \sigma)$.

Para ello prefijamos el nivel de significación del test $\alpha \in (0, 1)$ (por ejemplo, $\alpha = 0.05$) y observamos que, si H_0 fuera cierta,

$$\frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}. \quad (1)$$

Por otra parte, está claro que deberíamos sospechar que H_0 es falsa (y, por tanto, H_1 es cierta) cuando \bar{x} resulte estar “suficientemente alejada” de μ_0 . El resultado (1) nos ayuda a decidir, de un modo racional, que es lo que significa “suficientemente alejada”.

Ejemplo (cont.): En efecto, dada una muestra x_1, \dots, x_n , parece muy natural decidir que H_0 es falsa cuando tengamos

$$\left| \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right| > t_{n-1; \alpha/2}. \quad (2)$$

ya que, en este caso, tenemos una muestra que sería “muy rara” si realmente H_0 fuera cierta.

Obsérvese que, de todos modos, hay una probabilidad α , prefijada, de cometer un error de tipo I (rechazar H_0 siendo cierta).

Análogamente, si el problema hubiera sido contrastar

$$H_0 : \mu \geq \mu_0 \text{ frente a } H_1 : \mu < \mu_0,$$

un criterio razonable para rechazar H_0 con un nivel de significación α sería

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} < -t_{n-1; \alpha}. \quad (3)$$

Ejemplo (cont.): En el ejemplo del envasado de aceite $\bar{x} = 95$, $s^2 = 1.1$, $n = 5$. Por tanto,

$$\left| \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right| = 10.66$$

y como $10.66 > t_{4;0.025} = 2.776445$, $H_0 : \mu = 100$ se rechaza al nivel de significación $\alpha = 0.05$ y, dado que $10.66 > t_{4;0.005} = 4.604095$, también se rechaza al nivel 0.01.

Sin embargo, supongamos que hubiéramos obtenido $\bar{x} = 98$, $s^2 = 1.1$, $n = 5$. Entonces

$$\left| \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \right| = 4.264014$$

y la hipótesis H_0 se rechazaría al nivel $\alpha = 0.05$ pero NO se rechazaría al nivel 0.01.

Ejemplo (cont.): En el ejemplo de las botellas, supongamos que queremos contrastar

$$H_0 : \mu \geq 100 \text{ frente a } H_1 : \mu < 100.$$

Entonces el criterio para rechazar H_0 con un nivel de significación α sería

$$\frac{\bar{x} - 100}{s/\sqrt{5}} < -t_{4;\alpha}.$$

Supongamos que hubiéramos obtenido $\bar{x} = 98$. Entonces

$\frac{\bar{x} - 100}{s/\sqrt{5}} = -4.2640$. Como $-t_{4;0.01} = -3.7469$, la hipótesis nula

$H_0 : \mu \geq 100$ se rechaza al nivel 0.01 (y también por supuesto, al nivel 0.05, ya que $-t_{4;0.05} = -2.1318$).

Algunas consecuencias y observaciones:

El anterior ejemplo es un sólo un caso particular de contraste de hipótesis, pero nos permite extraer algunas consecuencias y dar algunas definiciones generales sobre la metodología del contraste de hipótesis.

- ▶ **Asimetría de las hipótesis:** H_0 se acepta a menos que “se haya obtenido suficiente evidencia estadística en contra de ella”. Por esta razón, cuando H_0 se acepta no debe pensarse que “se ha demostrado su validez”. H_0 representa la hipótesis que estamos dispuestos a aceptar a menos que se obtengan fuertes indicios en contra.
- ▶ **Errores de tipo I y II:** En todo caso siempre hay una probabilidad positiva de cometer uno de los dos posibles errores: rechazar H_0 cuando es cierta (error de tipo I) o aceptar H_0 cuando es falsa (error de tipo II).

- ▶ **El nivel de significación:** Los tests usuales están contruidos de modo que la máxima probabilidad de cometer el error de tipo I está acotada por un valor prefijado α , el nivel de significación (no confundir con el nivel de confianza de los intervalos).
- ▶ La decisión de rechazar o aceptar H_0 **depende del nivel de significación elegido**. Cuánto más pequeño es α más “conservador” se hace el test a favor de H_0 , es decir, que para aceptar H_1 cuando α es muy pequeño, debemos tener “muchas evidencias estadísticas”.
- ▶ Cuando se toma una decisión (aceptar o rechazar H_0) **debe indicarse siempre el nivel de significación del test que se ha utilizado**.
- ▶ **Cuando se acepta H_0** no debe pensarse que se ha demostrado H_0 sino que “no se ha encontrado suficiente evidencia empírica (al nivel prefijado α) en contra de H_0 ”.

- ▶ Cuando se acepta H_1 debe recordarse también que la interpretación correcta es que “los datos obtenidos proporcionan suficiente evidencia estadística al nivel α para aceptar H_1 ”. Una muestra o un nivel de significación diferentes podrían conducir a conclusiones distintas.
- ▶ **Dualidad con los intervalos de confianza:** en algunos casos de hipótesis nula simple (i.e. del tipo $H_0 : \theta = \theta_0$) el test usual rechaza H_0 (al nivel de significación α) si y sólo si el intervalo de nivel de confianza $1 - \alpha$ no contiene al valor θ_0 .

Ejemplo: Si $X \sim N(\mu, \sigma)$ la región de rechazo

$$R = \left\{ (x_1, \dots, x_n) : |\bar{x} - \mu_0| \geq t_{n-1; \alpha/2} \frac{s}{\sqrt{n}} \right\}$$

del contraste

$$\begin{aligned} H_0 : \mu &= \mu_0 \\ H_1 : \mu &\neq \mu_0 \end{aligned} \quad \alpha$$

equivale a

$$R = \{ (x_1, \dots, x_n) : \mu_0 \notin IC_{1-\alpha}(\mu) \}.$$

- ▶ **Estadístico del contraste:** Habitualmente la región de rechazo se expresa en términos de un estadístico del contraste (en el ejemplo anterior $t = (\bar{X} - \mu_0)/(S/\sqrt{n})$), cuya distribución bajo H_0 o para el valor del parámetro que marca la frontera entre H_0 y H_1 es conocida al menos asintóticamente.

La hipótesis nula se rechaza cuando el estadístico del contraste toma un valor poco probable bajo la supuesta distribución de probabilidad que tendría si H_0 fuera cierta.

A continuación reunimos algunos de los contrastes básicos más habituales.

Contrastes para la media de una distribución

En cada caso se rechaza H_0 cuando $(x_1, \dots, x_n) \in R$.

- **Distribución normal con varianza conocida**

Sea $X \sim N(\mu, \sigma)$ con σ conocido.

$$H_0 : \mu = \mu_0 \quad R = \left\{ (x_1, \dots, x_n) : \left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \right| \geq z_{\alpha/2} \right\}$$
$$= \{ (x_1, \dots, x_n) : |z| \geq z_{\alpha/2} \}$$

$$H_0 : \mu \leq \mu_0 \quad R = \{ (x_1, \dots, x_n) : z \geq z_{\alpha} \}$$

$$H_0 : \mu \geq \mu_0 \quad R = \{ (x_1, \dots, x_n) : z \leq z_{1-\alpha} \}$$

donde $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ es el *estadístico del contraste*.

- **Distribución normal con varianza desconocida**

Sea $X \sim N(\mu, \sigma)$ con σ desconocido.

$$H_0 : \mu = \mu_0 \quad R = \{(x_1, \dots, x_n) : |t| \geq t_{n-1; \alpha/2}\}$$

$$H_0 : \mu \leq \mu_0 \quad R = \{(x_1, \dots, x_n) : t \geq t_{n-1; \alpha}\}$$

$$H_0 : \mu \geq \mu_0 \quad R = \{(x_1, \dots, x_n) : t \leq t_{n-1; 1-\alpha}\}$$

siendo $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ el estadístico del contraste.

Cómo hacer un contraste de la t con R:

```
help(t.test)
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"), mu = 0,
       paired = FALSE, var.equal = FALSE, conf.level =
       0.95,...)
```

Ejemplo: Se certifica que un material estándar de referencia de un suelo contiene 94.6 ppm de un contaminante orgánico. Un análisis repetido arrojó los siguientes resultados: 98.6, 98.4, 97.2, 94.6 y 96.2 ppm. A un nivel de significación $\alpha = 0.05$ ¿hay suficiente evidencia estadística para concluir que los resultados difieren del valor esperado?. Si se disminuye α a 0.01, ¿se rechazaría H_0 ?

```
X = c(98.6, 98.4, 97.2, 94.6, 96.2)
t.test(X,alternative="two.sided",mu=94.6)
```

```
One Sample t-test
```

```
data: X
```

```
t = 3.2421, df = 4, p-value = 0.03161
```

```
alternative hypothesis: true mean is not equal to 94.6
```

```
95 percent confidence interval:
```

```
 94.94468 99.05532
```

```
sample estimates:
```

```
mean of x
```

```
 97
```

R no parece dar una solución al problema del contraste. ¿O sí?

Presentemos el concepto de p-valor.

- **Tests de nivel aproximado α (muestras grandes) para la media de cualquier distribución**

Sea X con $\mathbb{E}(X) = \mu < \infty$.

$H_0 : \mu = \mu_0$, frente a $H_1 : \mu \neq \mu_0$

$$R = \{(x_1, \dots, x_n) : |z| > z_{\alpha/2}\}$$

$H_0 : \mu \leq \mu_0$, frente a $H_1 : \mu > \mu_0$

$$R = \{(x_1, \dots, x_n) : z > z_{\alpha}\}$$

$H_0 : \mu \geq \mu_0$, frente a $H_1 : \mu < \mu_0$

$$R = \{(x_1, \dots, x_n) : z < -z_{\alpha}\}$$

siendo $z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ el estadístico del contraste.

- **Tests de nivel aproximado α (muestras grandes) para el parámetro p en una Bernoulli**

Sea $X \sim \text{Bernoulli}(p)$. Para el contraste

$$H_0 : p = p_0, \text{ frente a } H_1 : p \neq p_0,$$

la región de rechazo es

$$R = \{(x_1, \dots, x_n) : |z| > z_{\alpha/2}\},$$

donde $z = \frac{\bar{x} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ es el estadístico del contraste. Los tests con hipótesis H_0 y H_1 unilaterales tienen regiones de rechazo análogas.

En el formulario que se puede descargar de la página web hay una lista de las regiones críticas correspondientes a los contrastes de uso más frecuente.

El concepto de p-valor

Dado un test, definido para todos los niveles de significación posibles, se define el **p-valor**, para unos datos prefijados, como el **ínfimo de los valores α para los cuales se rechaza la hipótesis nula a un nivel de significación α .**

$$P(x_1, \dots, x_n) = \inf\{\alpha : H_0 \text{ es rechazada al nivel } \alpha\}.$$

Cuánto más pequeño es el p-valor, más evidencia estadística aportan los datos a favor de H_1 .

El p-valor se puede interpretar como la probabilidad (bajo H_0) de obtener un valor “al menos tan raro” como el obtenido.

Los programas informáticos que realizan contrastes de hipótesis (R, Excel, Matlab, . . .) no realizan el contraste para un nivel de significación α , sino que directamente nos dan el p-valor del contraste.

Ejemplo: Los siguientes datos son la cantidad de sodio (en ppm) presente en 20 muestras de análisis séricos (Fuente de los datos: Andrews and Herzberg 1985):

140	143	141	137	132	157	143	149	118	145
138	144	144	139	133	159	141	124	145	139

¿Hay evidencia de que el nivel medio de sodio en sangre es superior a 140 ppm? ¿Y de que sea inferior a 145?

Contrastes para la varianza de una normal

Sea X_1, \dots, X_n una muestra de $X \sim N(\mu, \sigma)$ con σ desconocido.

$$H_0 : \sigma = \sigma_0 \quad R = \left\{ \frac{(n-1)s^2}{\sigma_0^2} \notin (\chi_{n-1;1-\alpha/2}^2, \chi_{n-1;\alpha/2}^2) \right\}$$
$$= \{ \sigma_0^2 \notin IC_{1-\alpha}(\sigma^2) \}$$

$$H_0 : \sigma \leq \sigma_0 \quad R = \left\{ \frac{(n-1)s^2}{\sigma_0^2} \geq \chi_{n-1;\alpha}^2 \right\}$$

$$H_0 : \sigma \geq \sigma_0 \quad R = \left\{ \frac{(n-1)s^2}{\sigma_0^2} \leq \chi_{n-1;1-\alpha}^2 \right\}$$

El estadístico del contraste

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma_0^2} \quad (\text{X-squared en R})$$

sigue una distribución χ_{n-1}^2 si $\sigma^2 = V(X)$ es igual a σ_0^2 .

Para hacer este contraste con R, instalar el paquete TeachingDemos y usar la función `sigma.test`.

Ejemplo:

```
X = c(98.6, 98.4, 97.2, 94.6, 96.2)
sigma.test(X)
```

One sample Chi-squared test for variance

```
data: X
X-squared = 10.96, df = 4, p-value = 0.05403
alternative hypothesis: true variance is not equal to 1
95 percent confidence interval:
 0.9835518 22.6250622
sample estimates:
var of X
 2.74
```

```
sigma.test(X,sigma=2,alternative="greater")
```

One sample Chi-squared test for variance

```
data: X
X-squared = 2.74, df = 4, p-value = 0.6022
alternative hypothesis: true variance is greater than 4
95 percent confidence interval:
 1.155176      Inf
sample estimates:
var of X
 2.74
```

Contrastes para dos muestras

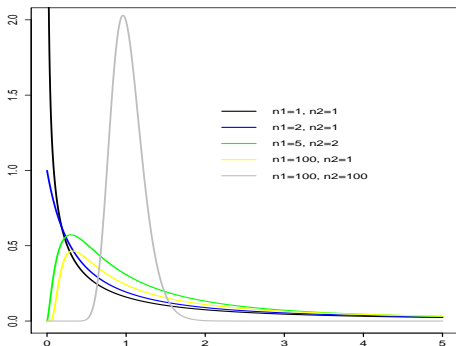
El objetivo es comparar el mismo parámetro en dos poblaciones diferentes.

Contrastes con dos poblaciones normales independientes

Sean X_1, \dots, X_{n_1} e Y_1, \dots, Y_{n_2} muestras aleatorias independientes de $X \sim N(\mu_1, \sigma_1)$ e $Y \sim N(\mu_2, \sigma_2)$ respectivamente (σ_1 y σ_2 desconocidas). X e Y son v.a. independientes.

Para contrastar la hipótesis de **homocedasticidad** (igualdad de **varianzas**) de dos poblaciones normales presentamos una nueva distribución auxiliar.

Sean Q_1 y Q_2 v.a. independientes con distribuciones $\chi_{n_1}^2$ y $\chi_{n_2}^2$, respectivamente. La distribución de $\frac{Q_1/n_1}{Q_2/n_2}$ se denomina **F de Fisher-Snedecor con n_1 y n_2 grados de libertad**, F_{n_1, n_2} .



Si s_1^2 , s_2^2 son las varianzas (insesgadas) de dos muestras independientes de tamaño n_1 y n_2 extraídas, respectivamente, de dos poblaciones $N(\mu_1, \sigma_1)$ y $N(\mu_2, \sigma_2)$, se tiene

$$\frac{(n_1 - 1)s_1^2}{\sigma_1^2} \sim \chi_{n_1-1}^2, \quad \frac{(n_2 - 1)s_2^2}{\sigma_2^2} \sim \chi_{n_2-1}^2.$$

Por tanto, bajo $H_0 : \sigma_1 = \sigma_2$, el estadístico del contraste $F = s_1^2/s_2^2$ sigue una distribución F_{n_1-1, n_2-1} . De este resultado se derivan los tests para comparar σ_1 y σ_2 .

$$H_0 : \sigma_1 = \sigma_2 \quad R = \left\{ F \notin (F_{n_1-1; n_2-1; 1-\alpha/2}, F_{n_1-1; n_2-1; \alpha/2}) \right\}$$
$$= \left\{ 1 \notin \text{IC}_{1-\alpha} \left(\frac{\sigma_1^2}{\sigma_2^2} \right) \right\}$$

$$H_0 : \sigma_1 \leq \sigma_2 \quad R = \{ F > F_{n_1-1; n_2-1; \alpha} \}$$

$$H_0 : \sigma_1 \geq \sigma_2 \quad R = \{ F < F_{n_1-1; n_2-1; 1-\alpha} \}$$

Uno de los tests más usuales es el de **igualdad de medias** para dos poblaciones normales “homocedásticas”, es decir, con

$$\sigma_1 = \sigma_2 = \sigma:$$

Se puede probar que, bajo $H_0 : \mu_1 = \mu_2$,

$$\frac{\bar{X} - \bar{Y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

y, por tanto, una región crítica al nivel α es

$$R = \left\{ |\bar{x} - \bar{y}| > t_{n_1+n_2-2; \alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$$

siendo

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

la varianza combinada (*pooled variance*), una estimación de σ^2 .

Ejemplo: Un microbiólogo desea averiguar si hay diferencia en el tiempo que tarda en producir yogur utilizando dos tipos de bacterias: *lactobacillus acidophilus* (A) y *bulgaricus* (B). Se prepararon siete remesas de yogur con cada tipo de lactobacilo. A continuación se muestra el tiempo (en horas) hasta que se produjo cada remesa:

							Cultivo A						
6.8	6.3	7.4	6.1	8.2	7.3	6.9							
							Cultivo B						
6.1	6.4	5.7	5.5	6.9	6.3	6.7							

Suponiendo que la distribución de ambos conjuntos de observaciones se puede considerar normal, contrastar la hipótesis de homocedasticidad (igualdad de varianzas) y la hipótesis de igualdad de medias.

Ejemplo (cont.):

```
A = c(6.8, 6.3, 7.4, 6.1, 8.2, 7.3, 6.9)
```

```
B = c(6.1, 6.4, 5.7, 5.5, 6.9, 6.3, 6.7)
```

```
var.test(A,B,ratio=1,alternative="two.sided",conf.level  
=0.9)
```

F test to compare two variances

data: A and B

F = 1.9814, num df = 6, denom df = 6, p-value = 0.4259

alternative hypothesis: true ratio of variances is not
equal to 1

90 percent confidence interval:

0.462521 8.487957

sample estimates:

ratio of variances

1.981378

Ejemplo (cont.):

```
t.test(A,B,alternative="two.sided",conf.level=0.95,var.
      equal=TRUE)
```

Two Sample t-test

data: A and B

t = 2.3375, df = 12, p-value = 0.03755

alternative hypothesis: true difference in means is not
equal to 0

95 percent confidence interval:

0.05238216 1.49047498

sample estimates:

mean of x mean of y

7.000000 6.228571

Caso de muestras emparejadas:

Surge en aquellas situaciones con $n_1 = n_2$ en que X_i e Y_i no son independientes (porque corresponden, por ejemplo, a mediciones sobre el mismo individuo antes y después de un tratamiento).

Se reducen a problemas de una muestra para la muestra de diferencias $D_i = X_i - Y_i$.

Ejemplo: Para evaluar el efecto del ejercicio físico en el control del colesterol se midió la concentración de colesterol en 15 individuos antes (X) y después (Y) de un programa de ejercicio aeróbico. Se analizan los datos con R de la manera siguiente:

```
x = c(265,240,258,295,251,245,287,314,260,279,283,240,238,225,247)
y = c(229,231,227,240,238,241,234,256,247,239,246,218,219,226,233)
t.test(x,y,paired=T)
```

Paired t-test

```
data: x and y
t = 5.4659, df = 14, p-value = 8.316e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 16.32430 37.40904
sample estimates:
mean of the differences
      26.86667
```

Construcción de tests. Test de cociente de verosimilitudes

Un método general y ampliamente utilizado para construir regiones de rechazo es el de razón de verosimilitudes.

Sea $f(\cdot; \boldsymbol{\theta})$ donde $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k) \in \Theta \subset \mathbb{R}^k$.

Dada una muestra $\mathbf{x} = (x_1, \dots, x_n)$, sea $L_n(\boldsymbol{\theta}; \mathbf{x}) = \prod_{i=1}^n f(x_i; \boldsymbol{\theta})$ la función de verosimilitud.

Consideremos un contraste de hipótesis paramétricas general

$$H_0 : \boldsymbol{\theta} \in \Theta_0$$

$$H_1 : \boldsymbol{\theta} \in \Theta \setminus \Theta_0.$$

El estadístico del *contraste de razón de verosimilitudes* es

$$\Lambda_n = \Lambda_n(\mathbf{x}) := \frac{\sup_{\theta \in \Theta_0} L_n(\theta; \mathbf{x})}{\sup_{\theta \in \Theta} L_n(\theta; \mathbf{x})} = \frac{\sup_{\theta \in \Theta_0} L_n(\theta; \mathbf{x})}{L_n(\hat{\theta}_n; \mathbf{x})},$$

siendo $\hat{\theta}_n$ el e.m.v. de θ .

El contraste de razón de verosimilitudes tiene como región de rechazo $R = \{\mathbf{x} : \Lambda_n(\mathbf{x}) \leq k_\alpha\}$, donde $\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(R) = \alpha$.

Puede ser complicado determinar el cuantil k_α . Veamos un resultado asintótico que facilita esta labor.

Consideremos el problema de contrastar

$$H_0 : \theta_i = c_i, \text{ para } i = 1, \dots, r \text{ (con } r \leq k)$$

$$H_1 : \theta_i \neq c_i \text{ para algún } i = 1, \dots, r$$

Entonces $\Theta_0 = \{\theta \in \Theta : \theta = (c_1, \dots, c_r, \theta_{r+1}, \dots, \theta_k)\}$.

TEOREMA.- *Supongamos que*

- (i) *El emv $\hat{\theta}_n$ es estimador consistente (en probabilidad) del parámetro θ .*
- (ii) *Para todo x , la función $\log f(x; \theta)$ tiene derivadas parciales terceras (respecto a las componentes θ_j de θ) continuas.*
- (iii) *En las integrales que involucran a la función $f(x; \theta)$ se pueden permutar las derivadas con el signo integral.*
- (iv) *La matriz de información de Fisher*

$$I(\theta) = -\mathbb{E}_{\theta} \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(X; \theta) \right)_{1 \leq i, j \leq k}$$

es invertible para cada θ .

Entonces, bajo H_0 ,

$$-2 \log \Lambda_n \xrightarrow{d} \chi_r^2. \quad (4)$$

Aplicación a tests de bondad de ajuste

Sea X una v.a. discreta que toma los valores a_1, \dots, a_k .

Denotemos $p_i = \mathbb{P}(X = a_i)$. Supongamos que se desea contrastar

$$H_0 : p_i = p_{i0}, \quad i = 1, \dots, k$$

basado en una muestra x_1, \dots, x_n . Obsérvese que, en este caso, con la notación del teorema, $r = k - 1$ porque cuando se fijan $k - 1$ probabilidades p_i queda fijada la probabilidad restante.

Por tanto, H_0 se rechaza al nivel α cuando

$$-2 \log \Lambda_n > \chi_{k-1; \alpha}^2,$$

Aquí el numerador de Λ_n es

$$\frac{n!}{O_1! \dots O_k!} p_{10}^{O_1} \dots p_{k0}^{O_k},$$

siendo $O_j = \#\{i : x_i = a_j\}$ las “frecuencias observadas” de los distintos valores de la variable [nótese que, bajo H_0 , (O_1, \dots, O_k) tiene distribución multinomial $\mathcal{M}(n; p_{10}, \dots, p_{k0})$]. El denominador de Λ_n es

$$\frac{n!}{O_1! \dots O_k!} \left(\frac{O_1}{n}\right)^{O_1} \dots \left(\frac{O_k}{n}\right)^{O_k}.$$

Sustituyendo en Λ_n es inmediato ver que el estadístico de contraste se puede expresar en la forma

$$-2 \log \Lambda_n = 2 \sum_{j=1}^k O_j \log \left(\frac{O_j}{e_j}\right),$$

donde $e_j = np_{j0}$, $j = 1, \dots, k$ son las “frecuencias esperadas (bajo H_0)” de los distintos valores de la variable en una muestra de tamaño n .

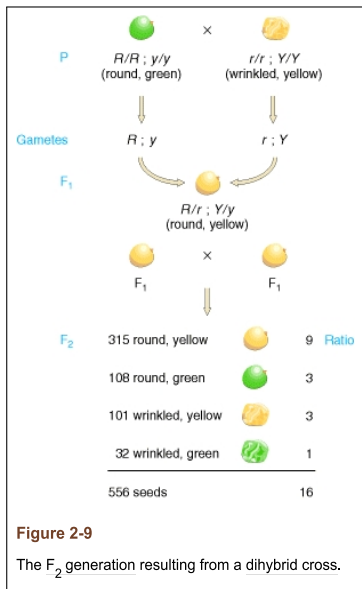
Un ejemplo clásico: el experimento de Mendel

En el famoso experimento de Mendel se cruzaron plantas de guisantes con fenotipo rugoso-amarillo con otras de fenotipo liso-verde. En la segunda generación se podían observar cuatro fenotipos (liso-amarillo, rugoso-amarillo, liso-verde, rugoso-verde) cuyas respectivas probabilidades, según la teoría de la herencia mendeliana, debían ser

$$p_{10} = \frac{9}{16}, p_{20} = \frac{3}{16}, p_{30} = \frac{3}{16}, p_{40} = \frac{1}{16}.$$

Observados $n = 556$ guisantes en la segunda generación del experimento se obtuvieron los siguientes números de guisantes con estos fenotipos:

$$O_1 = 315, O_2 = 101, O_3 = 108, O_4 = 32.$$



¿Proporcionan estos resultados alguna evidencia en contra de la teoría mendeliana?

Aplicamos el test para contrastar $H_0 : p_1 = \frac{9}{16}, \dots, p_4 = \frac{1}{16}$:

$$e_1 = 556 \cdot \frac{9}{16} = 312.75, \quad e_2 = e_3 = 556 \cdot \frac{3}{16} = 104.25, \quad e_4 = 556 \cdot \frac{1}{16} = 34.75$$

En definitiva, el test de cociente de verosimilitudes compara las O_i con las e_i y rechaza la hipótesis nula cuando hay “demasiadas diferencias” entre ellas. Esto se hace formalmente mediante el estadístico

$$-2 \log \Lambda_n = 2 \sum_{i=1}^k O_i \log \left(\frac{O_i}{e_i} \right) = 0.4754.$$

El p-valor (calculado a partir de la distribución χ_3^2) es 0.9281 lo que, por supuesto, no indica ninguna evidencia estadística en contra de H_0 .

Hay una controversia clásica en la historia de la ciencia en el sentido de que los resultados de Mendel eran “demasiado buenos”, es decir, había demasiada concordancia entre las O_i y las e_i (por ejemplo, R.A. Fisher era de esta opinión; ver su artículo de 1936, “Has Mendel’s work been rediscovered?”, en *The Annals of Science*).

Se ha sugerido que este supuesto “exceso de concordancia” podía deberse a un “sesgo de repetición” (*confirmation bias*) producido por la repetición de los resultados hasta que las O_i concordasen fuertemente con las e_i . También se ha conjeturado que algún ayudante de Mendel pudo actuar con “exceso de celo” manipulando los resultados. En todo caso, las ideas básicas de Mendel eran acertadas y han tenido una influencia decisiva.

Constrastes uniformemente más potentes

Se dice que un test con función de potencia β_n es **uniformemente más potente** (UMP) dentro de una clase $\mathcal{B}_{n,\alpha}$ de tests de nivel α basados en muestras de tamaño n cuando

$$\beta_n(\theta) \geq \tilde{\beta}_n(\theta), \quad \forall \theta \in \Theta_1$$

siendo $\tilde{\beta}_n$ la función de potencia de cualquier otro test de la clase $\mathcal{B}_{n,\alpha}$.

Esta condición es tan fuerte que los test UMP no existen en muchas situaciones reales, pero merece la pena plantearse cuándo podemos garantizar su existencia.

El lema de Neyman-Pearson

Se considera el problema de hipótesis nula y alternativa simples

$$H_0 : \theta = \theta_0 \text{ frente a } H_1 : \theta = \theta_1.$$

Denotemos $f_n(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$.

Dado $\alpha \in (0, 1)$, supongamos que la región

$$R^* = \left\{ (x_1, \dots, x_n) : \frac{f_n(x_1, \dots, x_n; \theta_1)}{f_n(x_1, \dots, x_n; \theta_0)} > k \right\}$$

verifica $\mathbb{P}_{\theta_0}(R^*) = \alpha$. Entonces

$$\mathbb{P}_{\theta_1}(R^*) \geq \mathbb{P}_{\theta_1}(R),$$

siendo R la región crítica de cualquier otro test tal que $\mathbb{P}_{\theta_0}(R) \leq \alpha$.
En otras palabras, R^* es el test óptimo de nivel α para el problema considerado.

Demostración del lema de Neyman-Pearson: Denotemos

$$\mathbf{x} = (x_1, \dots, x_n)$$

$$\mathbb{P}_{\theta_1}(R^*) - \mathbb{P}_{\theta_1}(R) = \int_{R^* \cap R^c} f_n(\mathbf{x}; \theta_1) d\mathbf{x} - \int_{R^{*c} \cap R} f_n(\mathbf{x}; \theta_1) d\mathbf{x},$$

pero, por definición de R^* ,

$$\int_{R^* \cap R^c} f_n(\mathbf{x}; \theta_1) d\mathbf{x} \geq k \int_{R^* \cap R^c} f_n(\mathbf{x}; \theta_0) d\mathbf{x}$$

y también

$$\int_{R^{*c} \cap R} f_n(\mathbf{x}; \theta_1) d\mathbf{x} \leq k \int_{R^{*c} \cap R} f_n(\mathbf{x}; \theta_0) d\mathbf{x}.$$

Por lo tanto

$$\begin{aligned} \mathbb{P}_{\theta_1}(R^*) - \mathbb{P}_{\theta_1}(R) &\geq k \left[\int_{R^* \cap R^c} f_n(\mathbf{x}; \theta_0) d\mathbf{x} - \int_{R^{*c} \cap R} f_n(\mathbf{x}; \theta_0) d\mathbf{x} \right] \\ &= k \left[\int_{R^*} f_n(\mathbf{x}; \theta_0) d\mathbf{x} - \int_R f_n(\mathbf{x}; \theta_0) d\mathbf{x} \right] \\ &= k [\mathbb{P}_{\theta_0}(R^*) - \mathbb{P}_{\theta_0}(R)] \geq 0. \end{aligned}$$

□

Referencias

Casella, G., Berger, R.L. (2002). *Statistical Inference*. Second Edition. Duxbury. Thomson Learning. Capítulo 8.

Knight, K. (2000). *Mathematical Statistics*. Chapman & Hall/CRC. Capítulo 7.