Tema 7. Constraste de hipótesis

Estadistica Básica- Grado en Matemáticas

José Jaime Noguera Noguera
josnoguera@denia.uned.es

9 de diciembre de 2022

CONTENIDOS

Introducción

Los test de hipótesis se utilizan para tomar decisiones acerca de características poblacionales.

Una hipótesis estadística es cualquier afirmación, verdadera o falsa realizada sobre alguna característica desconocida de la población. Dicha hipótesis puede referirse a:

- Valor de un parámetro desconocido: **contraste paramétrico**.
- Forma de la función de cuantía o de densidad de la población: **constraste no paramétrico**.

En este tema se estudiarán los contrastes paramétricos y supondremos que conocemos la función de cuantía o de densidad de la población.

Paso 1

Formular la hipótesis nula H_0 y la alternativa H_1. Deben ser mutuamente excluyentes y complementarias. Hay de varios tipos

I. $H_0 : \theta = \theta_0, \ H_1 : \theta \neq \theta_0$

II. $H_0 : \theta \geq \theta_0, \ H_1 : \theta < \theta_0$

III. $H_0 : \theta \leq \theta_0, \ H_1 : \theta > \theta_0$

IV. $H_0 : \theta_1 \leq \theta \leq \theta_2, \ H_1 : \theta < \theta_1$ ó $\theta > \theta_2$

Paso 2

Determinar el estadístico apropiado para realizar el test. Los más típicos son:

- Si contrastamos una media, μ y la población es $N(\mu, \sigma)$ con σ conocida:

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \to N(0,1)$$

- Si contrastamos una media, μ y la población es $N(\mu, \sigma)$ con σ desconocida:

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \to t_{n-1} \quad \left(n \geq 30,\ Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \to N(0,1)\right)$$

- Si contrastamos una proporción y n es grande:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \to N(0,1)$$

Paso 3

Seleccionar el **nivel de significación**, α. Esto lo establece el problema:

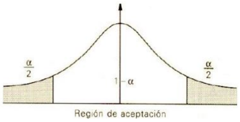
- Es la probabilidad de rechazar la hipótesis nula cuando ésta es cierta.
- Normalmente se toma α como $0{,}1\,;0{,}05\,;0{,}01$.

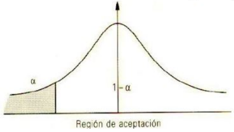
Paso 4

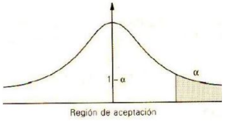
Determinar la **región crítica o de rechazo de H_0**:

- Es la parte más importante.
- Se determina en función de α, de la distribución muestral del estadístico y del tipo de contraste.

$\left.\begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{array}\right\}$	$\left.\begin{array}{l} H_0 : \theta \geq \theta_0 \\ H_1 : \theta < \theta_0 \end{array}\right\}$	$\left.\begin{array}{l} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{array}\right\}$







Paso 5

Calcular el valor del estadístico al aplicarlo a la muestra.

Paso 6

- Si el valor del estadístico cae en la región crítica, se rechaza H_0.
- Si el valor del estadístico cae fuera de la región crítica, se acepta H_0.
- Finalmente de interpretan los resultados.

p-valor

Alternativamente podemos tomar la decisión utilizando el concepto del ***p*-valor**, que representa el mínimo nivel de significación necesarion para rechazar H_0.

- Contraste no significativo: $p - valor > \alpha$, no rechazamos H_0
- Contraste significativo: $p - valor \leq \alpha$, se rechaza H_0

En muchos ejercicios no se nos da el valor de α y debemos decidir qué hipótesis escogemos en base al $p - valor$:

- Si $p - valor \geq 0'2 \Rightarrow$ aceptamos H_0.
- Si $p - valor \leq 0'01 \Rightarrow$ rechazamos H_0.
- Si $0'02 < p - valor < 0'2 \Rightarrow$ no hay una decisión clara.

Errores

Tipos de Errores

Si aceptamos H_0 siendo falsa se comete un error de Tipo II, mientras que si se rechaza H_0 siendo verdadera se comete un error de Tipo I.

- El **nivel de significación**, α, nos da la probabilidad de rechazar la hipótesis nula cuando es cierta, error Tipo I.
- La probabilidad de aceptar la hipótesis nula cuando es cierta será $1 - \alpha$ (coeficiente de confianza).
- Si denotamos por β la probabilidad de cometer un error Tipo II (no rechazar H_0 siendo falsa), entonces la probabilidad de rechazar H_0 siendo falsa es $1 - \beta$, cantidad que se conoce como **potencia del contrate**.

Nos interesa un nivel de significación pequeño y una potencia del contraste elevada.

Formulación de la hipótesis nula y alternativa

Como norma general, se debe tener en cuenta que aquello que queramos demostrar debe ir siempre a la hipótesis alternativa. Por otra parte si lo que queremos demostrar está en la hipótesis alternativa, el error de equivocarnos lo tendremos medido ya que sería el nivel de significación.

Ejemplo 1

Según un estudio publicado por una agencia de medios, menos del 15 % de los usuarios de internet y móvil utiliza nuestra app o nuestra página web para reservar los hoteles. Con la información que tenemos creemos que ese dato es mayor, por lo que decidimos realizar un estudio con el objeto de contrarrestar dicha información. Tomamos una muestra de 400 potenciales usuarios, de los cuales 80 contestaron que eran usuarios habituales de nuestra web/app. Realizar un contraste de hipótesis con nivel de significación $\alpha = 0'05$.

Ejemplo 1

Como es una proporción el estadístico apropiado será:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \to N(0,1)$$

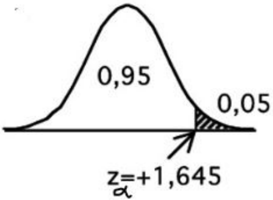
Además conocemos $\hat{p} = \frac{80}{400} = 0'2$ y $p_0 = 0'15$.

El test de hipótesis a plantear es: $\left.\begin{array}{l} H_0 : p \leq p_0 \\ H_1 : p > p_0 \end{array}\right\}$

El valor experimental del estadístico es $z_{exp} = \frac{0'2 - 0'15}{\sqrt{\frac{0'15 \cdot 0'85}{400}}} \approx 2'8$

Ejemplo 1

Al 5 % de significación y dado el contraste que hemos planteado, debemos buscar el valor crítico z_α que deja a la derecha un 5 % de área, es decir $P(Z > z_\alpha) = 0'05$ o bien $P(Z \leq z_\alpha) = 0'95$, es decir:



Como $z_{exp} = 2'8 > 1'645 = z_\alpha$, cae en la región crítica o de rechazo de H_0.

Por tanto, al 5 % de significación, con los datos de la muestra, hay evidencia para rechazar la hipótesis nula.

Ejemplo 1

Las tablas de Adenda nos facilitan la resolución de estos ejercicios:

Población $X \to B(1,\ p)$

$$\frac{\hat{p} - p}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}} \approx N(0,\ 1) \qquad I = \left[\hat{p} \mp z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

Acep.	$H_0 : p = p_0$ $H_1 : p \neq p_0$	$H_0 : p \leqslant p_0$ $H_1 : p > p_0$	$H_0 : p \geqslant p_0$ $H_1 : p < p_0$
H_0	$\dfrac{\lvert\hat{p} - p_0\rvert}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \leqslant z_{\alpha/2}$	$\dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \leqslant z_{\alpha}$	$\dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \geqslant z_{1-\alpha}$
H_1	$\dfrac{\lvert\hat{p} - p_0\rvert}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} > z_{\alpha/2}$	$\dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} > z_{\alpha}$	$\dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} < z_{1-\alpha}$

Ejemplo 1

Imaginemos que ahora nos piden que realicemos el contraste a un nivel de significación del 1 % y que utilicemos el $p - valor$.

Para nuestro contraste, el $p - valor$ es el área que deja a la derecha z_{exp}:

$$P(Z > z_{exp}) = P(Z > 2'8) = 0'0026$$

Como el $p - valor = 0'0026 < 0'01 = \alpha$ el contraste es significativo y se rechaza H_0.

Ejemplo 2

¿Qué pasaría si planteamos el test intercambiando H_0 y H_1?

Supongamos que el test lo planteamos como:

$$\left.\begin{array}{l} H_0 : p \geq p_0 \\ H_1 : p < p_0 \end{array}\right\}$$

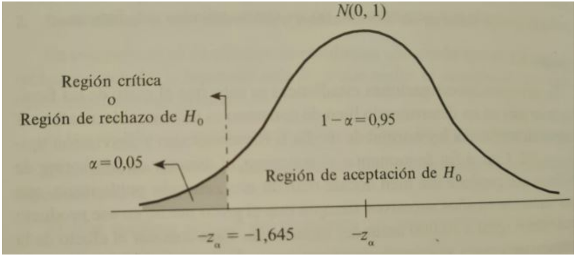
El estadístico es el mismo, al igual que p_0 y el valor del estadístico según la muestra elegida. Lo que cambia es la región de crítica. En este caso buscamos el valor crítico z_α que deja a la izquierda un 5 % de área, es decir $P(Z \leq -z_\alpha) = 0'05$ o bien $P(Z \geq -z_\alpha) = 0'95$, o con la notación de las tablas de test de hipótesis:

$$P(Z \geq z_{1-\alpha}) = 0'95$$

Aunque este valor no lo tenemos en la tabla de la Normal y podemos deducirlo de la primera expresión: $P(Z \leq -z_\alpha) = 0'05$.

Ejemplo 2

La situación será



En las tablas obtenemos $-z_\alpha = -1'645$
Si $z_{exp} < -1'645$ se rechaza H_0 y si $z_{exp} \geq -1'645$ se acepta. En nuestro caso:

$$z_{exp} = 2'8 \geq -1'645$$

Se acepta la H_0.

Ejemplo 2

La cotización de las acciones de un banco durante el pasado año se ha observado que se distribuye normalmente con desviación típica $\sigma = 800$ unidades monetarias. Se selecciona una muestra aleatoria de la cotización alcanzada en 100 días, obteniendo como cotización media $\bar{x} = 18800$ unidades monetarias y queremos contrastar la hipótesis de que, a la vista del valor que toma la media de la muestra, nosotros creemos que la cotización media de esa acción es $\mu = 19000$ unidades monetarias. Realiza un contraste de hipótesis con nivel de significación $\alpha = 0'05$.

Ejemplo 2

La población será $N(\mu, 800)$ y el contraste:

$$\left.\begin{array}{l} H_0 : \mu = \mu_0 = 19000 \\ H_1 : \mu \neq \mu_0 = 19000 \end{array}\right\}$$

El estadístico de contraste es $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ que toma el valor:

$$Z_{exp} = \frac{18800 - 19000}{800/\sqrt{100}} = -2'5$$

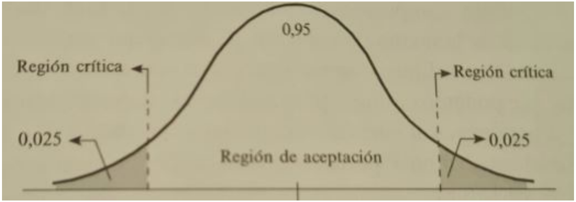
En este caso se acepta H_0 si:

$$-z_{\alpha/2} = -1'96 \leq z_{exp} \leq z_{\alpha/2} = 1'96$$

o bien como dice Adenda:

$$\frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} \leq z_{\alpha/2}$$

Ejemplo 2



Dado que $z_{exp} = -2'5 < -1'96$ se rechaza H_0.

Si se hace con módulos:

$$z_{epx} = 2'5 > 1'96$$

Ejemplo de examen 2014/15 segunda semana

El Rectángulo de Oro es un rectángulo en donde el cociente Q entre el lado menor y el lado mayor es igual a $1/(0'5(\sqrt{5}+1)) = 0'618$ (razón áurea), utilizada por ejemplo por los griegos en el Partenón y hoy en día utilizada en las tarjetas de visita. Se cree que las empresas con tarjetas de visita de valor Q mayor que la razón áurea, son conservadoras en sus negocios, mientras que empresas con tarjetas de visita de valores Q menor que la razón áurea son agresivas. Con esta idea se eligieron al azar cinco compañías del Ibex 35 consideradas como conservadoras obteniéndose los valores Q:

$$0'627, 0'690, 0'620, 0'622, 0'632$$

y cinco compañías consideradas como agresivas en las que se observaron los valores Q siguientes:

$$0'555, 0'601, 0'590, 0'570, 0'617$$

Suponiendo que los valores Q siguen una distribución normal, ¿es significativamente mayor la media de valores Q de las empresas del primer grupo (conservadoras) que las del segundo (agresivas)?

Ejemplo de examen 2014/15 segunda semana

Este problema es similar al **Ejemplo 7.20** del libro.
Lo que nos pide es contrastar las medias de dos poblaciones. Planteemos el contraste. Si consideramos que las empresas del primer grupo (conservadoras) tienen media μ_1 y las del grupo 2 (agresivas) tienen media μ_2, lo que queremos contrastar es:

$$\left.\begin{array}{l} H_0 : \mu_1 \leq \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{array}\right\}$$

Vamos a Adenda y buscamos el contraste que se ajuste al problema planteado. En este caso, el problema indica que son poblaciones normales: "INFERENCIAS SOBRE LAS MEDIAS DE DOS POBLACIONES NORMALES INDEPENDIENTES" (Pág. 18). Vemos que tenemos dos opciones, según si desconocemos las varianzas o no. En nuestro caso las varianzas son desconocidas y nuestras muestras son pequeñas. Por tanto el apartado a considerar será:

Ejemplo de examen 2014/15 segunda semana

σ_1 y σ_2 desconocidas. Muestras pequeñas

(a) $\sigma_1 = \sigma_2$

$$\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{v} \to t_{n_1 + n_2 - 2}$$

$$I = \left[\bar{x}_1 - \bar{x}_2 \mp t_{n_1 + n_2 - 2;\, \alpha/2} \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

Acep.	$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$	$H_0 : \mu_1 \leqslant \mu_2$ $H_1 : \mu_1 > \mu_2$	$H_0 : \mu_1 \geqslant \mu_2$ $H_1 : \mu_1 < \mu_2$
H_0	$\dfrac{\lvert \bar{x}_1 - \bar{x}_2 \rvert}{v} \leqslant t_{n_1 + n_2 - 2;\, \alpha/2}$	$\dfrac{\bar{x}_1 - \bar{x}_2}{v} \leqslant t_{n_1 + n_2 - 2;\, \alpha}$	$\dfrac{\bar{x}_1 - \bar{x}_2}{v} \leqslant t_{n_1 + n_2 - 2;\, 1 - \alpha}$
H_1	$\dfrac{\lvert \bar{x}_1 - \bar{x}_2 \rvert}{v} > t_{n_1 + n_2 - 2;\, \alpha/2}$	$\dfrac{\bar{x}_1 - \bar{x}_2}{v} > t_{n_1 + n_2 - 2;\, \alpha}$	$\dfrac{\bar{x}_1 - \bar{x}_2}{v} < t_{n_1 + n_2 - 2;\, 1 - \alpha}$

(b) $\sigma_1 \neq \sigma_2$

$$\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\cdots} \approx t_f \qquad I = \left[\bar{x}_1 - \bar{x}_2 \mp t_{f;\, \alpha/2} \sqrt{\frac{S_1^2}{\cdots} + \frac{S_2^2}{\cdots}}\right]$$

Ejemplo de examen 2014/15 segunda semana

Debemos pues ver si podemos considerar las varianzas iguales o no. Para ello debemos recurrir de nuevo a Adenda al apartado: INFERENCIAS SOBRE LAS VARIANZAS DE DOS POBLACIONES NORMALES INDEPENDIENTES

Como no conocemos las medias de las poblaciones estamos en el caso:

18 FÓRMULAS Y TABLAS ESTADÍSTICAS

μ_1 y μ_2 desconocidas

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \to F_{n_1-1,\,n_2-1} \qquad I = \left[\frac{S_1^2/S_2^2}{F_{n_1-1,\,n_2-1:\,\alpha/2}},\ \frac{S_1^2/S_2^2}{F_{n_1-1,\,n_2-1:\,1-\alpha/2}}\right]$$

Acep.	$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$	$H_0: \sigma_1^2 \leqslant \sigma_2^2$ $H_1: \sigma_1^2 > \sigma_2^2$	$H_0: \sigma_1^2 \geqslant \sigma_2^2$ $H_1: \sigma_1^2 < \sigma_2^2$
H_0	$\frac{S_1^2}{S_2^2} \in [F_{n_1-1,\,n_2-1:1-\alpha/2},\ F_{n_1-1,\,n_2-1:\alpha/2}]$	$\frac{S_1^2}{S_2^2} \leqslant F_{n_1-1,\,n_2-1:\alpha}$	$\frac{S_1^2}{S_2^2} \geqslant F_{n_1-1,\,n_2-1:1-\alpha}$
H_1	$\frac{S_1^2}{S_2^2} \notin [F_{n_1-1,\,n_2-1:1-\alpha/2},\ F_{n_1-1,\,n_2-1:\alpha/2}]$	$\frac{S_1^2}{S_2^2} > F_{n_1-1,\,n_2-1:\alpha}$	$\frac{S_1^2}{S_2^2} < F_{n_1-1,\,n_2-1:1-\alpha}$

Ejemplo de examen 2014/15 segunda semana

Con lo cual, si

$$\frac{S_1^2}{S_2^2} \in \left[F_{n_1-1,n_2-1;1-\frac{\alpha}{2}} \, , \, F_{n_1-1,n_2-1;\frac{\alpha}{2}}\right]$$

podremos aceptar la igualdad de varianzas.

Calculamos las medias muestrales y las cuasivarianzas muestrales:

$$\left.\begin{array}{l} n_1 = 5, \; \bar{x}_1 = 0'6382, \; S_1^2 = 0'0008602 \\ n_2 = 5, \; \bar{x}_1 = 0'5866, \; S_1^2 = 0'0006043 \end{array}\right\} \Rightarrow \frac{S_1^2}{S_2^2} = 1'423465$$

En este caso desconocemos el nivel de significación. Como en nuestra tabla de Adenda el valor que producirá el intervalo más pequeño és $0'1$, vamos a ver cuál sería nuestro intervalo de aceptación con $\frac{\alpha}{2} = 0'1 \rightarrow \alpha = 0'2$:

$$\left[F_{4,4;1-0'1} \, , \, F_{4,4;0'1}\right] = \left[\frac{1}{4'1073}, 4'1073\right] = [0'24347, 4'1073]$$

donde hemos utilizado que $F_{n,m;\beta} = \frac{1}{F_{m,n;1-\beta}}$.

Ejemplo de examen 2014/15 segunda semana

Si considerásemos $\alpha = 0,1 \Rightarrow \frac{\alpha}{2} = 0'05$ obtendríamos:

$$\left[F_{4,4;0'95}\ ,\ F_{4,4;0'05}\right] = \left[\frac{1}{6'3883}, 6'3883\right] = [0'1565, 6'3883]$$

es decir un intervalo más amplio.
Dado que $\frac{S_1^2}{S_2^2} = 1'423465$ cae dentro del intervalo de aceptación, podemos asumir que las varianzas son iguales.
Si nos preguntásemos a partir de qué valor de α aceptamos la hipótesis nula de igualdad de varianzas obtendríamos el $p - valor$ será $0'7405646$, lo cual justifica aceptación de igualdad de varianzas para valores elevados de α.

$p - valor$

El anterior valor, se ha obtenido con R:
2*(1-pf(1.423465,4,4))=0'7405646. Mediante las tablas que disponemos únicamente podemos deducir que $p - valor > 0'2$

Ejemplo de examen 2014/15 segunda semana

Volvamos pues al contraste que nos ocupa:

$$\left.\begin{array}{l} H_0 : \mu_1 \leq \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{array}\right\}$$

Se acepta H_0 si

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \leq t_{n_1+n_2-2;\alpha}$$

Haciendo los cálculos:

$$\frac{0'6382 - 0'5866}{\sqrt{\frac{(5-1)0'0008602^2+(5-1)0'0006063^2}{5+5-2}}\sqrt{\frac{1}{5}+\frac{1}{5}}} = 3'015$$

Ejemplo de examen 2014/15 segunda semana

Una vez más el problema no nos informa acerca del nivel de significación, con lo cual debemos recurrir al $p - valor$. A partir de las tablas observamos que

$$P\{t_8 > 2'896\} = 0'01$$
$$P\{t_8 > 3'449\} = 0'005$$

es decir, que dado que $2'896 < 3'015 < 3'449$, entonces

$$0,005 < p - valor < 0,01$$

Como el $p - valor$ es pequeño, podemos rechazar la hipótesis nula y quedarnos con la alternativa, es decir, que la media de los valores de Q es significativamente mayor en las empresas conservadoras.

Ejemplo de examen 2014/15 segunda semana con R

Si quisiésemos resolver esto último con R utilizaríamos:

```
> t.test(x,y,alternative="greater",var.equal=T)

        Two Sample t-test

data:  x and y
t = 3.015, df = 8, p-value = 0.008343
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.0197751       Inf
sample estimates:
mean of x mean of y
   0.6382    0.5866
```

Ejemplo de examen 2014/15 segunda semana

Si quisiésemos además comprobar la igualdad de varianzas con R deberíamos comprobar si 1 pertenece al intervalo de aceptación:

$$\left[\frac{S_1^2/S_2^2}{F_{n_1-1,n_2-1;\frac{\alpha}{2}}}, \frac{S_1^2/S_2^2}{F_{n_1-1,n_2-1;1-\frac{\alpha}{2}}}\right]$$

```
> var.test(x,y,ratio=1,alternative="two.sided",conf.level=0.8)

	F test to compare two variances

data:  x and y
F = 1.4235, num df = 4, denom df = 4, p-value = 0.7406
alternative hypothesis: true ratio of variances is not equal to 1
80 percent confidence interval:
 0.3465738 5.8465267
sample estimates:
ratio of variances
          1.423465
```

Vemos que 1 sí pertenece al intervalo de aceptación. El $p-valor$ es el mismo que en nuestra versión, al igual que el valor de $F = \frac{S_1^2}{S_2^2}$. Los extremos del intervalo de aceptación también pueden deducirse de nuestros resultados 'a mano': $[0'24347, 4'1073]$

Bibliografía

Algunos ejemplos aparecen en el libro ”Estadística Económica y Empresarial”de Casas, Vázquez, Zamora. Ed. Ramón Areces.