Ejercicios Tema 4 - Contraste hipótesis. Taller 3

Ricardo Alberich, Juan Gabriel Gomila y Arnau Mir

Curso completo de estadística inferencial con R y Python

Contenidos

1 Contraste hipótesis taller 3.

1.1 Librerías y datos necesarios

Para este taller necesitaremos los siguientes paquetes: `faraway, nortest, car` si no los tenéis instalados podéis ejecutar lo siguiente:

```
install.packages("faraway")
install.packages("nortest")
install.packages("car")
```

Para utilizarlos, deberéis cargarlos ejecutando las siguientes instrucciones:

```r
library("faraway")
library("nortest")
library("car")
```

2 Contrastes de dos parámetros.

Comparación de medias.

Los siguientes problemas tratan de contrastes de parámetros entre dos muestras Para cada uno de los enunciados:

1. Contrastar contra una de los dos tipos de hipótesis unilaterales o bilaterales.
2. Calcular también el intervalo de confianza para la diferencia o el cociente de los parámetros según el contraste sea bilateral o unilateral. Tomar finalmente la decisión más correcta.
3. Calcular todos los test e intervalos de confianza para $\alpha = 0.05$.
4. Calcular el p-valor en cada caso.

2.1 Ejercicio 1

Para comparar la producción media de dos procedimientos de fabricación de cierto producto se toman dos muestras, una con la cantidad producida durante 25 días con el primer método y otra con la cantidad producida durante 16 días con el segundo método. Por experiencia se sabe que la varianza del primer procedimiento es $\sigma_1^2 = 12$ y al del segundo $\sigma_2^2 = 10$. De las muestras obtenemos que $\overline{X}_1 = 136$ para el primer procedimiento y $\overline{X}_2 = 128$ para el segundo. Si μ_1 y μ_2 son los valores esperados para cada uno de los procedimientos, calcular un intervalo de confianza para $\mu_1 - \mu_2$ al nivel 99%. sol1$\{(5.2989, 10.7011)\}$

2.1.1 Solución

Estamos en el caso de un contraste de comparación de medias de muestras independientes y en el teórico caso de que las varianzas son conocidas.

Contrastaremos

$$\left\{\begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu \neq \mu 2 \end{array}\right.$$

El estadístico de contraste es

El **estadístico de contraste** toma el valor:

$$z_0 = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{136 - 128}{\sqrt{\frac{3.4641016^2}{25} + \frac{3.1622777^2}{16}}} = 7.61.$$

El estadístico sigue, aproximadamente, una distribución normal. La región crítica del contraste al nivel se significación $\alpha = 0.05$ es rechazar H_0 si $z_0 < z_{\alpha/2}$ o $z_0 > z_{1\alpha/2}$. Con nuestros datos

$$z_0 = 7.61 \not< z_{\alpha/2} = z_{0.05} = -1.959964 \text{ o } z_0 = 7.61\ z_{1-\alpha/2} = 1.959964$$

o lo que es lo mismo rechazamos H_0 si $|z_0| = 7.61 > z_{1-\alpha/2} = 1.959964$ lo que en este caso es cierto.

Los cuantiles los hemos calculado con

```
alpha=0.05
qnorm(alpha/2) # o tambien -qnorm(1-alpha/2)
```

```
## [1] -1.959964
```

```
qnorm(1-alpha/2)
```

```
## [1] 1.959964
```

El p-valor de este contraste para la alternativa bilateral es p-valor$=2 \cdot P(Z > |z_0|)$ donde Z es una normal estándar $N(\mu = 0, \sigma = 1)$. Podemos calcularlo con el código

En nuestro caso

$$2 \cdot P(Z > |z_0|) = 2 \cdot P(Z > |7.61|) = P(Z > 7.61) = 2 \cdot (1 - P(Z \leq 7.61)),$$

lo podemos calcular con R

```
z0
```

```
## [1] 7.61
```

```
2*(1-pnorm(abs(z0)))
```

```
## [1] 0.00000000000002731149
```

Es un p-valor muy pequeño lo que confirma que hay evidencias para rechazar las hipótesis nula: el rendimiento de los dos métodos de fabricación no tiene la misma media.

2.2 Ejercicio 2

Estamos interesados en comparar la vida media, expresada en horas de dos tipos de componentes electrónicos. Para ello se toma una muestra de cada tipo y se obtiene:

Tipo	tamaño	$\overline{x}$	$\tilde{s}$
1	50	1260	20
2	100	1240	18

Calcular un intervalo de confianza para $\mu_1 - \mu_2$ (μ_1 esperanza del primer grupo y μ_2 esperanza del segundo grupo) al nivel 98% Suponer si es necesario las poblaciones aproximadamente normales. % sol1$\{(12.19, 27.81)\}$

2.2.1 Solución

En este caso tenemos dos muestras independientes de tamaños $n1 = 50$, $n2 = 100$ y estadísticos $\overline{x}_1 = 1260$, $\overline{x}_2 = 1240$, $\tilde{s}_1 = 20$ y $\tilde{s}_2 = 18$.

Cargamos los datos en R

```
n1=50
n2=120
media1=1260
media2=1240
desv_tipica1=20
desv_tipica2=18
f0=desv_tipica1^2/desv_tipica2^2 # estadístico de contraste
f0
```

```
## [1] 1.234568
```

Tenemos pues, dos muestras independientes de tamaños muestrales y las varianzas desconocidas. Haremos un t-test pero tenemos dos casos: varianzas desconocidas iguales y varianzas desconocidas distintas. Primero haremos un test para saber si las varianzas son iguales o distintas.

El contraste es

$$\left\{\begin{array}{l} H_0 : \sigma_1 = \sigma_2 \\ H_1 : \sigma_1 \neq \sigma_2 \end{array}\right.$$

Se emplea el siguiente **estadístico de contraste**:

$$F = \frac{\widetilde{S}_1^2}{\widetilde{S}_2^2}$$

que, si las dos poblaciones son normales y la hipótesis nula $H_0 : \sigma_1 = \sigma_2$ es cierta, tiene distribución F de Fisher con grados de libertad $n_1 - 1$ y $n_2 - 1$.

En nuestro caso $f_0 = \frac{\tilde{s}_1^2}{\tilde{s}_1^2} = \frac{20^2}{18^2} = 1$

Resolveremos calculando el p-valor del contraste que este caso es

$$\begin{array}{l} \min\{2 \cdot P(F_{n_1-1,n_2-1} \leq f_0), 2 \cdot P(F_{n_1-1,n_2-1} \geq f_0)\} \\ = \min\{2 \cdot P(F_{50-1,100-1} \leq 1.2345679), 2 \cdot P(F_{50-1,100-1} \geq 1.2345679)\} \end{array}$$

calcularemos el p-valor con R

```
n1=50
n2=100
desv_tipica1=20
desv_tipica2=18
f0=desv_tipica1^2/desv_tipica2^2 # estadístico de contraste
f0
```

```
## [1] 1.234568
```

```
2*pf(f0,n1-1,n2-1,lower.tail = TRUE)# lower.tail = TRUE es el valor por defecto
```

```
## [1] 1.625115
```

```
2*pf(f0,n1-1,n2-1,lower.tail = FALSE)# o tambien 2*(1-pf(f0,n1-1,n2-1)
```

```
## [1] 0.3748846
```

```
pvalor=min(2*pf(f0,n1-1,n2-1,lower.tail = TRUE),2*pf(f0,n1-1,n2-1,lower.tail = FALSE))
pvalor
```

```
## [1] 0.3748846
```

El p-valor es alto así que no podemos rechazar la hipótesis nula; consideraremos las varianzas iguales.

Así pues vamos a contrastar la igualdad de medias contra que son distintas:

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

El estadístico de contraste sigue una ley t de Student con $n_1 + n_2 - 2$ grados de libertad su fórmula es

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \cdot \frac{((n_1-1)\widetilde{S}_1^2 + (n_2-1)\widetilde{S}_2^2)}{(n_1+n_2-2)}}},$$

en nuestro caso vale

$$t_0 = \frac{1260 - 1240}{\sqrt{\left(\frac{1}{50} + \frac{1}{100}\right) \cdot \frac{((50-1)\cdot 20^2 + (100-1)\cdot 18^2)}{(n_1+n_2-2)}}} = 7.7263624$$

con R es

```
n1=50
n2=120
media1=1260
media2=1240
desv_tipica1=20
desv_tipica2=18
t0= (media1-media2)/(sqrt((1/n1+1/n2)*((n1-1)*desv_tipica1+(n2-1)*desv_tipica2^2)/(n1+n2-2)))
t0
```

```
## [1] 7.745321
```

el p-valor para la alternativa bilateral es $2 \cdot P(t_{N1+n_2-2} > |t_0|) = 2 \cdot (1 - P(t_{N1+n_2-2} > |t_0|))$, con R es

```
t0
```

```
## [1] 7.745321
```

```
abs(t0)
```

```
## [1] 7.745321
```

```
n1
```

```
## [1] 50
```

```
n2
```

```
## [1] 120
```

```
pvalor= 2*(1-pt(abs(t0),df=n1+n2-2))
pvalor
```

```
## [1] 0.0000000000008566481
```

EL p-valor es extremadamente pequeño hay evidencias en contra de la hipótesis nula de igualdad de medias contra la hipótesis alternativa de que son distintas.

El intervalo de confianza al nivel $1-\alpha=0.95$ es

$$\left(\overline{x}_1-\overline{x}_2-t_{n_1+n_2-2,1-\frac{\alpha}{2}}\cdot\frac{(n_1-1)\cdot\widetilde{S}_1^2+(n_2-1)\cdot\widetilde{S}_2^2)}{(n_1+n_2-2)},\overline{x}_1-\overline{x}_2+t_{n_1+n_2-2,1-\frac{\alpha}{2}}\cdot\frac{((n_1-1)\cdot\widetilde{S}_1^2+(n_2-1)\cdot\widetilde{S}_2^2)}{(n_1+n_2-2)}\right)$$

lo calculamos con R

```
n1=50
n2=120
media1=1260
media2=1240
desv_tipica1=20
desv_tipica2=18
alpha=0.05
qt(1-alpha/2,df=n1+n2-2)
```

```
## [1] 1.974185
```

```
IC=c(media1-media2- qt(1-alpha/2,df=n1+n2-2)
     *sqrt((1/n1+1/n2)*((n1-1)*desv_tipica1+(n2-1)*desv_tipica2^2)/(n1+n2-2)),
     media1-media2+ qt(1-alpha/2,df=n1+n2-2)
     *sqrt((1/n1+1/n2)*((n1-1)*desv_tipica1+(n2-1)*desv_tipica2^2)/(n1+n2-2)))
IC
```

```
## [1] 14.90225 25.09775
```

La diferencia de medias $\mu_1-\mu_2$ verdadera se encontrará en el intervalo $(14.902251, 25.097749)$ al nivel de confianza del 95%. La $media_1$ es claramente más grande que la dos, en al menos 14

2.3 Ejercicio 3

Para reducir la concentración de ácido úrico en la sangre se prueban dos drogas. La primera se aplica a un grupo de 8 pacientes y la segunda a un grupo de 10. Las disminuciones observadas en las concentraciones de ácido úrico de los distintos pacientes expresadas en tantos por cien de concentración después de aplicado el tratamiento son:

droga 1	20	12	16	18	13	22	15	20		
droga 2	17	14	12	10	15	13	9	19	20	11

Calcular un intervalo de confianza para la diferencia de medias entre la primera droga y la segunda al nivel del 99%.

Suponer que las reducciones de ácido úrico siguen una distribución normal son independientes y de igual varianza. Ídem pero suponiendo que las varianza son distintas.%sol1$\{(-2.09, 8.09)\}$

2.4 Ejercicio 4

Para comparar la dureza media de dos tipos de aleaciones (tipo 1 y tipo 2) se hacen 5 pruebas de dureza con la de tipo 1 y 7 con la de tipo 2. Obteniéndose los resultados siguientes:

$$\overline{X}_1 = 18.2, \quad S_1 = 0.2 \text{ y}$$

$$\overline{X}_2 = 17.8; \quad S_2 = 0.5$$

Suponer que la población de las durezas es normal y que las desviaciones típicas no son iguales. Hacer lo mismo si las varianzas son distintas. % , buscar un intervalo de confianza para $\mu_1 - \mu_2$ con probabilidad 0.95 % sol1$\{(0.314, 0.486)\}$

2.5 Ejercicio 5

Se encuestó a dos muestras independientes de internautas, una en Menorca y otra en Mallorca, sobre si utilizaban telefonía por Internet. La encuesta de Menorca tuvo un tamaño $n_1 = 500$ y 100 usuarios mientras que en Mallorca se encuestaron a $n_2 = 750$ y se obtuvo un resultado de 138 usuarios.

2.6 Ejercicio 6

Se pregunta a un grupo de 100 personas elegido al azar asiste a una conferencia sobre tecnologías de la comunicación. Antes de la conferencia se les pregunta si consideran a Internet peligrosa, después de la conferencia se les vuelve a preguntar cual es su opinión. Los resultados fueron los siguientes:

		Después	
		Sí es peligrosa	No es peligrosa
Antes	Sí es peligrosa	50	30
	No es peligrosa	5	15

2.7 Ejercicio 7

Tenemos 10 ordenadores, deseamos optimizar su funcionamiento. Con este fin se piensa en ampliar su memoria. Se les pasa una prueba de rendimiento antes y después de ampliar la memoria. Los resultados fueron:

{

	Ordenador									
Muestra/Tiempo	1	2	3	4	5	6	7	8	9	10
Antes ampliación	98.70	100.48	103.75	114.41	97.82	91.13	85.42	96.8	107.76	112.94
Después ampliación	99.51	114.44	108.74	97.92	103.54	104.75	109.69	90.8	110.04	110.09

}

2.8 Ejercicio 8

Las siguientes muestras provienen de dos poblaciones independientes y supuestamente normales. Se desea comparar la igualdad de sus medias, pero antes debemos contrastar si podemos o no aceptar que sus varianza son iguales o distintas. Se pide hacer el contraste de las medias en el caso en que se se decida aceptar varianzas iguales o distintas al nivel de significación $\alpha = 0.05$.

Contrastar también la hipótesis de igualdad de medias en el otro caso (es decir si se decide varianzas distintas contrastar para iguales y viceversa).