UNIVERSIDAD COMPLUTENSE MADRID

Máster EMOS

2022

Entrega IESIM

"Práctica Intervalos de Confianza y Contrastes de Hipótesis"

Elsa Bueno Alonso y María Alejandra Ponce López

Profesora

Alba María Franco Pereira

12 de DICIEMBRE de 2022

Ejercicio 2 *Una empresa de estudios de mercado está interesada en conocer la proporción de hogares en los que se verá un determinado acontecimiento deportivo. Sabiendo que la proporción muestral es* $\hat{p} = 0{,}5$*, ¿qué muestra de hogares se deberá tomar si se desea que su estimación tenga un márgen de error de* $\pm 0{,}02$*, con un nivel de confianza del* $95{,}5\,\%$*?*

Resultado Como el nivel de confianza es 95.5 %, eso significa que $1 - \alpha = 0{,}955$, por tanto $\alpha = 0{,}045$, entonces $\frac{\alpha}{2} = 0{,}0225$. Luego $Z_{\frac{\alpha}{2}}$ es:

```
> qnorm(0.0225)

[1] -2.004654
```

Como R lee cuantiles a la derecha, en lugar de IC que usualmente se leen a la izquierda, el valor de $Z_{\frac{\alpha}{2}}$ es de

$$Z_{\frac{\alpha}{2}} = 2{,}005$$

El intervalo de confianza para una proporción p es:

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

Por tanto, si queremos calcular el tamaño muestral para que la estimación tenga un margen de error de $\pm 0{,}02$, entonces se deberá calcular lo siguiente:

$$2{,}005 * \sqrt{\frac{0{,}5 * 0{,}5}{n}} < 0{,}02$$

$$\sqrt{\frac{0{,}5 * 0{,}5}{n}} > \frac{0{,}02}{2{,}005}$$

$$n > \frac{0{,}5 * 0{,}5}{\left(\frac{0{,}02}{2{,}005}\right)^2}$$

```
> n=(0.5*0.5)/(0.02/qnorm(0.0225))^2
> n

[1] 2511.65
```

Solución Se necesita obtener un tamaño muestral de $n = 2512$ para garantizar que, con un margen de error de $\pm 0{,}02$, un $50\,\%$ de hogares vean un acontecimiento deportivo.

__Ejercicio 3__ Se recibe un envío de latas de conserva de las que se afirma que el peso medio es de 1000gr. Examinada una muestra de 5 latas, se obtiene un peso medio de 995gr. Con una cuasivarianza de $S^2 = 19{,}6$ y al nivel de confianza del 95 %, ¿se puede aceptar que el peso medio es de 1000gr?

Resultado Como el nivel de confianza es 95 %, eso significa que $1 - \alpha = 0{,}95$, por tanto $\alpha = 0{,}05$, entonces $\frac{\alpha}{2} = 0{,}025$

El IC para media μ y varianza desconocida σ^2 es:

$$\overline{X} \pm t_{n-1,\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$$

Donde $t_{n-1,\frac{\alpha}{2}} = t_{5,0{,}025}$

```
> qt(0.025,4)

[1] -2.776445
```

Por tanto, $t_{4,0{,}025} = 2{,}78$

Como nos dan la cuasivarianza muestral, la $S = \sqrt{S^2}$:

```
> sqrt(19.6)

[1] 4.427189
```

Y, $S = \sqrt{S} = 4{,}43$. El IC queda de la siguiente forma.

$$995 \pm 2{,}776 \frac{4{,}43}{\sqrt{(5)}}$$

```
> IC<-c(995+qt(0.025,4)*(4.43/sqrt(5)),995-qt(0.025,4)*(4.43/sqrt(5)))
> IC

[1]  989.4994 1000.5006
```

Solución Como en el intervalo de confianza encontramos la media poblacional, se puede afirmar a un nivel de confianza del 95 % que el peso medio es de 1000gr.

Ejercicio 4 *Un dentista afirma que el 40 % de los niños de 10 años presentan indicios de caries dental. Tomada una muestra de 100 niños, se observó que 36 presentaban indicios de caries. Contrastar la hipótesis del dentista para un nivel de confianza del 90 %.*

Resultado La muestra sigue una distribución binomial donde $Bin(100.p)$. Dados los datos, sabemos que $m = 100$ y podemos obtener el valor $\hat{p}$ para la proporción muestral:

$$\hat{p} = \frac{n}{m} = \frac{36}{100}$$

```
> 36/100
```

```
[1] 0.36
```

Y por tanto, $\hat{p} = 0{,}36$.

Para contrastar la hipótesis del dentista, hacemos un contraste de hipótesis de proporción p:

$$H_0 : p = p_0 = 0{,}4$$

$$H_1 : p \neq p_0 = 0{,}4$$

El estadístico Z es:

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

```
> Z=(0.36-0.40)/sqrt((0.36*(1-0.36))/100)
> Z
```

```
[1] -0.8333333
```

Y la región crítica es:

$$\{z : |z| > z_{\frac{\alpha}{2}}\} = \{z : 0{,}83 > z_{0{,}05}\}$$

```
> qnorm(0.05)
```

```
[1] -1.644854
```

Por tanto, $0{,}833 \not> 1{,}645$.

Solución Como el estadístico se encuentra en la región de aceptación, no tenemos razones suficientes para rechazar la hipótesis nula, es decir, podemos afirmar que el 40 % de los niños de 10 años tienen indicios de caries.

Ejercicio 5 *La concentración media de CO_2 en el aire en una zona no es habitualmente mayor a 355 p.p.m.v. Se sospecha que esta concentración es mayor en la capa de aire más próxima a la superficie. Para contrastar esta hipótesis, se analiza el aire en 20 puntos elegidos al azar en una misma altura próxima al suelo. De los cálculos resulta una* $\overline{X} = 580 p.p.m.v$ *y una* $S_c = 180$. *Suponiendo que las mediciones siguen aproximadamente una normal, ¿proporcionan los datos suficiente evidencia estadística al nivel de significación de 0.01 a favor de la hipótesis de que la concentración es mayor cerca del suelo? Indicar razonadamente si el p-valor es mayor que 0.01.*

Resultado El contraste de hipótesis unilateral para la media poblacional con varianza desconocida es:

$$H_0 : \mu = 335$$

$$H_1 : \mu > 335$$

El estadístico T es:

$$T = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

```
> T=(580-355)/(180/sqrt(20))
> T

[1] 5.59017
```

Y la región crítica es:

$$\{t : T > t_{n-1,\alpha}\} = \{t : 75 > t_{19,0,01}\}$$

```
> qt(0.01,19)

[1] -2.539483
```

Por tanto, $t_{n-1,\alpha} = t_{19,0,01} = 2{,}5394$

Solución Como el estadístico se encuentra en la región de rechazo, tenemos evidencias suficientes para rechazar H_0. Podemos afirmar que, cerca del suelo, la concentración media de CO_2 es $> 355 p.p.m.v.$

El p-valor es menor que el nivel de significación. Se rechaza H_0 si el p-valor asociado al resultado observado es igual o menor que el nivel de significación: $p - valor \leq \alpha$

Ejercicio 6 *Se desea comparar la proporción de viviendas de Extremadura y en Galicia. Se hace un muestreo en las dos comunidades con los siguientes resultados:*

- *Extremadura: de 500 viviendas elegidas al azar, 300 disponían de calefacción.*
- *Galicia: de 1000 viviendas elegidas al azar, 680 disponían de calefacción.*

¿Hay suficiente evidencia estadística para un nivel de confianza del 95 % que es menor la proporción de viviendas con calefacción en Extremadura que en Galicia?

Resultado Primero obtenemos las proporciones muestrales de ambos sitios:

$$\hat{p}_1 = \frac{300}{500} \quad \hat{p}_2 = \frac{680}{1000}$$

```
> p1<-300/500
> p1

[1] 0.6

> p2<-680/1000
> p2

[1] 0.68
```

El contraste de hipótesis unilateral para igualdad de proporciones p es:

$$H_0 : p_1 = p_1$$

$$H_1 : p_1 < p_2$$

El estadístico Z es:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}}$$

```
> Z=(p1-p2)/sqrt(((p1*(1-p1))/500)+((p2*(1-p2))/1000))
> Z

[1] -3.028913
```

Y la región crítica es:

$$\{z : Z > z_\alpha\} = \{z : -3{,}03 < z_{0,05}\}$$

```
> qnorm(0.05)
```

```
[1] -1.644854
```

Por tanto, $-3{,}03 < -1{,}645$.

Solución Como el estadístico se encuentra en la región de rechazo, hay evidencia suficiente para RH_0, por lo que podemos determinar que en Extremadura hay una menor proporción de hogares con calefacción que en Galicia.

Ejercicio 7 *La duración media de una muestra de bombillas es* $\overline{X} = 1250h$ *con una cuasidesviación tipica muestral* $S_x = 115$. *Se cambia el filamento por uno nuevo y de una muestra de 12 bombillas se obtuvo una duración media de* $\overline{Y} = 1343h$, *con una* $S_Y = 106$. *¿Puede aceptarse que las varianzas, antes y después del cambio de filamento son iguales?¿Bajo que hipótesis?¿Ha aumentado la dirección media de la bombilla?*

Resultado En primer lugar, bajo hipótesis de normalidad, debemos determinar si las varianzas de ambas situaciones son iguales o no, para lo que vamos a utilizar un contraste de hipótesis para la igualdad de varianzas. Este contraste es:

$$H_0 : \sigma_1 = \sigma_2$$

$$H_1 : \sigma_1 \neq \sigma_2$$

El estadístico F es:

$$F = \frac{S_x^2}{S_y^2}$$

```
> s_x<-115
> s_y<-106
> F=((s_x)^2)/((s_y)^2)
> F

[1] 1.17702
```

Y la región crítica es:

$$\{F : F < F_{n-1,m-1;1-\frac{\alpha}{2}}; F > F_{n-1,m-1;\frac{\alpha}{2}}\}$$

Como es un contraste bilateral y la distribución de F de Snedecor no es simétrica, necesitamos conocer ambos valores críticos:

```
> n_x<-10
> m_y<-12
> alpha<-0.05
> v_critico_1=qf(alpha/2,n_x-1,m_y-1 ,lower.tail = T)
> v_critico_1
```

```
[1] 0.2556189

> v_critico_2= qf(1-(alpha/2),n_x-1,m_y-1 ,lower.tail = T)
> v_critico_2

[1] 3.587899
```

Por tanto, 0,2556 < 1,17702 < 3,587.

Solución Como el estadístico se encuentra en la región de aceptación, no tenemos razones suficientes para rechazar la hipótesis nula, por lo que determinamos que la varianza de la duración de las bombillas es la misma antes y después del cambio de filamento.

Una vez determinada la igualdad de varianzas, vamos a determinar si la duración media de la bombilla se ha incrementado como consecuencia del cambio de filamento. Para ello, vamos a realizar otro contraste de hipótesis, aunque este es un contraste unilateral para la igualdad de medias con varianzas iguales (demostrado en el apartado anterior) y desconocidas. Este contraste es el siguiente:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 < \mu_2$$

El estadístico T es:

$$T = \frac{\overline{X} - \overline{Y}}{s_c * \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

Donde S_c es:

$$S_c = \sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}}$$

```
> media_x<-1250
> media_y<-1343
> s_x<-115
> s_y<-106
> n_x=10
```

```
> m_y=12
> S_c=sqrt(((n_x-1)*((s_x)^2)+(m_y-1)*((s_y)^2))/(n_x+m_y-2))
> T=(media_x-media_y)/(S_c*sqrt((1/n_x)+(1/m_y)))
> T

[1] -1.972028
```

Y la región crítica es:

$$\{T : T < t_{n+m-2;\alpha}\}$$

Como es un contraste unilateral, solo necesitamos saber el cuantil que deja α por debajo:

```
> n_x=10
> m_y=12
> alpha=0.05
> v_critico= qt((alpha),n_x+m_y-2,lower.tail = T)
> v_critico

[1] -1.724718
```

Por tanto, $-1{,}972 < -1{,}7247$.

Solución Como el estadístico se encuentra en la región de rechazo, tenemos razones suficientes para rechazar la hipótesis nula. Por lo que determinamos que el número medio de horas de duración de las bombillas se ha incrementado tras el cambio de filamento.

Ejercicio 8 *Un médico ha medido la tasa de hipotiroidismo congénito en una muestra de la población infantil española, obteniendo un valor de un caso por cada 2800 nacidos. Determinar el tamaño mínimo muestral para que el error relativo en la estimación sea inferior a 0.001 con un nivel de confianza del 95 %*

Resultado Como el nivel de confianza es 95 %, eso significa que $1 - \alpha = 0{,}95$, por tanto $\alpha = 0{,}05$, entonces $\frac{\alpha}{2} = 0{,}025$. Luego $Z_{\frac{\alpha}{2}}$ es:

```
> qnorm(0.025)
```

```
[1] -1.959964
```

Como R lee cuantiles a la derecha, en lugar de IC que usualmente se leen a la izquierda, el valor de $Z_{\frac{\alpha}{2}}$ es de

$$Z_{\frac{\alpha}{2}} = 1{,}96$$

El intervalo de confianza para una proporción p es:

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

Por tanto, si queremos calcular el tamaño muestral para que la estimación tenga un margen de error de $\pm 0{,}001$, entonces se deberá calcular lo siguiente:

$$1{,}96 * \sqrt{\frac{(\frac{1}{2800}) * (1 - \frac{1}{2800})}{n}} < 0{,}001$$

$$\sqrt{\frac{(\frac{1}{2800}) * (1 - \frac{1}{2800})}{n}} > \frac{0{,}001}{1{,}96}$$

$$n > \frac{\frac{1}{2800} * (1 - \frac{1}{2800})}{\left(\frac{0{,}001}{1{,}96}\right)^2}$$

```
> p=1/2800
> n=(p*(1-p))/(0.001/qnorm(0.025))^2
> n
```

```
[1] 1371.46
```

Solución Se necesita obtener un tamaño muestral de $n = 1372$ para garantizar que el margen de error sea como minimo de $\pm 0{,}001$.

Ejercicio 9 *A las personas que sufren de tensión alta se les recomienda seguir una dieta libre de sal. Queremos realizar un estudio para comprobar si esta dieta es efectivamente ventajosa. Para el estudio se escogió una muestra de personas y se tomó la tensión antes de la dieta y dos semanas después. Los resultados obtenidos fueron:*

Antes	93	106	87	92	102	95	88	110
Después	92	102	89	92	101	96	88	105

Suoponiendo que la tensión antes y después sigue una distribución normal, calcular un intervalo de confianza 0.95 para la diferencia de medias.

Resultado Debido a que para el estudio se utilizaron los mismos sujetos, se trata de datos interdependientes y por tanto el IC para la diferencia de medias será para datos apareados:

$$\overline{d} \pm t_{n-1;\frac{\alpha}{2}} \frac{S_d}{\sqrt{n}}$$

Donde:

$$\overline{d} = \frac{1}{n}\sum_{i=1}^{n} D_i \quad S_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{d})^2}$$

```
> antes=c(93, 106, 87, 92, 102, 95, 88, 110)
> despues=c(92, 102, 89, 92, 101, 96, 88, 105)
> d_i=antes - despues
> sumd<-sum(d_i)
> n<-8
> dmean=(1/n)*(sumd)
> S2_d=(1/(n-1))*sum((d_i-dmean)^2)
> S_d=sqrt(S2_d)
> IC<-c(dmean-qnorm(0.975,7)*(S_d/sqrt(8)),dmean+qnorm(0.975,7)*(S_d/sqrt(8)))
> IC

[1] -6.572552  8.572552
```

Solución Podemos afirmar que, debido a que el IC contempla el cero, la medias son iguales y por tanto no se puede afirmar que la dieta libre de sal sea ventajosa.

Ejercicio 10 *Se analizan dos catalizadores para determinar la forma en la que afectan el rendimiento promedio de un proceso químico. De manera específica, el catalizador 1 es el que se está empleando en este momento pero el catalizador 2 también es aceptable. Como el catalizador 2 es más caro, solo interesará emplearlo siempre que aumente el rendimiento promedio del proceso. Se hace una prueba piloto, y los rendimientos obtenidos en porcentaje son los siguientes:*

Catalizador 1	91.50	94.18	92.18	95.39	91.79	89.07	94.72	89.21
Catalizador 2	89.19	90.95	90.46	93.21	97.19	97.04	91.07	92.75

1. *Contrastar al 10 % de significación si la variabilidad de ambos catalizadores se puede considerar la misma.*

   Resultado En primer lugar, debemos determinar si la variabilidad de ambos catalizadores es la misma o no, para lo que vamos a utilizar un contraste de hipótesis para la igualdad de varianzas. Este contraste es:

$$H_0 : \sigma_1 = \sigma_2$$

$$H_1 : \sigma_1 \neq \sigma_2$$

   El estadístico F es:

$$F = \frac{S_x^2}{S_y^2}$$

```
> cataliz1=c(91.50, 94.18,92.18,95.39,91.79,89.07,94.72,89.21)
> cataliz2=c(89.19, 90.95, 90.46, 93.21, 97.19, 97.04, 91.07, 92.75)
> s_x<-var(cataliz1)
> s_y<-var(cataliz2)
> F=(s_x)/(s_y)
> F

[1] 0.6390651
```

   Y la región crítica es:

$$\{F : F < F_{n-1,m-1;1-\frac{\alpha}{2}}; F > F_{n-1,m-1;\frac{\alpha}{2}}\}$$

Como es un contraste bilateral y la distribución de F de Snedecor no es simétrica, necesitamos conocer ambos valores críticos:

```
> n_x<-8
> m_y<-8
> alpha=0.1
> v_critico_1= qf(alpha/2,n_x-1,m_y-1 ,lower.tail = T)
> v_critico_1

[1] 0.2640582

> v_critico_2= qf(1-(alpha/2),n_x-1,m_y-1 ,lower.tail = T)
> v_critico_2

[1] 3.787044
```

Por tanto, 0,264 < 0,639 < 3,787.

Solución Como el estadístico se encuentra en la región de aceptación, no tenemos razones suficientes para rechazar la hipótesis nula, por lo que determinamos que la variabilidad de ambos catalizadores es la misma.

2. *Contrastar si al 5 % interesa emplear el catalizador 2.*

Resultado Una vez determinada la igualdad de varianzas, vamos a determinar si es recomendable el uso del catalizador 2. Esto se dará cuando el rendimiento medio del proceso químico usando este nuevo catalizador sea superior al obtenido mediante el catalizador 1. Para ello, vamos a realizar otro contraste de hipótesis, aunque este es un contraste unilateral para la igualdad de medias con varianzas iguales (demostrado en el apartado anterior) y desconocidas. Este contraste es el siguiente:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 < \mu_2$$

El estadístico T es:

$$T = \frac{\overline{X} - \overline{Y}}{s_c \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

Donde S_c es:

$$S_c = \sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}}$$

```
> cataliz1=c(91.50, 94.18,92.18,95.39,91.79,89.07,94.72,89.21)
> cataliz2=c(89.19, 90.95, 90.46, 93.21, 97.19, 97.04, 91.07, 92.75)
> media_x<-mean(cataliz1)
> media_y<-mean(cataliz2)
> s_x<-var(cataliz1)
> s_y<-var(cataliz2)
> n_x=8
> m_y=8
> S_c=sqrt((((n_x-1)*s_x)+((m_y-1)*s_y))/(n_x+m_y-2))
> T=(media_x-media_y)/(S_c*sqrt((1/n_x)+(1/m_y)))
> T

[1] -0.3535909
```

Y la región crítica es:

$$\{T : T < t_{n+m-2;\alpha}\}$$

Como es un contraste unilateral, solo necesitamos saber el cuantil que deja α por debajo:

```
> n_x<-8
> m_y<-8
> alpha=0.05
> v_critico= qt((alpha),n_x+m_y-2)
> v_critico

[1] -1.76131
```

Por tanto, $-1{,}76 < -0{,}35$.

Solución Como el estadístico se encuentra en la región de aceptación, no tenemos razones suficientes para rechazar la hipótesis nula, por lo que determinamos que NO es recomensable el uso del catalizador 2 al no incrementar el rendimiento promedio del proceso quimico.

Ejercicio 12 *En un país se desea investigar la proporción de electores que prefiere celebrar las elcciones en día festivo, preguntando a una muestra de ciudadanos con derecho a voto.*

1. *¿Qué número de entrevistas habría que realizar para estimar la proporción, con un error no superior a 0.04 y con una confianza del 95 %?*

   Resultado Como el nivel de confianza es 95 %, eso significa que $1-\alpha = 0{,}95$, por tanto $\alpha = 0{,}05$, entonces $\frac{\alpha}{2} = 0{,}025$. Luego $Z_{\frac{\alpha}{2}}$ es:

   ```
   > qnorm(0.025)
   ```

   ```
   [1] -1.959964
   ```

   Como R lee cuantiles a la derecha, en lugar de IC que usualmente se leen a la izquierda, el valor de $Z_{\frac{\alpha}{2}}$ es de

   $$Z_{\frac{\alpha}{2}} = 1{,}96$$

   El intervalo de confianza para una proporción p es:

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

   Por tanto, si queremos calcular el tamaño muestral para que la estimación tenga un margen de error de $\pm 0{,}04$, entonces se deberá calcular lo siguiente:

   $$1{,}96 * \sqrt{\frac{0{,}5 * 0{,}5}{n}} < 0{,}04$$

   $$\sqrt{\frac{0{,}5 * 0{,}5}{n}} > \frac{0{,}04}{1{,}96}$$

   $$n > \frac{0{,}5 * 0{,}5}{\left(\frac{0{,}04}{1{,}96}\right)^2}$$

   ```
   > n=(0.5*0.5)/(0.04/qnorm(0.025))^2
   > n
   ```

   ```
   [1] 600.2279
   ```

   Solución Se necesita obtener un tamaño muestral de $n = 601$ para garantizar que, con un margen de error de $\pm 0{,}04$, un 50 % de las personas entrevistadas prefiera votar en día lectivo.

2. *Elegida una muestra de 500 ciudadanos, 350 de ellos prefirieron celebrar las elecciones el día festivo. Con una confianza del 95 %, ¿entre qué valores se encuentra comprendida la proporción de electores en ese país que prefieren un día festivo?*

Resultado Como hemos comentado en el apartado anterior, el intervalo de confianza para una proporción p es:

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

```
> alpha<-0.05
> n<-500
> p_muestral=350/500
> cuantil= qnorm(1-(alpha/2),mean=0, sd=1,lower.tail = T)
> amplitud=cuantil*sqrt(p_muestral*(1-p_muestral)/n)
> lim_inf=p_muestral-amplitud
> lim_inf

[1] 0.7401673

> lim_sup=p_muestral+amplitud
> lim_sup

[1] 0.6598327
```

Solución Con un 95 % de confianza, la proporción de electores que prefiere votar en día festivo se encuentra entre el 65.98 % y el 74 % de la población de ese país.