Práctica 7 R-Commander

NOMBRE: *Jaime Osés Azcona*

1 - Intervalos de confianza y contrastes de hipótesis para dos poblaciones.

1. Hojas de problemas de clase (ejercicio 1): Un fabricante de envases para alimentos congelados está valorando utilizar un nuevo tipo de material. La resistencia del material es muy importante, y se sabe que sigue una distribución normal. Se sabe que desviación de la resistencia es 1 psi(libras/pulgada^2)para los dos tipos de material. A partir de la muestra aleatoria de 10 envases fabricados con el plástico nuevo y otra de 12 con el clásico, se obtiene $\bar{X}_1$ = 162,5 y $\bar{X}_2$ = 155 respectivamente. La compañía no adoptará el material nuevo a menos que su resistencia media supere a la del material clásico en al menos 7 psi. Con base en la información muestral, se quiere decidir si deberá utilizarse el nuevo material:

a) Construye un intervalo de confianza y establece conclusiones al 95 %.

Para calcular dicho intervalo podemos hacerlo manualmente con la siguiente fórmula:

$$IC = \bar{X}_1 - \bar{X}_2 \pm Z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\ .$$

```
> qnorm(c(0.975), mean=0, sd=1, lower.tail=TRUE)
[1] 1.959964
```

$$IC = \bar{X}_1 - \bar{X}_2 \pm 0{,}8392$$

IC (95%) = (6.6608, 8.3392)

b) Efectúa el contraste con α = 0,05 que plantea como hipótesis alternativa que la diferencia entre resistencias medias del material nuevo y el clásico no llega a ser 7.

El contraste será el siguiente:

$H_0 : \mu_1 - \mu_2 \geq 7$

$H_1 : \mu_1 - \mu_2 < 7$

Para calcular dicho contraste tendremos que hallar la Zobs y Z_α.

$$Zobs = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(162{,}5 - 155) - 7}{\sqrt{\frac{1^2}{10} + \frac{1^2}{12}}} = 1{,}1677$$

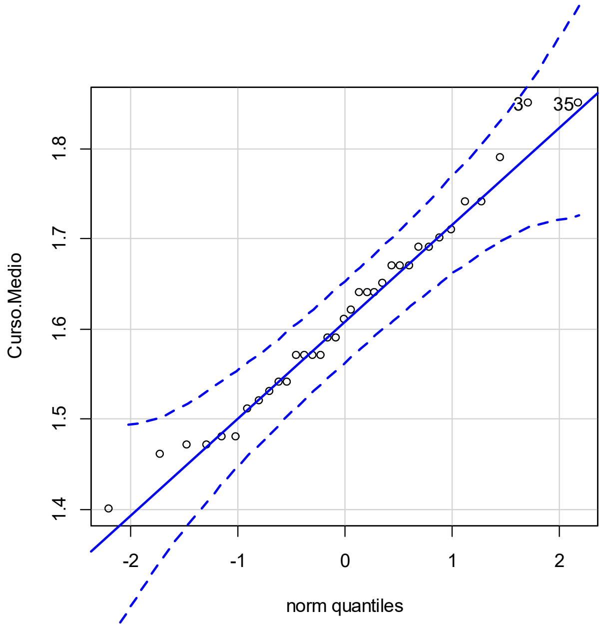
```
> qnorm(c(0.05), mean=0, sd=1, lower.tail=TRUE)
[1] -1.644854
```

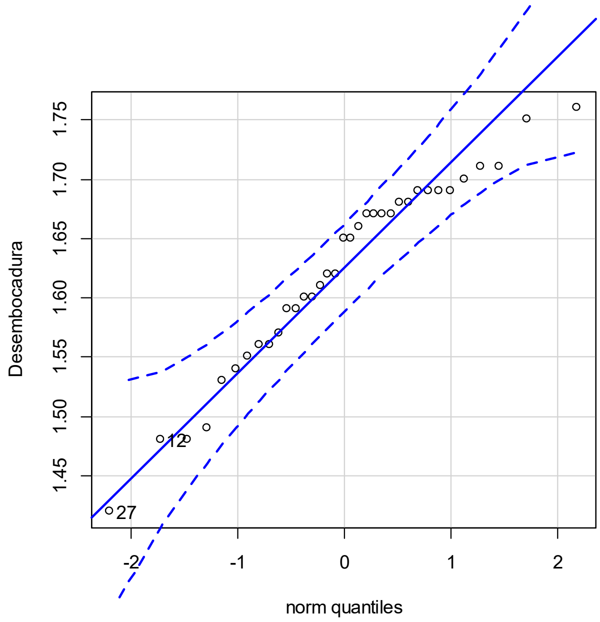
Como Zobs está a la derecha de Z_α no se rechaza la hipótesis nula.

2. Se está haciendo un estudio sobre la cantidad de mercurio en los peces de un río. En esta parte del estudio quiere compararse la cantidad de mercurio en dos zonas del río, el curso medio y la desembocadura.
Para ello se ha tomado una muestra de 35 peces en cada zona de forma independiente; los resultados son:

Curso Medio	**1,64 1,67 1,85 1,57 1,59 1,61 1,53 1,40 1,70 1,48 1,46 1,74 1,67 1,57 1,65 1,48 1,47 1,64 1,79 1,69 1,54 1,71 1,57 1,51 1,54 1,52 1,57 1,67 1,47 1,64 1,74 1,62 1,69 1,59 1,85**
Desembocadura	**1,56 1,55 1,69 1,67 1,60 1,68 1,65 1,59 1,75 1,49 1,69 1,48 1,62 1,48 1,70 1,65 1,67 1,69 1,76 1,59 1,61 1,67 1,69 1,53 1,57 1,62 1,42 1,71 1,54 1,71 1,56 1,67 1,68 1,60 1,66**

a) Realiza un estudio de normalidad.
Para comprobar si los datos siguen una distribución normal, iremos al menú **"Gráficas-Gráfica de comparación de cuantiles"** y seleccionaremos las variables que habremos introducido previamente con los datos de las tablas.





Como podemos observar en la gráfica, la mayoría de los puntos están muy próximos a la línea central, a excepción de los de las colas que están ligeramente separados. Sin embargo estos últimos entran dentro de los limites establecidos, por tanto podemos afirmar que los datos siguen una distribución normal.

b) Estudia mediante un contraste si hay igualdad de varianzas.
Asumiremos un nivel de confianza del 95%.
El contraste será el siguiente:
$H_0 : \sigma_1 / \sigma_2 = 1$
$H_1 : \sigma_1 / \sigma_2 \neq 1$
Para calcular dicho contraste tendremos que hallar la Fobs, $F_{1-\alpha/2}$ y $F_{\alpha/2}$.
Para ello tendremos que calcular las varianzas de los datos de la tabla. Esto lo haremos con un resumen numérico.

```
                  mean         sd
Curso.Medio   1.612286 0.10960050
Desembocadura 1.622857 0.08169424
```

$$\text{Fobs} = \frac{S_1^2 \cdot \sigma_2^2}{S_2^2 \cdot \sigma_1^2} = \frac{S_1^2}{S_2^2} = 1{,}7999$$

```
> qf(c(0.025), df1=34, df2=34, lower.tail=TRUE)
[1] 0.5047652

> qf(c(0.975), df1=34, df2=34, lower.tail=TRUE)
[1] 1.981119
```

Como F_{obs} está entre $F_{1-\alpha/2}$ y $F_{\alpha/2}$ no se rechaza la hipótesis nula.

c) Construye un intervalo de confianza para la diferencia de medias teniendo en cuenta el apartado anterior y establece conclusiones al 95 %.

Teniendo en cuenta el apartado anterior, podremos decir que las varianzas son iguales.
Para calcular dicho intervalo podemos hacerlo manualmente con la siguiente fórmula:

$$IC = \bar{X}_1 - \bar{X}_2 \pm t_{68,\alpha/2} \cdot S \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$. Donde $S = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} = 0{,}096666$

```
> qt(c(0.975), df=68, lower.tail=TRUE)
[1] 1.995469
```

$$IC = \bar{X}_1 - \bar{X}_2 \pm 0{,}04611$$

IC (95%) = (-0.05668, 0.035539)

d) Efectúa el contraste con α = 0,05 que plantea como hipótesis alternativa que no hay diferencia entre de la cantidad media de mercurio.

El contraste será el siguiente:

$H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 \neq 0$

Para calcular dicho contraste tendremos que hallar la tobs, $t_{68,1-\alpha/2}$ y $t_{68,\alpha/2}$.

$$\text{tobs} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{1{,}612286 - 1{,}622857}{\sqrt{\frac{34 \cdot 0{,}1096^2 + 34 \cdot 0{,}08169^2}{68}} \cdot \sqrt{\frac{2}{35}}} = -0{,}4575$$

```
> qt(c(0.025), df=68, lower.tail=TRUE)
[1] -1.995469

> qt(c(0.975), df=68, lower.tail=TRUE)
[1] 1.995469
```

Como t_{obs} está entre $t_{68,1-\alpha/2}$ y $t_{68,\alpha/2}$ no se rechaza la hipótesis nula.

3. Se quiere estudiar si en los embalses de agua el nivel de Chl es el mismo en la cabecera que en la cola.
Para ello se han tomado muestras en 10 embalses; el resultado ha sido:

Embalse	1	2	3	4	5	6	7	8	9	10
Cabecera	11,6	29,2	16,7	37,8	33,9	12,2	5,9	18,5	20,4	10,1
Cola	34,5	11,8	21,6	40,2	34,1	12,5	12,9	24,5	21,7	18,6

a) Construye un intervalo de confianza y establece conclusiones al 99 %.
Calculamos las varianzas de los datos de la tabla. Esto lo haremos con un resumen numérico.

```
          mean        sd  n
Cabecera 19.63 10.722982 10
Cola     21.24  8.132814 10
```

Para calcular dicho intervalo podemos hacerlo manualmente con la siguiente fórmula:

IC = $\bar{X}_1 - \bar{X}_2 \pm t_{w,\alpha/2} \cdot \sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$. Donde w = $\dfrac{(S_1^2/n_1)^2+(S_2^2/n_2)^2}{\dfrac{(S_1^2/n_1)^2}{n_1-1}+\dfrac{(S_2^2/n_2)^2}{n_2-1}}$ = 8,99999 ≈ 9

```
> qt(c(0.995), df=9, lower.tail=TRUE)
[1] 3.249836
```

IC = $\bar{X}_1 - \bar{X}_2$ ± 13,83

IC (99%) = (-15.44, 12.22)

b) Efectúa el contraste con α = 0,05 que plantea como hipótesis alternativa que la diferencia entre el nivel medio de Clh en la cola del embalse es superior en al menos 2 unidades al nivel en la cabecera.
El contraste será el siguiente:
$H_0 : \mu_2 - \mu_1 \leq 2$
$H_1 : \mu_2 - \mu_1 > 2$
Para calcular dicho contraste tendremos que hallar la tobs, $t_{9,1-\alpha}$.

Tobs = $\dfrac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$ = $\dfrac{21{,}21-19{,}63-2}{\sqrt{\frac{10{,}72^2}{10}+\frac{8{,}13^2}{10}}}$ = -0,0215

```
> qt(c(0.95), df=9, lower.tail=TRUE)
[1] 1.833113
```

Como t_{obs} está a la izquierda de $t_{9,1-\alpha}$ no se rechaza la hipótesis nula.

4. Con los datos de galtonR.RData, crea un nuevo factor de modo que si la edad es menor o igual que 23 se asocie a una nivel del factor denominado "joven" y para el resto de edades en otro nivel de factor denominado "mayor". Se quiere comparar la proporción de casados entre los jóvenes y mayores.

a) Construye un intervalo de confianza para la diferencia de proporciones y establece conclusiones al 99 %.

Podríamos hallar el intervalo de confianza a mano, pero al ser una gran cantidad de datos y tenerlos ya metidos en R, lo realizaremos con el programa.

Iremos al menú **"Estadísticos-Proporciones-...para dos muestras..."** y seleccionaremos las 2 variables que queremos comparar e indicaremos el nivel de confianza (0,99).

```
data:  .Table
X-squared = 12.004, df = 1, p-value = 0.0005309
alternative hypothesis: two.sided
99 percent confidence interval:
 -0.19520635 -0.04350893
sample estimates:
    prop 1     prop 2
0.04861111 0.16796875
```

IC (99%) = (-0.1952, -0.435)

b) Efectúa el contraste con α = 0,05 que plantea que no hay diferencia en la proporción de casados entre los jóvenes y mayores.

El contraste que queremos realizar es el siguiente:

$H_0 : \pi_1 - \pi_2 = 0$

$H_1 : \pi_1 - \pi_2 \neq 0$

Para calcular dicho contraste haremos lo mismo que en el apartado anterior. Iremos al menú **"Estadísticos-Proporciones-...para dos muestras..."** y seleccionaremos las 2 variables que queremos comparar e indicaremos el nivel de confianza (0,95).

```
Percentage table:
       estado
jm         C    S Total Count
  joven  4.9 95.1   100   144
  mayor 16.8 83.2   100   256

        2-sample test for equality of proportions without continuity correction

data:  .Table
X-squared = 12.004, df = 1, p-value = 0.0005309
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1770714 -0.0616439
sample estimates:
    prop 1     prop 2
0.04861111 0.16796875
```

Como podemos observar, el p-value es menor que el nivel de significación, por tanto rechazamos la hipótesis nula.