

TAREA 6

Álvaro Morales Sánchez – 18240

EJERCICIO 1 ($\alpha = 0.05$)

1. Utilizando RStudio se obtiene:

```
datos=read.table('quimicos.txt', header=T)
attach(datos)

datos$temp=factor(datos$temp)
datos$labor=factor(datos$labor)

datos$temp=factor(datos$temp, labels=c('I','II','III','IV'))
datos$labor=factor(datos$labor, labels=c('A','B','C'))

> modelo=aov(datos$y2 ~ datos$temp)
> anova(modelo)
Analysis of Variance Table

Response: datos$y2
      Df Sum Sq Mean Sq F value Pr(>F)
datos$temp  3  93.062   31.021   0.8889 0.4872
Residuals  8 279.172   34.897
> model.tables(modelo,'means')
Tables of means
Grand mean

39.92453

      datos$temp
datos$temp
      I      II     III     IV
41.91 42.91 39.14 35.74
```

Con estos resultados se obtiene el siguiente modelo:

$$y_2(\text{rendimiento químico}) = \mu_i + u$$

Siendo $\mu = 39.92453$.

Y las distintas μ_i :

- $\mu_I = 41.91$
- $\mu_{II} = 42.91$
- $\mu_{III} = 39.14$
- $\mu_{IV} = 35.74$

Además, suponemos que se cumplen las hipótesis de normalidad, homocedasticidad e independencia.

2. El **intervalo de confianza** resultante para la **varianza** es: [15.92148, 128.0782]

```

> chia=qchisq(0.025,8, lower.tail=T)
> chib=qchisq(0.975, 8, lower.tail=T)
> Linferior=8*34.897/chib
> Lsuperior=8*34.897/chia
> Linferior
[1] 15.92148
> Lsuperior
[1] 128.0782

```

3. A través de RStudio se obtiene:

```

> anova(modelo)
Analysis of Variance Table

Response: datos$y2
      Df Sum Sq Mean Sq F value Pr(>F)
datos$temp  3  93.062   31.021   0.8889 0.4872
Residuals   8 279.172   34.897

```

El **p-valor** es grande (**0.4872**) para cualquier α , por lo que se aceptará la hipótesis nula. Esta indica que la media de los rendimientos a distintas temperaturas es la misma, por lo que la **temperatura no** tendrá efecto **significativo**.

4. Mediante RStudio:

```

> modelo2=aov(y2 ~ temp+labor)
> anova(modelo2)
Analysis of Variance Table

Response: y2
      Df Sum Sq Mean Sq F value    Pr(>F)
temp    3  93.062   31.021   140.26 6.046e-06 ***
labor    2 277.845  138.923   628.14 1.074e-07 ***
Residuals  6   1.327    0.221
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> model.tables(modelo2,'means')
Tables of means
Grand mean

39.92453

temp
temp
  I    II   III   IV
41.91 42.91 39.14 35.74

labor
labor
  A    B    C
43.85 42.78 33.15
> model.tables(modelo2,'effects')
Tables of effects

temp
temp
  I    II   III   IV
1.985 2.987 -0.782 -4.189

labor
labor
  A    B    C
3.925 2.851 -6.777

```

El modelo obtenido tendrá la siguiente forma:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + u_{ijk}$$

Siendo i: I, II, III, IV (las distintas temperaturas).

Siendo j: A, B, C (los distintos laboratorios).

K será el número de muestras de cada caso.

La media μ del modelo resulta ser: **39.92453**.

Donde α_i , estimaciones de las temperaturas según los distintos laboratorios, o media de las columnas:

```
temp
      I      II      III      IV
1.985  2.987 -0.782 -4.189
```

Donde β_j , estimaciones para los laboratorios según las distintas temperaturas, o media de las filas:

```
labor
      A      B      C
3.925  2.851 -6.777
```

No hay $\alpha\beta_{ij}$ ya que no hay interacción.

Y por último **u** resulta ser **0.221**.

5. Del apartado anterior deducimos:

Los parámetros α_i , son las estimaciones de las temperaturas según los distintos laboratorios, o media de las columnas:

```
temp
      I      II      III      IV
1.985  2.987 -0.782 -4.189
```

Los parámetros β_j , son las estimaciones para los laboratorios según las distintas temperaturas, o media de las filas:

```
labor
      A      B      C
3.925  2.851 -6.777
```

Por último, **no hay $\alpha\beta_{ij}$** ya que no hay interacción.

6. La tabla obtenida mediante RStudio del nuevo modelo resulta ser:

```
> anova(modelo2)
Analysis of Variance Table

Response: y2
      Df Sum Sq Mean Sq F value    Pr(>F)
temp    3  93.062   31.021   140.26 6.046e-06 ***
labor    2 277.845  138.923   628.14 1.074e-07 ***
Residuals 6   1.327    0.221
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

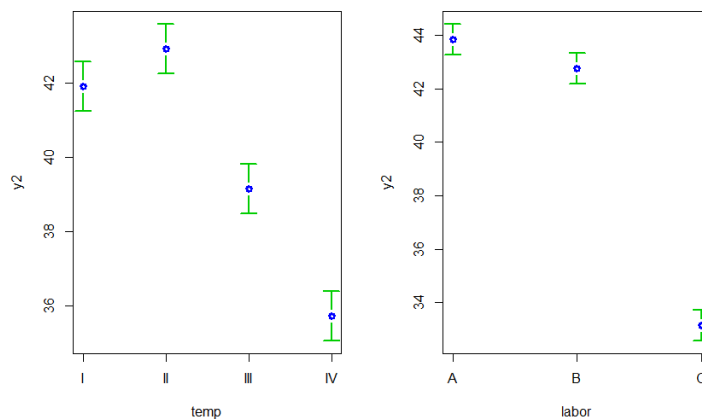
7. Observando la tabla anterior se llega a la siguiente conclusión:

- La **temperatura** es **significativa**, su p-valor es 6.046e-06. Al ser este valor tan pequeño se rechaza la hipótesis nula, esto significa que las medias de los rendimientos a distintas temperaturas varían. Cosa que en el análisis de un

factor anterior resultaba ser no significativo, por lo que la temperatura junto a la influencia del laboratorio sí es significativa, pero sin este no es significativa.

- El **laboratorio** es **significativo**, su p-valor es 1.047e-07. Es un valor muy pequeño por lo que se rechaza la hipótesis nula, resultando la media de los rendimientos distinta dependiendo del laboratorio.

8. Para conocer las mejores condiciones se analizan los gráficos de los IC para las medias:



Observando el gráfico de la izquierda podemos indicar que la **temperatura II** es la que influye de manera más positiva, tiene una media de rendimiento más alta. También se muestra que la temperatura I solapa con la temperatura II, por lo que logra también el rendimiento más alto, aunque algo menor.

En el caso de la gráfica de la derecha vemos que el **laboratorio A** tiene un rendimiento más elevado que el resto, siendo la mejor condición. Pero al solapar con el laboratorio B éste también puede dar una máxima respuesta, aunque algo menos elevada.

```
> predict(modelo2, interval='confidence')
      fit      lwr      upr
1  45.83432 45.02063 46.64801
2  44.76043 43.94674 45.57413
3  35.13238 34.31869 35.94608
4  46.83669 46.02300 47.65038
5  45.76281 44.94912 46.57650
6  36.13476 35.32107 36.94845
7  43.06754 42.25385 43.88123
8  41.99365 41.17996 42.80735
9  32.36561 31.55191 33.17930
10 39.66065 38.84695 40.47434
11 38.58676 37.77307 39.40046
12 28.95871 28.14502 29.77241
```

De la tabla anterior sacamos que para esa condición el intervalo es: [46.02300 , 47.65038] (valores para el fit más alto).

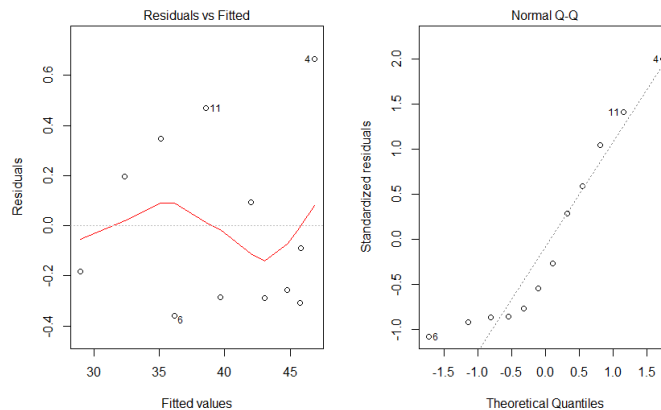
9. El **intervalo de confianza** resultante para la **varianza** es: [0.09176867, 1.07165]

```

> chia=qchisq(0.025,6, lower.tail=T)
> chib=qchisq(0.975, 6, lower.tail=T)
> LInferior=6*0.221/chib
> LSuperior=6*0.221/chia
> LInferior
[1] 0.09176867
> LSuperior
[1] 1.07165

```

10. La diagnosis del modelo es:



Del gráfico de la izquierda podemos observar que se cumple la hipótesis de **homocedasticidad**, ya que el intervalo más ancho entre el más estrecho es menor que 3. A simple vista no se observa, pero con ese breve cálculo podemos decir que tiene homocedasticidad.

En el de la derecha vemos que no sigue la forma de la recta por lo que podríamos concluir que no cumple la hipótesis de **normalidad**.

EJERCICIO 2 ($\alpha = 0.05$)

1. Utilizando RStudio se obtiene:

```

datos=read.table('bacterias.csv', sep=";", header=T)
attach(datos)

datos$conc=factor(datos$conc)
datos$temp=factor(datos$temp)

```

```

> anova(modelo)
Analysis of Variance Table

Response: datos$bacterias
      Df Sum Sq Mean Sq F value    Pr(>F)    
datos$conc  4  41386 10346.5   5.1907 0.001014 **
Residuals 70 139531  1993.3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.tables(modelo, 'means')
Tables of means
Grand mean

82.28

      datos$conc
datos$conc
      0.6      0.8      1      1.2      1.4
51.87  63.20  81.47 117.53  97.33

```

Con estos resultados se obtiene el siguiente modelo:

$$y(\text{cultivo de una bacteria}) = \mu_i + u$$

Siendo $\mu = 82.28$.

Y las distintas μ_i :

- $\mu_{0.6} = 51.87$
- $\mu_{0.8} = 63.20$
- $\mu_1 = 81.47$
- $\mu_{1.2} = 117.53$
- $\mu_{1.4} = 97.33$

Además, suponemos las hipótesis de normalidad, homocedasticidad y linealidad.

2. La **varianza** estimada es **1993.3**.

El **intervalo de confianza** resultante para la **varianza** es: [**1468.389** , **2861.73**]

```

> chia=qchisq(0.025,70, lower.tail=T)
> chib=qchisq(0.975, 70, lower.tail=T)
> LIinferior=70*1993.3/chib
> LSuperior=70*1993.3/chia
> LIinferior
[1] 1468.389
> LSuperior
[1] 2861.73

```

3. Volviendo a la tabla de antes:

```

> anova(modelo)
Analysis of Variance Table

Response: datos$bacterias
      Df Sum Sq Mean Sq F value    Pr(>F)    
datos$conc  4  41386 10346.5   5.1907 0.001014 **
Residuals 70 139531  1993.3
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

El **p-valor** es pequeño (**0.001014**) para cualquier α , por lo que se rechazará la hipótesis nula. Esta indica que la media de las bacterias con distintas

concentraciones no es la misma, por lo que la **concentración** tendrá efecto **significativo**.

4. Observando las medias:

```
> model.tables(modelo, 'means')
Tables of means
Grand mean

82.28

datos$conc
datos$conc
 0.6    0.8    1    1.2    1.4
51.87 63.20 81.47 117.53 97.33
```

Podemos ver que para la **concentración 1.2 %** se obtiene el valor más alto.

También se debe observar si para el intervalo de 1.2% solapa con alguno de los otros. Para ello estudiamos la siguiente tabla:

```
> TukeyHSD(modelo)
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = datos$bacterias ~ datos$conc)

$`datos$conc`
      diff      lwr      upr    p adj
0.8-0.6  11.33333 -34.316306  56.98297 0.9568895
1-0.6    29.60000 -16.049640  75.24964 0.3729457
1.2-0.6  65.66667  20.017027 111.31631 0.0012927
1.4-0.6  45.46667  -0.182973  91.11631 0.0514313
1-0.8    18.26667 -27.382973  63.91631 0.7952196
1.2-0.8  54.33333   8.683694  99.98297 0.0116452
1.4-0.8  34.13333 -11.516306  79.78297 0.2343661
1.2-1    36.06667  -9.582973  81.71631 0.1872816
1.4-1    15.86667 -29.782973  61.51631 0.8663001
1.4-1.2 -20.20000 -65.849640  25.44964 0.7286650
```

Comparando el p.adj con α para todos los valores con respecto a 1.2% podemos ver que el intervalo de 1.2 solapa únicamente con 1% (padj = 0.1872816) y 1.4% (padj = 0.7286650). Esto significa que aunque el valor más alto se logre con la concentración del 1.2%, con las concentraciones de 1 y 1.4 % se logran valores muy próximos pero algo menores.

5. Para calcular **el intervalo de confianza** entre las diferencias de medias entre el 1% y el 1.2% se realiza:

```
> LimS=(117.53-81.47)+(1.994*44.64*sqrt(2/15))
> LimI=(117.53-81.47)-(1.994*44.64*sqrt(2/15))
> LimS
[1] 68.56265
> LimI
[1] 3.557355
```

El **intervalo de confianza** para $\mu_1 - \mu_{1.2} = [3.557355 , 68.56265]$

6. Mediante RStudio:

```

> anova(modelo2)
Analysis of Variance Table

Response: datos$bacterias
          Df Sum Sq Mean Sq  F value    Pr(>F)    
datos$conc    4  41386   10347   32.6287 1.906e-14 ***
datos$temp    2 110444   55222  174.1466 < 2.2e-16 ***
datos$conc:datos$temp  8  10061    1258    3.9661 0.0008024 ***
Residuals    60  19026     317                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model.tables(modelo2,'means')
Tables of means
Grand mean

82.28

      datos$conc
      0.6      0.8      1      1.2      1.4
51.87  63.20  81.47 117.53  97.33

      datos$temp
      27      35      43
28.40 114.84 103.60

      datos$conc:datos$temp
              datos$temp
datos$conc 27      35      43
      0.6  13.8  78.8  63.0
      0.8  21.0  86.2  82.4
      1    30.8 100.6 113.0
      1.2  43.0 172.8 136.8
      1.4  33.4 135.8 122.8
> model.tables(modelo2,'effects')
Tables of effects

      datos$conc
      datos$conc
      0.6      0.8      1      1.2      1.4
-30.41 -19.08  -0.81  35.25  15.05

      datos$temp
      datos$temp
      27      35      43
-53.88  32.56  21.32

      datos$conc:datos$temp
              datos$temp
datos$conc 27      35      43
      0.6  15.813  -5.627 -10.187
      0.8  11.680  -9.560  -2.120
      1    3.213  -13.427  10.213
      1.2 -20.653  22.707  -2.053
      1.4 -10.053   5.907   4.147

```

El modelo obtenido tendrá la siguiente forma:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + u_{ijk}$$

Siendo i: 0.6, 0.8, 1.0, 1.2, 1.4 % (las distintas concentraciones).

Siendo j: 27, 35, 43 (las distintas temperaturas).

K será el número de muestras de cada caso.

La media μ del modelo resulta ser: **82.28**.

Donde α_i , estimaciones de las concentraciones según las distintas temperaturas, o media de las columnas:

```

datos$conc
      0.6      0.8      1      1.2      1.4
-30.41 -19.08  -0.81  35.25  15.05

```


Donde β_j , estimaciones para las temperaturas según las distintas concentraciones, o media de las filas:

```
datos$temp
      27      35      43
-53.88  32.56  21.32
```

Para $\alpha\beta_{ij}$ ya que hay interacción, resultan las siguientes interacciones entre concentración y temperatura.

```
      datos$temp
datos$conc 27      35      43
      0.6  15.813 -5.627 -10.187
      0.8  11.680 -9.560 -2.120
      1     3.213 -13.427  10.213
      1.2 -20.653  22.707 -2.053
      1.4 -10.053   5.907  4.147
```

Y por último μ resulta ser **317**.

7. El **valor** que predice el modelo para concentración 1.2% y temperatura 35°C es **172.8**.

Obtenido de:

```
> model.tables(modelo2, 'means')
Tables of means
Grand mean
82.28

datos$conc
datos$conc
  0.6  0.8  1  1.2  1.4
51.87 63.20 81.47 117.53 97.33

datos$temp
datos$temp
  27  35  43
28.40 114.84 103.60

datos$conc:datos$temp
      datos$temp
datos$conc 27      35      43
      0.6  13.8  78.8  63.0
      0.8  21.0  86.2  82.4
      1    30.8 100.6 113.0
      1.2  43.0 172.8 136.8
      1.4  33.4 135.8 122.8
```

8. La tabla se ha mostrado anteriormente, de nuevo:

```
> anova(modelo2)
Analysis of Variance Table

Response: datos$bacterias
      Df Sum Sq Mean Sq F value    Pr(>F)
datos$conc      4  41386   10347  32.6287 1.906e-14 ***
datos$temp      2 110444   55222 174.1466 < 2.2e-16 ***
datos$conc:datos$temp  8  10061    1258   3.9661 0.0008024 ***
Residuals     60  19026     317
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

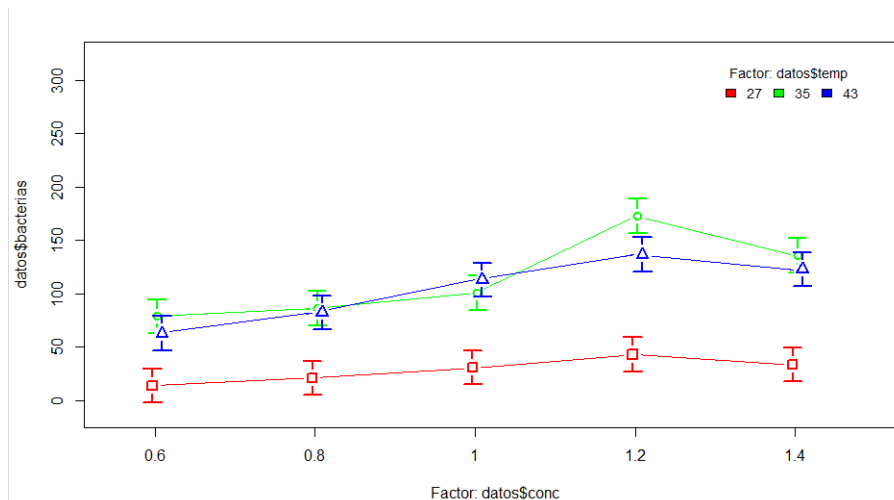
Analizando la tabla sabemos:

- La **concentración** es **significativa**, su p-valor es 1.906e-14. Al ser este valor tan pequeño se rechaza la hipótesis nula, esto significa que las medias de las bacterias a distintas concentraciones varían.

- El **laboratorio** es **significativo**, su p-valor es menor que 2.2e-16. Es un valor muy pequeño por lo que se rechaza la hipótesis nula, resultando la media de las bacterias distinta dependiendo de la temperatura.
- La **interacción** es **significativa**, su p-valor es 0.0008024. Hay interacción y efecto entre la concentración y la temperatura.

9. Mediante RStudio y la función InterIC.r:

```
> interIC(modelo2,'datos$conc','datos$temp')
  datos$conc datos$temp media 2.5% 97.5%
1          0.6         27  13.8  -2.13 29.73
2          0.6         35  78.8  62.87 94.73
3          0.6         43  63.0  47.07 78.93
4          0.8         27  21.0   5.07 36.93
5          0.8         35  86.2  70.27 102.13
6          0.8         43  82.4  66.47 98.33
7          1         27  30.8  14.87 46.73
8          1         35 100.6  84.67 116.53
9          1         43 113.0  97.07 128.93
10         1.2         27  43.0  27.07 58.93
11         1.2         35 172.8 156.87 188.73
12         1.2         43 136.8 120.87 152.73
13         1.4         27  33.4  17.47 49.33
14         1.4         35 135.8 119.87 151.73
15         1.4         43 122.8 106.87 138.73
```



Como se puede observar en este último gráfico para una concentración de 1.2 % y una temperatura de 35 °C se obtiene el mayor número de bacterias. En la tabla correspondiente a esta función también vemos que para esas condiciones se obtiene la media más alta.

Analizando la siguiente tabla de la comparación Tukey veremos si solapa con algún otro intervalo.

```
> TukeyHSD(modelo2)
Tukey multiple comparisons of means
 95% family-wise confidence level

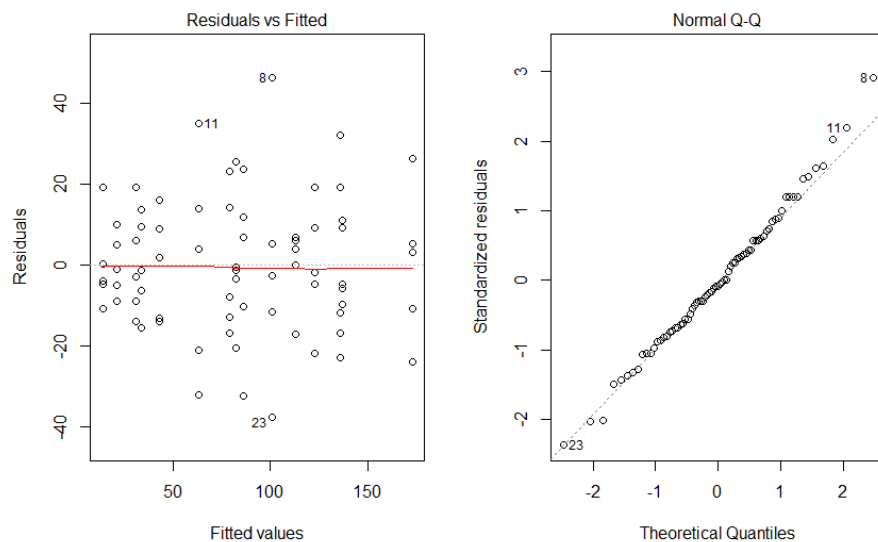
Fit: aov(formula = datos$bacterias ~ datos$conc * datos$temp)

$`datos$conc`
      diff          lwr          upr    p adj
0.8-0.6  11.33333   -6.95414189  29.620809 0.4161377
1-0.6    29.60000   11.31252478  47.887475 0.0002493
1.2-0.6  65.66667   47.37919145  83.954142 0.0000000
1.4-0.6  45.46667   27.17919145  63.754142 0.0000000
1-0.8    18.26667   -0.02080855  36.554142 0.0503989
1.2-0.8  54.33333   36.04585811  72.620809 0.0000000
1.4-0.8  34.13333   15.84585811  52.420809 0.0000204
1.2-1     36.06667   17.77919145  54.354142 0.0000067
1.4-1     15.86667   -2.42080855  34.154142 0.1187756
1.4-1.2  -20.20000  -38.48747522  -1.912525 0.0232744

$`datos$temp`
      diff          lwr          upr    p adj
35-27   86.44    74.3358  98.5441973 0.0000000
43-27   75.20    63.0958  87.3041973 0.0000000
43-35  -11.24   -23.3442   0.8641973 0.0739983
```

El p adj es 0.00 para las condiciones mencionadas, por lo que son las únicas para alcanzar el mayor número de bacterias, no solapan con otras condiciones.

10. La diagnosis del modelo resulta:



De la gráfica de la izquierda concluimos que el modelo cumple la hipótesis de homocedasticidad. Se ve claramente una anchura homogénea a lo largo de ésta. Además, se puede comprobar con el breve cálculo de dividir el intervalo más ancho entre el más estrecho y ver que la diferencia sea menor que 3, como ocurre en este caso.

En la segunda gráfica vemos como los valores tienden a la forma de la recta, de lo que deducimos que cumple la hipótesis de normalidad.