4 ENCRIGE

D. Daza

junio 07, 2023 | 15:34:14 | CDT

Table of Contents

[1 Pruebas ANOVA 1](#_Toc137045003)

[1.1 ANOVAS normales 1](#_Toc137045004)

[1.2 ANOVAS log 1](#_Toc137045005)

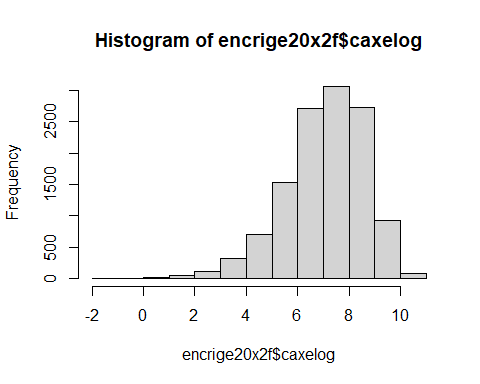
# 1 Pruebas ANOVA

Transformación de la variable

library(MASS)  
library(moments)  
  
# Calculamos el sesgo  
sesgocaxe <- skewness(encrige20x2f$ca\_xe)  
  
# Imprimimos el valor de sesgo  
print(sesgocaxe)

[1] 3.053058

# Aplicamos el logaritmo  
encrige20x2f$caxelog <- log(encrige20x2f$ca\_xe)  
  
# Visualizar la distribución de tu variable transformada con un histograma  
hist(encrige20x2f$caxelog)



# 1.1 ANOVAS normales

ANOVA por tamaño, 1

anova\_1 <- aov(ca\_xe ~ tamano\_2, data = encrige20x2f)  
  
summary(anova\_1)

Df Sum Sq Mean Sq F value Pr(>F)   
tamano\_2 3 9.994e+09 3.331e+09 245.9 <2e-16 \*\*\*  
Residuals 12237 1.658e+11 1.355e+07   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

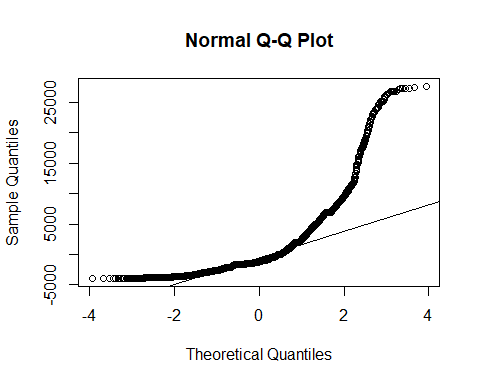
tab\_anova1 <- summary(anova\_1)[[1]] %>%  
 as.data.frame() %>%  
 kable("pipe", caption = "Resultados del análisis ANOVA 1") %>%  
 kable\_styling()  
  
tab\_anova1

Resultados del análisis ANOVA 1

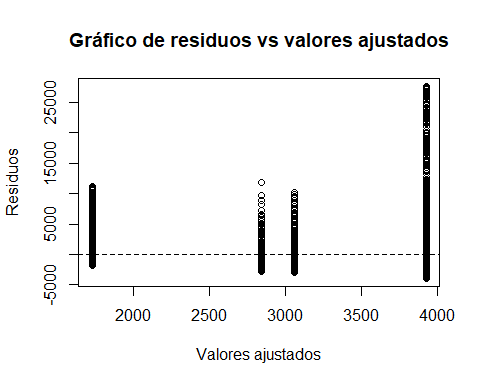
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| tamano\_2 | 3 | 9993996845 | 3331332282 | 245.8675 | 0 |
| Residuals | 12237 | 165802760548 | 13549298 | NA | NA |

Validación de los supuestos del ANOVA 1

###################VISUAL  
  
# Comprobación de la normalidad de los residuos con un gráfico Q-Q  
qqnorm(residuals(anova\_1))  
qqline(residuals(anova\_1))



# Comprobación de la homogeneidad de las varianzas con un gráfico de residuos vs valores ajustados  
plot(anova\_1$fitted.values, anova\_1$residuals,   
 xlab="Valores ajustados",   
 ylab="Residuos",  
 main="Gráfico de residuos vs valores ajustados")  
abline(h = 0, lty = 2) # Agrega una línea horizontal en y = 0



###################ESTADÍSTICA  
  
# Comprobar la normalidad de los residuos con la prueba de Shapiro-Wilk  
#shapiro.test(residuals(anova\_1))  
  
# Comprobar la homogeneidad de las varianzas con la prueba de Levene  
leveneTest(anova\_1)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)   
group 3 181.22 < 2.2e-16 \*\*\*  
 12237   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Preparación de datos para los ANOVA por sector

# Crear dataframes separados para cada sector  
encrige20x2f\_comercio <- subset(encrige20x2f, Sector == "Comercio")  
encrige20x2f\_manufacturas <- subset(encrige20x2f, Sector == "Manufacturas")  
encrige20x2f\_servicios <- subset(encrige20x2f, Sector == "Servicios")

ANOVA sector comercio, 2

anova\_2 <- aov(ca\_xe ~ tamano\_2, data = encrige20x2f\_comercio)  
  
summary(anova\_2)

Df Sum Sq Mean Sq F value Pr(>F)   
tamano\_2 3 9.303e+09 3.101e+09 146 <2e-16 \*\*\*  
Residuals 4811 1.022e+11 2.124e+07   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

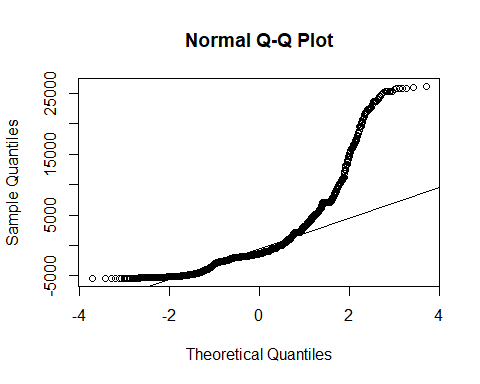
tab\_anova2 <- summary(anova\_2)[[1]] %>%  
 as.data.frame() %>%  
 kable("pipe", caption = "Resultados del ANOVA: Comercio") %>%  
 kable\_styling()  
  
tab\_anova2

Resultados del ANOVA: Comercio

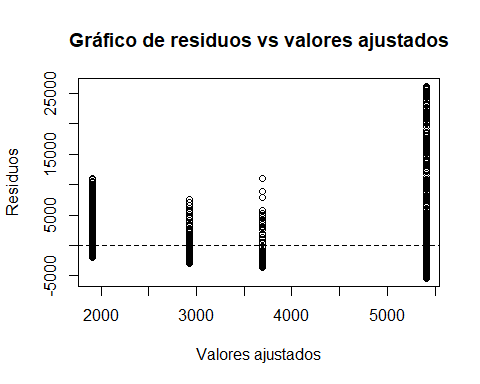
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| tamano\_2 | 3 | 9303180740 | 3101060247 | 145.9938 | 0 |
| Residuals | 4811 | 102190652889 | 21241042 | NA | NA |

Validación de los supuestos del ANOVA 2

# Comprobación de la normalidad de los residuos con un gráfico Q-Q  
qqnorm(residuals(anova\_2))  
qqline(residuals(anova\_2))



# Comprobación de la homogeneidad de las varianzas con un gráfico de residuos vs valores ajustados  
plot(anova\_2$fitted.values, anova\_2$residuals,   
 xlab="Valores ajustados",   
 ylab="Residuos",  
 main="Gráfico de residuos vs valores ajustados")  
abline(h = 0, lty = 2) # Agrega una línea horizontal en y = 0



# Comprobar la normalidad de los residuos con la prueba de Shapiro-Wilk  
shapiro.test(residuals(anova\_2))

Shapiro-Wilk normality test  
  
data: residuals(anova\_2)  
W = 0.76825, p-value < 2.2e-16

# Comprobar la homogeneidad de las varianzas con la prueba de Levene  
leveneTest(anova\_2)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)   
group 3 151.21 < 2.2e-16 \*\*\*  
 4811   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ANOVA sector manufacturas, 3

anova\_3 <- aov(ca\_xe ~ tamano\_2, data = encrige20x2f\_manufacturas)  
  
summary(anova\_3)

Df Sum Sq Mean Sq F value Pr(>F)   
tamano\_2 3 9.222e+08 307383628 42.57 <2e-16 \*\*\*  
Residuals 1961 1.416e+10 7221017   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

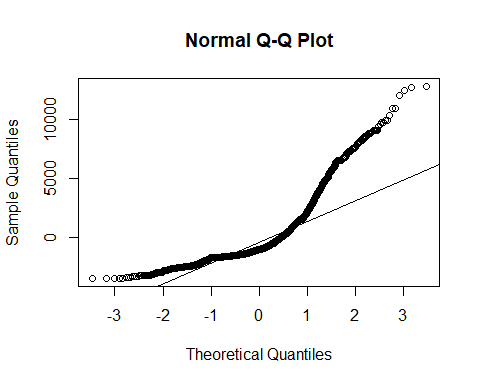
tab\_anova3 <- summary(anova\_3)[[1]] %>%  
 as.data.frame() %>%  
 kable("pipe", caption = "Resultados del ANOVA: Manufacturas") %>%  
 kable\_styling()  
  
tab\_anova3

Resultados del ANOVA: Manufacturas

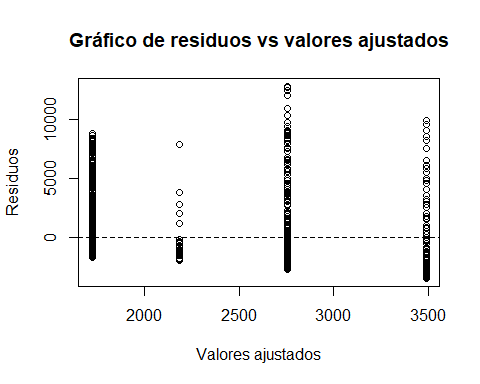
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| tamano\_2 | 3 | 922150883 | 307383628 | 42.56791 | 0 |
| Residuals | 1961 | 14160414012 | 7221017 | NA | NA |

Validación de los supuestos del ANOVA 3

# Comprobación de la normalidad de los residuos con un gráfico Q-Q  
qqnorm(residuals(anova\_3))  
qqline(residuals(anova\_3))



# Comprobación de la homogeneidad de las varianzas con un gráfico de residuos vs valores ajustados  
plot(anova\_3$fitted.values, anova\_3$residuals,   
 xlab="Valores ajustados",   
 ylab="Residuos",  
 main="Gráfico de residuos vs valores ajustados")  
abline(h = 0, lty = 2) # Agrega una línea horizontal en y = 0



# Comprobar la normalidad de los residuos con la prueba de Shapiro-Wilk  
shapiro.test(residuals(anova\_3))

Shapiro-Wilk normality test  
  
data: residuals(anova\_3)  
W = 0.80115, p-value < 2.2e-16

# Comprobar la homogeneidad de las varianzas con la prueba de Levene  
leveneTest(anova\_3)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)   
group 3 18.442 8.47e-12 \*\*\*  
 1961   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ANOVA sector servicios, 4

anova\_4 <- aov(ca\_xe ~ tamano\_2, data = encrige20x2f\_servicios)  
  
summary(anova\_4)

Df Sum Sq Mean Sq F value Pr(>F)   
tamano\_2 3 2.612e+09 870736520 108.3 <2e-16 \*\*\*  
Residuals 5457 4.387e+10 8039988   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

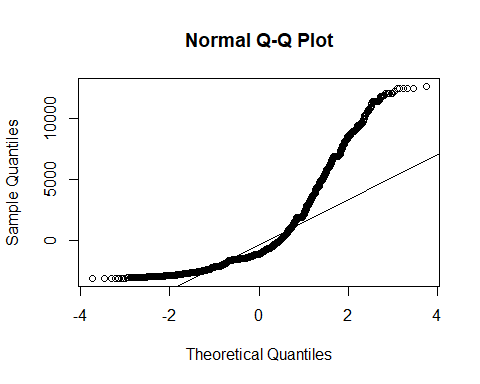
tab\_anova4 <- summary(anova\_4)[[1]] %>%  
 as.data.frame() %>%  
 kable("pipe", caption = "Resultados del ANOVA: Servicios") %>%  
 kable\_styling()  
  
tab\_anova4

Resultados del ANOVA: Servicios

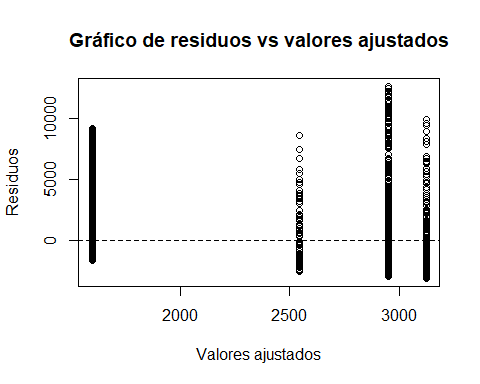
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| tamano\_2 | 3 | 2612209559 | 870736520 | 108.3007 | 0 |
| Residuals | 5457 | 43874215066 | 8039988 | NA | NA |

Validación de los supuestos del ANOVA 4

# Comprobación de la normalidad de los residuos con un gráfico Q-Q  
qqnorm(residuals(anova\_4))  
qqline(residuals(anova\_4))



# Comprobación de la homogeneidad de las varianzas con un gráfico de residuos vs valores ajustados  
plot(anova\_4$fitted.values, anova\_4$residuals,   
 xlab="Valores ajustados",   
 ylab="Residuos",  
 main="Gráfico de residuos vs valores ajustados")  
abline(h = 0, lty = 2) # Agrega una línea horizontal en y = 0



# Comprobar la normalidad de los residuos con la prueba de Shapiro-Wilk  
#shapiro.test(residuals(anova\_4))  
  
# Comprobar la homogeneidad de las varianzas con la prueba de Levene  
leveneTest(anova\_4)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)   
group 3 44.739 < 2.2e-16 \*\*\*  
 5457   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# 1.2 ANOVAS log

ANOVA 1 Log

anova\_1log <- aov(caxelog ~ tamano\_2, data = encrige20x2f)  
  
summary(anova\_1log)

Df Sum Sq Mean Sq F value Pr(>F)   
tamano\_2 3 3534 1178.0 572.2 <2e-16 \*\*\*  
Residuals 12237 25191 2.1   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

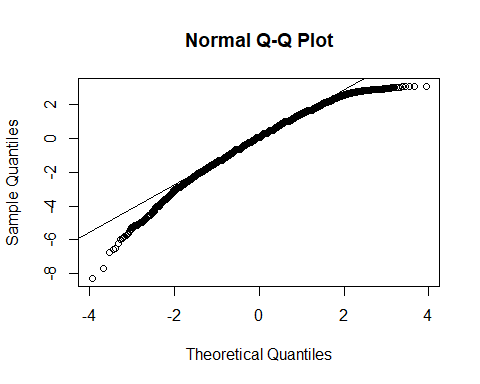
tab\_anova1log <- summary(anova\_1log)[[1]] %>%  
 as.data.frame() %>%  
 kable("pipe", caption = "Resultados del análisis ANOVA 1 Log") %>%  
 kable\_styling()  
  
tab\_anova1log

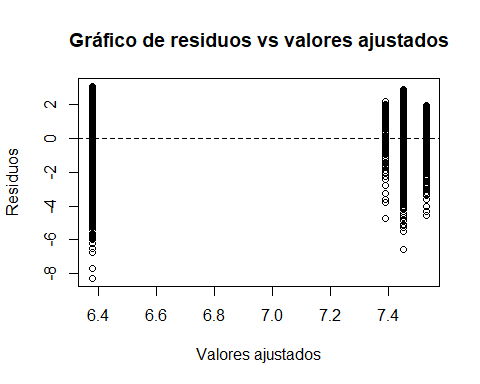
Resultados del análisis ANOVA 1 Log

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| tamano\_2 | 3 | 3533.899 | 1177.966243 | 572.2148 | 0 |
| Residuals | 12237 | 25191.190 | 2.058608 | NA | NA |

Validación de los supuestos del ANOVA 1 log

###################VISUAL  
  
# Comprobación de la normalidad de los residuos con un gráfico Q-Q  
qqnorm(residuals(anova\_1log))  
qqline(residuals(anova\_1log))





###################ESTADÍSTICA  
  
# Comprobar la normalidad de los residuos con la prueba de Shapiro-Wilk  
#shapiro.test(residuals(anova\_1log))  
  
# Comprobar la homogeneidad de las varianzas con la prueba de Levene  
leveneTest(anova\_1log)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)   
group 3 234.83 < 2.2e-16 \*\*\*  
 12237   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ANOVA sector comercio, 2 log

anova\_2log <- aov(caxelog ~ tamano\_2, data = encrige20x2f\_comercio)  
  
summary(anova\_2log)

Df Sum Sq Mean Sq F value Pr(>F)   
tamano\_2 3 1135 378.3 189.9 <2e-16 \*\*\*  
Residuals 4811 9582 2.0   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

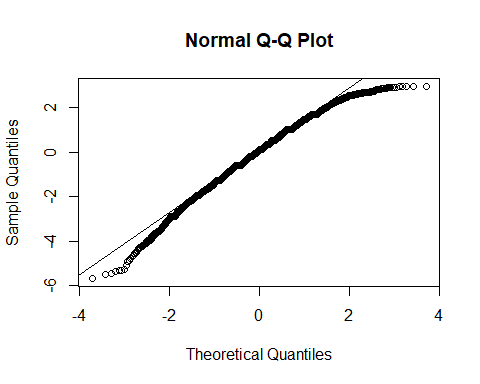
tab\_anova2log <- summary(anova\_2log)[[1]] %>%  
 as.data.frame() %>%  
 kable("pipe", caption = "Resultados del ANOVA: Comercio Log") %>%  
 kable\_styling()  
  
tab\_anova2log

Resultados del ANOVA: Comercio Log

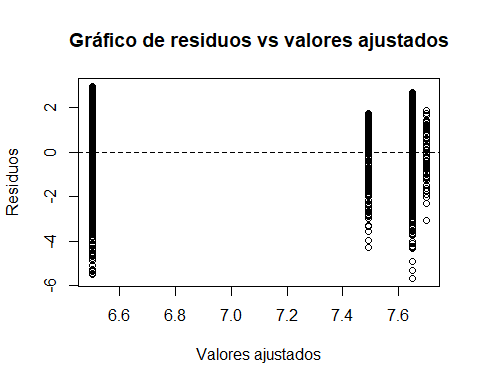
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| tamano\_2 | 3 | 1134.811 | 378.270207 | 189.9171 | 0 |
| Residuals | 4811 | 9582.382 | 1.991765 | NA | NA |

Validación de los supuestos del ANOVA 2 log

# Comprobación de la normalidad de los residuos con un gráfico Q-Q  
qqnorm(residuals(anova\_2log))  
qqline(residuals(anova\_2log))



# Comprobación de la homogeneidad de las varianzas con un gráfico de residuos vs valores ajustados  
plot(anova\_2log$fitted.values, anova\_2log$residuals,   
 xlab="Valores ajustados",   
 ylab="Residuos",  
 main="Gráfico de residuos vs valores ajustados")  
abline(h = 0, lty = 2) # Agrega una línea horizontal en y = 0



# Comprobar la normalidad de los residuos con la prueba de Shapiro-Wilk  
shapiro.test(residuals(anova\_2log))

Shapiro-Wilk normality test  
  
data: residuals(anova\_2log)  
W = 0.98797, p-value < 2.2e-16

# Comprobar la homogeneidad de las varianzas con la prueba de Levene  
leveneTest(anova\_2log)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)   
group 3 97.339 < 2.2e-16 \*\*\*  
 4811   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ANOVA sector manufacturas, 3 log

anova\_3log <- aov(caxelog ~ tamano\_2, data = encrige20x2f\_manufacturas)  
  
summary(anova\_3log)

Df Sum Sq Mean Sq F value Pr(>F)   
tamano\_2 3 459 153.08 66.32 <2e-16 \*\*\*  
Residuals 1961 4526 2.31   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

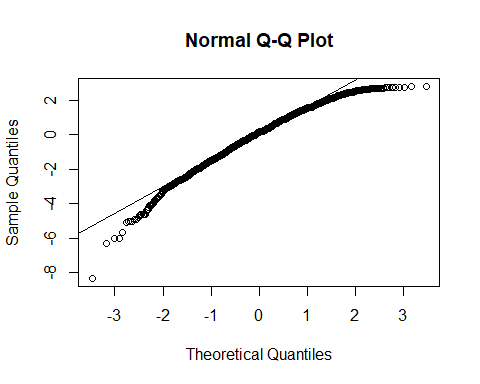
tab\_anova3log <- summary(anova\_3log)[[1]] %>%  
 as.data.frame() %>%  
 kable("pipe", caption = "Resultados del ANOVA: Manufacturas Log") %>%  
 kable\_styling()  
  
tab\_anova3log

Resultados del ANOVA: Manufacturas Log

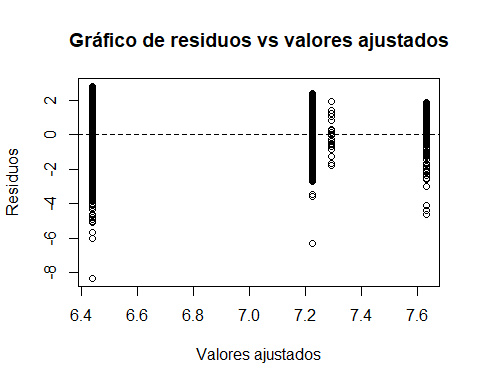
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| tamano\_2 | 3 | 459.2288 | 153.076262 | 66.32002 | 0 |
| Residuals | 1961 | 4526.2737 | 2.308146 | NA | NA |

Validación de los supuestos del ANOVA 3 log

# Comprobación de la normalidad de los residuos con un gráfico Q-Q  
qqnorm(residuals(anova\_3log))  
qqline(residuals(anova\_3log))



# Comprobación de la homogeneidad de las varianzas con un gráfico de residuos vs valores ajustados  
plot(anova\_3log$fitted.values, anova\_3log$residuals,   
 xlab="Valores ajustados",   
 ylab="Residuos",  
 main="Gráfico de residuos vs valores ajustados")  
abline(h = 0, lty = 2) # Agrega una línea horizontal en y = 0



# Comprobar la normalidad de los residuos con la prueba de Shapiro-Wilk  
shapiro.test(residuals(anova\_3log))

Shapiro-Wilk normality test  
  
data: residuals(anova\_3log)  
W = 0.97514, p-value < 2.2e-16

# Comprobar la homogeneidad de las varianzas con la prueba de Levene  
leveneTest(anova\_3log)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)   
group 3 26.958 < 2.2e-16 \*\*\*  
 1961   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ANOVA sector servicios, 4 log

anova\_4log <- aov(caxelog ~ tamano\_2, data = encrige20x2f\_servicios)  
  
summary(anova\_4log)

Df Sum Sq Mean Sq F value Pr(>F)   
tamano\_2 3 1811 603.5 301.9 <2e-16 \*\*\*  
Residuals 5457 10908 2.0   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

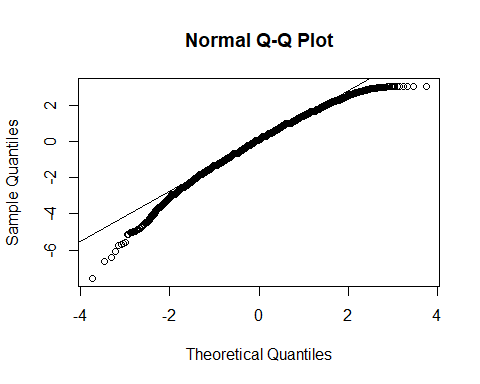
tab\_anova4log <- summary(anova\_4log)[[1]] %>%  
 as.data.frame() %>%  
 kable("pipe", caption = "Resultados del ANOVA: Servicios Log") %>%  
 kable\_styling()  
  
tab\_anova4log

Resultados del ANOVA: Servicios Log

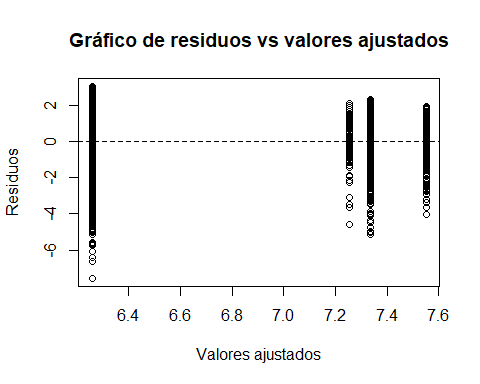
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| tamano\_2 | 3 | 1810.647 | 603.548905 | 301.9392 | 0 |
| Residuals | 5457 | 10908.043 | 1.998908 | NA | NA |

Validación de los supuestos del ANOVA 4 log

# Comprobación de la normalidad de los residuos con un gráfico Q-Q  
qqnorm(residuals(anova\_4log))  
qqline(residuals(anova\_4log))



# Comprobación de la homogeneidad de las varianzas con un gráfico de residuos vs valores ajustados  
plot(anova\_4log$fitted.values, anova\_4log$residuals,   
 xlab="Valores ajustados",   
 ylab="Residuos",  
 main="Gráfico de residuos vs valores ajustados")  
abline(h = 0, lty = 2) # Agrega una línea horizontal en y = 0



# Comprobar la normalidad de los residuos con la prueba de Shapiro-Wilk  
#shapiro.test(residuals(anova\_4log))  
  
# Comprobar la homogeneidad de las varianzas con la prueba de Levene  
leveneTest(anova\_4log)

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)   
group 3 130.81 < 2.2e-16 \*\*\*  
 5457   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1