4 ENCRIGE

D. Daza

junio 05, 2023 | 16:30:09 | CDT

Table of Contents

[1. ENCRIGE 1](#_Toc136877306)

[2. Cargas administrativas 1](#_Toc136877307)

[2.1 Estadística descriptiva general 1](#_Toc136877308)

[2.2 Estadística descriptiva por tamaño, y sector y tamaño 5](#_Toc136877309)

[2.3 Pruebas ANOVA 7](#_Toc136877310)

# 1. ENCRIGE

Ahora es momento de trabajar con la base de datos de la Encuesta Nacional de Calidad Regulatoria e Impacto Gubernamental en las Empresas (ENCRIGE) 2020.

# 2. Cargas administrativas

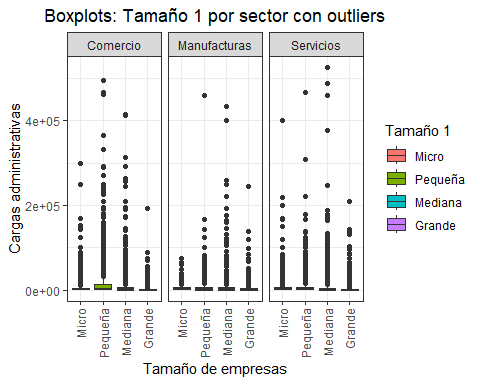
## 2.1 Estadística descriptiva general

#Estadística descriptiva  
  
#Tabla 1  
  
tab1 <- ds\_encrige20x1f %>%  
 kable(caption = "1. Estadística descriptiva, datos brutos ",  
 format.args = list(decimal.mark = ".", big.mark = ","),  
 format = "pandoc", align = "c")  
  
tab1

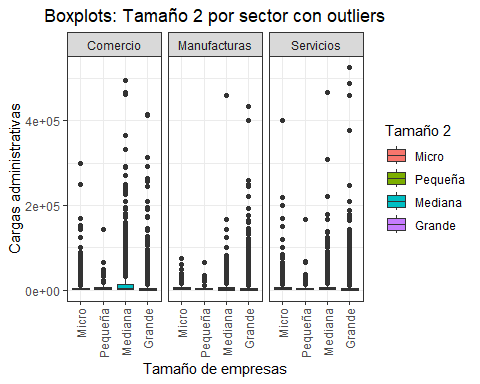
1. Estadística descriptiva, datos brutos

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No. obs. | Media | Desv. estándar | Mínimo | Q1 | Q3 | Máximo |
| 14,109 | 8,832.6 | 25,928.3 | 0.15 | 571.43 | 6,000 | 526,315.8 |

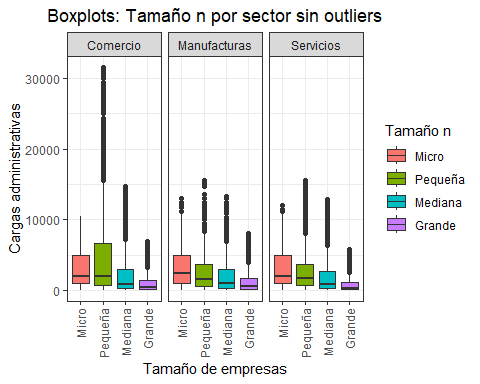
Graficación de boxplot por sector y tamaño para identificar outliers  
fibp\_sizexsecto\_s1



fibp\_sizexsecto\_s2



Graficación de boxplot por sector y tamaño SIN outliers.

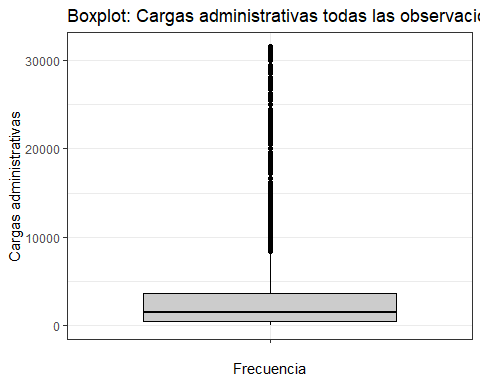


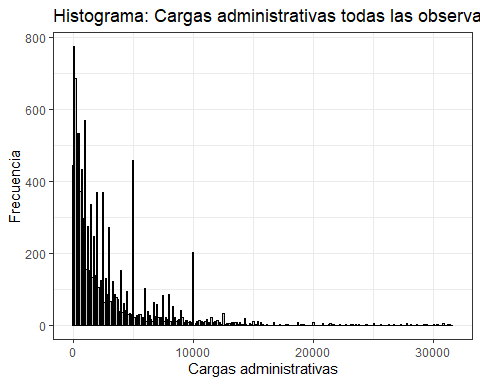
Estadística descriptiva sin outliers: todas las observaciones

2. Estadística descriptiva para todas las observaciones

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No. obs. | Media | Desv. estándar | Mínimo | Q1 | Q3 | Máximo |
| 12,217 | 2,812.03 | 3,813.74 | 0.15 | 481.93 | 3,597.12 | 31,578.95 |

Graficación sin outliers: todas las observaciones





## 2.2 Estadística descriptiva por tamaño, y sector y tamaño

Estadística descriptiva por tamaño de empresa

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| tamano\_1 | n | mean | sd | min | q1 | q3 | max |
| Micro | 4178 | 3004.444 | 2740.116 | 14.285714 | 1000.00000 | 5000.000 | 13000.000 |
| Pequeña | 3507 | 3929.447 | 5520.604 | 2.439024 | 666.66667 | 4472.222 | 31578.947 |
| Mediana | 2903 | 2178.302 | 2933.938 | 1.190476 | 272.72727 | 2830.189 | 14772.727 |
| Grande | 1629 | 1042.217 | 1503.439 | 0.148258 | 96.15385 | 1293.900 | 8045.977 |

Sector y tamaño

Sector comercio: Estadística descriptiva

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| tamano\_1 | n | mean | sd | min | q1 | q3 | max |
| Micro | 1991 | 2947.284 | 2674.773 | 25.000000 | 1000.00000 | 4900.000 | 10500.000 |
| Pequeña | 1434 | 5408.178 | 7384.983 | 7.142857 | 717.53247 | 6644.737 | 31578.947 |
| Mediana | 1051 | 2325.487 | 3144.782 | 2.764977 | 289.15831 | 2989.543 | 14772.727 |
| Grande | 347 | 1097.939 | 1592.810 | 2.702703 | 92.09522 | 1370.015 | 6906.077 |

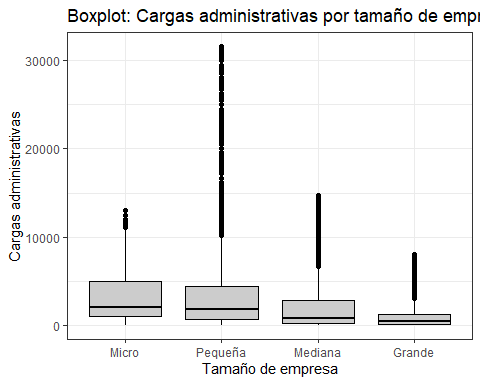
Manufacturas: Estadística descriptiva

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| tamano\_1 | n | mean | sd | min | q1 | q3 | max |
| Micro | 343 | 3342.974 | 3119.023 | 20.000000 | 1000.0000 | 5000.000 | 13000.000 |
| Pequeña | 471 | 2756.520 | 3188.320 | 2.439024 | 506.4103 | 3636.364 | 15517.241 |
| Mediana | 568 | 2252.636 | 2875.416 | 4.687500 | 306.7485 | 2885.132 | 13235.294 |
| Grande | 582 | 1279.499 | 1726.097 | 0.148258 | 144.6100 | 1649.616 | 8045.977 |

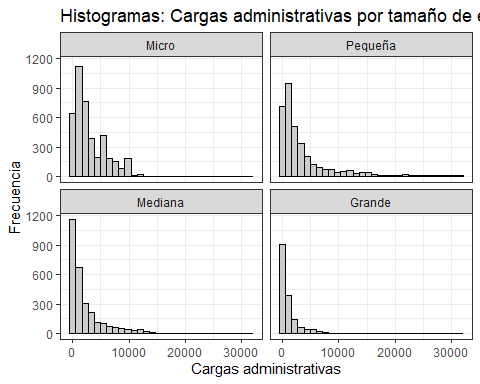
Servicios: Estadística descriptiva

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| tamano\_1 | n | mean | sd | min | q1 | q3 | max |
| Micro | 1844 | 3003.1907 | 2731.592 | 14.2857143 | 1000.00000 | 5000.000 | 12000.000 |
| Pequeña | 1602 | 2950.6376 | 3408.802 | 8.8235294 | 683.23864 | 3636.364 | 15555.556 |
| Mediana | 1284 | 2024.9420 | 2771.104 | 1.1904762 | 240.29207 | 2661.352 | 12931.034 |
| Grande | 700 | 817.3121 | 1197.105 | 0.2720348 | 82.71391 | 1038.427 | 5836.576 |

Graficación de boxplots por tamaño de empresa únicamente



Gráficación de histogramas por tamaño de empresa únicamente



## 2.3 Pruebas ANOVA

ANOVA por tamaño, 1

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

Df Sum Sq Mean Sq F value Pr(>F)   
tamano\_1 3 1.080e+10 3.601e+09 263.5 <2e-16 \*\*\*  
Residuals 12213 1.669e+11 1.366e+07   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

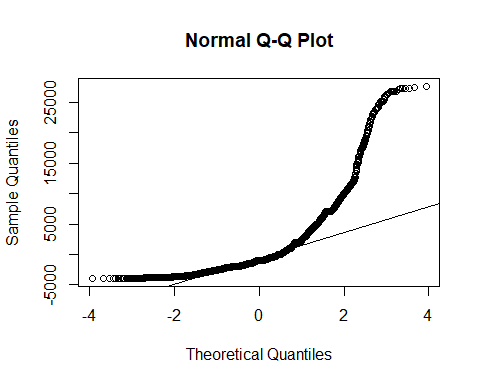
tab\_anova1 <- summary(anova\_1)[[1]] %>%  
 as.data.frame() %>%  
 kable("pipe", caption = "Resultados del análisis ANOVA 1") %>%  
 kable\_styling()  
  
tab\_anova1

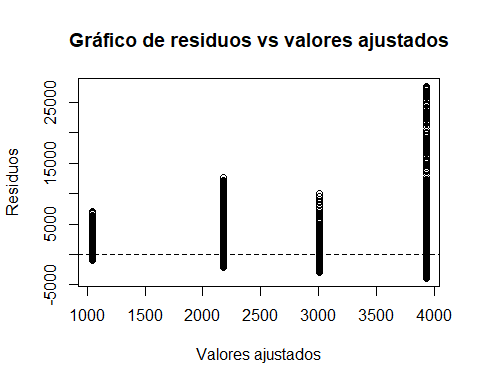
Resultados del análisis ANOVA 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| tamano\_1 | 3 | 10801885924 | 3600628641 | 263.5179 | 0 |
| Residuals | 12213 | 166874714521 | 13663696 | 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))





###################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 3 204.11 < 2.2e-16 \*\*\*  
 12213   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ANOVA sector comercio, 2

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

Df Sum Sq Mean Sq F value Pr(>F)   
tamano\_1 3 9.243e+09 3.081e+09 143.2 <2e-16 \*\*\*  
Residuals 4819 1.037e+11 2.151e+07   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

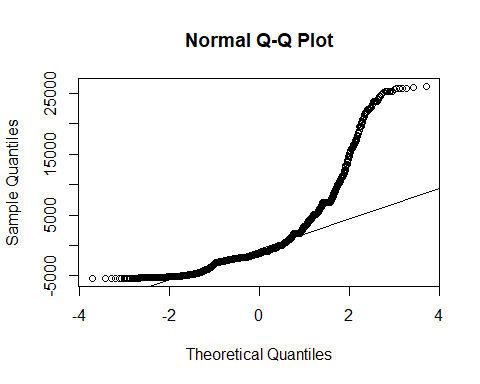
tab\_anova2 <- summary(anova\_2)[[1]] %>%  
 as.data.frame() %>%  
 kable("pipe", caption = "Resultados del ANOVA: Comercio") %>%  
 kable\_styling()  
  
tab\_anova2

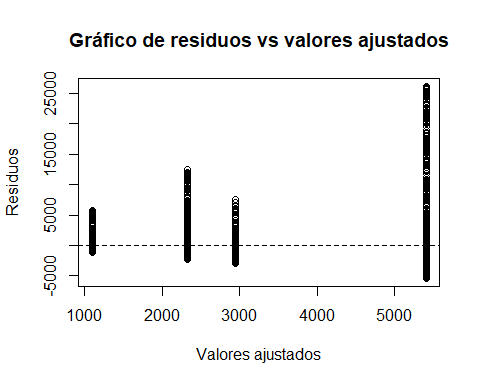
Resultados del ANOVA: Comercio

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| tamano\_1 | 3 | 9242922460 | 3080974153 | 143.2408 | 0 |
| Residuals | 4819 | 103652147422 | 21509057 | 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))





# 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 = 0.76947, 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 3 149.32 < 2.2e-16 \*\*\*  
 4819   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ANOVA sector manufacturas, 3

anova\_3 <- aov(ca\_xe ~ tamano\_1, data = encrige20x2f\_manufacturas)  
  
summary(anova\_3)

Df Sum Sq Mean Sq F value Pr(>F)   
tamano\_1 3 1.077e+09 359160003 48.47 <2e-16 \*\*\*  
Residuals 1960 1.452e+10 7410109   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

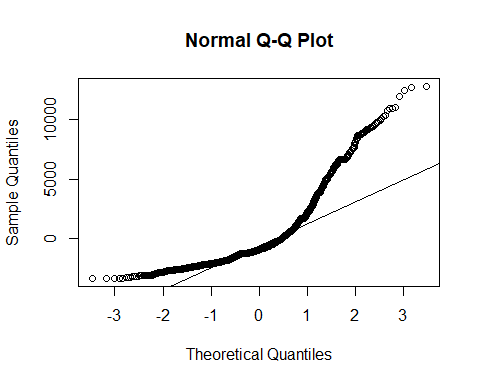
tab\_anova3 <- summary(anova\_3)[[1]] %>%  
 as.data.frame() %>%  
 kable("pipe", caption = "Resultados del ANOVA: Manufacturas") %>%  
 kable\_styling()  
  
tab\_anova3

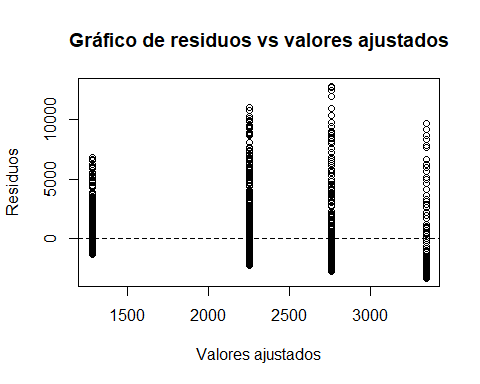
Resultados del ANOVA: Manufacturas

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| tamano\_1 | 3 | 1077480009 | 359160003 | 48.46892 | 0 |
| Residuals | 1960 | 14523814616 | 7410109 | 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))





# 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 = 0.80553, 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 27.371 < 2.2e-16 \*\*\*  
 1960   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

ANOVA sector servicios, 4

anova\_4 <- aov(ca\_xe ~ tamano\_1, data = encrige20x2f\_servicios)  
  
summary(anova\_4)

Df Sum Sq Mean Sq F value Pr(>F)   
tamano\_1 3 3.060e+09 1.020e+09 128.1 <2e-16 \*\*\*  
Residuals 5426 4.321e+10 7.963e+06   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

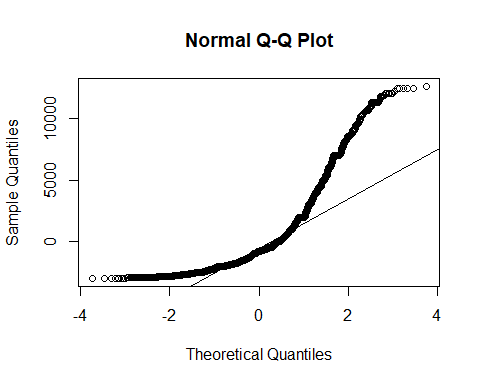
tab\_anova4 <- summary(anova\_4)[[1]] %>%  
 as.data.frame() %>%  
 kable("pipe", caption = "Resultados del ANOVA: Servicios") %>%  
 kable\_styling()  
  
tab\_anova4

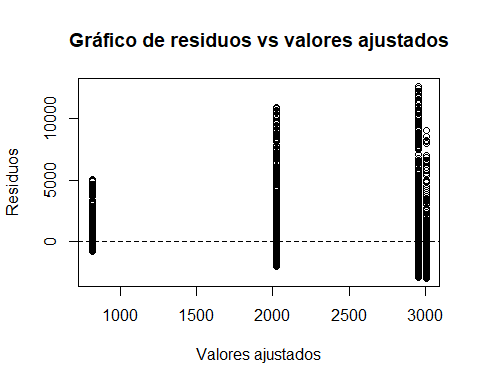
Resultados del ANOVA: Servicios

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
| tamano\_1 | 3 | 3060497885 | 1020165962 | 128.1077 | 0 |
| Residuals | 5426 | 43209122562 | 7963347 | 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))





# 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 71.804 < 2.2e-16 \*\*\*  
 5426   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

|  |
| --- |
| **Estadística descriptiva final** |
| Table: Estadística descriptiva: cargas administrativas |
| No. obs. Media Desv. estándar Mínimo Q1 Q3 Máximo |

42 2,720.27 530.67 1,210.51 2,423.59 3,091.57 4,023.31