

Tarea 2 Metodos no Parametricos

Rudy Miranda

julio, 2023

Índice

1. Introducción	1
2. Origen vs Tiempo	1
2.1. Analisis Descriptivo	1
3. Tipo vs Capitalización	2
3.1. Analisis Descriptivo	2
4. Test de Mann-Whitney	2
4.1. Hipótesis	2
4.2. Estadístico de Prueba y desarrollo del test	3
4.2.1. Definiciones	3
4.3. Region de Rechazo	3
4.4. Evaluación y Conclusión	3
5. Test Kruskal-Wallis	4
5.1. Hipótesis	4
5.2. Estadístico de prueba y desarrollo del test	4
5.2.1. Definiciones:	4
6. Evaluación y Conclusión	4

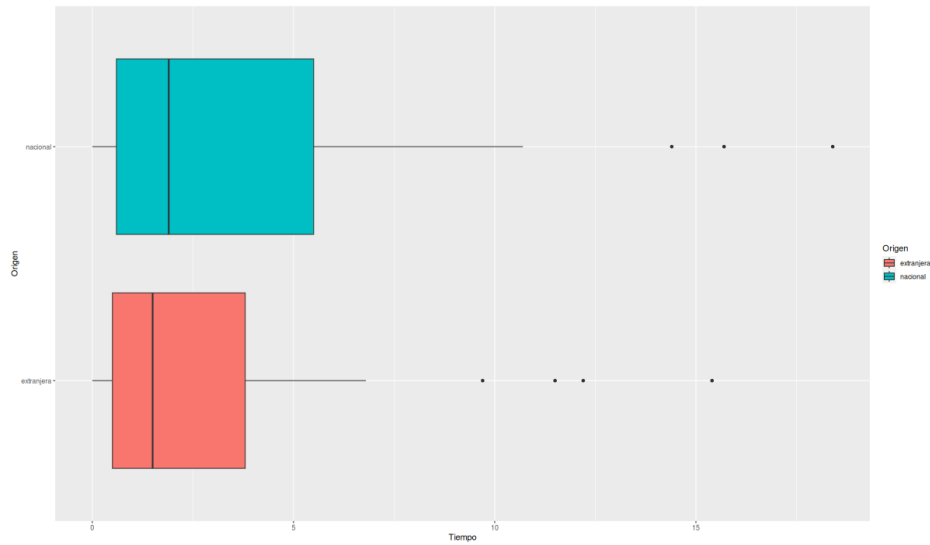
1. Introducción

Se nos presenta el caso donde somos contratados por un instituto de estudio financieros para analizar los datos provenientes de 100 empresas del mercado bursatil y realizar comparaciones mediante disitintos test no parametricos.

2. Origen vs Tiempo

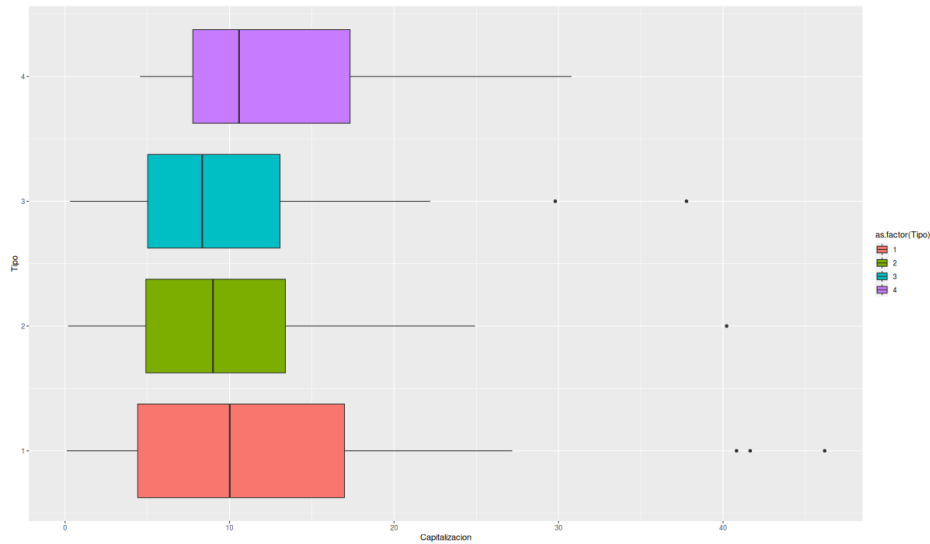
mediante este boxplot podemos hacernos una idea de las caractisticas de los datos. Apoyaremos esto con un analisis descreptivo de los estadísticos mas comunes.

2.1. Analisis Descriptivo



Origen	media	dest	min	Q1	Q2	Q3	max
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
extranjera	2.75	3.25	0	0.5	1.5	3.8	15.4
nacional	3.78	4.49	0	0.6	1.9	5.5	18.4

3. Tipo vs Capitalización



3.1. Analisis Descriptivo

4. Test de Mann-Whitney

4.1. Hipótesis

$$H_0 : P(X_{\text{Nacional}} > X_{\text{Extranjera}}) = P(X_{\text{Nacional}} < X_{\text{Extranjera}}) \quad (1)$$

$$H_1 : \text{no ocurre } H_0 \quad (2)$$

Tipo	media	dest	min	Q1	Q2	Q3	max
<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	13.2	12.2	0.1	4.42	10.0	17.0	46.2
2	10.1	8.80	0.18	4.91	8.99	13.4	40.2
3	10.4	8.52	0.3	5.02	8.33	13.1	37.8
4	13.4	7.39	4.55	7.76	10.6	17.3	30.8

4.2. Estadístico de Prueba y desarrollo del test

El estadístico de prueba para este caso es:

$$U = \min U_1, U_2 = 1115,5 \quad (3)$$

donde:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 = 1115,5 \quad (4)$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2 = 1303,5 \quad (5)$$

4.2.1. Definiciones

- R_1 : suma de rangos para el grupo 1 = 2164.5
- R_2 : suma de rangos para el grupo 2 = 2885.5
- n_1 y n_2 : numero de observaciones para cada grupo = 41 y 59

En el caso de que la cantidad de observaciones sea mayor a 20(en total), podemos aproximar a una normal estandar la siguiente estandarización del estadístico

$$Z = \frac{U - \mu_U}{\sigma_U} \approx N(0,1), \text{ donde } \mu_U = \frac{n_1 n_2}{2} \text{ y } \sigma_U = \sqrt{\frac{n_1 n_2 (n+1)}{12} - \frac{n_1 n_2 \sum_{k=1}^K t_k^3 - t_k}{12n(n-1)}} \quad (6)$$

Notar que se usa la expresión de σ_U para el caso donde hay valores repetidos. Los resultados para cada una de las expresiones fueron los siguientes,

$$\mu_U = 1209,5, \quad \sigma_U = 142,56, \quad Z = -0,66 \Rightarrow p - \text{value} = 0,51 \quad (7)$$

Para una visión mas detallada revisar script de R .

4.3. Region de Rechazo

Mediante el comando `wilcox.test` podemos obtener el intervalo de confianza pra un nivel de confianza dado, y con ello obtener la region de rechazo como su complemento.

```
wilcox.test(Tiempo ~ Origen, df, alternative = "two.sided", mu = 0,
paired = FALSE, conf.int = 0.95)
```

dando como resultado un intervalo de confianza:

$$IC_{0,95} = (-1,4, 0,4) \quad (8)$$

4.4. Evaluación y Conclusión

La evaluación del test se realizo en el punto 4.2, obteniendo un *valor-p* = 0,51, con el no rechazamos la hipótesis nula. Por lo anterior podemos afirmar que no hay una diferencia significativa entre el origen de las empresas con respecto a su longevidad.

5. Test Kruskal-Wallis

5.1. Hipótesis

En este punto buscamos ver si incide el tipo de empresa en su capitalización. Para este fin, usaremos el test de Kruskal-Wallis, con las siguientes hipótesis

$$H_0 : \text{Las empresas de tipo 1, 2, 3, 4 son iguales con respecto a su capitalización.} \quad (9)$$

$$H_1 : \text{Al menos un tipo de empresa es diferente en cuanto a su capitalización.} \quad (10)$$

5.2. Estadístico de prueba y desarrollo del test

Primero que nada, al igual que con el test de Mann-Whitney, tenemos dos casos; uno para el caso de empates y para la ausencia de estos.

En nuestro análisis exploratorio nos damos cuenta de que la cantidad de empates en este caso son solo dos entre 100 posibles. Por ello usaremos el estadístico de prueba para el caso en el que no hay empates, en busca de facilitar el test, y debido que para tan bajo ratio de empates tampoco marcaría una gran diferencia.

El estadístico en cuestión es:

$$H = \frac{12}{N(N+1)} \left(\sum_{i=1}^g n_i \bar{r}_i^2 \right) - 3(N+1) \quad (11)$$

5.2.1. Definiciones:

- g : cantidad de grupos = 4
- n_i : número de observaciones en cada uno de los grupos = 30, 25, 27 y 18
- N : número total de observaciones de la muestra, $N = n_1 + n_2 + \dots + n_g = 100$
- \bar{r}_i : rango promedio de todas las observaciones del grupo i -ésimo = 52, 45.3, 46.4 y 61.4

Reemplazando los datos, obtenemos el valor de nuestro estadístico $H = 4.02$, y dado que para cada grupo hay más de 5 observaciones podemos ocupar la distribución χ_{g-1}^2 , y con ello obtener el valor p

$$p = P(\chi_3^2 > 4.02) = 0.26 \quad (12)$$

6. Evaluación y Conclusión

Una vez obtenidos tanto el valor del estadístico, como el valor p , determinamos que no se rechaza la hipótesis nula y por lo tanto no hay ninguno de los 4 grupos que se diferencie significativamente del resto.