

Trabajo Grupal #2

Jesus M. Huayhua Flores

2024-05-08

Librerías a utilizar

```
library(DescTools)
library(readxl)
library(ggplot2)
```

Lectura del archivo

```
data <- read_xlsx(path="Sistema.xlsx")
head(data)
```

```
## # A tibble: 6 x 2
##   Sistema Tiempo
##   <chr>      <dbl>
## 1 Windows    28
## 2 Windows    27
## 3 Windows    25
## 4 Windows    21
## 5 Windows    23
## 6 MacOS     18
```

Pregunta01

Defina las variables que se deben involucrar en el estudio y establezca la hipótesis nula y alternativa que estén en conformidad con el objetivo del estudio.

Solucion

- X_1 Tiempo de realización de una tarea informática, que una persona consigue usando Windows.
- X_2 Tiempo de realización de una tarea informática, que una persona consigue usando Apple.
- X_3 Tiempo de realización de una tarea informática, que una persona consigue usando Android.

$$H_o : \mu_1 = \mu_2 = \mu_3 = \mu$$
$$H_1 : \exists i | \mu_i \neq \mu$$

Pregunta02

Considerando de que participaron 15 operadores, los cuales fueron divididos aleatoriamente en 3 grupos de 5 cada uno: Windows, MacOS y Android ¿qué se puede decir sobre las 3 muestras de operadores que formaron los 3 grupos?

Solución

Los 15 operadores fueron seleccionados aleatoriamente en grupos de 5 cada uno. Por lo tanto, los datos de cada grupo no están influenciados por alguno de los otros grupos.

Así mismo, se puede afirmar que las 3 muestras de los operadores son independientes y representativas de la población de operadores.

Pregunta 03

Verifique si es posible aplicar una prueba de hipótesis paramétrica (establezca y haga las pruebas para los contrastes de hipótesis que deben ser chequeados)

Solución

Realizamos la prueba de Shapiro-Wilk, para encontrar los valores de *p-value* de cada uno de los grupos.

```
windows <- data$Tiempo[data$Sistema == "Windows"]
macos <- data$Tiempo[data$Sistema == "MacOS"]
android <- data$Tiempo[data$Sistema == "Android"]
# Prueba de Shapiro
shapiro.test(windows)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  windows
## W = 0.96222, p-value = 0.8234
```

```
shapiro.test(macos)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  macos
## W = 0.94273, p-value = 0.6853
```

```
shapiro.test(android)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  android
## W = 0.86757, p-value = 0.2567
```

Como podemos observar los *p-value* son:

$$p - value_{windows} = 0.8234$$

$$p - value_{macos} = 0.6853$$

$$p - value_{android} = 0.2567$$

Estos valores *p* representan la probabilidad de observar los datos de cada grupo si la distribución verdadera de los datos fuera normal.

Así mismo, todos los valores *p* fueron mayor a $\alpha = 0.05$. Por lo tanto, en ninguno de los grupos se rechaza la normalidad.

Podemos asumir, que se sigue una distribución normal para poder aplicar la prueba paramétrica ANOVA.

```
bartlett.test(list(windows, macos, android))
```

```
##
## Bartlett test of homogeneity of variances
##
## data: list(windows, macos, android)
## Bartlett's K-squared = 0.23827, df = 2, p-value = 0.8877
```

Dado que el $p = 0.8877 > \alpha = 0.05$, no contamos con las suficientes pruebas para rechazar la hipótesis nula de homogeneidad de varianzas. Entonces, se puede afirmar que las varianzas entre los 3 grupos analizados son aproximadamente iguales.

Pregunta 04

Realice la prueba de hipótesis que sea apropiada para responder al objetivo del estudio. No se olvide que, si es el caso, debe reescribir su hipótesis nula y alternativa. ¿Cuáles son sus conclusiones?

Solución

Dado que en la pregunta anterior verificamos que se cumplen los supuestos para aplicar una prueba ANOVA. Procederemos para verificar si existen diferencias significativas entre los sistemas operativos.

```
anova <- aov(c(windows, macos, android) ~ rep(c("Windows", "MacOs", "Android"), each=5))
summary(anova)
```

```
##                                     Df Sum Sq Mean Sq F value
## rep(c("Windows", "MacOs", "Android"), each = 5)  2  115.7    57.87    9.234
## Residuals                                     12    75.2     6.27
##                                     Pr(>F)
## rep(c("Windows", "MacOs", "Android"), each = 5) 0.00373 **
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El valor $p_{asociado} = 0.00373 < \alpha = 0.05$, se puede concluir que si existe una diferencia significativa en los tiempos de realización entre los sistemas operativos.

Pregunta 05

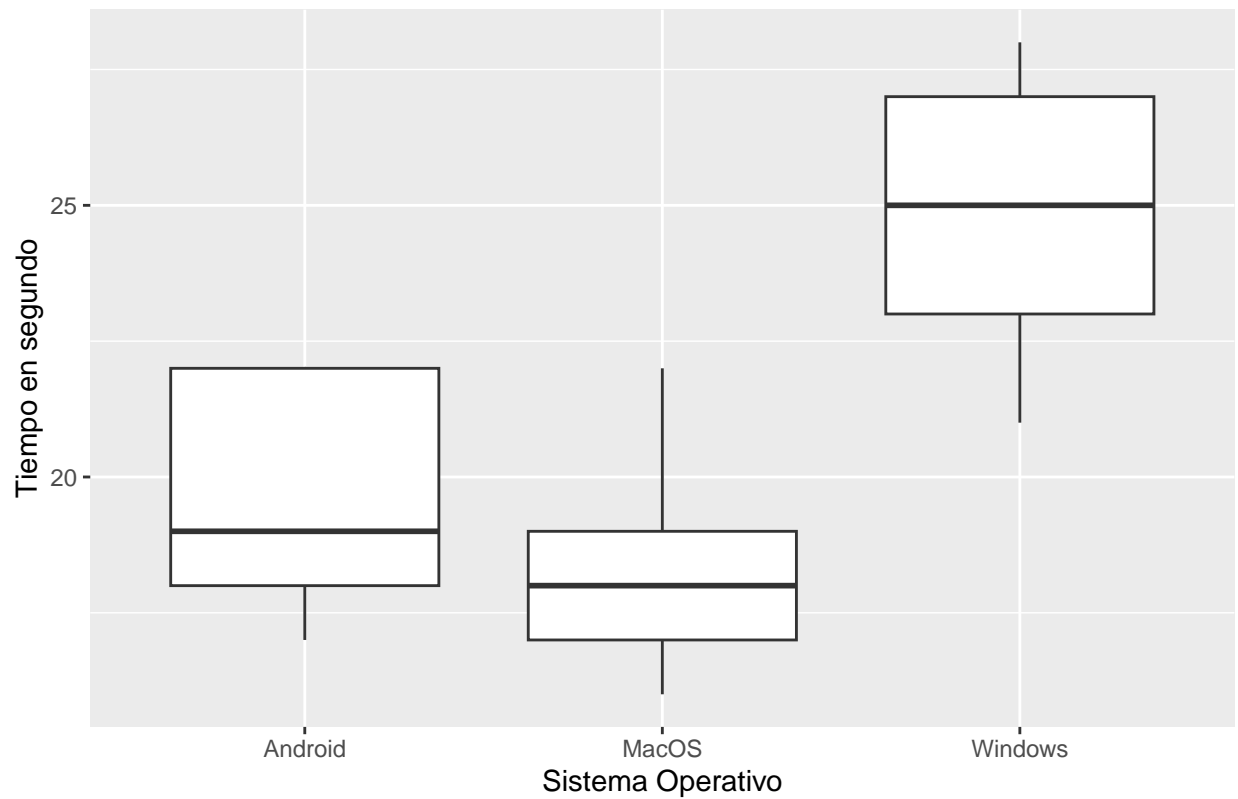
Descubra, mediante un método gráfico, cuál sistema es el más eficiente para realizar la tarea.

Solución

```
datos <- data.frame(
  sistema = rep(c("Windows", "MacOS", "Android"), each = 5),
  tiempo = c(windows, macos, android)
)

# Crear el diagrama de caja con ggplot2 y asignar colores a cada grupo
ggplot(datos, aes(x = sistema, y = tiempo)) +
  geom_boxplot() +
  labs(title = "Diagrama de Caja para los sistemas operativos",
       x = "Sistema Operativo",
       y = "Tiempo en segundo")
```

Diagrama de Caja para los sistemas operativos



Del gráfico, podemos ver que la media de MacOS es menor en comparación con los demás sistemas operativos. Así mismo, que es windows con la mayor media y el que demora una mayor cantidad de tiempo.