

Bioestadística_RLadiesGTO

Lavinia Lavin

2025-02-20

Vamos a comenzar con la carga de la base de datos para poder visualizar las tablas con las que vamos a trabajar

```
library(readxl)
Copia_de_Oxygen_consumption <- read_excel("Copia de Oxygen consumption.xls")
View(Copia_de_Oxygen_consumption)
```

Vamos a adjuntar la base de datos con la que vamos a trabajar

```
names<-Copia_de_Oxygen_consumption
names
attach(Copia_de_Oxygen_consumption)
```

vamos a sacar los primeros datos de nuestro set de datos

```
consumption<-Copia_de_Oxygen_consumption$consumption
mean(consumption)
var(consumption)
sd(consumption)
```

#Con un histograma nos podemos dar una idea de como estan distribuidos nuestros datos

```
hist(consumption, col= "red")
```

#Para comprobarlo vamos a realizar una prueba de normalidad: Shapiro-Wilk donde los valores criticos son: Si p-value <0.05= Los datos no son normales Si p-Value >0.05 Los datos son normales (pruebas parametricas)

```
shapiro.test(consumption)
```

#Procesos basicos para los analisis de datos, a partir de la carga de datos que estamos usando

```
qqnorm(consumption, col="blue")
```

#Vamos a instalar las paqueterias basicas para el analisis de datos estadisticos

```
install.packages("modeest")
library(modeest)
```

#Este comando nos sirve para: obtener la moda de nuestros datos

```
mlv(consumption, method="mfv")
```

#con el paquete plotrix podemos ser graficos de cajas y bigotes (boxplot)

```
install.packages("plotrix")
library(plotrix)
```

#Vamos a observar cuantos datos tenemos en esta variable

```
length(consumption)
```

#Vamos a llamar a un objeto d para determinar que solo queremos trabajar con los datos de consumo de oxigeno

```
d<-Copia_de_Oxygen_consumption
```

#Como nuestros datos siguen una distribucion normal podemos aplicar una prueba de T de Student:si p-Value <0.05 rechazamos la Ho, hay evidencia de que las medias son diferentes Si p-Value >0.05 No se rechaza la Ho, no hay suficiente evidencia para afirmar que las medias sean diferentes

```
t.test(d)
```

#Podemos jugar con nuestros datos, separandolos por características, en este caso le damos un valor discreto (1 o 2), con la funcion subset podemos agruparlos por sus características

```
sp1<-subset(d, species==1)
```

```
sp2<-subset(d, species==2)
print(sp1)
print(sp2)
```

#vamos a generar nuestro primer boxplot con las variables que tiene nuestro set de datos

```
boxplot(consumption~species, col="green", main="Oxygen consumption per species")
```

#vamos a obtener la Chi2 (X2) donde los valores criticos son: Si p-Value es < 0.05 se rechaza la Ho, hay evidencia que las variables estan asociadas Si p-Value >0.05 No se rechaza la Ho, no hay suficiente evidencia para decir que las variables estan relacionadas

```
library(stats)
chisq.test(consumption)
```

#Vamos a generar una prueba de F de Fisher donde los valores criticos son: Si p-Value < 0.05 se rechaza la Ho, hay una diferencia significativa entre medias o varianzas Si p-Value >0.05 NO se rechaza la Ho, no hay suficiente evidencia para afirmar que las varianzas o medias sean diferentes #Primero suponemos un alpha de 0.05

```
alfa <- 0.05
n <- 143
d<-47
valor_critico_F <- qf(1 - alfa, df1 = n, df2 = d)
print(valor_critico_F)
```

#Vamos a hacer una ANOVA, si p-Value < 0.05, se rechaza la Ho, al menos un grupo tiene una media diferente Si p-Value > 0.05 No se rechaza la Ho, no hay suficiente evidencia de diferencias entre grupos. Con este set de datos:

```
anova<-aov(species~consumption)
summary(anova)
```