

Bioestadística_RLadiesGTO

Lavinia Lavin

2025-02-20

Primero vamos a comenzar con la carga de nuestros archivos, para hacer esto necesitamos dirigir la sesión a nuestro directorio donde tenemos los archivos. Una vez elegido nuestro directorio seleccionamos en cargar Excel, en este caso vamos a trabajar con la siguiente tabla que solo tiene una hoja, podemos seleccionar con que hoja queremos trabajar

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

En seguida vamos a pasar a analizar la naturaleza de nuestros datos. Primero vamos a ver con que variables cuenta nuestra base de datos, para esto usamos el comando “names”. Después adjuntamos la tabla para asegurarnos de que vamos a trabajar con este set y no con otro con la función “attach”

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

Vamos a crear un objeto con los datos de consumo de oxígeno para esto usamos el signo de pesos \$ que infiere que estamos trabajando con la columna seleccionada del set de datos. Vamos a obtener los primeros datos de nuestro set siendo “mean” el promedio “var” la varianza y “sd” el error estándar, podemos observar que la varianza está por arriba del promedio por lo cual podemos inferir que los datos están dispersos de acuerdo a la media

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

Para corroborar esto, podemos visualizar nuestros datos con la función “hist”, en este caso parece que nuestros datos siguen la distribución de la campana sin embargo podemos notar que están un poco sesgados a la izquierda

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

Para comprobar si nuestros datos EN REALIDAD siguen una distribución normal hacemos una prueba de normalidad y como nuestros datos son menores de 50 elegimos una prueba de normalidad Shapiro-Wilk, teniendo un p value mayor a 0.05 sabemos que nuestros datos pueden comportarse como datos paramétricos. Si p-Value >0.05 Los datos son normales (pruebas paramétricas)

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

Pasamos a nuestro grafico de QQ o cuantil-cuantil, si obtenemos una ligera curvatura en la grafica nos dará señales de que posiblemente los datos no sean tan normales como pensamos, en este caso observamos una ligera curvatura por lo tanto podemos inferir que posiblemente estén sesgados

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

Vamos a instalar las paqueterias basicas para el analisis de datos estadisticos Con esta información que ya tenemos donde: 1 nuestros datos pueden ser normales, pero están sesgados a pesar de esto, sabemos que existe varianza mas alta que el promedio Podemos pasar al verdadero análisis estadístico ya que sabemos la naturaleza de nuestros datos.

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

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

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

Con el paquete plotrix podemos obtener nuestros boxplots, pero primero, si queremos saber la longitud de nuestros datos o cuantos datos hay en la variable consumption utilizamos la funcion “length”

```
length(consumption)
```

Para pasar a nuestro primer análisis tenemos que llamar a un objeto con nuestros datos totales, en este caso lo llamaremos d

```
d<-Copia_de_Oxygen_consumption
```

Para realizar la primera prueba estadística básica haremos una T de Student simplemente con el comando “t.test” de nuestra variable d, obtenemos una p menor del 0.05 por lo cual podemos inferir que las medias de nuestros datos son diferentes :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)
```

Tambien podemos hacer subsets o sub conjuntos de nuestros datos, en este caso queremos dividirlos por especies, tenemos dos especie 1 y 2, al aplicar este tratamiento podemos observar solo los datos que se contengan en este sub set

```
sp1<-subset(d, species==1)  
sp2<-subset(d, species==2)  
print(sp1)  
print(sp2)
```

Vamos a pasar a nuestro segundo plot para visualizar los datos que tenemos, en este caso llamamos a la función `boxplot` y hacemos un comparativo entre el consumo de oxígeno y las dos especies para eso usamos el operador `o tilde`, para determinar que estamos buscando una relación entre estas dos variables, el color se lo damos con la función `col` y el título se lo podemos editar con el comando `main` y los nombres de los ejes se editan con `xlab` y `ylab`

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

Llamamos a nuestra librería básica de stats esta nos va a permitir hacer los análisis estadísticos básicos, vamos a comenzar con la chi cuadrada, la función básica `chisq.test` para nuestros datos de consumo de oxígeno, esto nos va a dar un p value menor del 0.05 muy apenas, por lo cual podríamos inferir que estos datos están asociados Si p-Value es < 0.05 se rechaza la H_0 , hay evidencia que las variables están asociadas Si p-Value > 0.05 No se rechaza la H_0 , no hay suficiente evidencia para decir que las variables están relacionadas

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

Después pasaremos a la prueba de F de Fisher que nos permite comparar varianzas entre datos, en este caso suponemos una probabilidad estadística con un 95% de confianza de que nuestros datos, esto significa que: si repitiéramos el experimento muchas veces, el 95% de las veces el intervalo de confianza contendría el verdadero valor poblacional. En otras palabras suponemos un término denominado alfa, donde alfa sería de 0.05, con una n imaginaria de 143 y de grados de libertad de 47 Para realizar esto primero tenemos que crear el objeto que va a contener el cálculo matemático de la siguiente manera: Para ver los resultados solo damos al comando `print` de nuestro objeto que creamos con el cálculo matemático, el valor que nos arroja que el valor crítico calculado es de 1.5, esto nos dice que si el valor estadístico calculado es mayor que este significa que al menos un grupo es significativamente diferente, si es menor que 1.51 indica que no hay diferencias significativas entre los grupos. Con esta función podemos evitarnos buscar en tablas para calcular el valor de F Si p-Value < 0.05 se rechaza la H_0 , hay una diferencia significativa entre medias o varianzas Si p-Value > 0.05 NO se rechaza la H_0 , no hay suficiente evidencia para afirmar que las varianzas o medias sean diferentes

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

Finalmente haremos un intento de anova, lo que se tiene que hacer es crear un objeto con la función `aov` en seguida y con la tilde determinamos que datos queremos comparar, en este caso las especies con el consumo de oxígeno si p-Value < 0.05 , se rechaza la H_0 , al menos un grupo tiene una media diferente Si p-Value > 0.05 No se rechaza la H_0 , no hay suficiente evidencia de diferencias entre grupos. Con este set de datos:

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

Obtenemos muchos valores en este caso, donde Df son los grados de libertad, sum sq es la suma de los cuadrados, Mean sq es la suma de cuadrados entre los grados de libertad, F value nos da el valor estadístico de Fisher y finalmente el Pr o valor asociado a F (P value) mide la probabilidad de obtener un F tan extremo si la hipótesis nula fuera cierta, en este caso obtuvimos un P-value mayor de 0.05 por lo cual podemos inferir que no hay suficiente evidencia de diferencias entre los grupos.