

Laboratorio N. 1

1. Tema: Estadística descriptiva.

2. Objetivos:

- Desarrollar habilidades para el manejo y análisis de datos a través del software R.
- Argumentar soluciones a problemas propuestos a partir del comportamiento de los datos.
- Ayudar a desarrollar un pensamiento que reconoce el elemento de la variabilidad.

3. Pregunta problematizadora:

Se cuenta con dos drogas diferentes para tratar una enfermedad. Se seleccionaron 390 personas para participar en un estudio a quienes se les proporcionó durante un periodo de tres meses una de las dos drogas. Al final del estudio se cuenta con información acerca del éxito (E) o el fracaso (F) de la droga. Con la información suministrada en la base de datos (tratamiento), se desea dar respuesta a la pregunta: ¿Cuál de las drogas parece ser más efectiva para el tratamiento de la enfermedad?

4. Resultados de aprendizaje:

- 4.1. Diferencia variables categóricas de variables cuantitativas (discretas y continuas). Clase magistral 1: duración de 3 horas.
- 4.2. Analiza adecuadamente la información de variables cualitativas o categóricas. Clase magistral 1: duración de 3 horas.
- 4.3. Representa de manera tabular y gráfica (diagrama de barras, diagrama circular) variables categóricas. Clase magistral 1: duración de 3 horas.
- 4.4. Análisis adecuadamente la información de variables cuantitativas. Clase magistral 2: duración de 3 horas.
- 4.5. Representa de forma tabular y gráfica (histogramas, polígonos, diagrama de cajas) variables cuantitativas. Clase magistral 2: duración de 3 horas.
- 4.6. Calcula e interpreta medidas de tendencia central (media, mediana, moda), medidas o índices de posición (percentiles, cuartiles) y medidas o índices numéricos de dispersión o volatilidad (rango, varianza, desviación estándar, coeficiente de variación). Clase magistral 3: duración de 3 horas.

5. Recursos necesarios:

Software R.

RStudio o Rstudio Cloud (editores recomendados para trabajar con R).

6. Fórmulas en R

Operadores aritméticos:

+
-
*
/
^

Vectores

x[1] # primer elemento
x[1:5] # subvector que contiene los cinco primeros elementos.

Matrices y data frames

m[4,] # cuarta fila
m[,3] # tercera columna
dfr[dfr\$var<=30,] # data frame parcial
subset(dfr, var<=30) # mismo resultado del anterior

Importar datos a R:
read.csv()

Interacción con el objeto creado:

str()
dim()
names()
head()
tail()

Creación de tablas de frecuencias:

length()
table()
prop.table()
sum()
cumsum()
seq()
cut()

Creación de gráficos:

par(mfrow=c(2,4))
barplot()
pie()
hist()
boxplot()

Estadística descriptiva:

```
min()  
max()  
mean()  
median()  
sd()  
var()  
quantile()
```

Operadores lógicos:

```
|  
&  
<  
<=  
>  
>=  
==  
!=
```

7. Preguntas a responder por los estudiantes:

1. Importar la base de datos que se encuentra en el archivo tratamiento.csv.
2. Crear una nueva variable que tenga dos categorías. Una para los menores o iguales a 30 años y otra para los mayores a 30 años.
3. Clasificar cada variable según su naturaleza y según su nivel de medición, incluyendo la variable creada en el punto anterior.
4. Realizar análisis de frecuencias unidimensionales para cada variable.
5. Calcular todas las medidas descriptivas (tendencia central, dispersión y posición) para la variable Edad. Interpretar a la luz del problema.
6. Construir un histograma para la variable Edad que tenga exactamente 6 intervalos. ¿Qué se puede decir de la distribución de la variable?
7. Construir un diagrama de caja y bigotes comparativo de la Edad según la variable Resultado. Comente cualquier característica interesante.
8. Construir una tabla de frecuencias relativas cruzada para las variables Droga y Resultado. Presente los resultados totalizados por la variable Droga. ¿Cuál de las drogas parece ser más efectiva para el tratamiento de la enfermedad?
9. Realice el mismo procedimiento del punto anterior, pero esta vez separando por las dos categorías de la variable creada en el punto 2. ¿La conclusión es la misma a la que se llegó en el punto anterior?, ¿Qué puede usted concluir?