

R

- Como empezar
 - Mirar directorio con `getwd()`
 - Cambiar directorio desde Session en caso de que no sea el deseado
 - Meter los datos en el directorio
 - Guardar los datos mediante:
 - `datos<-read.table("datos.txt", header=TRUE, sep = "\t", quote = "\"", dec = ".")`
- Funciones
 - `c(X:Y)`
 - Crea una lista con todos los números entre X y Y.
 - `seq(X,Y, length.out=Z)`
 - Crea una lista de Z elementos entre X y Z.
 - `seq(X, by=Y,length.out=Z)`
 - Crea una lista de Z elementos empezando en X con distancia Y.
 - `rep(X,Y)`
 - Repite X elemento Y veces.
 - `summary()`
 - Muestra la información
 - `sum()`
 - Sumatorio
 - `which(X)`
 - Devuelve la posición de X
 - `data.frame()`
 - Crea un data frame
 - `colnames()`
 - Permite cambiar el nombre de las columnas
 - `subset()`
 - Obtiene los elementos de una secuencia que cumplan una condición
 - `subset(df, Grado=="M")`
 - `table(dato)`
 - Calcula **frecuencia absoluta** de unos datos
 - `prop.table(dato)`
 - Calcula la **frecuencia relativa** (el dato es la frecuencia absoluta)
 - `as.vector()`
 - Convertimos en un vector. Sirve para quitar los nombre de las columnas
 - `cumsum()`
 - Sumatorio. Sirve para calcular **frecuencias acumuladas**
 - `barplot()`
 - Diagrama de barras
 - `barplot(tabla.f$Fr.Abs)`
 - `barplot(tabla.f$Fr.Abs,xlab="Grado",ylab="Frecuencia Absoluta", main = "Diagrama de barras", col = "LightBlue", border = "Black",names.arg = c("EAut","Elec","Mec"))`
 - `pie()`
 - Diagrama de sectores
 - `pie(fabs,main="Diagrama de sectores")`
 - `plot()`
 - Poligono de frecuencias
 - `plot(fabs, type="b",col="red",lwd=3,xlab="Grado",ylab="Frecuencia absoluta",main="Poligono de frecuencias")`
 - `stem()`
 - Diagrama de tallos y hojas
 - `stem(datos$Altura,0.25,85)`
 - `length(X)`
 - Numero de elementos en X
 - `sqrt(X)`
 - Raiz cuadrada de X
 - `round(X)`
 - Redonde X
 - `min, max()`
 - Calculan el maximo y minimo de unos datos
 - `L<-min(datos$Altura);H<-max(datos$Altura);L;H;A<-(H-L)/k;A`
 - Sirven para calcular los limites
 - `lim<-seq(L,H,A)`
 - `cut()`
 - Contra entre limites
 - `cut(datos$Altura,lim,include.lowest = TRUE,right = FALSE)`
 - `sort()`
 - Ordena los datos

- `mean(X, na.rm=TRUE)`
 - Da la media de X
- `meadian(X, na.rm=TRUE)`
 - Da la mediana de X
- `moda`
 - Para calcular la moda
 - `as.numeric(names(which.max(table(X))))`
- `var()`
 - Calcula su quasivarianza
- Varianza
 - Para calcular la varianza real es:
 - `n=length(altura)`
 - `var=((n-1)/n)*cvar;var`

- EJEMPLOS:

#EJERCICIO 2. DATOS CATEGORICOS

#a) OBTENER LA TABLA DE FRECUENCIAS ABSOLUTAS Y RELATIVAS PARA LAS TRES DIFERENTES CATEGORIAS DE LA VARIABLE GRADO

```
fabs<-table(datos$Grado);fabs #Frecuencia Absoluta
frel<-prop.table(fabs);frel #Frecuencia Relativa
frel<-as.vector(frel);frel #Quitamos el nombre de las columnas de frecuencia relativa
tabla.f<-data.frame(fabs,frel);tabla.f #Creamos una tabla de frecuencias
colnames(tabla.f)<-c("Grado","Fr.Abs","Fr.Rel");tabla.f #Cambiamos sus nombres
```

#b) AMPLIAR LA TABLA DE FRECUENCIAS ANTERIOR CON LAS FRECUENCIAS ACUMULAS (ABSOLUTA Y RELATIVA)

```
tabla.f$Fr.Abs.Acum<-cumsum(tabla.f$Fr.Abs);tabla.f #Sumo consecutivamente las frecuencias absolutas y las guardo en una nueva variable
tabla.f$Fr.Rel.Acum<-cumsum(tabla.f$Fr.Rel);tabla.f #Sumo consecutivamente las frecuencias relativas y las guardo en una nueva variable
```

#c) TRAZAR UN DIAGRAMA DE BARRAS DE FRECUENCIAS ABSOLUTAS Y OTRO DE FRECUENCIAS RELATIVAS

```
barplot(tabla.f$Fr.Abs) #Forma mas basica de hacer un diagrama de barras
barplot(tabla.f$Fr.Abs,xlab="Grado",ylab="Frecuencia Absoluta", main = "Diagrama de barras", col = "LightBlue", border = "Black",names.arg = c("EAut","Elec","Mec"))
text(c(0.7,1.9,3.1),tabla.f$Fr.Abs/2,labels=tabla.f$Fr.Abs,col = "red")
#FALTA FR.REL
```

#e) REPRESENTAR LAS FRECUENCIAS ABSOLUTAS DE LA VARIABLE EN UN DIAGRAMA DE SECTORES

```
pie(fabs,main="Diagrama de sectores")
```

#f) TRAZAR UN POLIGONO DE FRECUENCIAS ABSOLUTAS Y OTRO DE FRECUENCIAS RELATIVAS

```
plot(fabs, type="l",col="red",lwd=3,xlab="Grado",ylab="Frecunecia absoluta",main="Poligono de frecuencias")
plot(fabs, type="b",col="red",lwd=3,xlab="Grado",ylab="Frecunecia absoluta",main="Poligono de frecuencias")
text(c(1,2,3),tabla.f$Fr.Abs-5,labels=tabla.f$Fr.Abs,col = "red")
```

#EJERCICIO 3. DATOS NUMERICOS

#a) DETERMINAR EL NUMERO DE INTERVALOS DE CLASE Y SU AMPLITUD

```
n<-length(datos$Altura);n
round(sqrt(n))
nclass.Sturges(datos$Altura)
#Hay que entender los dos ultimos valores como recomendaciones
k<-10
L<-min(datos$Altura);H<-max(datos$Altura);L;H;A<-(H-L)/k;A #CALCULAMOS LOS INTERVALOS
diff(range(datos$Altura))/k #Otra Opcion
```

#b) OBTENER LA TABLA DE FRECUENCIAS CON LOS VALORES DE LA VARIABLE AGRUPADOS EN INTERVALOS DE CLASE

```
lim<-seq(L,H,A);lim #Los limites puestos como un Array
cut(datos$Altura,lim,include.lowest = TRUE,right = FALSE)
int.clase<-cut(sort(datos$Altura),lim,include.lowest = TRUE,right = FALSE);int.clase #Ordena
fabsA=table(int.clase);fabsA #Frecuencia Absoluta
frelA=prop.table(fabsA);frelA #Frecuencia Relativa
fabsAcum=cumsum(fabsA);fabsAcum #Frecuencia Absoluta Acumulada
frelAcum=cumsum(frelA);frelAcum #Frecuencia Relativa Acumulada
```

```

intervalos=levels(int.clase);intervalos
t.frec<-data.frame(intervalos,fabsA,frelA,fabsAcum,frelAcum);t.frec

#c) REPRESENTAR UN DIAGRAMA DE TALLO Y HOJAS
stem(datos$Altura)
stem(datos$Altura,2)
stem(datos$Altura,0.25,85)

#d) TRAZAR EL HISTOGRAMA DE LAS FRECUENCIAS ABSOLUTAS DE LOS VALORES DE LA VARIABLE AGRUPADOS EN INTERVALOS DE CLASE
h=hist(datos$Altura,lim,include.lowest=TRUE, right=TRUE, plot=FALSE)
str(h)

hist(datos$Altura,breaks=lim,xaxt="n",right=TRUE,freq=TRUE,col="light blue",main="Histograma frecuencias absolutas", xlab = "Intervalos de clase", ylab="Frecuencias absolutas")
axis(1,at=lim)

=====

# EJERCICIO 4. ESTADISTICOS DE POSICION CENTRAL

#a) OBTENER LAS MEDIAS DE LAS VARIABLES

#Guardamos todos los valores del df en una variable
altura=datos$Altura;altura
peso=datos$Peso
edad=datos$Edad

# Hacemos la media aritmetica
mean(altura, na.rm=TRUE) #MEDIA DE ALTURA
mean(peso, na.rm=TRUE) #MEDIA DE PESO
mean(edad, na.rm=TRUE) #MEDIA DE EDAD

# Redondeamos con round() en caso de necesario

#b) REPETIR EL APARTADO ANTERIOR HACIENDO USO DE LA FUNCION APPLY()
apply(datos[1:3],2,FUN=mean,na.rm=TRUE)

#c) CALCULAR DE NUEVO LAS TRES MEDIAS ANTERIORES ELIMINANDO EL 5% DE LOS VALORES

#MAS EXTREMOS DE CADA UNO DE LAS VARIABLES
apply(datos[1:3],2,FUN=mean,trim=0.05,na.rm=TRUE)

#d) OBTENER LAS MEADIANAS DE LAS VARIABLES
median(altura,na.rm=TRUE)

#e) REPETIR EL APARTADO ANTERIOR HACIENDO USO DE LA FUNCION APPLY()
apply(datos[1:3],2,FUN=median,na.rm=TRUE)

#f) OBTENER LA MODA DE LAS VARIABLES ALTURA (CUANTITATIVA) Y POBLACION (CUALITATIVA)
as.numeric(names(which.max(table(altura))))

=====

#EJERCICIO 5. ESTADISTICOS DE DISPERSION

# a) OBTENER EL RECORRIDO DE LAS VARIABLES
#METODO1
diff(range(altura))
diff(range(peso))
diff(range(edad))

# b) CALCULAR SU VARIANZA
cvar=apply(datos[1:3],2,FUN=var,na.rm=TRUE);cvar #ESTO ES LA QUASIVARIANZA QUE LA PROPIA VARIANZA

```

```

n=length(altura)
var=((n-1)/n)*cvar;var #ESTO SI ES LA VARIANZA
# c) DETERMINAR SU DESVIACION TIPICA
desviacion=sqrt(var);desviacion
c.desviacion=sqrt(cvar);c.desviacion
# d) CALCULAR EL COEFICIENTE DE VARIACION DE PEARSON, ¿CUAL DE LAS TRES VARIABLES PRESENTA MAYOR DISPERSION?
cv=function(x)
{
  n=length(x)
  sd.x=sqrt((n-1)/n)*sd(x)
  sd.x/abs(mean(x))*100
}
cv(altura)

=====

#EJERCICIO 6. ESTADISTICOS DE POSICION NO CENTRAL
#a) CALCULAR LOS SIGUIENTES CUANTILES: C0.10, C0.25, C0.50, C0.75
quantile(altura,0.10,na.rm=TRUE) #CUANTIL 0.10
quantile(altura,0.25,na.rm=TRUE) #CUANTIL 0.25
quantile(altura,0.50,na.rm=TRUE) #CUANTIL 0.50
quantile(altura,0.75,na.rm=TRUE) #CUANTIL 0.75
#b) COMPRAR LOS RESULTADOS OBTENIDOS CON LA TABLA DE FRECUENCIAS DE LA VARIABLE
cumsum(prop.table(table(altura)))
#c) OBTENER TODOS LOS DECILES
quantile(altura,seq(0,1,0.10),na.rm=TRUE) #DECILES (10)
#d) HALLAR EL RANGO INTERCUARTILICO
IQR(altura)

=====

#EJERCICIO 7. NUMEROS RESUMEN
#a) DETERMINAR LOS CINCO NUMEROS RESUMEN DE LAS VARIABLES JUNTO CON SU MEDIA
fivenum(altura)
summary(altura)
#b) OBTENER LOS NUMEROS RESUMEN DE LA VARIABLE ALTRUA CLASIFICADOS SEGUN LOS NIVELES DE LA VARIABLE GRADO
by(altura, datos$Grado,FUN=summary)

=====

# EJERCICIO 8. DIAGRAMAS DE CAJA
#a) TRAZAR UN DIAGRAMA DE CAJA BASICO
boxplot(altura)
#b) REPETIR EL APARTADO ANTERIOR DANDO COLOR A LA CAJA Y AÑADIENDO ALGUN TITUTLO AL GRAFICO
boxplot(altura,main="Boxplot de Altura",ylab="Metros",col="Light Blue", notch=TRUE)
#c) ETIQUETAR EN EL GRAFICO LOS VALORES REPRESENTADOS EN EL DIAGRAMA (HACER USO DE LA INFORMACION ADICIONAL QUE FACILITA BOXPLOT)
b<-boxplot(altura,main="Boxplot de Altura",ylab="Metros",col="Light Blue", notch=FALSE)
b$stats #NOS DA LOS CUANTILES Y LOS EXTREMOS DE LOS BIGOTES
b$out #NOS DA LOS DATOS ATIPIICOS

```

#d) REPRESENTAR EL BOXPLOT EN FORMA HORIZOTAL

```
boxplot(altura,horizontal=TRUE,main="Boxplot de Altura",ylab="Metros",col="Light Blue", notch=FALSE)
```

#e) TRAZAR LOS DIAGRAMAS DE CAJA PARA LA VARIABLE NUMERICA ALTURA CLASIFICADOS SEGUN LOS NIVELES DEL FACTOR GRADO

```
boxplot(altura~datos$Grado,horizontal=TRUE,main="Boxplot de Altura",ylab="Metros",col="Light Blue", notch=FALSE)
```

=====

EJERCICIO 9. ESTADISTICOS DE FORMA

#a) DETERMINAR LA SIMETRIA DE SUS DISTRIBUCIONES

```
skewness(altura) #ASIMETRIA
```

#b) OBTENER LA CURTOSIS DE SUS DISTRIBUCIONES

```
kurtosis(altura) #CURTOSIS
```