PROGRAMA R

En R usamos <- para hacer asignaciones. De este modo, podemos asignar el valor 3 a la variable radio

```
variable<-3
```

Los nombres de las variables pueden incluir letras, números, puntos y guiones bajos. Deben empezar siempre con una letra o un punto y si empiezan con un punto, a este no puede seguirle un número.

Para colocar comentarios:

```
# Este es un comentario
```

Para obtener la documentación de una función, escribimos el ? antes de su nombre y lo ejecutamos. También podemos usar la función help(), con el nombre de la función. Los dos procedimientos siguientes son equivalentes.

```
?mean()
```

```
help("mean")
```

Puedes encontrar cuál es tu directorio de trabajo con la función getwd(). Sólo tienes que escribir la función en la consola y ejecutarla.

```
getwd()
```

Para cambiar la dirección se puede usar.

```
setwd("C:\otro_directorio") #cambiar directorio
```

Los scripts son documentos de texto con la extensión de archivo .R, por ejemplo mi_script.R. Estos archivos son iguales a cualquier documento de texto, pero R los puede leer y ejecutar el código que contienen.

Aunque R permite el uso interactivo, es recomendable que guardes tu código en un archivo .R, de esta manera puedes usarlo después y compartirlo con otras personas. En realidad, en proyectos complejos, es posible que sean necesarios múltiples scripts para distintos fines. Podemos abrir y ejecutar scripts en R usando la función source(), dándole como argumento la ruta del archivo .R en nuestra computadora, entre comillas.

Por ejemplo.

```
source("C:/Mis scripts/mi_script.R")
```

Se pueden abrir documentos de varios tipos donde se encuentren datos:

Para documento tipo .csv:

```
datos<-read.csv(file="bank.csv",sep=";")
```

Estimación puntual de estadísticos descriptivos usando R

Medidas de tendencia central: cálculo de la media, mediana y moda

El cálculo de la media es sencillo en cuanto a los códigos y aritmética necesarios. En las dos siguientes líneas de comandos se observa el cálculo de una media simple y luego de una media recortada. En ambos casos se utiliza el comando mean, indicando la variable sobre la cual

se desea hacer el cálculo, acompañado del argumento `na.rm = TRUE` para excluir del cálculo a los casos dados como perdidos (valores NA que ya fueron codificados al preparar las variables). Para el caso de la media recortada, se agrega el argumento `trim`, que permite indicar la proporción de casos que se eliminan en cada extremo de la distribución. Así, se obtiene la estimación puntual de ambos estadísticos.

```
mean(datos$age, na.rm = TRUE)
```

```
mean(datos$age, na.rm=TRUE, trim=0.025)
```

Para el caso de la mediana, se trata de un comando similar. Se utiliza el comando `median`, indicando la variable sobre la cual se desea hacer el cálculo, acompañado del argumento `na.rm = TRUE` para excluir del cálculo a los casos dados como perdidos (valores NA que ya fueron codificados al preparar las variables). De esa forma se obtiene la estimación puntual de este estadístico.

```
median(datos$age, na.rm=TRUE)
```

Para el cálculo de la moda, debe haberse instalado previamente el paquete `modeest` y haberlo cargado en la sesión de trabajo; esto permitirá contar con una función que calculará de forma automática el o los valores más frecuentes de la distribución, herramienta que no existe en los paquetes básicos de R. El resultado de este comando indica el valor, o los valores, con más frecuencia dentro de la distribución de casos de la variable señalada.

```
#Ejecutar previamente "install.packages("modeest")"
```

```
library(modeest)
```

```
mfv(datos$age) #Indica el o los valores con más frecuencia
```

Medidas de posición: cálculo de frecuencias absolutas y relativas, cuantiles

El cálculo de tablas de frecuencias absolutas para una variable se efectúa mediante el comando `table`, indicando como argumento de la función la variable sobre la cual se ejecuta el cálculo.

```
tabla<-table(datos$age)
```

```
tabla1<-table(datos$loan)
```

Para facilitar la comparación con otros datos, estos resultados pueden expresarse como frecuencias relativas. Para el cálculo de frecuencias relativas, se usa el objeto `tabla` de frecuencias simples ya construido (`tabla`): como se ve en el ejercicio 7.5, sobre tal objeto, se ejecuta la función `prop.table` que construye una nueva tabla en la que cada celda es la división simple entre la cantidad de casos de la categoría y el total de casos.

```
prop.table(tabla) #tabla de frecuencias relativas porcentuales
```

```
round(prop.table(tabla)*100,2) #tabla con dos decimales
```

```
cumsum(tabla) #tabla de frecuencias acumuladas
```

```
cumsum(prop.table(tabla)) #tabla de frecuencias acumuladas relativas
```

Una última forma de describir la concentración de casos de una distribución es mediante los cuantiles. Tales estadísticos permiten describir la concentración de casos según cualquier posición relativa de los datos en la distribución, que resulte de interés.

```
quantile(datos$age,prob=c(0.25,0.5,0.75),na.rm=TRUE)
```

Medidas de dispersión: rango, varianza, desviación estándar y coeficiente de variación

```
range(datos$age,na.rm=TRUE) #rango
```

```
min(datos$age,na.rm=TRUE) #mínimo
```

```
max(datos$age,na.rm=TRUE) #máximo
```

```
max(datos$age,na.rm=TRUE)-min(datos$age,na.rm=TRUE)
```

Otras dos medidas de uso generalizado para describir la dispersión de una variable son la varianza y la desviación estándar. El cálculo de ambas medidas sigue la misma lógica en lo que refiere a la estructura de la función utilizada. Respectivamente, las funciones utilizadas son var para varianza y sd.

```
var(datos$age,na.rm=TRUE) #varianza
```

```
sd(datos$age,na.rm=TRUE) #desviación estándar
```

Coeficiente de variación

#Asegurarse de ejecutar previamente el comando
"install.packages("FinCal")

```
library(FinCal)
```

```
coefficient.variation(sd=sd(datos$age),avg=mean(datos$age))
```

TABLAS DE FRECUENCIA

names(datos): Nombre de las variables de un documento

as.data.frame(table(datos\$age)): Tabla de frecuencias absolutas

tabla2<-as.data.frame(table(datos\$age)): Tabla de frecuencias guardadas

transform(tabla1, Fre_Ac=cumsum(tabla1\$Freq), Fre_Re=round(prop.table(tabla1\$Freq), 3), Fre_Rel_Acu=cumsum(round(prop.table(tabla1\$Freq), 3))):
Tabla de frecuencias absolutas, absolutas acumuladas, relativas y relativas acumuladas