

UNIDAD 2: Práctica 06 - Análisis de datos categóricos

ESCALAS DE MEDICIÓN

Como la estadística analiza los datos y éstos son producto de las mediciones, necesitamos estudiar las escalas de medición. Este tema es de suma importancia, pues el tipo de escala de medición utilizado para reunir los datos ayuda a determinar el tipo de análisis a utilizar en los datos.

Existen cuatro clases de escalas que aparecen de manera común en las ciencias: nominal, ordinal, de intervalo y de razón. Ellas difieren en el número de atributos matemáticos que poseen.

Los tipos de datos univariados que vamos a analizar en esta práctica son:

Categóricos. Tienen la característica de que todos los miembros de una categoría se consideran iguales en lo que se refiere a ese tipo. Este tipo de datos se subdivide en nominales y ordinales

- **Nominales.** Los valores que pueden asumir sirven para clasificarlos pero no para ordenarlos. En caso de usarse números, sólo se adoptan como nombres o identificaciones.
- **Ordinales.** Los valores que puede asumir este tipo de datos son categorías que conllevan un juicio de valor que exige comparar a los diferentes elementos de la muestra con respecto a este tipo con el objeto de establecer un orden. Es decir, que los datos se organizan a través de las relaciones de igualdad, mayor o menor

1 ANÁLISIS ESTADÍSTICO DE DATOS CATEGÓRICOS

Ejemplo: Se realiza un estudio para conocer las preferencias sobre el tipo de gaseosa que se consume: "CC"=Coca Cola, "PC"=Pepsi Cola, "SC"=Salva Cola, para ello se toma una muestra aleatoria de 20 personas.

- 1) Activar el directorio de trabajo

```
getwd()
## [1] "C:/Users/user/OneDrive/Paquete R/PRACTICAS-S2"
setwd("C:/Users/user/OneDrive/Paquete R/PRACTICAS-S2")
```

- 2) Crear un nuevo script y llamarle Script06-DatosCategoricos

- 3) Crear un vector con el tipo de gaseosa y otro con la muestra generada aleatoriamente:

```
Tipo <- c("CC", "PC", "SC")
Tipo
```

```
## [1] "CC" "PC" "SC"

#crea un vector en las que contiene los tres tipos de refrescos
Consumo <- sample(Tipo, 20, replace=TRUE)
Consumo

## [1] "CC" "SC" "SC" "SC" "PC" "SC" "SC" "SC" "CC" "PC" "SC" "PC" "SC" "PC" "CC"
## [16] "PC" "SC" "PC" "PC" "CC"

#Genera una muestra de tamaño 20 obtenida de los elementos del vector Tipo
#y los elementos se seleccionan con reemplazamiento

#Suponiendo que se quiere editar o agregar datos
data.entry(Consumo)
```

4) Guarde el vector en un archivo de datos

```
#Guardar los datos en su directorio de trabajo
write(Consumo, "Consumo.txt")
```

5) Eliminar los objetos que existen en el espacio de trabajo (Workspace)

```
ls()

## [1] "Consumo" "Tipo"

rm(list=ls(all=TRUE))
ls()

## character(0)
```

6) Leer o recuperar el vector de datos o archivo de texto

```
Consumo<-scan("Consumo.txt", what = character(), na.strings = "NA",
flush=FALSE)
Consumo

## [1] "CC" "SC" "SC" "SC" "PC" "SC" "SC" "SC" "CC" "PC" "SC" "PC" "SC" "PC" "CC"
## [16] "PC" "SC" "PC" "PC" "CC"

ls()

## [1] "Consumo"

# Si el vector contiene caracteres se ocupa: what = character()
# na.strings = "NA", le indica a R que los valores faltantes son identificados con \NA"
```

7) Crear la tabla de distribución de frecuencias y proporciones

```
frec<-table(Consumo)
frec

## Consumo
## CC PC SC
## 4 7 9
```

```
prop<-table(Consumo)/length(Consumo)
prop

## Consumo
##   CC   PC   SC
## 0.20 0.35 0.45
```

Note que la salida por defecto no es para nada atractiva en comparación con el resto de paquetes estadísticos

En cambio, si estamos usando LATEX y queremos incorporar estos cuadros o cualquier otro podemos utilizar el comando `xtable(table(Consumo))` (NOTE QUE EL ARGUMENTO DEBE SER UN CUADRO), y con esto automáticamente se nos genera el código en LATEX y luego incorporarlo a nuestro informe, lo mejor de todo es que salida resultante es mucho más presentable.

8) Conocer un resumen de los datos

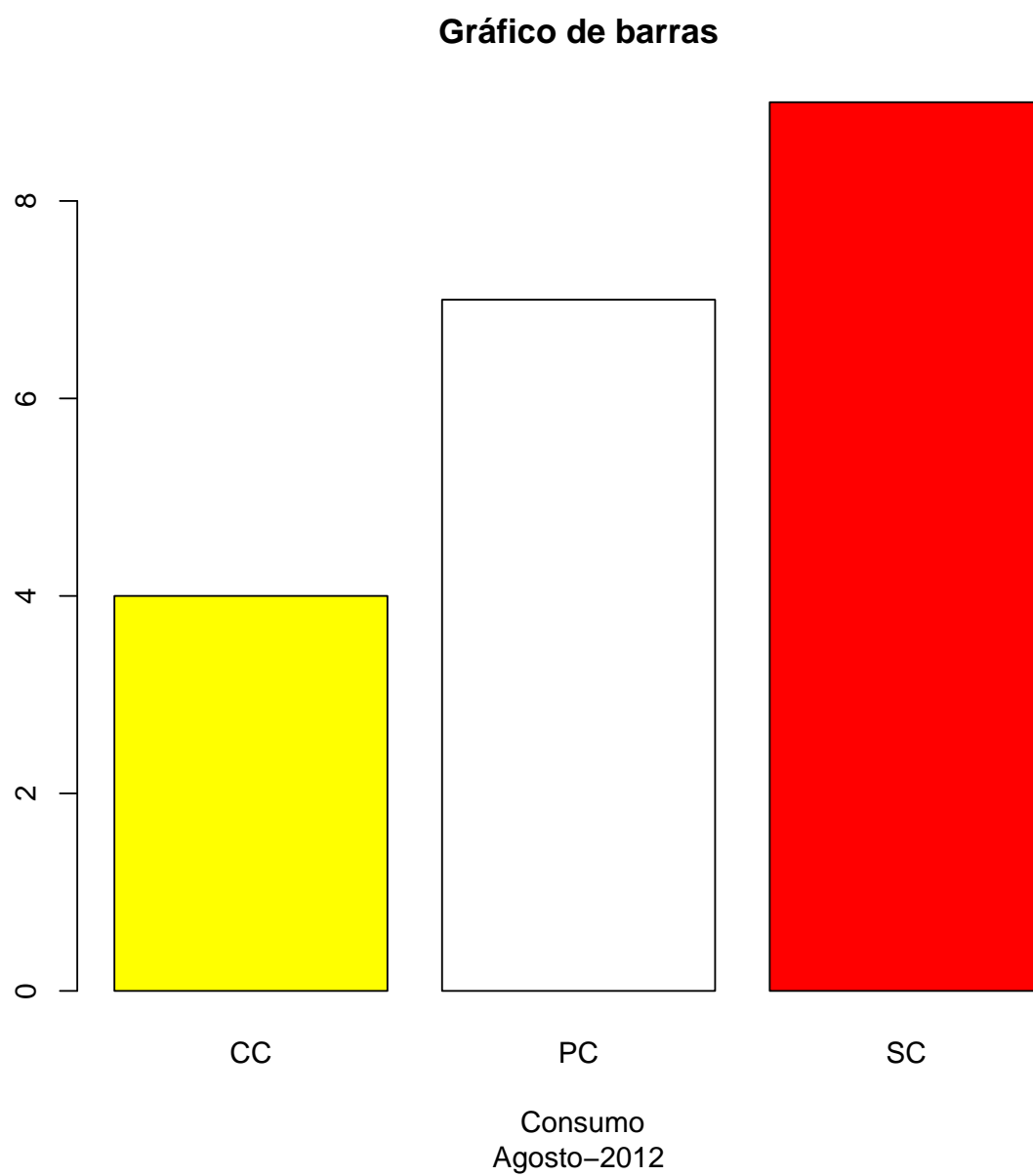
```
summary(Consumo)

##      Length      Class      Mode 
##          20 character character

# note que por tratarse de variables cualitativas únicamente muestra el número de elementos, y el tipo
```

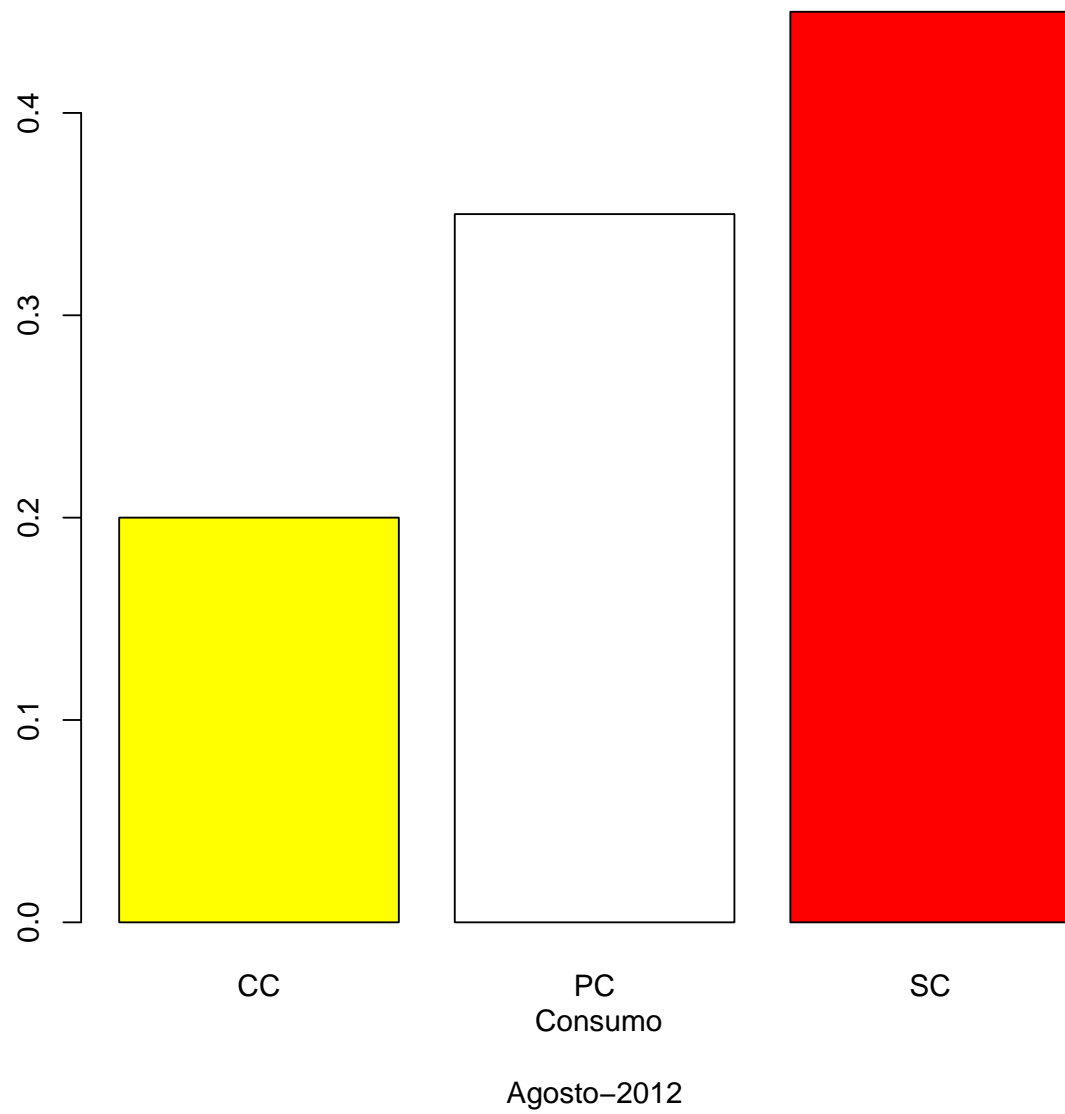
9) Realizar un gráfico de barras

```
# Para las frecuencias absolutas
barplot(frec, main="Gráfico de barras", xlab=" Consumo", col=c("yellow", "white", "red"), sub="Agosto")
```



```
# Para las frecuencias relativas  
barplot(prop, main="Gráfico de barras", xlab=" Consumo\n", col=c("yellow", "white","red"), sub="Agosto-2012")
```

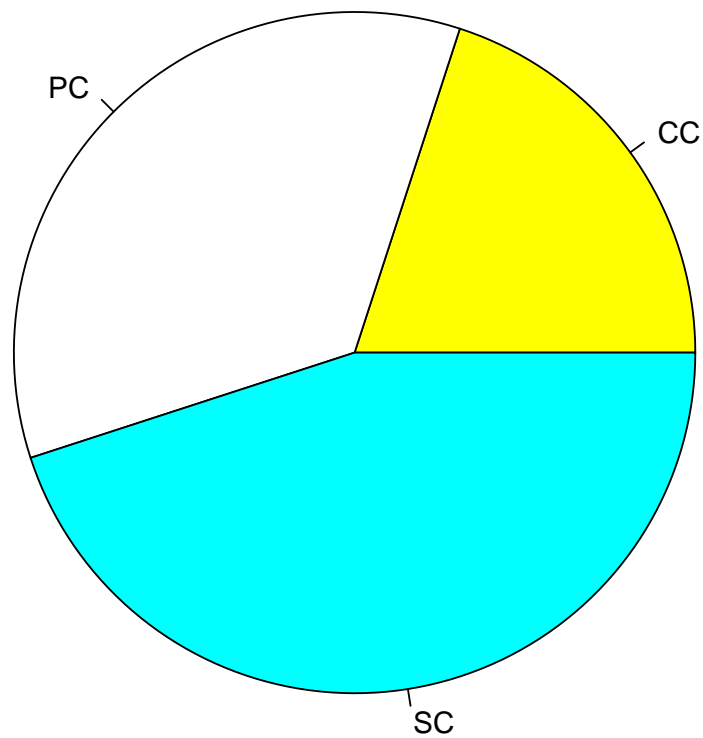
Gráfico de barras



10) Realizar un gráfico de pastel

```
pie(frec, main="Gráfico de pastel", xlab="Tipo de Consumo", col=c("yellow", "white", "cyan"), sub="Agosto-2012")
```

Gráfico de pastel

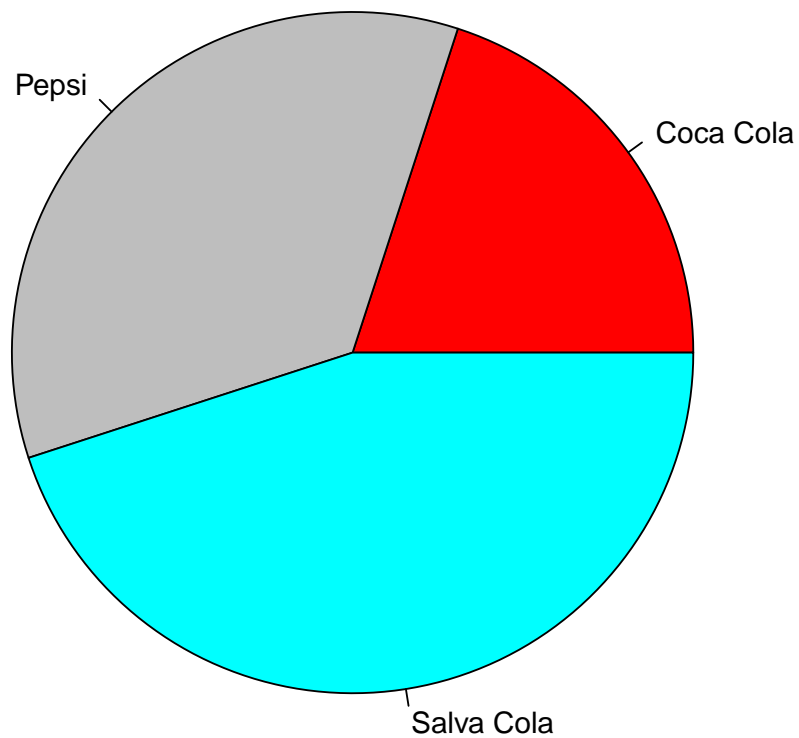


Tipo de Consumo
Agosto-2012

Se puede especificar nombres para las categorías y el color de los sectores

```
names(frec)=c("Coca Cola", "Pepsi", "Salva Cola")  
pie(frec, main="Gráfico de pastel", xlab=" Consumo", radius=0.8, col=c("red", "gray", "cyan"), sub="Agosto-2012")
```

Gráfico de pastel



Consumo
Agosto-2012

Los colores se asignan dependiendo del orden en que han sido especificados por `names()`

Note con la instrucción `radius` se especifica el tamaño de la figura, mientras más cerca de uno (uno de menos uno) se encuentre más grande será (el ángulo cambia).

11) Colocar valores numéricos en los sectores del gráfico

```
n <- length(frec)

hoja <- data.frame(frec)
hoja

##      Var1 Freq
## 1 Coca Cola   4
## 2 Pepsi      7
## 3 Salva Cola  9
```

```
etiq <- c(paste(hoja$Var1, "-", hoja$Freq))
etiq

## [1] "Coca Cola - 4" "Pepsi - 7"      "Salva Cola - 9"

pie(frec, main="Gráfico de pastel", labels=etiq, col=rainbow(n), border=TRUE)
```

Gráfico de pastel

