

# Análisis Estadístico con R

## Estadística descriptiva

CNE/ISCIH

# Estructura del curso

- ① Variable cuantitativa
- ② Variable cuantitativa agrupada
- ③ Variable cualitativa
- ④ Variable cualitativa agrupada

# Section 1

## Datos

## Carga de datos

```
rm(list=ls())  
  
setwd("/Users/pfernandezn/Desktop/datos")  
  
load("datos.curso1.RData")
```

## Section 2

### 1. Variable cuantitativa

## Estadísticos de tendencia central

```
mean(datos$"peso")
```

```
## [1] 69.94747
```

```
median(datos$"peso")
```

```
## [1] 69.75524
```

```
exp(mean(log(datos$"peso")))
```

```
## [1] 69.21468
```

## Medidas de posicion

```
quantile(datos$"peso")
```

```
##           0%          25%          50%          75%          100%  
## 56.56025 59.98709 69.75524 80.00352 82.28081
```

```
quantile(datos$"peso",prob=seq(0,1,1/4))
```

```
##           0%          25%          50%          75%          100%  
## 56.56025 59.98709 69.75524 80.00352 82.28081
```

```
quantile(datos$"peso",prob=seq(0,1,1/3))
```

```
##           0% 33.33333% 66.66667%          100%  
## 56.56025 60.39460 79.53670 82.28081
```

## Categorización de variables cuantitativas (funcion cut)

```
datos$"peso.gr" <- cut(datos$"peso",breaks=quantile(datos$"peso",  
  prob=seq(0,1,1/3)),right=TRUE, include.lowest=TRUE)  
  
table(datos$"peso.gr",exclude=NULL)
```

```
##  
## [56.6,60.4] (60.4,79.5] (79.5,82.3]  
##           67           66           67
```



## Medidas de dispersion

```
sd(datos$"peso")
```

```
## [1] 10.11981
```

```
var(datos$"peso")
```

```
## [1] 102.4105
```

```
IQR(datos$"peso")
```

```
## [1] 20.01642
```

## Medidas de dispersion

```
range(datos$"peso")
```

```
## [1] 56.56025 82.28081
```

```
min(datos$"peso")
```

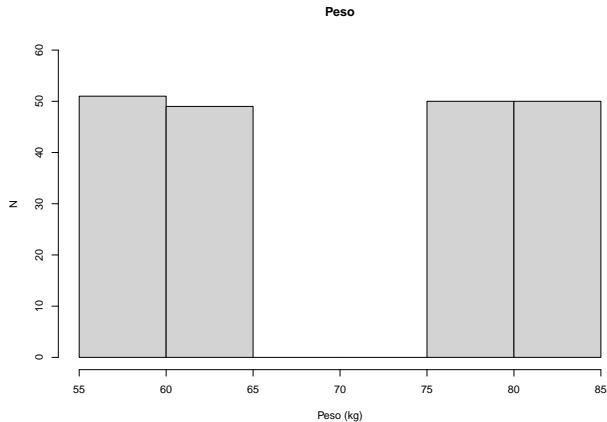
```
## [1] 56.56025
```

```
max(datos$"peso")
```

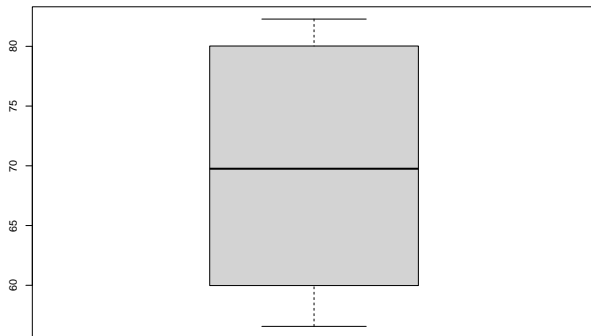
```
## [1] 82.28081
```

# Gráficos

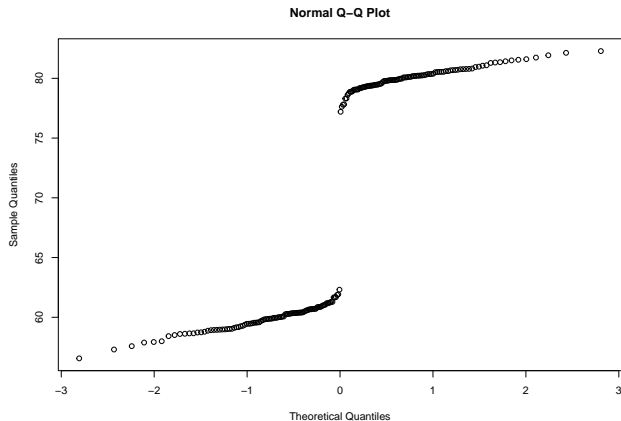
```
hist(datos$"peso",xlab="Peso (kg)",ylab="N",main="Peso",ylim=c(0,60))
```



```
boxplot(datos$"peso")
```



```
qqnorm(datos$"peso")
```



## Summary

```
summary(datos[,c(1:4)])
```

##	ID	edad	sexo	estado.civil
##	Min. : 1.00	Min. : 1.00	Length:200	Length:200
##	1st Qu.: 50.75	1st Qu.:21.00	Class :character	Class :character
##	Median :100.50	Median :42.00	Mode :character	Mode :character
##	Mean :100.50	Mean :42.17		
##	3rd Qu.:150.25	3rd Qu.:63.00		
##	Max. :200.00	Max. :85.00		

## Summary

```
library(crosstable)
library(knitr)
res<-crosstable(datos, c(peso))
kable(res)
```

.id	label	variable	value
peso	peso	Min / Max	56.6 / 82.3
peso	peso	Med [IQR]	69.8 [60.0;80.0]
peso	peso	Mean (std)	69.9 (10.1)
peso	peso	N (NA)	200 (0)

## Tabla

```
tabla<-data.frame(Estadistico=c("media","mediana","media.geometrica","cuart  
"coeficiente de variacion","desviación típica","minimo","maximo"),  
Edad=NA,Peso=NA,Altura=NA,  
stringsAsFactors=FALSE)
```

```
tabla[, "Edad"]<-c(mean(datos$edad), median(datos$edad),exp(mean(log(datos$e  
paste(round(quantile(datos$edad),digits=2),collapse=";"),(sd(datos$edad)/me  
sd(datos$edad),range(datos$edad)[1],range(datos$edad)[2])
```

```
tabla[, "Peso"]<-c(mean(datos$peso), median(datos$peso),exp(mean(log(datos$pe  
paste(round(quantile(datos$peso),digits=2),collapse=";"),(sd(datos$peso)/me  
sd(datos$peso),range(datos$peso)[1],range(datos$peso)[2])
```

```
tabla[, "Altura"]<-c(mean(datos$altura), median(datos$altura),  
exp(mean(log(datos$altura))),paste(round(quantile(datos$altura),digits=2),c  
c(sd(datos$altura)/mean(datos$altura))*100,sd(datos$altura),  
range(datos$altura)[1],range(datos$altura)[2])
```

```
library("openxlsx")  
write.xlsx(tabla,file="/Users/pfernandezn/Desktop/tabla_descriptiva1.xlsx")
```



## Section 3

### 2. Variable cuantitativa agrupada

## Una manera

```
datos.hombre<-datos[datos$"sexo"%in%"Hombre",]
```

```
mean(datos.hombre$"peso") # 79.991
```

```
## [1] 79.991
```

```
datos.mujer<-datos[datos$"sexo"%in%"Mujer",]
```

```
mean(datos.mujer$"peso") # 59.90394
```

```
## [1] 59.90394
```

## Otra manera

```
tapply(datos$"peso",datos$"sexo",mean)
```

```
##      Hombre      Mujer  
## 79.99100 59.90394
```

```
tapply(datos$"peso",datos$"sexo",mean)[1]
```

```
## Hombre  
## 79.991
```

```
tapply(datos$"peso",datos$"sexo",mean)[2]
```

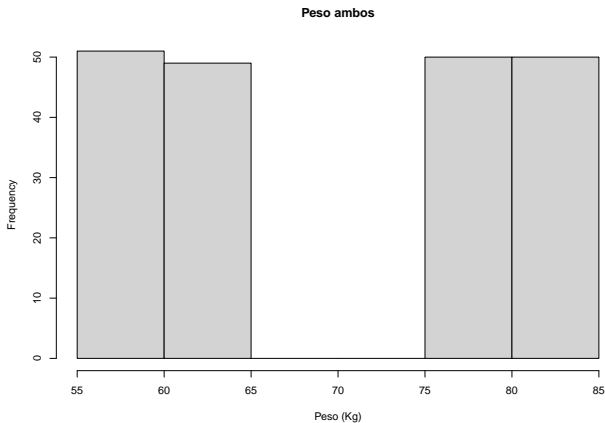
```
##      Mujer  
## 59.90394
```

## Otra manera

```
library(crosstable)
library(knitr)
res<-crosstable(datos, c(peso), by=sexo)
kable(res)
```

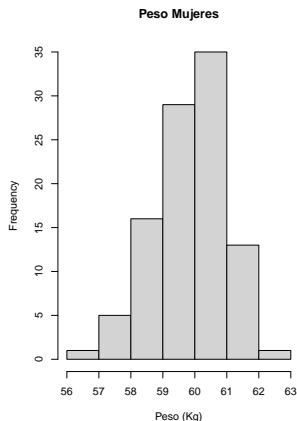
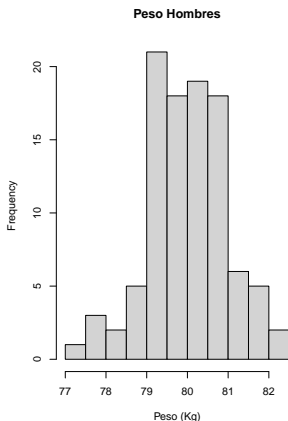
.id	label	variable	Hombre	Mujer
peso	peso	Min / Max	77.2 / 82.3	56.6 / 62.3
peso	peso	Med [IQR]	80.0 [79.4;80.6]	60.0 [59.1;60.7]
peso	peso	Mean (std)	80.0 (1.0)	59.9 (1.1)
peso	peso	N (NA)	100 (0)	100 (0)

```
hist(datos$"peso",main="Peso ambos",xlab="Peso (Kg)")
```

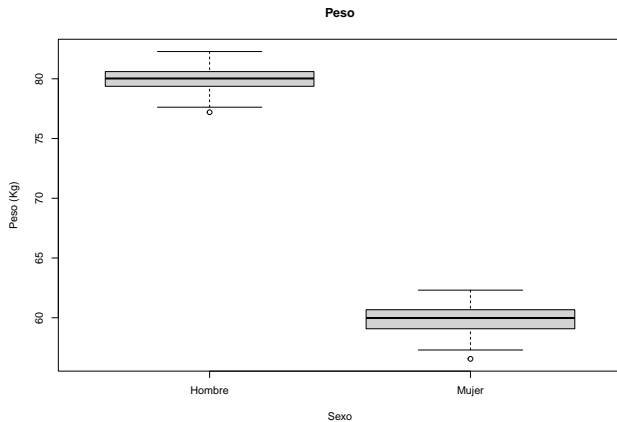


# Graficos

```
par(mfrow=c(1,2))  
hist(datos$"peso"[datos$"sexo"%in%"Hombre"],  
      main="Peso Hombres",xlab="Peso (Kg)")  
hist(datos$"peso"[datos$"sexo"%in%"Mujer"],  
      main="Peso Mujeres",xlab="Peso (Kg)")
```

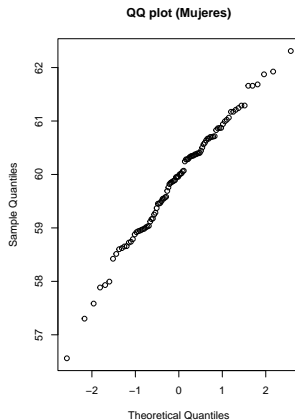
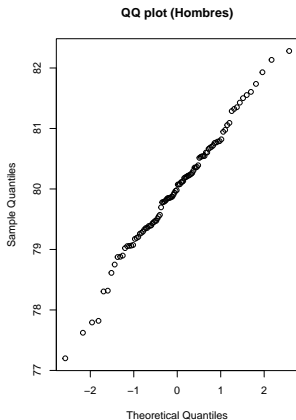


```
boxplot(datos$"peso" ~ datos$"sexo",  
        main="Peso", xlab="Sexo", ylab="Peso (Kg)")
```



# Graficos

```
par(mfrow=c(1,2))  
qqnorm(datos$"peso"[datos$"sexo"%in%"Hombre"],main="QQ plot (Hombres)")  
qqnorm(datos$"peso"[datos$"sexo"%in%"Mujer"],main="QQ plot (Mujeres)")
```





## Section 4

### 3. Variable cualitativa

## Frecuencias absolutas

```
table(datos$"estado.civil",exclude=NULL)
```

```
##  
##      Casado Divorciado   Soltero  
##          50         75         75
```

```
as.data.frame(table(datos$"estado.civil",exclude=NULL))
```

```
##      Var1 Freq  
## 1   Casado   50  
## 2 Divorciado  75  
## 3   Soltero  75
```

```
margin.table(table(datos$"estado.civil",exclude=NULL))
```

```
## [1] 200
```

## Frecuencias relativas

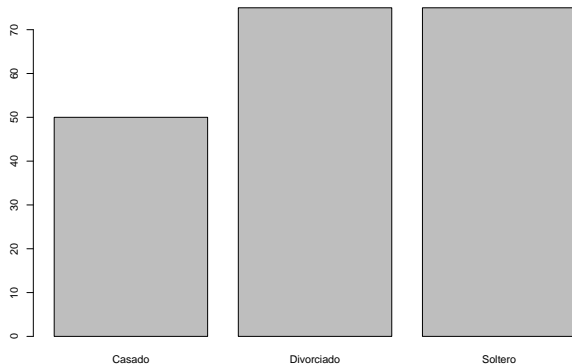
```
prop.table(table(datos$"estado.civil", exclude=NULL))
```

```
##  
##      Casado Divorciado      Soltero  
##      0.250      0.375      0.375
```

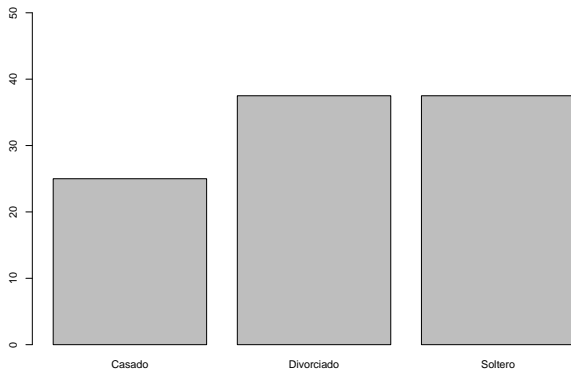
```
prop.table(table(datos$"estado.civil", exclude=NULL))*100
```

```
##  
##      Casado Divorciado      Soltero  
##      25.0      37.5      37.5
```

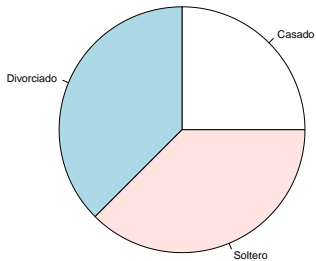
```
barplot(table(datos$"estado.civil"))
```



```
barplot(prop.table(table(datos$"estado.civil"))*100,ylim=c(0,50))
```



```
pie(table(datos$"estado.civil"))
```



## Section 5

### 4. Variable cualitativa agrupada

## Frecuencias absolutas

```
table(datos$"estado.civil",datos$"sexo",exclude=NULL)
```

```
##  
##           Hombre Mujer  
## Casado           30   20  
## Divorciado        39   36  
## Soltero           31   44
```

```
margin.table(table(datos$"estado.civil",datos$"sexo",exclude=NULL),1)
```

```
##  
## Casado Divorciado Soltero  
##      50         75       75
```

```
margin.table(table(datos$"estado.civil",datos$"sexo",exclude=NULL),2)
```

```
##  
## Hombre  Mujer  
##    100    100
```



## Frecuencias relativas

```
prop.table(table(datos$"estado.civil",datos$"sexo",exclude=NULL))*100
```

```
##  
##           Hombre Mujer  
## Casado         15.0  10.0  
## Divorciado     19.5  18.0  
## Soltero        15.5  22.0
```

## Frecuencias relativas

```
prop.table(table(datos$"estado.civil",datos$"sexo",exclude=NULL),1)*100
```

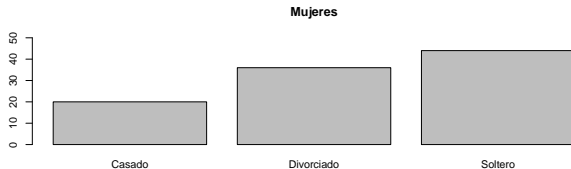
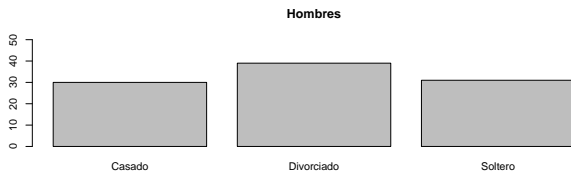
```
##  
##           Hombre  Mujer  
## Casado      60.00000 40.00000  
## Divorciado  52.00000 48.00000  
## Soltero     41.33333 58.66667
```

```
prop.table(table(datos$"estado.civil",datos$"sexo",exclude=NULL),2)*100
```

```
##  
##           Hombre  Mujer  
## Casado      30     20  
## Divorciado  39     36  
## Soltero     31     44
```

## Gráficos

```
par(mfrow=c(2,1))  
barplot(table(datos$"estado.civil"[datos$"sexo"%in%"Hombre"]),  
        main="Hombres",ylim=c(0,50))  
barplot(table(datos$"estado.civil"[datos$"sexo"%in%"Mujer"]),  
        main="Mujeres",ylim=c(0,50))
```



## Tabla frecuencias absolutas y relativas

```
frecu_abs<-as.data.frame(table(datos$"estado.civil",datos$"sexo"))
frecu_rela<-as.data.frame(prop.table(table(datos$"estado.civil",datos$"sexo")
totales_fila<-margin.table(table(datos$"estado.civil",datos$"sexo",exclude=
totales_colu<-margin.table(table(datos$"estado.civil",datos$"sexo",exclude=
total<-margin.table(table(datos$"estado.civil",datos$"sexo",exclude=NULL))
```

```
tabla_fre<-data.frame(Estado_civil=c(names(table(datos$"estado.civil")), "To
Hombre=c(paste(frecu_abs$Freq[1:3], "(", round(frecu_rela$Freq[1:3], 2), "%)", s
Mujer=c(paste(frecu_abs$Freq[4:6], "(", round(frecu_rela$Freq[4:6], 2), "%)", se
Total=c(totales_fila, total))
```

```
tabla_fre
```

##	Estado_civil	Hombre	Mujer	Total
## Casado	Casado	30(60%)	20(40%)	50
## Divorciado	Divorciado	39(52%)	36(48%)	75
## Soltero	Soltero	31(41.33%)	44(58.67%)	75
##	Total	100	100	200

## Summary

```
gmodels::CrossTable(datos$"estado.civil",datos$"sexo",  
  prop.r=TRUE, prop.c=TRUE,prop.chisq=FALSE)
```

```
##  
##  
##      Cell Contents  
## |-----|  
## |                      N |  
## |      N / Row Total |  
## |      N / Col Total |  
## |      N / Table Total |  
## |-----|  
##  
##  
## Total Observations in Table:  200  
##  
##  
##          | datos$sexo  
## datos$estado.civil |      Hombre |      Mujer | Row Total |  
## -----|-----|-----|-----|  
##          Casado |          30 |          20 |          50 |  
##          |          0.600 |          0.400 |          0.250 |
```

## Summary

```
library(crosstable)
crosstable(datos,c(estado.civil),by=sexo)%>%
  as_flextable(keep_id=TRUE)
```