OCE 313
TÉCNICAS DE ANÁLISIS NO PARAMÉTRICO

CLASE 3 - ANÁLISIS EXPLORATORIO DE DATOS

Dr. José Gallardo

marzo 2021

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

PLAN DE LA CLASE

- **ANÁLISIS EXPLORATORIO DE DATOS: ¿Qué? ¿Para qué? ¿Cómo?**
- ORIENTACIONES GENERALES
  - **Tipos de variables**
  - Variación de los datos
  - **Valores típicos y atípicos**
  - Covariación entre variables cuantitativas
  - **Covaruación entre variables cuantitativas y cualitativas**
  - Interacción de factores
  - **Tamaño de los efectos.**

ANÁLISIS EXPLORATORIO DE DATOS

¿Qué es?
Visualizar, transformar y modelar tus datos antés de hacer un análisis estadístico.

¿Para qué?
Lograr el mejor entendimiento posible de tus datos.

¿Cómo proceder?*
Ante todo, un método no formal, creativo e idealmente colaborativo.

* Sólo daremos sugerencias o aproximaciones de como proceder pues la naturaleza de los datos es a veces impredecible.

ORIENTACIONES GENERALES

1. ¿Qué tipos de variables tienen mis datos?

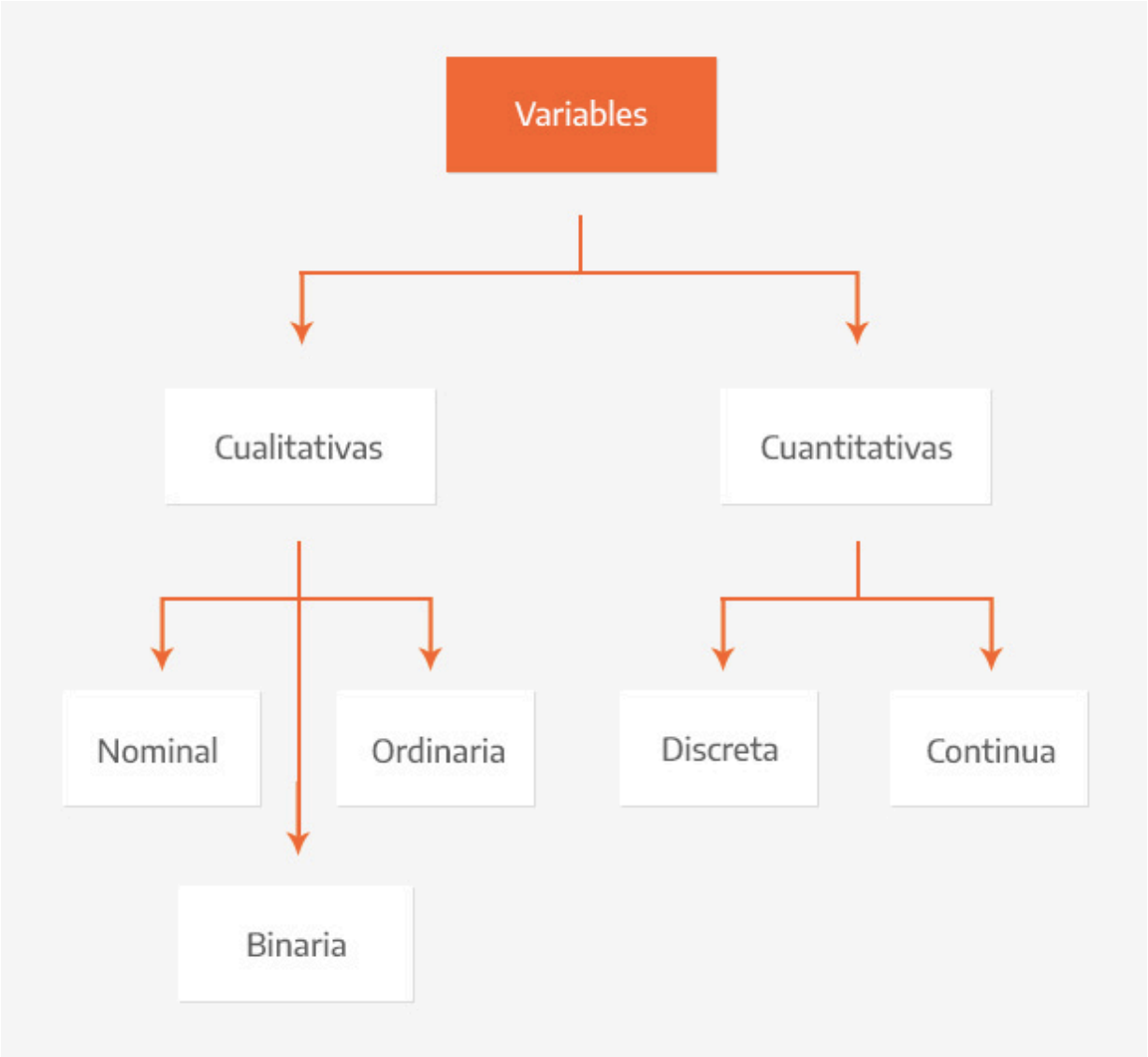
2. ¿Qué variación existe dentro de cada una de mis variables?

3. ¿Cuáles son los valores típicos o más comunes de mis datos?

4. ¿Existen valores atípicos, faltantes, errores de contenido o formato en mis datos?

5.¿Qué tipo de covariación ocurre entre mis diferentes variables?

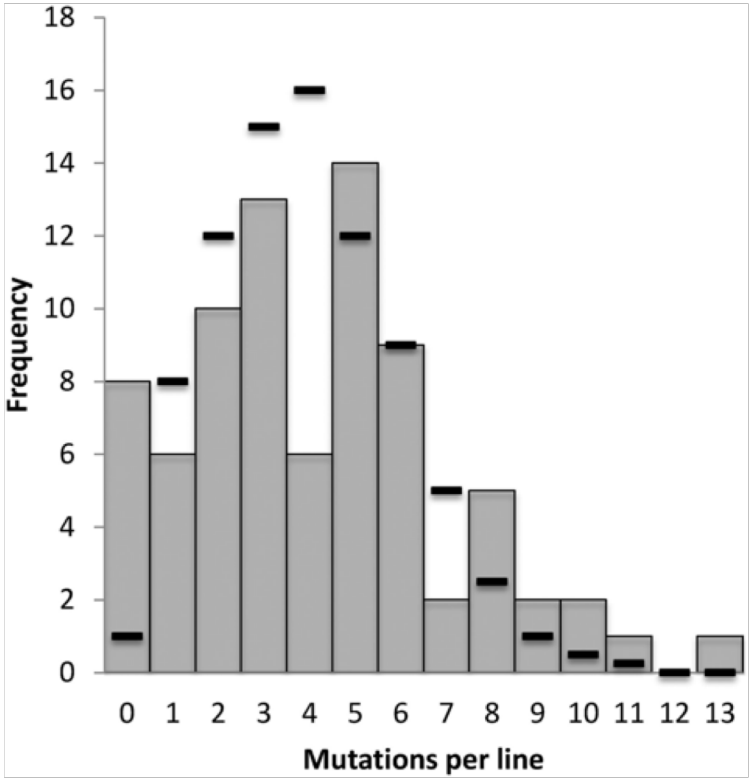
¿QUÉ TIPOS DE VARIABLES?

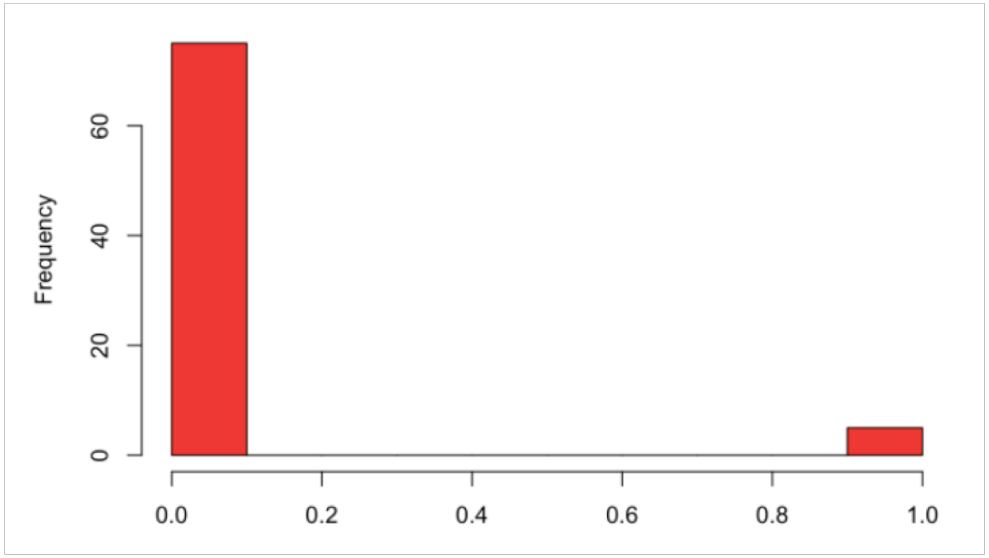


¿EXISTE VARIACIÓN?

 # Histogramas.

hist(dat, col="", main = "", xlab = "")

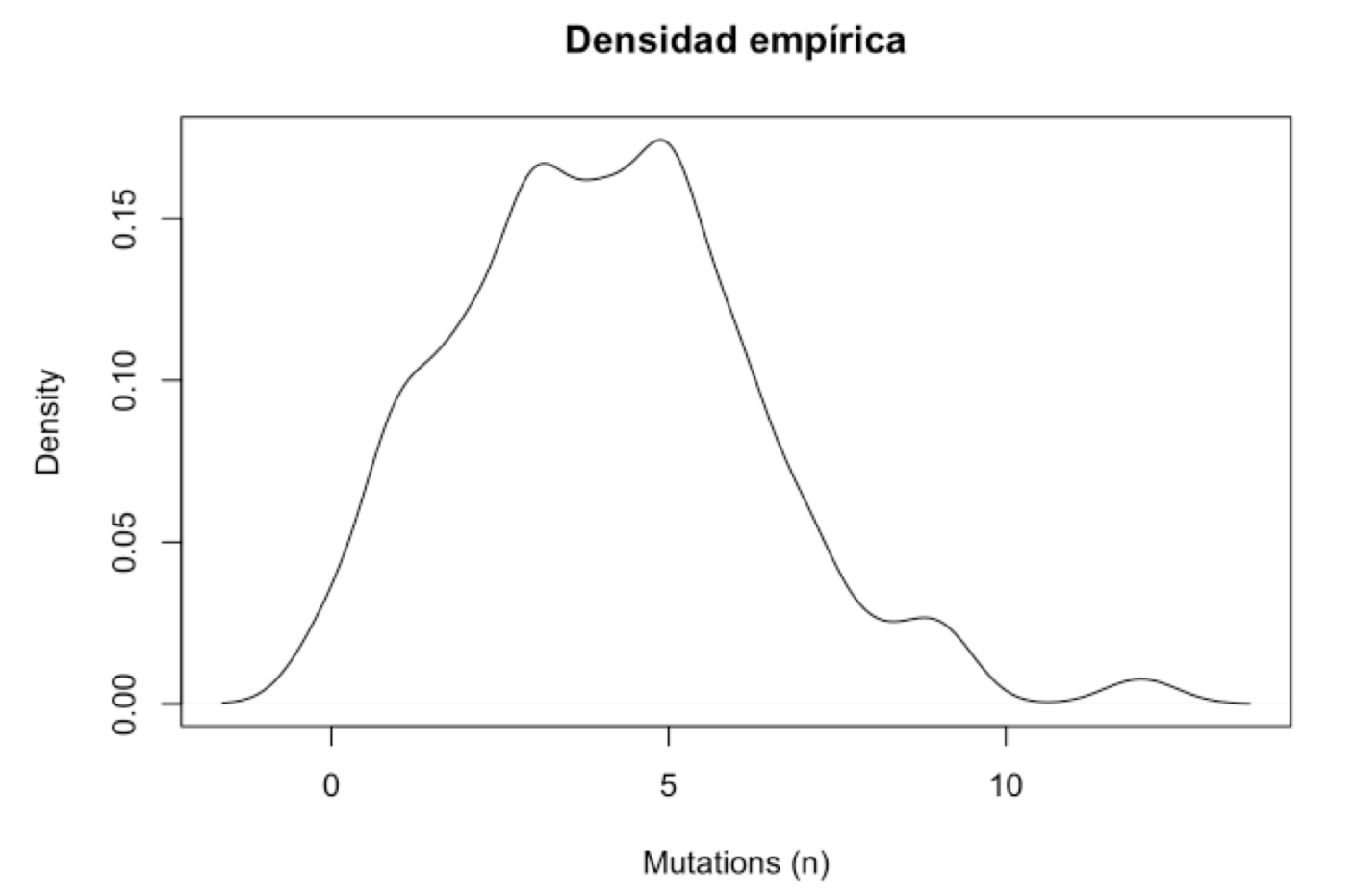




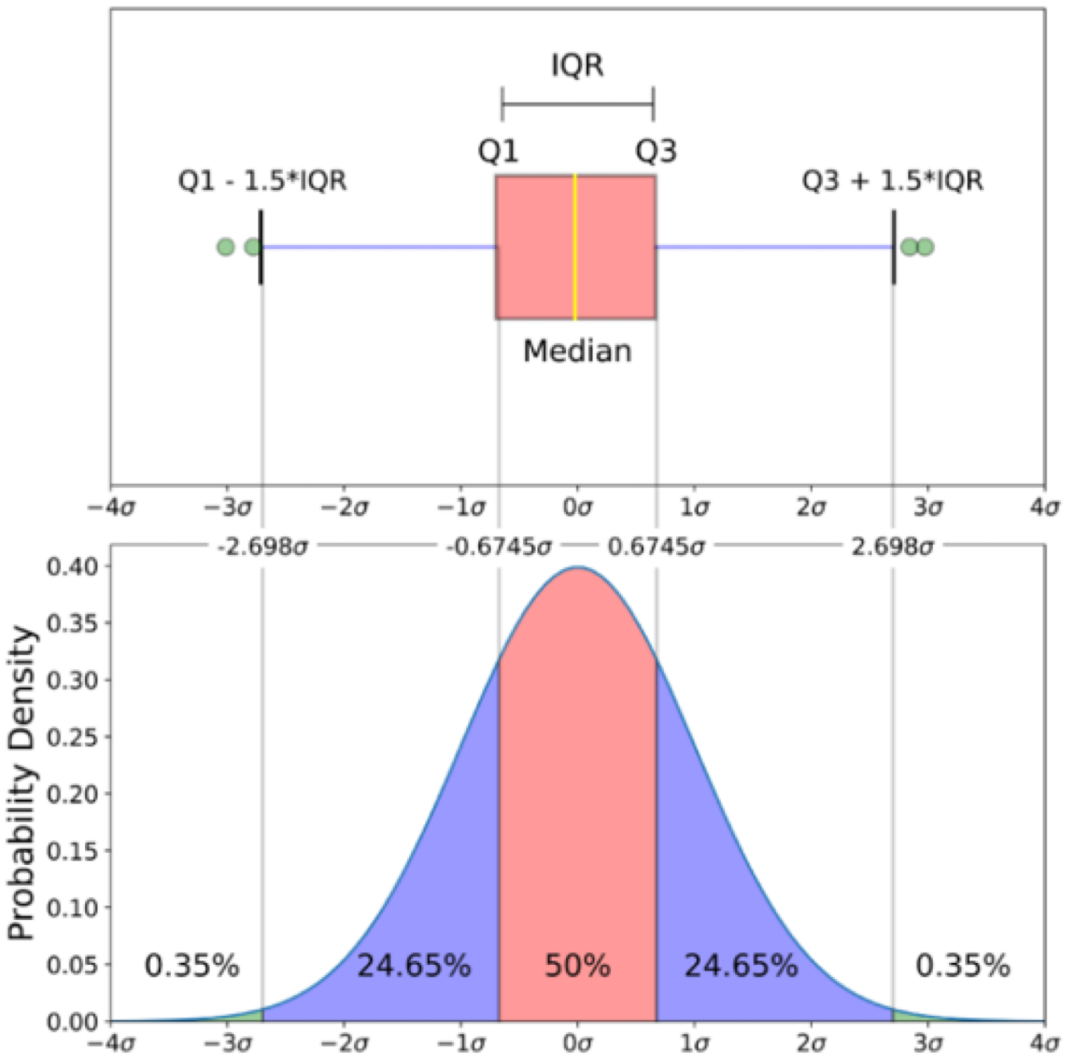
¿EXISTE VARIACIÓN?

 # Distribución de probabilidad empírica

```
plot(density(dat), main="", xlab="")
```



VALORES TÍPICOS Y ATÍPICOS

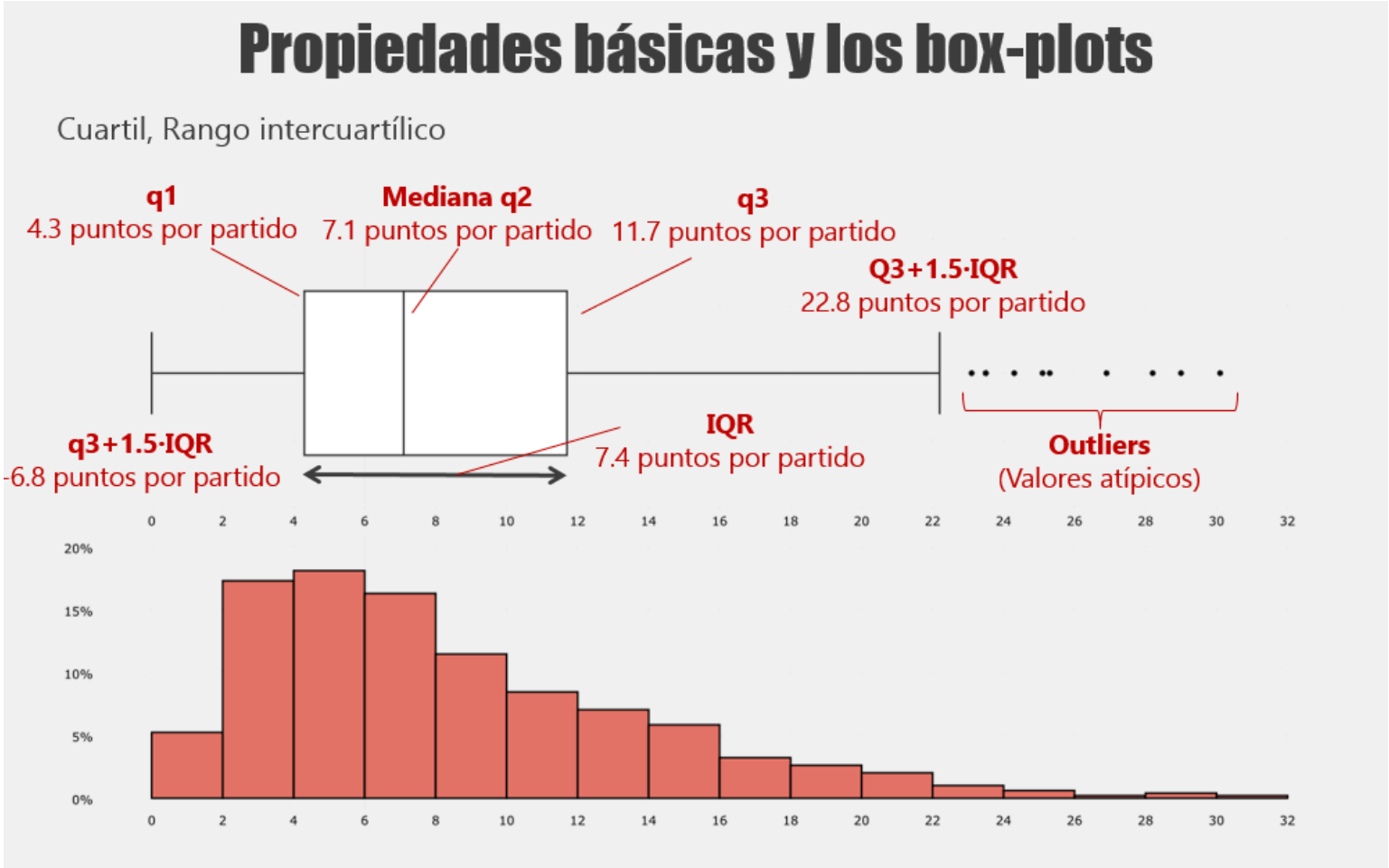


Comparison of a boxplot of a nearly normal distribution and a probability density function (pdf) for a normal distribution

```
# Gráfica de cajas y bigotes
boxplot(dat, col="",
main = "", ylab = "")
```

VALORES TÍPICOS Y ATÍPICOS

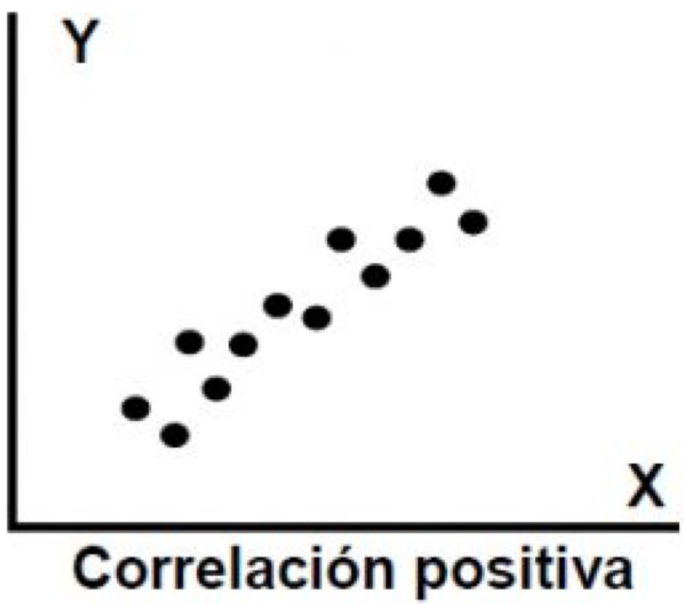
Gráfica de cajas y bigotes 2
(Otras distribuciones)

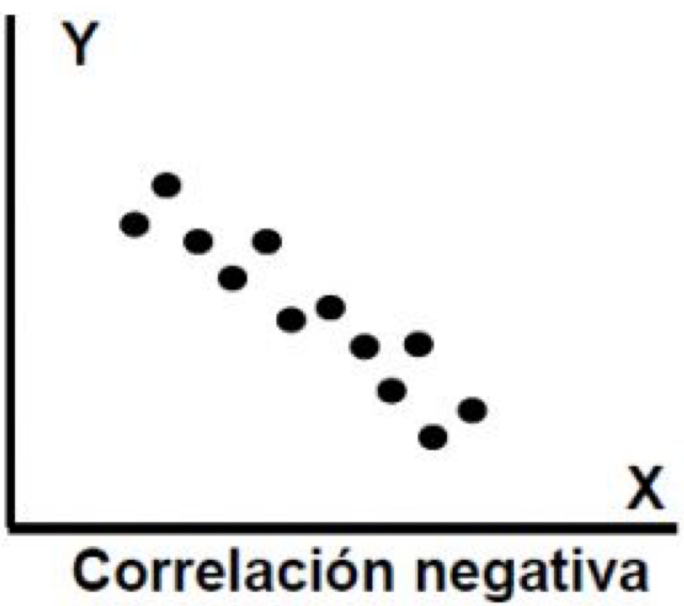


PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

EXISTE COVARIACIÓN ENTRE VARIABLES CUANTITATIVAS: SI

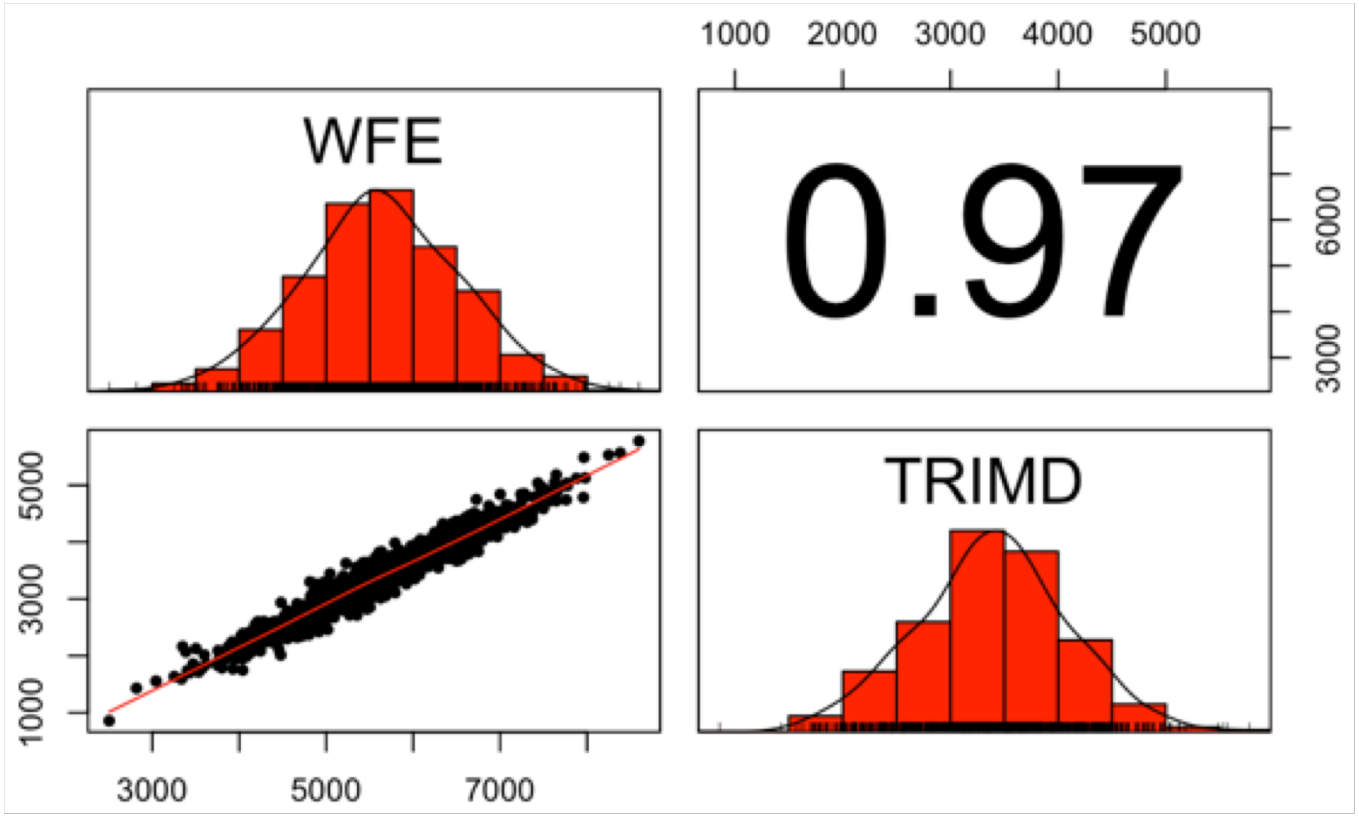
La correlación representa el comportamiento recíproco de dos variables *X* e *Y*. *La correlación no implica causalidad.*





EDA: CORRELACIÓN DE VARIABLES CUANTITATIVAS

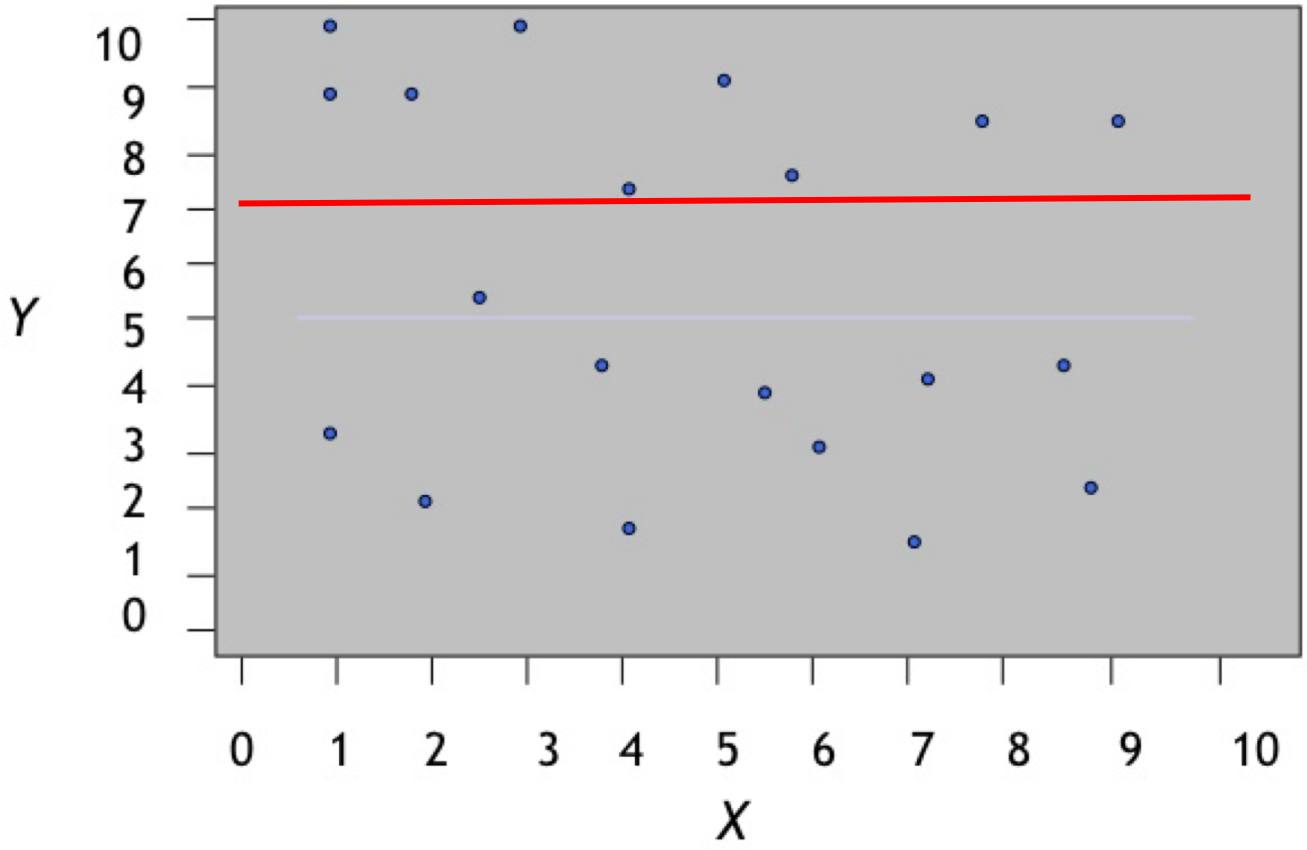
```
# Gráfica de correlación de variables cuantitativas
pairs.panels(dat[,], method = "pearson", hist.col = "red",  density = TRUE)
```



PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

EXISTE COVARIACIÓN ENTRE VARIABLES CUANTITATIVAS: NO

- **Dos variables X e Y también podrían no estar correlacionadas.**



COVARIACIÓN ENTRE CUANTITATIVAS Y CUALITATIVAS (FACTORES)

 # Tablas de 2 factores

table(factor a, factor b)

 # Estadístico y tablas

tapply(factor a, fn(x))

Table 3: Tabla de frecuencia feed x sexo

	casein	horsebean	linseed	meatmeal	soybean	sunflower
0	10	3	7	4	7	8
1	2	7	5	7	7	4

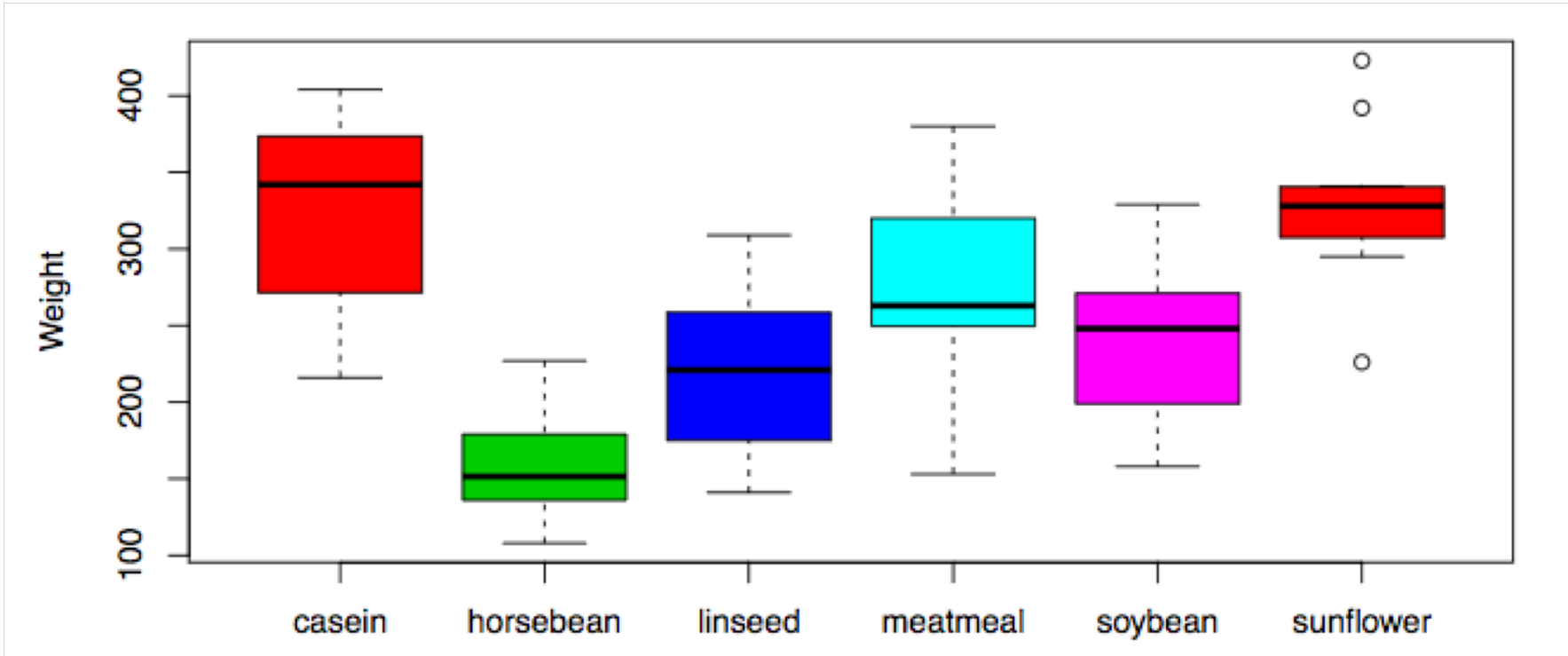
Table 4: Promedio de peso x feed

	Mean
casein	323.5833
horsebean	160.2000
linseed	218.7500
meatmeal	276.9091
soybean	246.4286
sunflower	328.9167

COVARIACIÓN ENTRE CUANTITATIVAS Y CUALITATIVAS (FACTORES)

 # Boxplot con factores

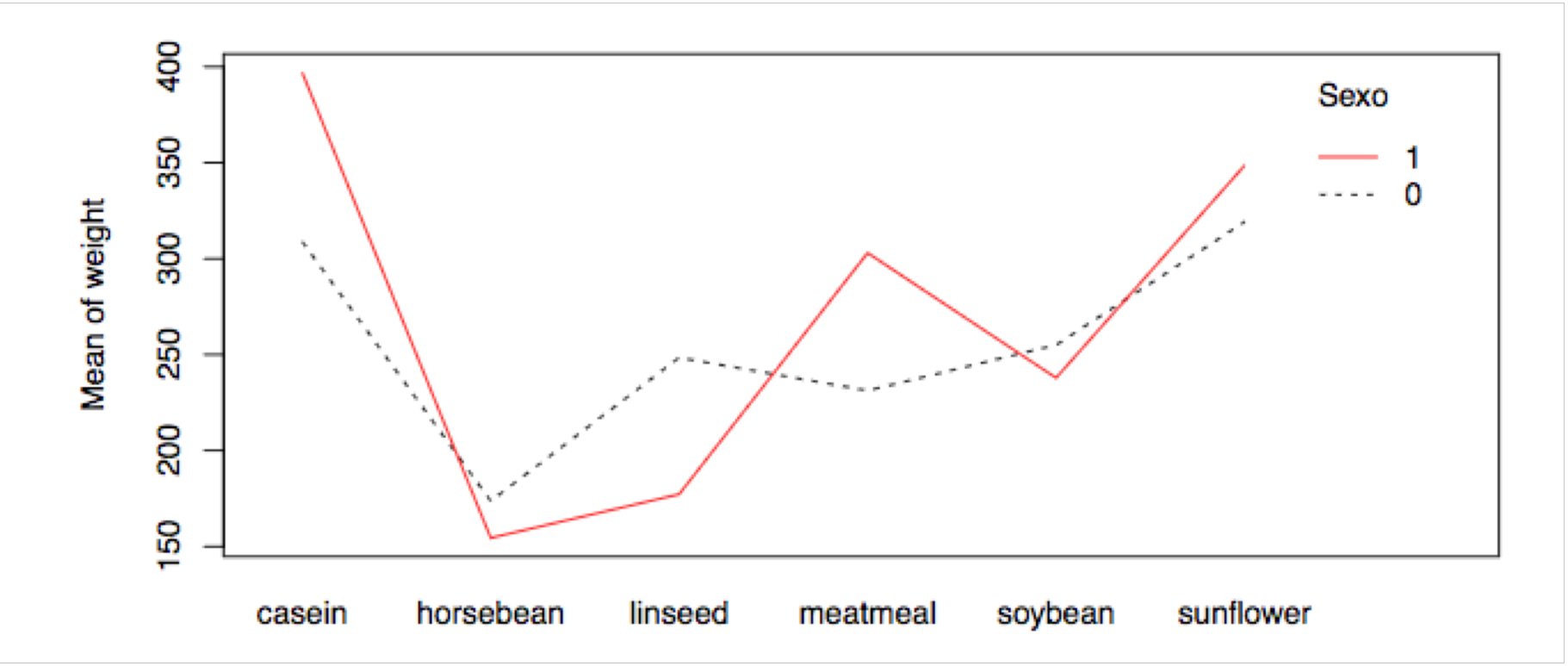
boxplot(variable ~ factor a)



INTERACCIÓN ENTRE FACTORES

Two-way Interaction Plot

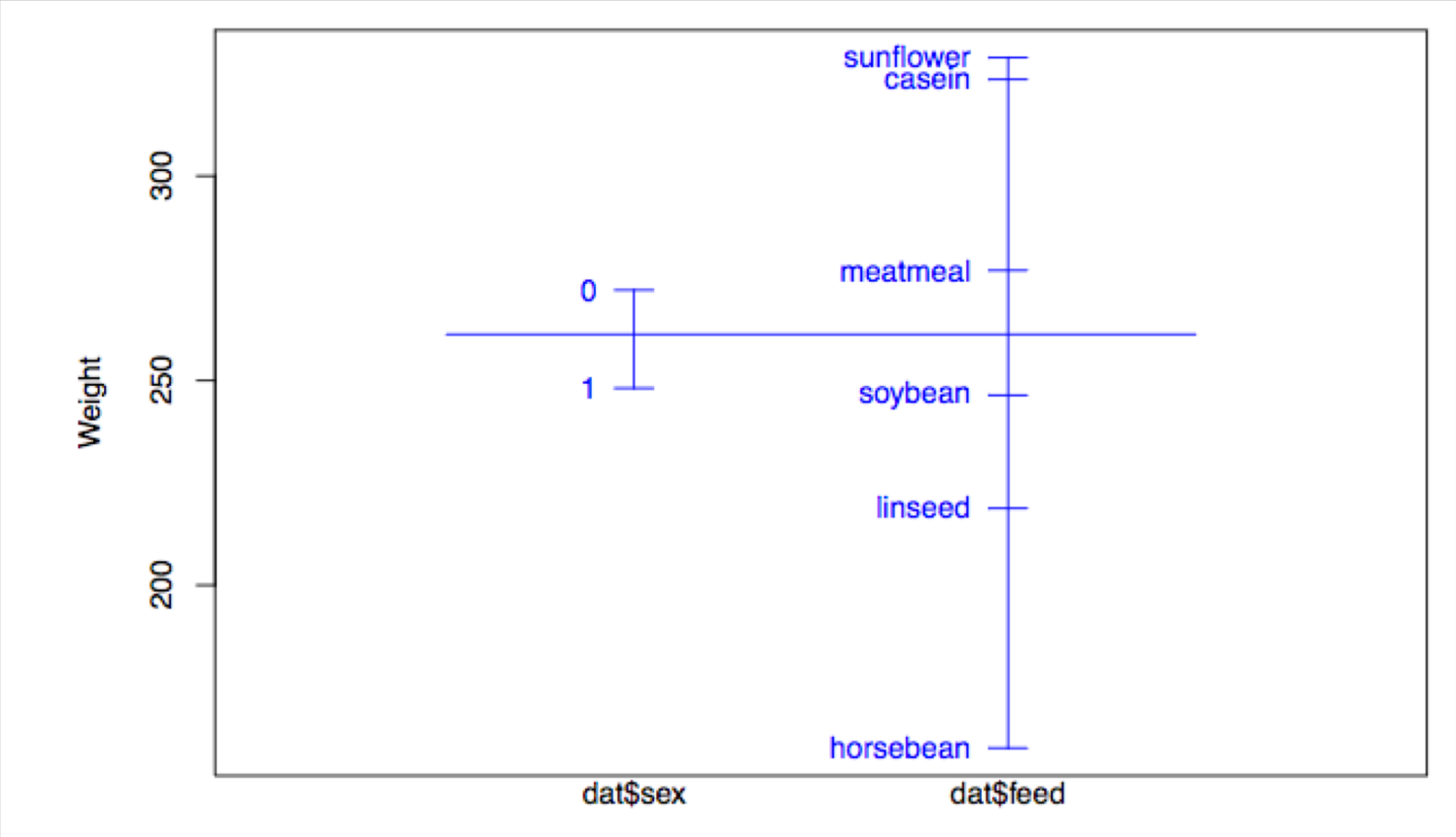
interaction.plot(factor a, factor b, variable)



TAMAÑO DE LOS EFECTOS

 # Plot Uni-bivariate Effects of a Design or Model

plot.design(variable~ factor a * factor b)



TRABAJO PRÁCTICO

1.- Desarrollar guía de repaso.
https://drive.google.com/file/d/1rF4QCAJ6kfmwZN_NHeLI3wOfjGbskFz_/view?usp=sharing

2.- Desarollar guía de ejercicios.
https://rstudio.cloud/spaces/128975/project/2373685

RESUMEN DE LO APRENDIDO

1. **Principios del análisis exploratorio de datos.**
2. Reconocer los tipos de variables.
3. **Investigar por variación de los datos.**
4. Reconocer covariación de los datos.