

# Estadística descriptiva con datos cuantitativos

Ramon Ceballos

30/1/2021

## Ejercicio de analisis de raw data de datos cuantitativos

Carga el conjunto de datos de la carpeta data llamado grades.txt en R y responde a las siguientes preguntas.

```
data = read.table("../../data/grades.txt", header = TRUE)
```

```
#Observo las primeras filas del DF  
head(data)
```

```
##      estudio      nota  
## 1 Industrial 4.811598  
## 2 Telematica 3.825744  
## 3 Telematica 3.518318  
## 4      Mates 0.000000  
## 5 Telematica 5.879675  
## 6 Informatica 3.554049
```

```
#Observo la estructura del DF  
str(data)
```

```
## 'data.frame':    200 obs. of  2 variables:  
## $ estudio: Factor w/ 4 levels "Industrial","Informatica",...: 1 4 4 3 4 2 3 2 4 4 ...  
## $ nota : num  4.81 3.83 3.52 0 5.88 ...
```

Preguntas de esta tarea

1. Calcula la media, la mediana y la moda redondeada a dos cifras decimales de las notas numéricas de los exámenes tanto de forma conjunta como por estudio universitario. ¿En qué grupo observamos la nota más alta? ¿Qué grupo está por encima de la media de estudiantes?

```
#CALCULOS PARA EL CONJUNTO DE NOTAS RECOGIDAS CON INDEPENDENCIA DE LA CARRERA SELECCIONADA  
  
c(MEDIA = round(mean(data$nota),2),  
  MEDIANA = round(median(data$nota),2),  
  MODA = round(as.numeric(names(which(table(data$nota)==max(table(data$nota))))),2))
```

```
## MEDIA MEDIANA MODA
## 3.87 3.50 0.00
```

*#CALCULOS PARA EL CONJUNTO DE NOTAS RECOGIDAS POR ESTUDIOS UNIVERSITARIOS*

*##MEDIA*

```
round(by(data$nota, data$estudio, FUN = mean),2)
```

```
## data$estudio: Industrial
## [1] 3.58
## -----
## data$estudio: Informatica
## [1] 3.95
## -----
## data$estudio: Mates
## [1] 4.22
## -----
## data$estudio: Telematica
## [1] 3.7
```

*##MEDIANA*

```
round(by(data$nota, data$estudio, FUN = median),2)
```

```
## data$estudio: Industrial
## [1] 3.44
## -----
## data$estudio: Informatica
## [1] 3.48
## -----
## data$estudio: Mates
## [1] 4.1
## -----
## data$estudio: Telematica
## [1] 3.41
```

*##MODA*

```
by(data$nota,data$estudio, FUN=function(x){as.numeric(names(which(table(x)==max(table(x))))))})
```

```
## data$estudio: Industrial
## [1] 0
## -----
## data$estudio: Informatica
## [1] 1.846116 1.932031 1.956344 2.336915 2.395989 2.453453 2.469777 2.770846
## [9] 2.784283 2.827685 2.961614 2.992383 3.131438 3.236063 3.291426 3.296116
## [17] 3.315346 3.321207 3.362891 3.427916 3.477870 3.481309 3.554049 3.735585
## [25] 3.765947 4.124979 4.141371 4.196208 4.359728 4.552850 4.992463 5.128825
## [33] 5.167346 5.172384 5.592137 5.833281 5.908514 6.102675 6.395469 7.823779
## [41] 8.515653
## -----
## data$estudio: Mates
## [1] 0.0000000 0.6077362 1.0207988 1.0511178 1.0835325 1.1711770 1.2725944
## [8] 1.6817515 1.7356666 1.8346297 1.9615499 1.9878330 2.2266540 2.2958132
## [15] 2.6127648 2.7005468 3.0351804 3.0503472 3.0863145 3.2975285 3.4136447
```

```
## [22] 3.4308250 3.4739754 3.4865372 3.5229780 3.7274243 3.7833108 3.8546601
## [29] 4.3480738 4.5030344 4.5135289 4.5865059 4.6216129 4.7779196 4.9673143
## [36] 5.0081123 5.2432285 5.6721144 5.7100921 5.7468063 5.7513194 5.9231460
## [43] 5.9601867 5.9846300 6.0077021 6.2688331 6.4494775 6.5550523 6.5969452
## [50] 6.6781263 6.7278867 7.8688089 7.9566733 8.2341875 8.4505200 8.5399794
## -----
## data$estudio: Telematica
## [1] 0.0000000 0.6577969 0.8850120 0.9764619 1.2279058 1.3817389 1.4085427
## [8] 1.6359643 1.8836089 2.0116603 2.0568973 2.2570168 2.2747866 2.3985057
## [15] 2.4073405 2.5494692 2.5650097 2.5783268 2.7227514 2.8632973 2.9054596
## [22] 2.9513959 2.9778022 2.9944822 3.0158789 3.0360135 3.0940300 3.3605270
## [29] 3.4090172 3.4109679 3.5127030 3.5183177 3.6391383 3.6553545 3.7008815
## [36] 3.7911751 3.8257442 3.9761653 4.1874499 4.5692315 4.6940934 4.7515590
## [43] 4.7709564 4.8313561 5.2121328 5.3975366 5.4485237 5.7368340 5.7934587
## [50] 5.8796750 6.0203200 6.2825232 6.3690967 6.5335745 6.5955539 7.1112415
## [57] 7.8587433 9.2904650
```

```
#Averiguar la nota más alta de cada estudio universitario
by(data$nota, data$estudio, FUN = max)
```

```
## data$estudio: Industrial
## [1] 10.02469
## -----
## data$estudio: Informatica
## [1] 8.515653
## -----
## data$estudio: Mates
## [1] 8.539979
## -----
## data$estudio: Telematica
## [1] 9.290465
```

2. ¿En qué grupo observamos más variación de notas? Justifica tu respuesta

```
#Averiguar que grupo tiene más variación de notas
by(data$nota, data$estudio, FUN = sd)
```

```
## data$estudio: Industrial
## [1] 2.010948
## -----
## data$estudio: Informatica
## [1] 1.537978
## -----
## data$estudio: Mates
## [1] 2.195236
## -----
## data$estudio: Telematica
## [1] 1.909706
```

3. ¿Hay mucha diferencia entre el grupo que has respondido en el ejercicio 1 y en el ejercicio 2? Intenta dar una explicación objetiva del suceso

```
summary(data[data$estudio=="Mates",])
```

```
##      estudio      nota
## Industrial : 0  Min.   :0.000
## Informatica: 0  1st Qu.:2.534
## Mates      :56  Median :4.101
## Telematica : 0  Mean    :4.215
##              3rd Qu.:5.932
##              Max.    :8.540
```

```
summary(data[data$estudio=="Industrial",])
```

```
##      estudio      nota
## Industrial :45  Min.   : 0.000
## Informatica: 0  1st Qu.: 2.622
## Mates      : 0  Median : 3.435
## Telematica : 0  Mean    : 3.577
##              3rd Qu.: 4.775
##              Max.    :10.025
```

```
summary(data[data$estudio=="Informatica",])
```

```
##      estudio      nota
## Industrial : 0  Min.   :1.846
## Informatica:41  1st Qu.:2.962
## Mates      : 0  Median :3.478
## Telematica : 0  Mean    :3.954
##              3rd Qu.:4.992
##              Max.    :8.516
```

```
summary(data[data$estudio=="Telematica",])
```

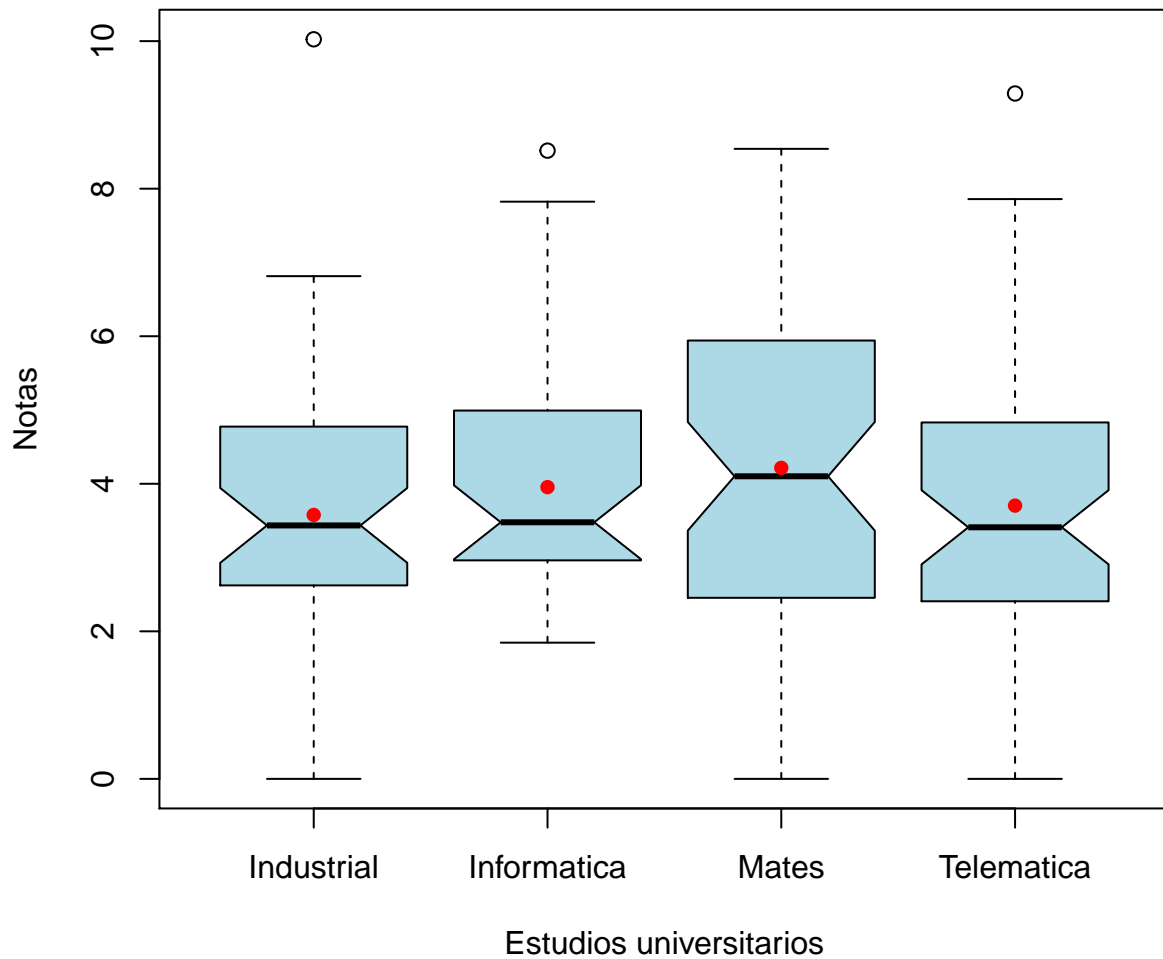
```
##      estudio      nota
## Industrial : 0  Min.   :0.000
## Informatica: 0  1st Qu.:2.443
## Mates      : 0  Median :3.410
## Telematica :58  Mean    :3.704
##              3rd Qu.:4.816
##              Max.    :9.290
```

4. Dibuja en un único gráfico los cuatro diagramas de caja, uno para cada grupo. Añade así mismo la nota media, pon nombre, título y color al gráfico para documentar el gráfico correctamente.

```
#Boxplot de notas para cada estudio universitario
boxplot(data$nota~data$estudio,
        ylab = "Notas",
        xlab = "Estudios universitarios",
        mean = "Boxplot de notas para cada estudio universitario",
        col = "lightblue",
        notch = TRUE)

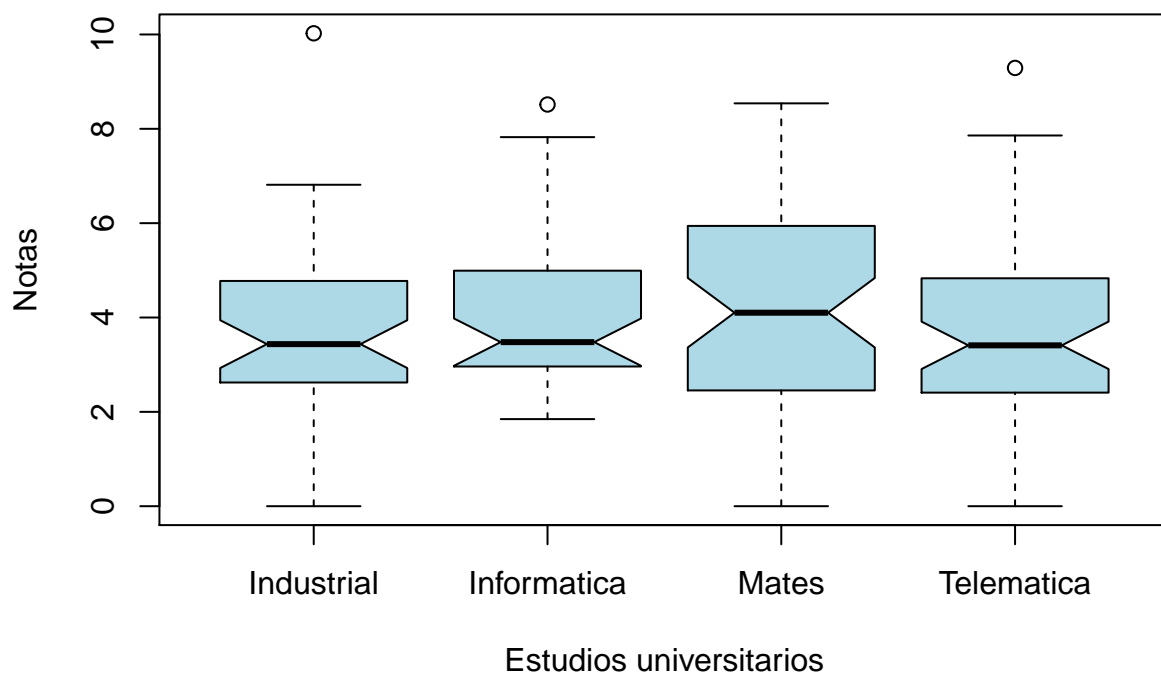
media = aggregate(data$nota~data$estudio, FUN =mean)

points (media, col = "red", pch =16)
```



5. ¿Observas algunos valores atípicos en el boxplot anterior? ¿A qué grupo pertenece?

```
str(boxplot(data$nota~data$estudio,
  ylab = "Notas",
  xlab = "Estudios universitarios",
  mean = "Boxplot de notas para cada estudio universitario",
  col = "lightblue",
  notch = TRUE))
```



```
## List of 6
## $ stats: num [1:5, 1:4] 0 2.62 3.44 4.77 6.81 ...
## $ n : num [1:4] 45 41 56 58
## $ conf : num [1:2, 1:4] 2.93 3.94 2.98 3.98 3.37 ...
## $ out : num [1:3] 10.02 8.52 9.29
## $ group: num [1:3] 1 2 4
## $ names: chr [1:4] "Industrial" "Informatica" "Mates" "Telematica"
```

6. Indica a partir del boxplot anterior en qué grupo observamos más variedad de notas.

7. En base a todo lo anterior, ¿puedes afirmar de forma estadística que el examen de carrera ha ido mejor en algún estudio que en los demás?