

MEDIDAS 3

Jocelyn Sánchez Campos

2023-10-19

R Markdown

Teoría de la medición

La escala de una medición es el nivel de medida de una variable. Se encarga de describir la naturaleza de la información contenida dentro de los números asignados a un objeto.

-Existen: **Cualitativas** y **Cuantitativas**

1.1 Cualitativa nominal: Las variables cualitativas nominales son categorías que no se pueden clasificar. 1.2 Cualitativa ordinal: presenta modalidades no numéricas, en las que existe un orden. Ejemplos: La nota en un examen: suspenso, aprobado, notable, sobresaliente. 2.1 Cuantitativa discreta: cuando se toman valores aislados. 2.2 Cuantitativa continua: cuando, entre dos valores cualesquiera, puede haber valores intermedios. Entre ellas hay **de intervalo** las cuales el cero no se considera como absoluto, es decir el 0 es inherente y las **de razón** donde el 0 si es absoluto ya que hay ausencia de la medición del objeto.

##Medidas de tendencia central

Son medidas estadísticas que pretenden resumir en un solo valor a un conjunto de valores. Representan un centro en torno al cual se encuentra ubicado el conjunto de los datos. Las medidas de tendencia central más utilizadas son: **media**, **mediana** y **moda**. **Media:** Es la medida equitativa de todas las observaciones (\bar{x}) Es para datos numéricos y distribuciones simétricas, es importante saber que es sensible a datos “atípicos”

Mediana: Es la medida que divide las observaciones a la mitad, es para datos ordinales o numéricos con distribución sesgada. Es ideal para datos atípicos Y su procedimiento es: 1. Los datos se ordenan de menor a mayor. 2. Se identifican los valores que se encuentran a la mitad del total de las observaciones. 3. Si el número es par, se seleccionan los dos valores y se divide entre dos. 3.1. Si el número es impar se selecciona el dato central.

Moda: Es la observación que tiene la frecuencia absoluta mayor (x) Se usa para distribuciones bimodales. Procedimiento. 1. Los datos se ordenan de menor a mayor. 2. Se identifican los valores que más se repiten.

##Medidas de posición

Son aquellas en donde pueden dividir los datos en dos partes iguales, llamada mediana, lo puedes dividir en cuatro partes iguales llamado cuartiles, en diez partes iguales llamados deciles y en percentiles dividir en 100 partes iguales. Se obtienen con la función **summary()** que ya está precargada en R.

##Medidas de dispersión

Son números que indican si una variable se mueve mucho, poco, más o menos que otra. La razón de ser de este tipo de medidas es conocer de manera resumida una característica de la variable estudiada. En este sentido, deben acompañar a las medidas de tendencia central. Dentro de esta categoría se encuentran:

Error Es la diferencia del valor de una variable con respecto a un valor fijo.

Varianza Es una medida de dispersión que representa la variabilidad de una serie de datos respecto a su media. Formalmente se calcula como la suma de los residuos al cuadrado divididos entre el total de observaciones.

Desviación estandar Es la raíz cuadrada de la varianza. Entre más dispersa está una distribución de datos, más grande es su desviación estándar.

Coficiente de variación Dispersión relativa de los datos. Permite el análisis de las desviaciones de los datos con respecto a la media y al mismo tiempo las dispersiones que tienen los datos dispersos entre sí.

Comenzamos exportando la matriz de datos “penguins”

Environment /Import dataset/from excel/ Browser/ seleccionar el archivo/ aceptar/ (visualizar)/ import

Instalación de la paqueteria

```
install.packages("readxl")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

Abrimos la paqueteria

```
library("readxl")
```

```
penguins<-read_excel("penguins.xlsx")
```

Exploración de la matriz

```
dim(penguins)
```

```
## [1] 344 9
```

```
str(penguins)
```

```
## tibble [344 x 9] (S3: tbl_df/tbl/data.frame)  
## $ ID : chr [1:344] "i1" "i2" "i3" "i4" ...  
## $ especie : chr [1:344] "Adelie" "Adelie" "Adelie" "Adelie" ...  
## $ isla : chr [1:344] "Torgersen" "Torgersen" "Torgersen" "Torgersen" ...  
## $ largo_pico_mm : num [1:344] 39.1 39.5 40.3 37.8 36.7 39.3 38.9 39.2 34.1 42 ...  
## $ grosor_pico_mm : num [1:344] 18.7 17.4 18 18.1 19.3 20.6 17.8 19.6 18.1 20.2 ...  
## $ largo_aleta_mm : num [1:344] 181 186 195 190 193 190 181 195 193 190 ...  
## $ masa_corporal_g: num [1:344] 3750 3800 3250 3700 3450 ...  
## $ genero : chr [1:344] "male" "female" "female" "female" ...  
## $ año : num [1:344] 2007 2007 2007 2007 2007 ...
```

```
colnames(penguins)
```

```
## [1] "ID" "especie" "isla" "largo_pico_mm"  
## [5] "grosor_pico_mm" "largo_aleta_mm" "masa_corporal_g" "genero"  
## [9] "año"
```

```
anyNA(penguins)
```

```
## [1] FALSE
```

```
##Tendencia central media y mediana usando summary
```

```
summary(penguins)
```

```
## ID especie isla largo_pico_mm  
## Length:344 Length:344 Length:344 Min. :32.10  
## Class :character Class :character Class :character 1st Qu.:39.20  
## Mode :character Mode :character Mode :character Median :44.45  
## Mean :43.92  
## 3rd Qu.:48.50
```

```
##                                     Max.    :59.60
## grosor_pico_mm largo_aleta_mm masa_corporal_g genero
## Min.    :13.10 Min.    :172.0 Min.    :2700 Length:344
## 1st Qu.:15.60 1st Qu.:190.0 1st Qu.:3550 Class :character
## Median :17.30 Median :197.0 Median :4050 Mode  :character
## Mean   :17.15 Mean   :200.9 Mean   :4202
## 3rd Qu.:18.70 3rd Qu.:213.2 3rd Qu.:4756
## Max.   :21.50 Max.   :231.0 Max.   :6300
## año
## Min.    :2007
## 1st Qu.:2007
## Median :2008
## Mean   :2008
## 3rd Qu.:2009
## Max.   :2009
```

2.- Moda Se descargará el paquete “modeest”

```
install.packages("modeest")
```

```
## Installing package into '/cloud/lib/x86_64-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

abrimos la libreria

```
library(modeest)
```

3.- Calculamos de la moda para la variable isla y largo del pico

```
mfv(penguins$isla) # categorica
```

```
## [1] "Biscoe"
```

```
mfv(penguins$largo_pico_mm) # numerica
```

```
## [1] 41.1
```

Medidas de posición

1.- Cuartiles (cuantiles)

```
summary(penguins)
```

```
##      ID      especie      isla      largo_pico_mm
## Length:344      Length:344      Length:344      Min.    :32.10
## Class :character Class :character Class :character 1st Qu.:39.20
## Mode  :character Mode  :character Mode  :character Median :44.45
##                                     Mean   :43.92
##                                     3rd Qu.:48.50
##                                     Max.   :59.60
## grosor_pico_mm largo_aleta_mm masa_corporal_g genero
## Min.    :13.10 Min.    :172.0 Min.    :2700 Length:344
## 1st Qu.:15.60 1st Qu.:190.0 1st Qu.:3550 Class :character
## Median :17.30 Median :197.0 Median :4050 Mode  :character
## Mean   :17.15 Mean   :200.9 Mean   :4202
## 3rd Qu.:18.70 3rd Qu.:213.2 3rd Qu.:4756
## Max.   :21.50 Max.   :231.0 Max.   :6300
## año
## Min.    :2007
```

```
## 1st Qu.:2007
## Median :2008
## Mean :2008
## 3rd Qu.:2009
## Max. :2009
```

Selección de una variable de la matriz de datos

```
largo_aleta_mm<-penguins$largo_aleta_mm
```

Pasamos a una tabla:

```
table(largo_aleta_mm)
```

```
## largo_aleta_mm
## 172 174 176 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194
## 1 1 1 4 1 5 7 3 2 7 9 7 16 6 7 23 13 7 15 5
## 195 196 197 198 199 200 201 202 203 205 206 207 208 209 210 211 212 213 214 215
## 17 10 10 8 6 4 6 4 5 3 1 2 8 5 14 2 7 6 6 12
## 216 217 218 219 220 221 222 223 224 225 226 228 229 230 231
## 8 6 5 5 8 5 7 2 3 4 1 4 2 7 1
```

2.- Quintil

```
quintil<-quantile(penguins[["largo_aleta_mm"]],
                  p=c(.20, .40, .60, .80))
```

Vamos a visualizar las variables

```
quintil
```

```
## 20% 40% 60% 80%
## 188 194 203 215
```

3.- Decil

```
decil<-quantile(penguins[["largo_aleta_mm"]],
                 p=c(.10, .20, .30, .40, .50, .60,
                     .70, .80, .90))
```

```
decil
```

```
## 10% 20% 30% 40% 50% 60% 70% 80% 90%
## 185 188 191 194 197 203 210 215 221
```

4.- Percentil

```
percentil<-quantile(penguins[["largo_aleta_mm"]],
                    p=c(.33, .66, .99))
```

```
percentil
```

```
## 33% 66% 99%
## 192 209 230
```

Valoramos la interpretación con los siguientes datos: Interpretacion: <192 = Bajo 192-209 = Intermedio >
209 = Alto

```
table(largo_aleta_mm)
```

```
## largo_aleta_mm
## 172 174 176 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194
```

```
## 1 1 1 4 1 5 7 3 2 7 9 7 16 6 7 23 13 7 15 5
## 195 196 197 198 199 200 201 202 203 205 206 207 208 209 210 211 212 213 214 215
## 17 10 10 8 6 4 6 4 5 3 1 2 8 5 14 2 7 6 6 12
## 216 217 218 219 220 221 222 223 224 225 226 228 229 230 231
## 8 6 5 5 8 5 7 2 3 4 1 4 2 7 1
```

##Medidas de dispersión

1.- Cálculo de la varianza (sólo para variables cuantitativas)

```
var(penguins$grosor_pico_mm)
```

```
## [1] 3.884256
```

2.- Cálculo de la desviación estándar

```
sd(penguins$grosor_pico_mm)
```

```
## [1] 1.970852
```

3.- Error

```
media_pico<-mean(penguins$largo_pico_mm)
error<-(penguins$largo_pico_mm-(media_pico))
error
```

```
## [1] -4.82412791 -4.42412791 -3.62412791 -6.12412791 -7.22412791
## [6] -4.62412791 -5.02412791 -4.72412791 -9.82412791 -1.92412791
## [11] -6.12412791 -6.12412791 -2.82412791 -5.32412791 -9.32412791
## [16] -7.32412791 -5.22412791 -1.42412791 -9.52412791 2.07587209
## [21] -6.12412791 -6.22412791 -8.02412791 -5.72412791 -5.12412791
## [26] -8.62412791 -3.32412791 -3.42412791 -6.02412791 -3.42412791
## [31] -4.42412791 -6.72412791 -4.42412791 -3.02412791 -7.52412791
## [36] -4.72412791 -5.12412791 -1.72412791 -6.32412791 -4.12412791
## [41] -7.42412791 -3.12412791 -7.92412791 0.17587209 -6.92412791
## [46] -4.32412791 -2.82412791 -6.42412791 -7.92412791 -1.62412791
## [51] -4.32412791 -3.82412791 -8.92412791 -1.92412791 -9.42412791
## [56] -2.52412791 -4.92412791 -3.32412791 -7.42412791 -6.32412791
## [61] -8.22412791 -2.62412791 -6.32412791 -2.82412791 -7.52412791
## [66] -2.32412791 -8.42412791 -2.82412791 -8.02412791 -2.12412791
## [71] -10.42412791 -4.22412791 -4.32412791 1.87587209 -8.42412791
## [76] -1.12412791 -3.02412791 -6.72412791 -7.72412791 -1.82412791
## [81] -9.32412791 -1.02412791 -7.22412791 -8.82412791 -6.62412791
## [86] -2.62412791 -7.62412791 -7.02412791 -5.62412791 -5.02412791
## [91] -8.22412791 -2.82412791 -9.92412791 -4.32412791 -7.72412791
## [96] -3.12412791 -5.82412791 -3.62412791 -10.82412791 -0.72412791
## [101] -8.92412791 -2.92412791 -6.22412791 -6.12412791 -6.02412791
## [106] -4.22412791 -5.32412791 -5.72412791 -5.82412791 -0.72412791
## [111] -5.82412791 1.67587209 -4.22412791 -1.72412791 -4.32412791
## [116] -1.22412791 -5.32412791 -6.62412791 -8.22412791 -2.82412791
## [121] -7.72412791 -6.22412791 -3.72412791 -2.52412791 -8.72412791
## [126] -3.32412791 -5.12412791 -2.42412791 -4.92412791 0.17587209
## [131] -5.42412791 -0.82412791 -7.12412791 -6.42412791 -5.82412791
## [136] -2.82412791 -8.32412791 -3.72412791 -6.92412791 -4.22412791
## [141] -3.72412791 -3.32412791 -11.82412791 -3.22412791 -6.62412791
## [146] -4.92412791 -4.72412791 -7.32412791 -7.92412791 -6.12412791
## [151] -7.92412791 -2.42412791 2.17587209 6.07587209 4.77587209
## [156] 6.07587209 3.67587209 2.57587209 1.47587209 2.77587209
```

```
## [161] -0.62412791  2.87587209 -3.02412791  5.07587209  1.57587209
## [166]  4.47587209  1.87587209  5.37587209 -1.92412791  5.27587209
## [171]  2.27587209  4.77587209  6.27587209  1.17587209  2.57587209
## [176]  2.37587209 -1.02412791  2.17587209  0.57587209  3.87587209
## [181]  4.27587209  6.07587209  3.37587209 -1.12412791  1.17587209
## [186] 15.67587209  5.17587209  4.47587209 -1.32412791  0.47587209
## [191]  0.07587209  4.77587209 -1.22412791  5.67587209  1.37587209
## [196]  5.67587209  6.57587209 -0.32412791  1.57587209  6.57587209
## [201]  0.97587209  1.27587209  2.67587209  4.57587209  1.17587209
## [206]  6.17587209  2.57587209  1.07587209 -0.12412791  1.57587209
## [211] -0.72412791  6.47587209  1.37587209  2.27587209  1.77587209
## [216] 10.37587209  1.87587209  5.87587209  2.27587209  5.57587209
## [221] -0.42412791  6.77587209  3.77587209  2.47587209  4.27587209
## [226]  2.57587209  2.47587209  4.67587209  3.57587209  7.17587209
## [231]  1.27587209  1.27587209  5.17587209  8.57587209  3.47587209
## [236]  6.07587209  0.97587209  6.87587209 -0.52412791  7.37587209
## [241]  3.57587209  8.17587209  3.57587209  8.27587209  1.57587209
## [246]  5.57587209  0.57587209  6.87587209  5.47587209  2.97587209
## [251]  4.47587209  7.17587209  4.57587209 11.97587209  3.27587209
## [256]  5.17587209  3.37587209  2.87587209 -2.22412791  9.47587209
## [261] -0.62412791  4.17587209  6.57587209  5.87587209 -0.42412791
## [266]  7.57587209  2.27587209 11.17587209  0.57587209  4.87587209
## [271]  3.27587209  6.87587209  2.87587209  6.47587209  1.27587209
## [276]  5.97587209  2.57587209  6.07587209  7.37587209  1.47587209
## [281]  8.77587209  1.27587209  2.17587209  7.37587209  2.07587209
## [286]  7.37587209  2.67587209  7.77587209  3.07587209  8.07587209
## [291]  1.97587209  6.57587209  6.37587209 14.07587209  2.47587209
## [296]  5.27587209 -1.52412791  4.57587209 -0.72412791  6.67587209
## [301]  2.77587209  8.07587209  6.57587209  5.57587209  2.47587209
## [306]  8.87587209 -3.02412791 10.27587209 -1.42412791  7.07587209
## [311]  5.77587209  3.57587209  3.67587209  8.07587209  2.97587209
## [316]  9.57587209  5.07587209  2.27587209  6.97587209  1.57587209
## [321]  6.97587209  6.87587209  6.17587209  5.07587209  7.57587209
## [326]  5.87587209  4.17587209  7.47587209  1.77587209  6.77587209
## [331] -1.42412791  8.27587209  1.27587209  5.37587209  6.27587209
## [336]  1.67587209  7.97587209  2.87587209  1.77587209 11.87587209
## [341] -0.42412791  5.67587209  6.87587209  6.27587209
```

4.- Coeficiente de variacion

```
CV<-sd(penguins$largo_pico_mm)/mean(penguins$largo_pico_mm)*100
CV
```

```
## [1] 12.44487
```

5.- Rango intercuartilico (IQR)

```
IQR(penguins$largo_pico_mm)
```

```
## [1] 9.3
```

6.- Rango

```
pico<-penguins$largo_pico_mm
rango<-max(pico)-min(pico)
rango
```

```
## [1] 27.5
```