REPORTE E-DESCRIPTIVA

Juan Llivisaca

2023-09-26

# BASE DE DATOS DE PRUEBA

# Cargar los datos  
library(readr)  
base <- read\_delim("C:/Users/Juan Llivisaca/Downloads/base.csv",   
 delim = ";", escape\_double = FALSE, trim\_ws = TRUE)

# Estadística descriptiva  
summary(base$edad)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 7.00 13.75 26.50 31.60 48.00 71.00 10

library(descriptr)  
ds\_auto\_summary\_stats(base,edad)

## -------------------------------- Variable: edad --------------------------------  
##   
## ------------------------------ Summary Statistics ------------------------------  
##   
## -------------------------------- Variable: edad --------------------------------  
##   
## Univariate Analysis   
##   
## N 30.00 Variance 389.09   
## Missing 10.00 Std Deviation 19.73   
## Mean 31.60 Range 64.00   
## Median 26.50 Interquartile Range 34.25   
## Mode 7.00 Uncorrected SS 27364.00   
## Trimmed Mean 30.78 Corrected SS 7392.80   
## Skewness 0.51 Coeff Variation 62.42   
## Kurtosis -0.87 Std Error Mean 4.41   
##   
## Quantiles   
##   
## Quantile Value   
##   
## Max 71.00   
## 99% 69.67   
## 95% 64.35   
## 90% 57.70   
## Q3 48.00   
## Median 26.50   
## Q1 13.75   
## 10% 10.60   
## 5% 7.00   
## 1% 7.00   
## Min 7.00   
##   
## Extreme Values   
##   
## Low High   
##   
## Obs Value Obs Value   
## 7 7 1 71   
## 17 7 12 64   
## 6 11 4 57   
## 9 11 20 52   
## 8 13 3 48   
##   
##   
##   
## NULL  
##   
##   
## ---------------------------- Frequency Distribution ----------------------------  
##   
## Variable: edad   
## |-----------------------------------------------------------------------|  
## | Bins | Frequency | Cum Frequency | Percent | Cum Percent |  
## |-----------------------------------------------------------------------|  
## | 7 - 19.8 | 6 | 6 | 30 | 30 |  
## |-----------------------------------------------------------------------|  
## | 19.8 - 32.6 | 5 | 11 | 25 | 55 |  
## |-----------------------------------------------------------------------|  
## | 32.6 - 45.4 | 3 | 14 | 15 | 70 |  
## |-----------------------------------------------------------------------|  
## | 45.4 - 58.2 | 4 | 18 | 20 | 90 |  
## |-----------------------------------------------------------------------|  
## | 58.2 - 71 | 2 | 20 | 10 | 100 |  
## |-----------------------------------------------------------------------|  
## | Missing | 10 | - | 33.33333 | - |  
## |-----------------------------------------------------------------------|  
## | Total | 30 | - | 100.00 | - |  
## |-----------------------------------------------------------------------|

# descriptr::ds\_summary\_stats(base,edad)  
  
# Imputación de datos  
  
base$edad1 <- ifelse(is.na(base$edad)== T, mean(base$edad), base$edad)  
ds\_auto\_summary\_stats(base,edad1)

## ------------------------------- Variable: edad1 -------------------------------  
##   
## ------------------------------ Summary Statistics ------------------------------  
##   
## ------------------------------- Variable: edad1 -------------------------------  
##   
## Univariate Analysis   
##   
## N 30.00 Variance 389.09   
## Missing 10.00 Std Deviation 19.73   
## Mean 31.60 Range 64.00   
## Median 26.50 Interquartile Range 34.25   
## Mode 7.00 Uncorrected SS 27364.00   
## Trimmed Mean 30.78 Corrected SS 7392.80   
## Skewness 0.51 Coeff Variation 62.42   
## Kurtosis -0.87 Std Error Mean 4.41   
##   
## Quantiles   
##   
## Quantile Value   
##   
## Max 71.00   
## 99% 69.67   
## 95% 64.35   
## 90% 57.70   
## Q3 48.00   
## Median 26.50   
## Q1 13.75   
## 10% 10.60   
## 5% 7.00   
## 1% 7.00   
## Min 7.00   
##   
## Extreme Values   
##   
## Low High   
##   
## Obs Value Obs Value   
## 7 7 1 71   
## 17 7 12 64   
## 6 11 4 57   
## 9 11 20 52   
## 8 13 3 48   
##   
##   
##   
## NULL  
##   
##   
## ---------------------------- Frequency Distribution ----------------------------  
##   
## Variable: edad1   
## |-----------------------------------------------------------------------|  
## | Bins | Frequency | Cum Frequency | Percent | Cum Percent |  
## |-----------------------------------------------------------------------|  
## | 7 - 19.8 | 6 | 6 | 30 | 30 |  
## |-----------------------------------------------------------------------|  
## | 19.8 - 32.6 | 5 | 11 | 25 | 55 |  
## |-----------------------------------------------------------------------|  
## | 32.6 - 45.4 | 3 | 14 | 15 | 70 |  
## |-----------------------------------------------------------------------|  
## | 45.4 - 58.2 | 4 | 18 | 20 | 90 |  
## |-----------------------------------------------------------------------|  
## | 58.2 - 71 | 2 | 20 | 10 | 100 |  
## |-----------------------------------------------------------------------|  
## | Missing | 10 | - | 33.33333 | - |  
## |-----------------------------------------------------------------------|  
## | Total | 30 | - | 100.00 | - |  
## |-----------------------------------------------------------------------|

base[c(2:5),1]

## # A tibble: 4 × 1  
## ...1  
## <dbl>  
## 1 2  
## 2 3  
## 3 4  
## 4 5

En esta base de datoe existen por ejemplo 10 Na, la media es de 31.6 y la mpda es de 48.