



Científico de Datos

Aliados:



Microsoft

Vigilada Mineducación



Advanced analytics for business

INTRODUCCIÓN A LA ESTADÍSTICA DESCRIPTIVA

Aliados:



Microsoft

Vigilada Mineducación



DEFINITIONS



ESTADÍSTICA
DESCRIPTIVA



ESTADÍSTICA
INFERENCIAL

TYPES OF VARIABLES

CATEGORICAL

1

Nominal

- Gender
- Color
- Blood type

2

Ordinal

- Education level (elementary,secondary,college)
- Pain level (mild,moderate,severe)
- Satisfaction level (dissatisfied,neutral,satisfied)

1

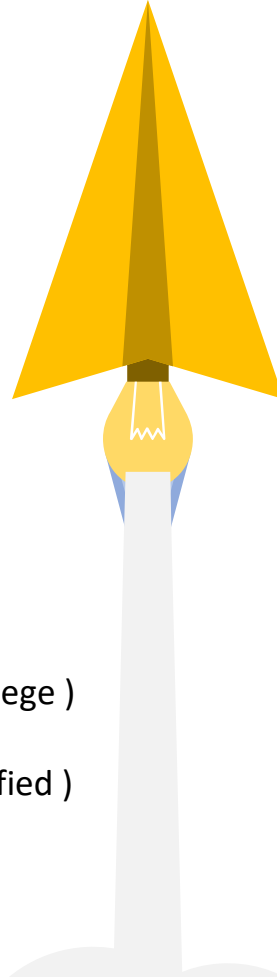
Discrete

- Number of students present
- Number of red marbles in a jar
- Number of red cars

2

Continuous

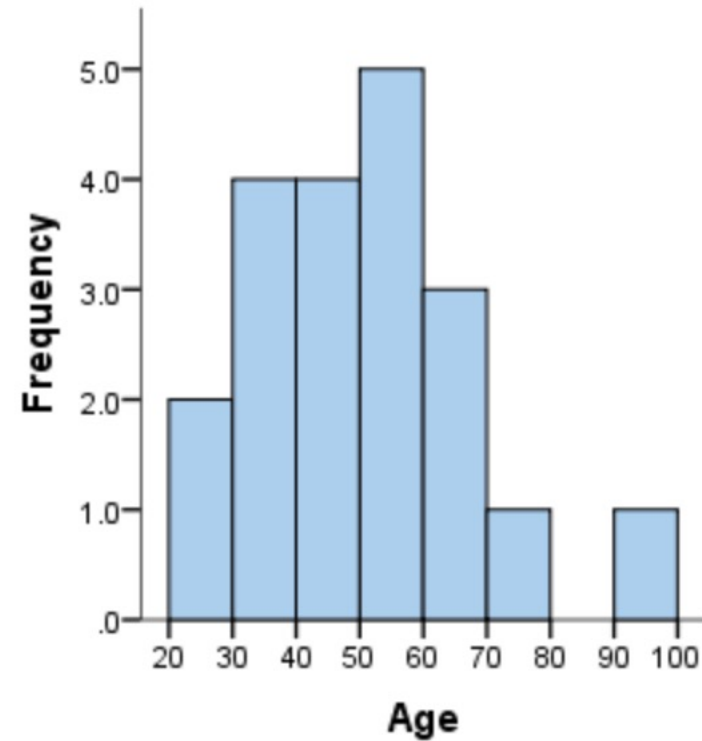
- Height of students in class
- Time it takes to get to school
- Distance traveled between classes



HISTOGRAMS



Shows frequency of
occurrence of data



MEASURES OF CENTRAL TENDENCY

MEAN

$$\mu = \frac{x_1 + \dots + x_n}{n}$$

2, 3, 3, 5, 8, 10, 11

$$(2+3+3+5+8+10+11)/7 = 6$$

MEDIA = 6

2, 2, 3, 3, 5, 7, 8, 130

$$(2+2+3+3+5+7+8+130)/8 = 20$$

MEDIA = 20

2, 3, 3, 5, 8, 10, 11

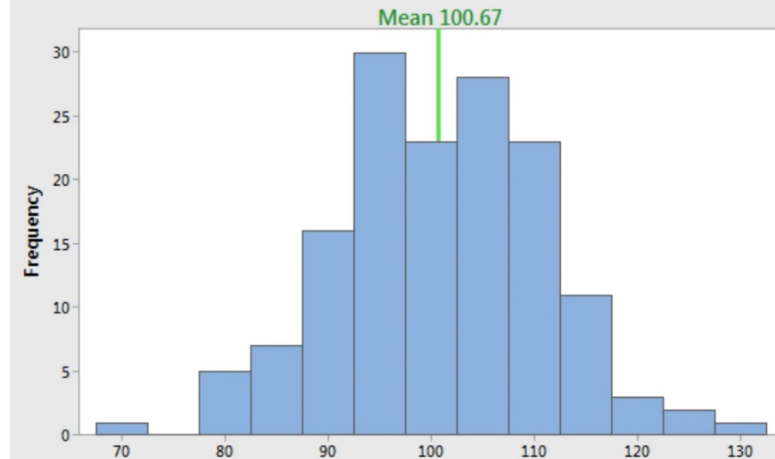
MED = 5

2, 2, 3, 3, 5, 7, 8, 130

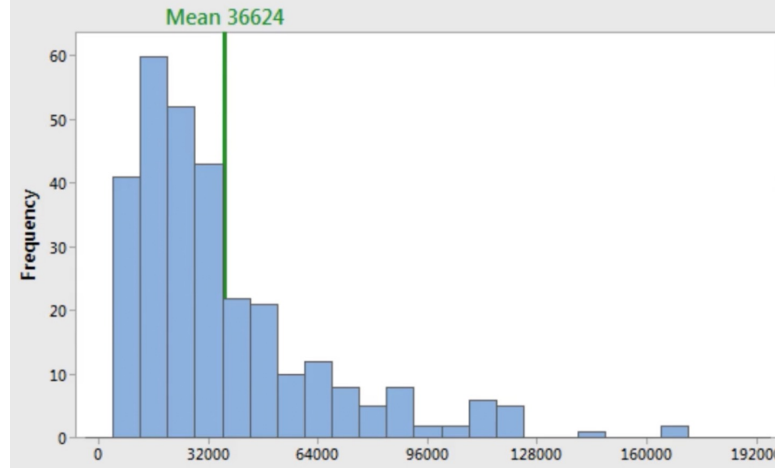
$$(3+5)/2 = 4$$

MED = 4

Histogram of Symmetric Continuous



Histogram of Skewed Continuous



MEASURES OF CENTRAL TENDENCY

MEDIAN

Is the middle value

1. Order your data from smallest to largest
2. Find the data point that has an equal amount of values above it, and below it

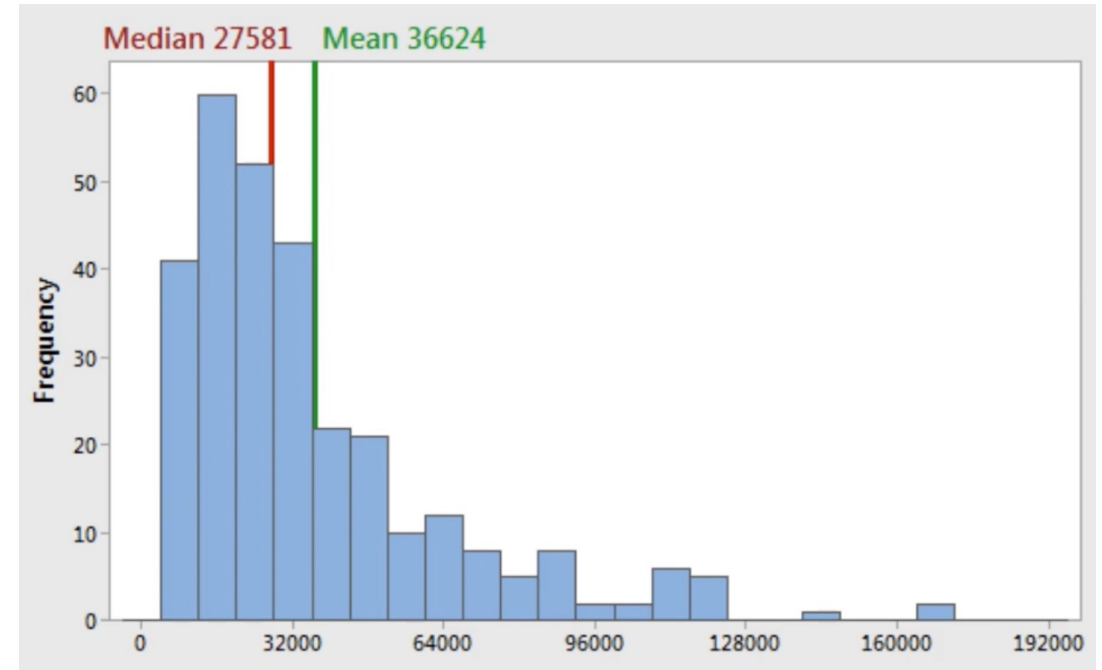
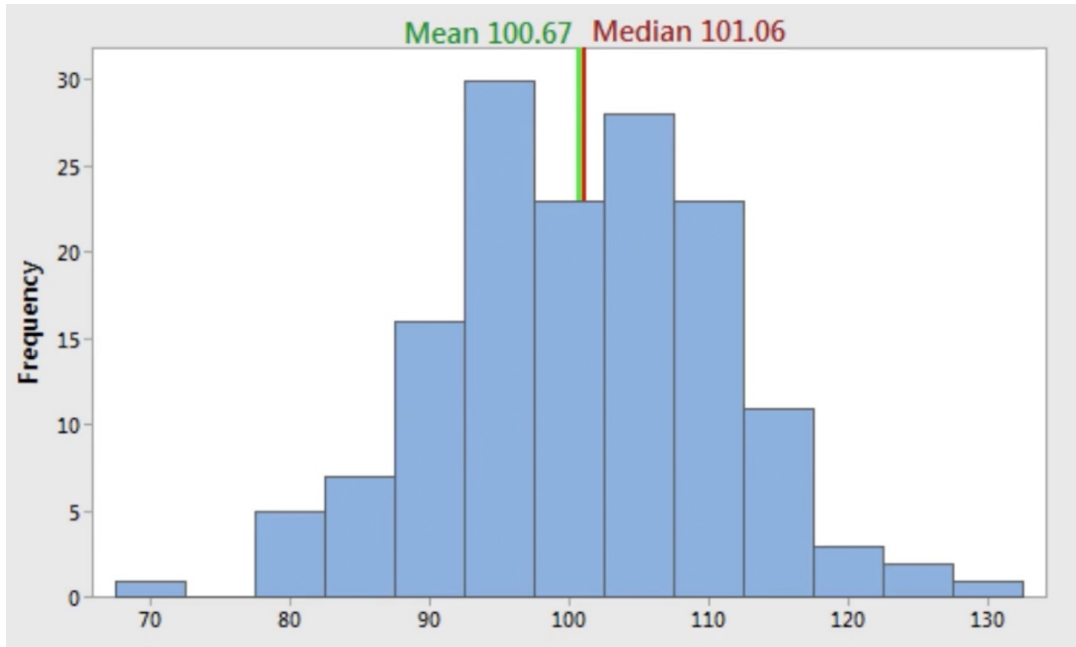


Outliers have a smaller effect on the median

Median Odd	Median Even
23	40
21	38
18	35
16	33
15	32
13	30
12	29
10	27
9	26
7	24
6	23
5	22
2	19
	17

28

MEASURES OF CENTRAL TENDENCY



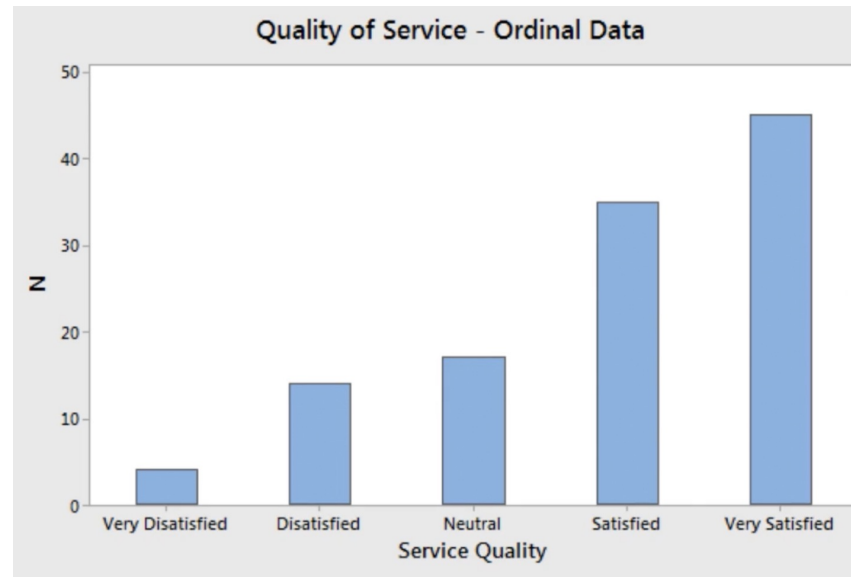
MEASURES OF CENTRAL TENDENCY

Mode
5
5
5
4
4
3
2
2
1

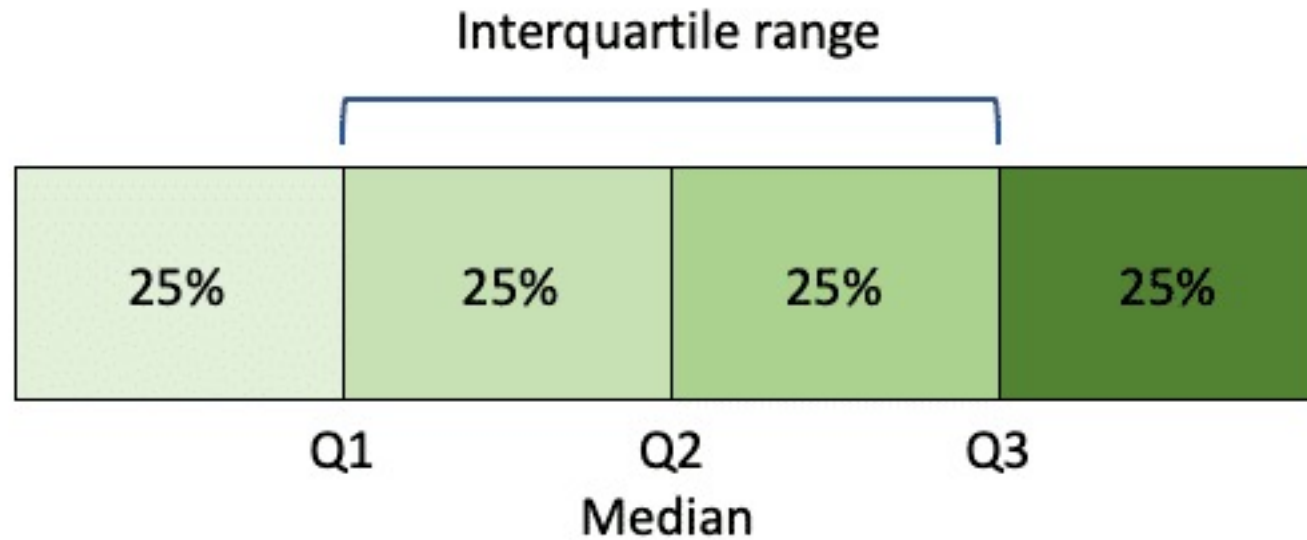
No Mode
122.275
109.085
103.079
102.691
98.228
96.221
94.724
92.619
89.483
75.762

MODE

Most frequent value in your categorical data

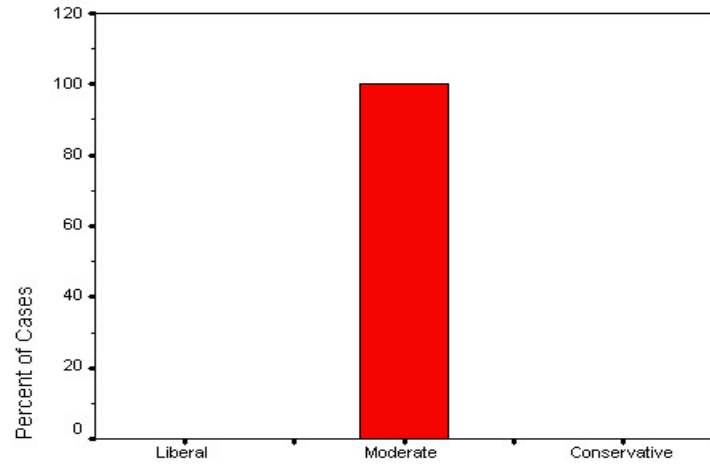


QUARTILE

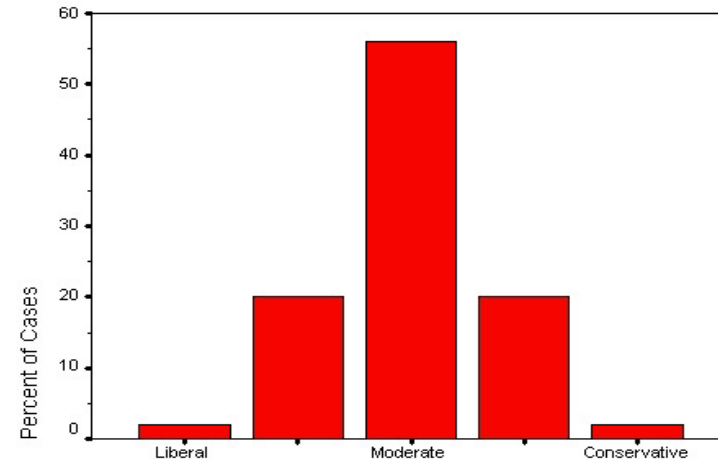


- Divides the number of data points into four parts
- Data must be ordered from smallest to largest to compute quartiles

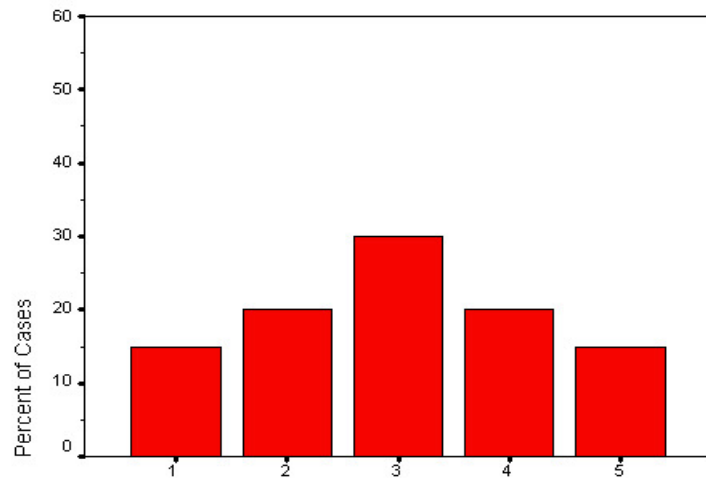
MEASURES OF DISPERSION



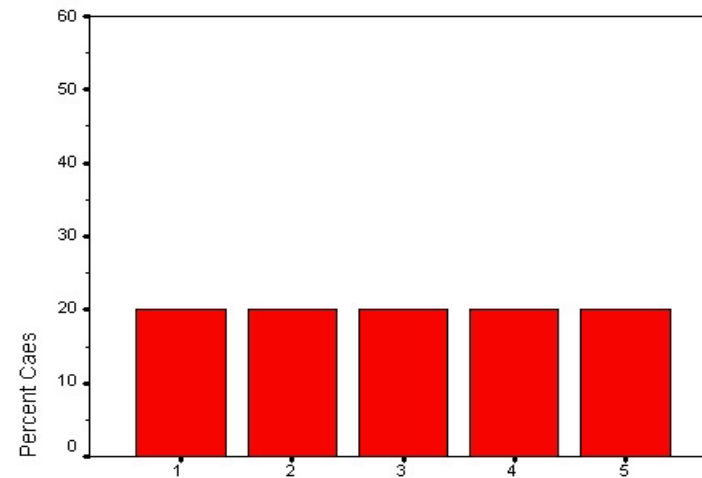
IDEOLOGY (CT = 3, D = 0)



IDEOLOGY (CT = 3, D = small [~normal])



IDEOLOGY (CT = 3, D = modest [~normal])



IDEOLOGY (CT = 3, D = large [uniform])

MEASURES OF DISPERSION

RANGE

Difference between the highest and lowest scores in a data set

23 56 45 65 69 55 62 54 85 25

$$(85 - 23 = 62)$$

MEASURES OF DISPERSION

VARIANCE

How far each observation falls from the mean of the distribution

$$S^2_x = \frac{\sum (x - \bar{x})^2}{N - 1}$$

MEASURES OF DISPERSION

STANDAR DEVIATION

$$Sx = \sqrt{S^2 x}$$

square root of the variance



DEMO

[See Jupyter Notebook](#)

Q&A

Aliados:



Microsoft

Vigilada Mineducación



Workshop #1

<https://github.com/santyjara/Data-science-workshops>

Lab capítulo #1

Primera actividad evaluativa

Aliados:



Microsoft

Vigilada Mineducación



¡Gracias!

Aliados:



Microsoft

Vigilada Mineducación

