Summary statistics

Summary Statistics

- Mean / Average
- Median
- Mode
- Percentiles/quantiles
- Variance/Standard deviation
- Interquartile range

DATA

Observations:

$$x_1, x_2, x_3, \ldots, x_n$$

Example:

A group of 15 people has the following ages:

26, 21, 31, 22, 19,

19, 20, 21, 20, 22,

25, 21, 24, 22, 21.

MEAN

The average observation

NHH
TECH3

Mean / Average

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

$$\bar{x} = \frac{1}{15}(26+21+31+\cdots+21) = 22.3$$

NHH
TECH3

MEDIAN

The middle observation

MEDIAN

Sort the observations
19 19 20 20 21 21 21 21 22 22 22 24 25 26 31

Find the middle observation
19 19 20 20 21 21 21 21 22 22 22 24 25 26 31

Median is 21

NHH
TECH3

MODE

The most common observation

Mode

Make a frequency table

Age	19	20	21	22	24	25	26	31
Frequency	2	2	4	3	1	1	1	1

Which observation is the most frequent?

Age	19	20	21	22	24	25	26	31
Frequency	2	2	4	3	1	1	1	1

Mode is 21

PERCENTILES

The αth observation ($\alpha \in [0,100]$)

α^{th} PERCENTILE

- Sort the observations

  19 19 20 20 21 21 21 21 22 22 22 24 25 26 31

- Divide the sorted observations into 100 equally sized bins

- The αth percentile is the number for which the proportion of observations below is $\alpha\%$, for $\alpha = 0,1,\ldots,100$.

- Many different formulas exists

- Quantile vs percentile

NHH
TECH3

MEDIAN

The 50th percentile

21

1ˢᵀ Qᴜᴀʀᴛɪʟᴇ

The 25ᵗʰ percentiles

20.5

NHH
TECH3

3ᴿᴰ Qᴜᴀʀᴛɪʟᴇ

The 75ᵗʰ percentiles

23

Minimum

The 0th percentiles

19

NHH
TECH3

Maximum

The 100th percentiles

31

VARIANCE

The average squared distance from the mean

Variance

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$s^2 = \frac{1}{15-1}[(26 - 22.3)^2 + (21 - 22.3)^2 + \cdots + (21 - 22.3)^2]$$

$$s^2 = 9.9 \quad years^2$$

NHH
TECH3

Standard deviation

Square root of the variance

STANDARD DEVIATION

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$s = \sqrt{\frac{1}{15-1}[(26-22.3)^2 + (21-22.3)^2 + \cdots + (21-22.3)^2]}$$

$s = 3.2$ years

NHH
TECH3

INTERQUARTILE RANGE

The difference between the 1st and 3rd quartile
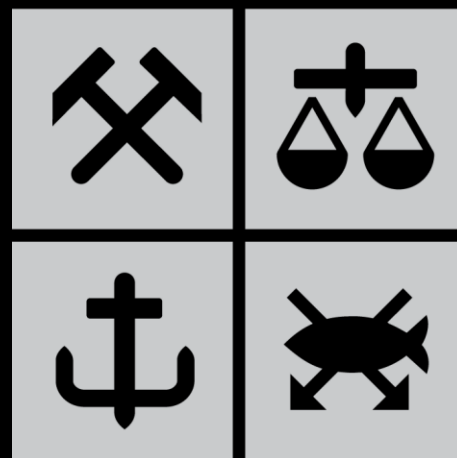
NHH
TECH3

INTERQUARTILE RANGE

$$q_1 = p_{25\%} = 20.5$$
$$q_3 = p_{75\%} = 23.0$$

The interquartile range is

$$q_3 - q_1 = 23 - 20.5 = 2.5$$

NHH
TECH3