

## How to interpret descriptive statistics?

### A) Measures of location

- **percentile:**

30<sup>th</sup> percentile=3, i.e. 30% of all students said that they had 3 family members or below

75<sup>th</sup> percentile=4, i.e. 75% of all students said that they had 4 family members or below

95<sup>th</sup> percentile=5, i.e. 95% of all students said that they had 5 family members or below

- **quartile:**

Q1=3 (25<sup>th</sup> percentile), i.e. 25% of all students said that they had 3 family members or below

Q2=4 (50<sup>th</sup> percentile; median), i.e. 50% of all students said that they had 4 family members or below

Q3=4 (75<sup>th</sup> percentile), i.e. 75% of all students said that they had 4 family members or below

Q4=5 (100<sup>th</sup> percentile; maximum value), i.e. all students said that they had 5 family members or below

### B) Measures of central tendency (centre)

- **mean** (arithmetic weighted):

$\bar{x} = 3.49$ , i.e. the average number of family members is 3 (4)

- **mode:**

$\hat{x} = 4$ , i.e. most of students said that they had 4 family members

- **median:**

$\tilde{x} = 4$ , i.e. half (50%) of students have 4 family members or below and half (50%) of students have 4 family members or above

### C) Measures of variability (spread)

- **IQR (interquartile range):**

IQR=1, it refers to spread of the middle 50% of data

- **range:**

R=4, i.e. the difference between the minimum and maximum number of family members is 4

- **variance** (sample):

$s^2 = 1.38$ , it can't be interpreted because it's expressed in units squared (in the same units as the statistical attribute is measured)

- **standard deviation** (sample):

$s = 1.17$ , i.e. 68% of students said that they had  $3(4) \pm 1$  family member(s) (mean  $\pm$  one standard deviation according to the rule of one sigma = empirical rule)

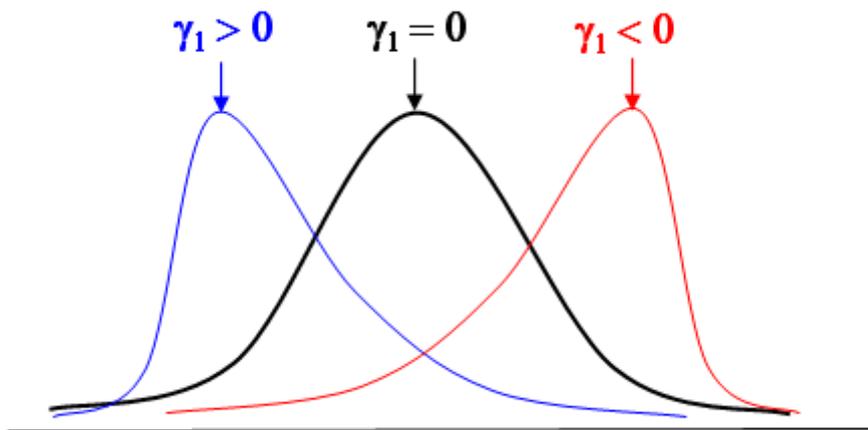
- **coefficient of variation:**

$CV=0.3369=33.7\%$ , i.e. the value of (sample) standard deviation is 33.7% of the value of (sample) mean, it refers to variability within the data set especially when we'd like to compare several data sets with different measurement units

**D) Measures of skewness (refer to the shape of distribution)**

- **coefficient of skewness:**

$\gamma_1=-0.38$ , i.e. data are negatively skewed (not a symmetrical distribution), tail of the distribution is to the left and mode is located to the right of mean



**E) Measures of kurtosis (refer to the shape of distribution)**

- **coefficient of kurtosis:**

$\gamma_2=-0.83$ , i.e. the shape of the distribution is flat (platykurtic)

