GDAŃSK UNIVERSITY OF TECHNOLOGY
FACULTY OF MANAGEMENT AND ECONOMICS

ESSENTIALS OF STATISTICS NO. III.

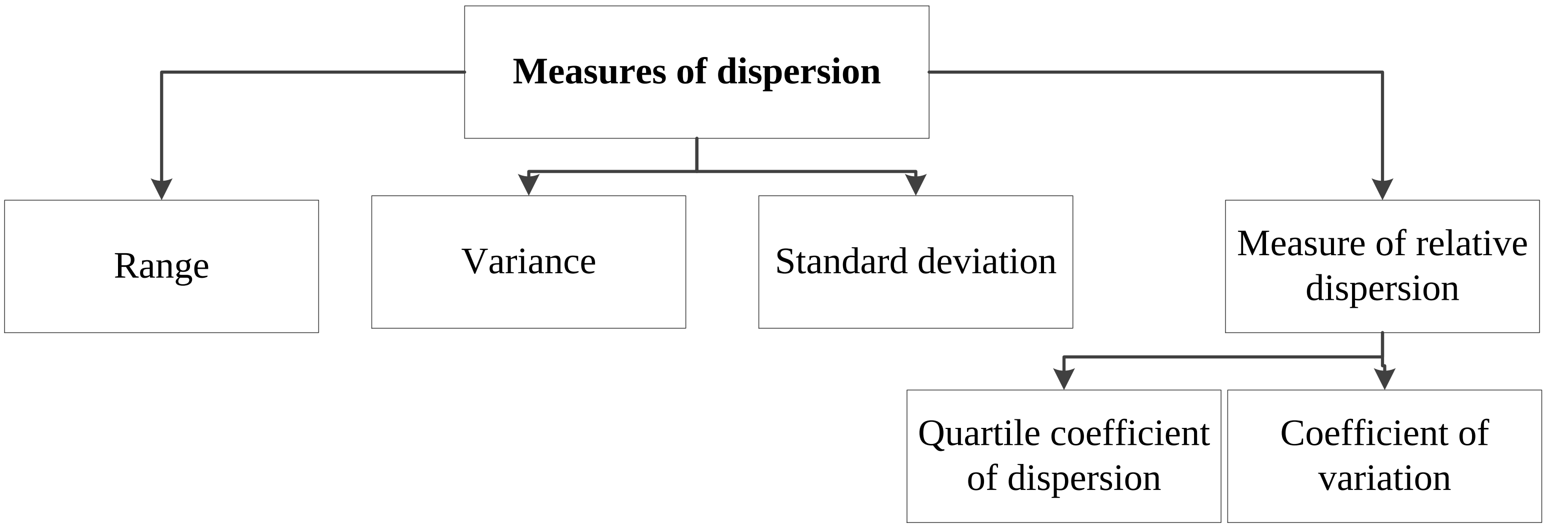
MEASURES OF DISPERSION

KAROLINA TURA, PHD
DEPARTMENT OF ECONOMIC SCIENCE

AGENDA

1. **Measures of dispersion**
    1. Range
    2. Variance
    3. Standard deviation
2. **Measures of relative dispersion**
    1. Coefficient of range
    2. Quartile coefficient of dispersion
    3. Coefficient of variation
3. **Rules**
    1. Chebyshev’s theorem
    2. Rule of the thumb
4. **Practice**

MEASURES OF DISPERSION



RANGE

The range of a set of a measurements is the numerical difference between the largest and the smallest observations

$$x_{\max} - x_{\min}$$

The inter-quartile range is the difference between the first and the third quartiles

$$Q_3 - Q_1$$

TASK 1.

15 students scored the points on the exam of Statististics:

1,2,3,4,4,5,5,5,6,6,6,6,7,8,8.

Find:

a) range,

b) inter-quartile range.

$$\boxed{\begin{aligned} &x_{\max} - x_{\min} = 8 - 1 = 7 \\ &Q_3 - Q_1 = 6 - 4 = 2 \end{aligned}}$$

COEFFICIENT OF RANGE

Coefficient of range

$$\frac{x_{\max} - x_{\min}}{x_{\max} + x_{\min}}$$

Quartile coefficient of dispersion

$$\frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Measures dispersion

Mainly used to compare the data sets

TASK 2.

15 students scored the points on the exam of Statististics:

1,2,3,4,4,5,5,5,6,6,6,6,7,8,8.

Find:

a) Coefficient of range,

b) Quartile coefficient of dispersion

$$\boxed{\begin{aligned} &\frac{x_{\max} - x_{\min}}{x_{\max} + x_{\min}} = \frac{8-1}{8+1} = \frac{7}{9} \\ &\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{6-4}{6+4} = 0.2 \end{aligned}}$$

VARIANCE

Variance

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

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 n_i$$ Frequency table

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(\dot{x}_i - \bar{x})^2 n_i$$ Frequency table with equal class intervals

Midpoint of each class

The variance is the average of all squared deviations from the mean

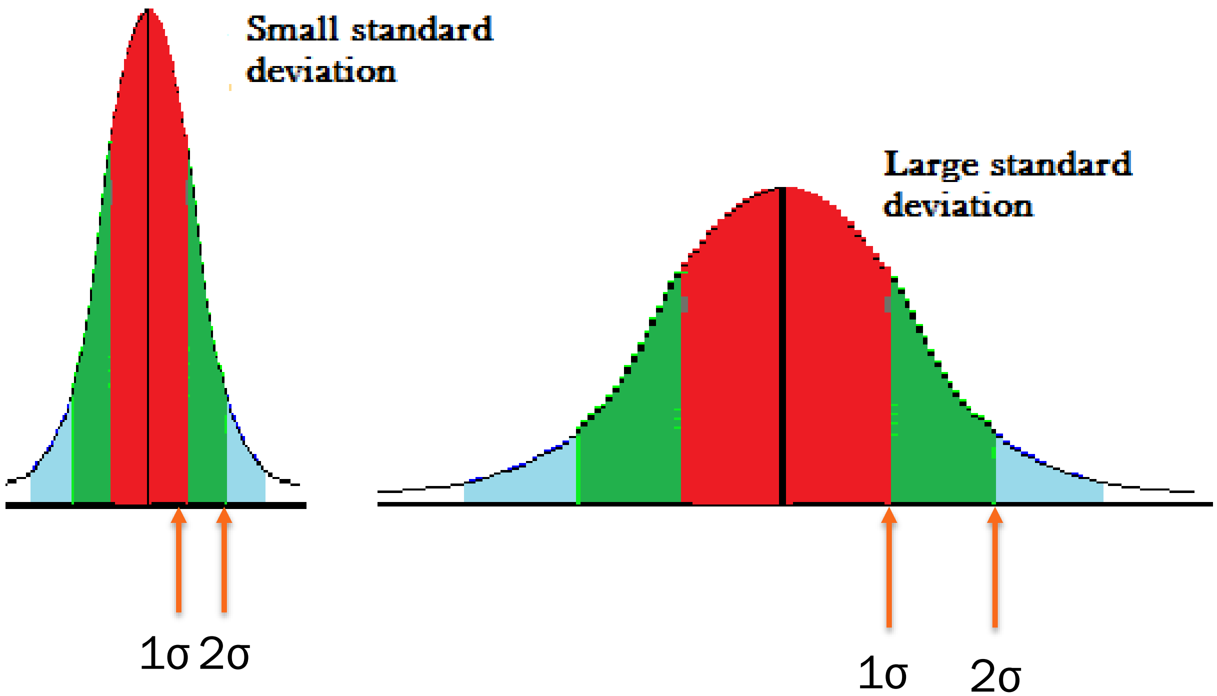
How to examine the degree of dispersion?

STANDARD DEVIATION

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

The standard deviation of a set of measurements is the positive square root of the variance of measurements.
It measures how concentrated the data are around the mean; the more concentrated, the smaller the standard deviation.
A small standard deviation means that the values in a statistical data set are close to the mean of the data set, on average, and a large standard deviation means that the values in the data set are farther away from the mean, on average.

STANDARD DEVIATION



S sample
σ population

COEFFICIENT OF VARIATION (RSD, CV)

$$CV = \frac{S}{\bar{x}} * 100$$

Relative standard deviation

V	Interpretation
$CV \leq 35\%$	Small dispersion.
$35\% < CV \leq 60\%$	Moderate dispersion
$60\% < CV \leq 75\%$	Large dispersion
$75\% < CV \leq 100\%$	Great dispersion

TASK 3. DETAILED FORM

15 students scored the points on the exam of Statististics:

1,2,3,4,4,5,5,5,6,6,6,6,7,8,8. The average is $\bar{x} = 5.07$.

Find and interpret:

a) variance,

b) standard deviation,

c) coefficient of variation.

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{56.93}{15} \approx 3.795$$

$$\sqrt{S^2} = S = \sqrt{3.795} \approx 1.95$$

$$CV = \frac{S}{\bar{x}} * 100\% = \frac{1.95}{5.07} * 100\% \approx 38,5\%$$

HINT

	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
	1,00	-4,07	16,54
	2,00	-3,07	9,40
	3,00	-2,07	4,27
	4,00	-1,07	1,14
	4,00	-1,07	1,14
	5,00	-0,07	0,00
	5,00	-0,07	0,00
	5,00	-0,07	0,00
	6,00	0,93	0,87
	6,00	0,93	0,87
	6,00	0,93	0,87
	6,00	0,93	0,87
	7,00	1,93	3,74
	8,00	2,93	8,60
	8,00	2,93	8,60
Sum	76,00	0	56,93

TASK 4. FREQUENCY TABLE

The number of hours (per week) which students spend on learning Statistics in 2014 is given in table. The average is $\bar{x} = 1.07$.

Find and interpret:

a) variance,

b) standard deviation,

c) coefficient of variation.

x Hours (per week)	n Frequency
0	25
1	54
2	11
3	9
4	1

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 n_i = \frac{80.51}{100} = 0.8051$$

$$\sqrt{S^2} = \sqrt{0.8051} \approx 0.897$$

$$CV = \frac{S}{\bar{x}} * 100\% = \frac{0.897}{1.07} \approx 84\%$$

HINT

	x_i	n	$x_i n$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n$
	0,00	25,00	0,00	-1,07	1,14	28,62
	1,00	54,00	54,00	-0,07	0,00	0,26
	2,00	11,00	22,00	0,93	0,86	9,51
	3,00	9,00	27,00	1,93	3,72	33,52
	4,00	1,00	4,00	2,93	8,58	8,58
Sum		100,00	107,00	4,65	21,62	80,51

TASK 5. FREQUENCY TABLE WITH INTERVALS

Observations of consumer credit borrowers were collected at the Bank X in 2008. Results were grouped in a series of observations given in the table below. The average is 11.48.

Credit [zł] x	Credit borrowers n
0-5	300
5-10	470
10-15	693
15-20	328
20-25	120
25-30	35

Find and interpret:

a) variance,

b) standard deviation,

c) coefficient of variation.

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(\dot{x}_i - \bar{x})^2 n_i = $$

$$= \frac{67800.22}{1946} \approx 34.84$$

$$S = \sqrt{S^2} = \sqrt{34.84} \approx 5.9$$

$$CV = \frac{S}{\bar{x}} * 100\% = $$

$$= \frac{5.9}{11.48} \approx 51.4\%$$

HINT

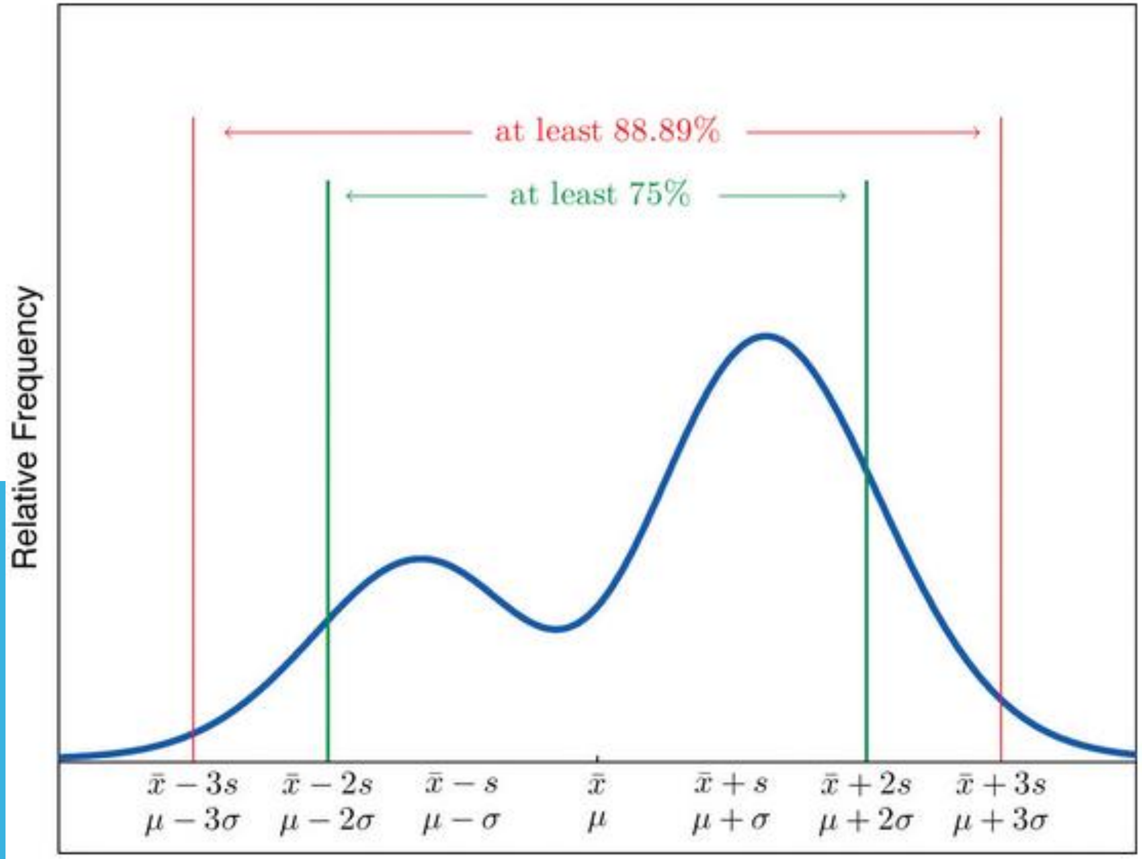
	x_i	n	$\dot{x}$	$\dot{x}_i n$	$\dot{x}_i - \bar{x}$	$(\dot{x}_i - \bar{x})^2$	$(\dot{x}_i - \bar{x})^2 n$
	0-5	300,00	2,50	750,00	-8,98	80,64	24192,12
	5-10	470,00	7,50	3525,00	-3,98	15,84	7444,99
	10-15	693,00	12,50	8662,50	1,02	1,04	721,00
	15-20	328,00	17,50	5740,00	6,02	36,24	11886,85
	20-25	120,00	22,50	2700,00	11,02	121,44	14572,85
	25-30	35,00	27,50	962,50	16,02	256,64	8982,41
Sum		1946,00		22340,00	21,12	511,84	67800,22

CHEBYSHEV'S THEOREM

Chebyshev's theorem

Given any set of measurements and a number $k \geq 1$, the fraction of these measurements that lie within k standard deviations of their mean is at least $1-1/k^2$

k	Interval	Fraction of measurements in interval
1	$(\overline{x} - s, \overline{x} + s)$	At least 0
2	$(\overline{x} - 2s, \overline{x} + 2s)$	At least 3/4
2,5	$(\overline{x} - 2.5s, \overline{x} + 2.5s)$	At least 21/25
3	$(\overline{x} - 3s, \overline{x} + 3s)$	At least 8/9



TASK 6. CHEBYSHEV'S THEOREM

The telephone-call durations have a mean of 10.26 and a standard deviation of 4.29. Giving nothing else about the distribution od the durations, Chebyshev's theorem asserts that at least ¾ (75%) of the durations lie in the interval :

…………………………………………………………………………………………

$$(\bar{x} - 2s, \bar{x} + 2s) = (10.26 - 2 * 4.29, 10.26 + 2 * 4.29) = (8.58, 11.94)$$

RULE OF THE THUMB

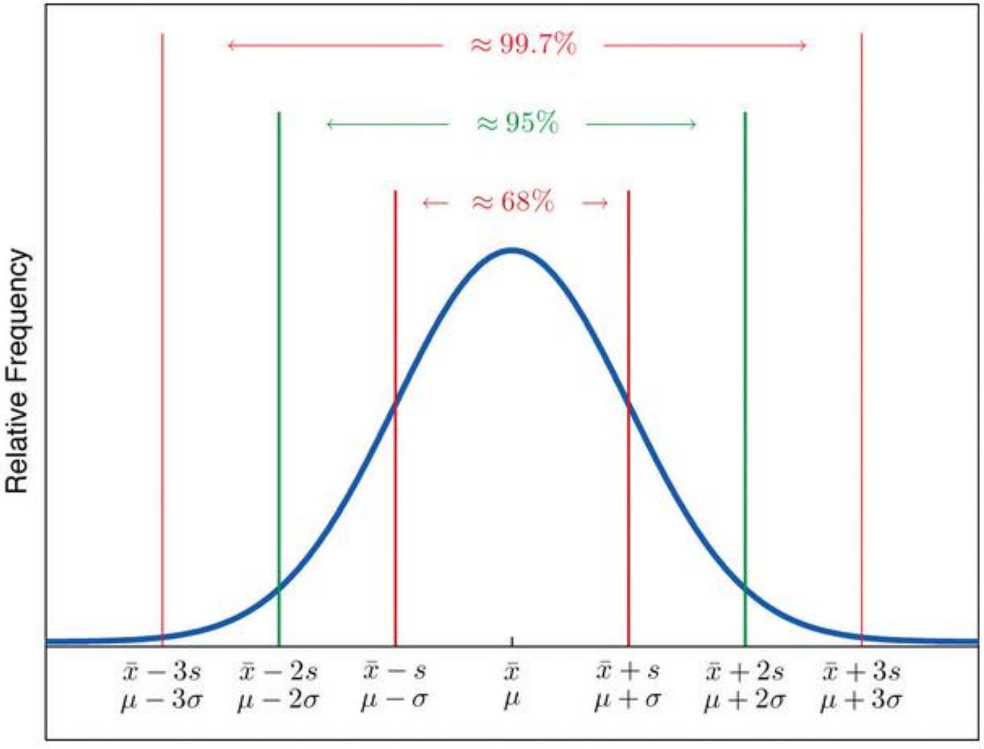
Empirical rule

If a sample of measurements has a mound shaped distribution, the interval:

a) $(\bar{x} - s, \bar{x} + s)$ contains approximately 68% of the measurements,

b) $(\bar{x} - 2s, \bar{x} + 2s)$ contains approximately 95% of the measurements,

c) $(\bar{x} - 3s, \bar{x} + 3s)$ contains virtually all of the measurements.



TASK 7. EMPIRICAL RULE

The telephone-call durations have a mean of 10.26 and a standard deviation of 4.29.

The sample of measurements has a mound shaped distribution. Empirical rule asserts that at least 95% of the durations lie in the interval :

……………………………………………………………………………………………

$$(\bar{x} - 2s, \bar{x} + 2s) = (10.26 - 2 * 4.29, 10.26 + 2 * 4.29) = (8.58, 11.94)$$

STATISTICA

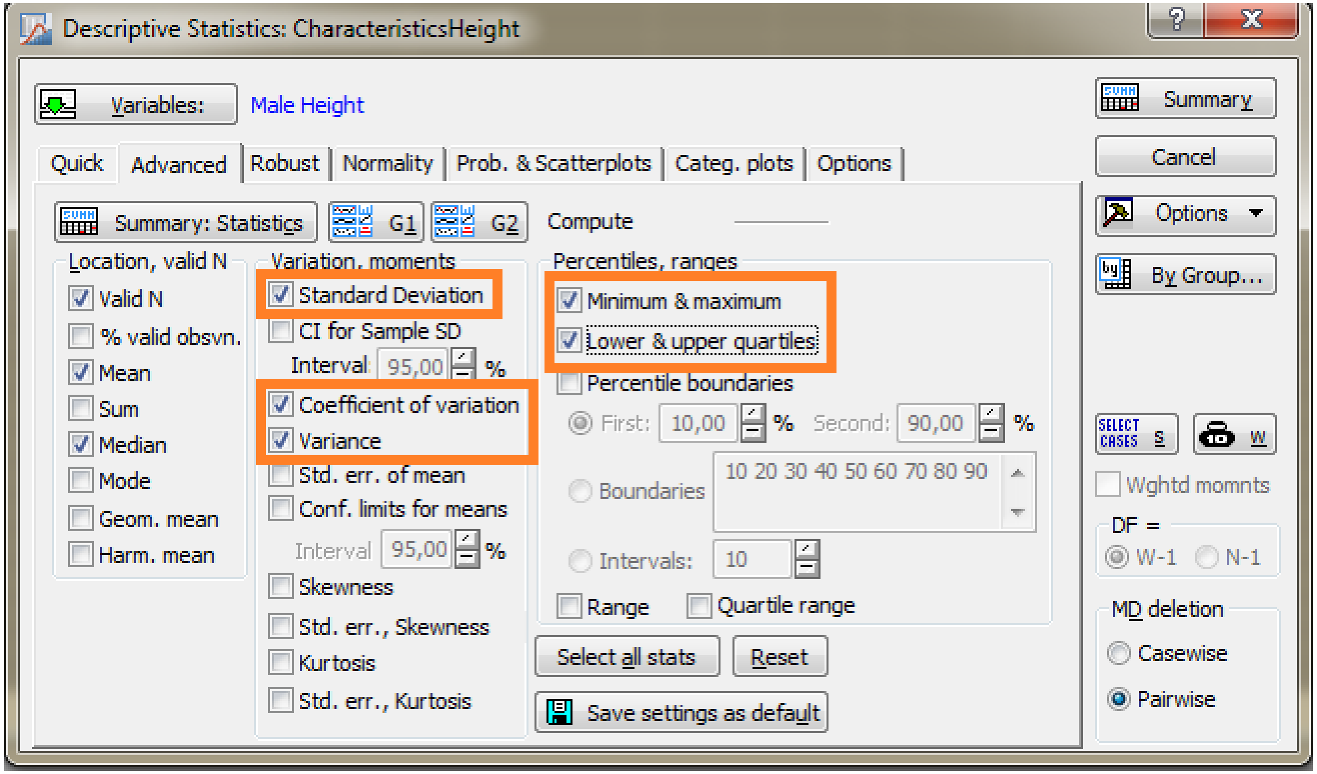
TASK 8.

Scientists examined height of randomly selected men from the city of Gdynia. On the basis of the data contained in the file CharacteristicsHeight.sta perform an analysis. Find and interpret:

a) range,
b) inter-quartile range,
c) variance,
d) standard deviation,
e) coefficient of variation.

Statistisc>Basic statistics/tables> Descriptive statistics

HINT



Variable	Descriptive Statistics (CharacteristicsHeight) Valid N	Mean	Median	Minimum	Maximum	Lower Quartile	Upper Quartile	Variance	Std.Dev.	Coef.Var.
Male Height	52	67,78846	68,00000	57,00000	74,00000	66,00000	70,00000	12,17006	3,488561	5,146247

TASK 9.

Scientists examined weight of randomly selected women from the city of Gdynia. On the basis of the data contained in the file CharacteristicsHeight.sta perform an analysis.

Find and interpret:

a) range,
b) inter-quartile range,
c) variance,
d) standard deviation,
e) coefficient of variation.

PREPARATION FOR THE NEXT CLASSES

McClave, J. T., Benson, P. G., Sincich, T. (2008) , *Statistics for Business & Economics*, Pearson Education Inc., New Jersey, p. 74-100;

Thank you for your attention

GDAŃSK UNIVERSITY OF TECHNOLOGY

FACULTY OF MANAGEMENT AND ECONOMICS