



**GDAŃSK UNIVERSITY
OF TECHNOLOGY**

FACULTY OF MANAGEMENT AND ECONOMICS

ESSENTIALS OF STATISTICS

CORRELATION COEFFICIENTS (QUANTITATIVE VARIABLES)

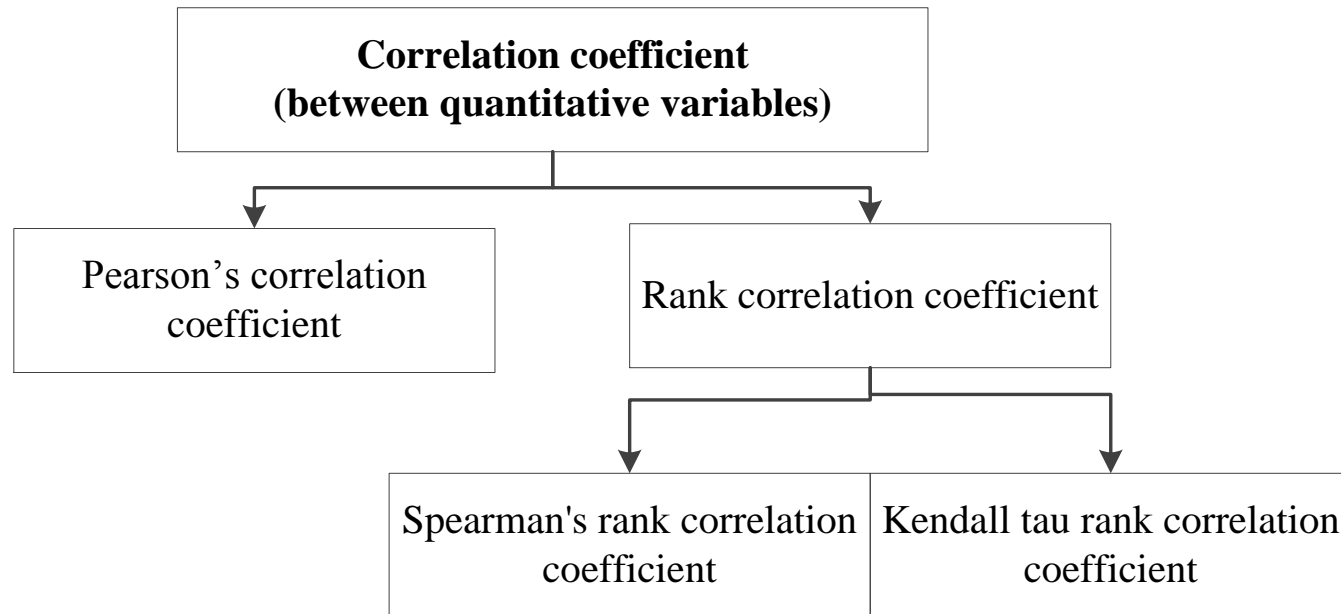
KAROLINA TURA, PHD

DEPARTMENT OF ECONOMIC SCIENCE

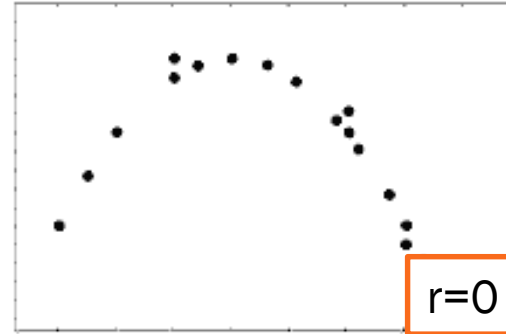
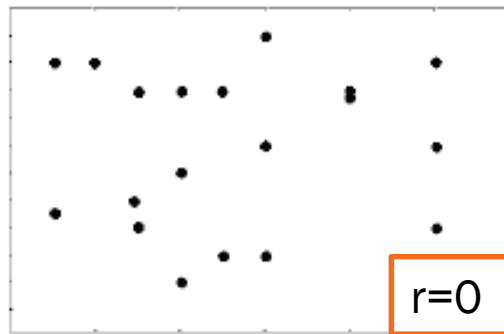
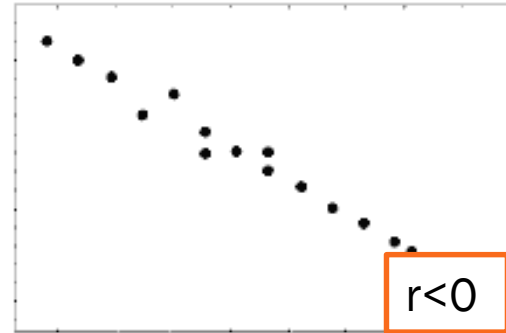
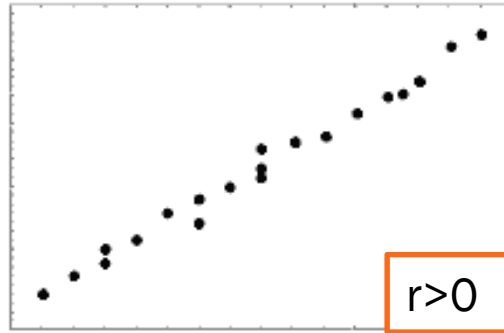
AGENDA

1. **Pearson's Correlation Coefficient**
2. **Rank Correlation Coefficients**
3. **Practice**

CORRELATION BETWEEN QUANTITATIVE VARIABLES



PEARSON'S CORRELATION COEFFICIENT (1)



PEARSON'S CORRELATION COEFFICIENT (2)

Covariance

A measure of how much two random variables change together

$$\text{cov}(X , Y) = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})$$

∨

$$\text{cov}(X , Y) = \frac{1}{n} \sum x_i y_i - (\bar{x}\bar{y})$$

Pearson's Correlation Coefficient

$$r_{XY} = \frac{\text{cov}(xy)}{S_x S_y} = r_{YX}$$

$$r_{XY} \in \langle -1, 1 \rangle$$

1. PEARSON'S CORRELATION COEFFICIENT

Direction	
$r_{xy} > 0$	Positive linear correlation
$r_{xy} = 0$	Lack of linear correlation
$r_{xy} < 0$	Negative linear correlation
$r_{xy} = 1 \vee r_{xy} = -1$	Perfect linear correlation
Strength	
$ r_{xy} < 0,2$	Lack of linear correlation
$0,2 \leq r_{xy} < 0,4$	Small linear correlation
$0,4 \leq r_{xy} < 0,7$	Moderate linear correlation
$0,7 \leq r_{xy} < 0,9$	Large linear correlation
$ r_{xy} \geq 0,9$	Great linear correlation

1. COEFFICIENT OF DETERMINATION R^2

$$R^2 = r_{xy}^2$$

$$R^2 \in \langle 0, 1 \rangle$$

The *coefficient of determination* is the ratio of the explained variation to the total variation

TASK 1. PEARSON'S CORRELATION COEFFICIENT

The following table shows the points scored on the exam in Statistics and the hours of studying which 10 students had spent before the exam. Create a scatter plot. Find and interpret the corresponding Pearson's correlation coefficient and coefficient of determination.

X	136	132	141	144	152	148	158	163	154	155
Y	12	4	7	11	8	5	14	12	9	7

HINT

Moderate positive linear correlation

i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	136	12	-12,3	3,1	151,29	9,61	-38,13
2	132	4	-16,3	-4,9	265,69	24,01	79,87
3	141	7	-7,3	-1,9	53,29	3,61	13,87
4	144	11	-4,3	2,1	18,49	4,41	-9,03
5	152	8	3,7	-0,9	13,69	0,81	-3,33
6	148	5	-0,3	-3,9	0,09	15,21	1,17
7	158	14	9,7	5,1	94,09	26,01	49,47
8	163	12	14,7	3,1	216,09	9,61	45,57
9	154	9	5,7	0,1	32,49	0,01	0,57
10	155	7	6,7	-1,9	44,89	3,61	-12,73
Sum	1483	89	\bar{x}	\bar{y}	890,1	96,9	127,3

$$r_{XY} = \frac{\text{cov}(X, Y)}{S_x S_y} = \frac{12,73}{9,43 * 3,11} = 0,43$$

$$\text{cov}(X, Y) = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}) = \frac{127,3}{10} = 12,73$$

$$S_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{890,1}{10} \approx 89 \quad S_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{96,9}{10} \approx 9,7$$

$$\sqrt{S_x^2} = S_x = \sqrt{89} \approx 9,43$$

$$\sqrt{S_y^2} = S_y = \sqrt{9,7} \approx 3,11$$

2. SPEARMAN'S RANK CORRELATION COEFFICIENT

$$R_{XY} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Difference between the ranks

$$R_{XY} \in \langle -1, 1 \rangle$$

TASK 2. SPEARMAN'S RANK CORRELATION COEFFICIENT

The following table shows the points scored on the exam in Statistics and the hours of studying which 10 students had spent before the exam.

Find and interpret the corresponding Spearman's correlation coefficient.

X	136	132	141	144	152	148	158	163	154	155
Y	12	4	7	11	8	5	14	12	9	7

HINT

i	x_i	y_i	$rankX$	$rankY$	d_i	d_i^2
1	136	12	9	2,5	6,5	42,25
2	132	4	10	10	0	0
3	141	7	8	7,5	0,5	0,25
4	144	11	7	4	3	9
5	152	8	5	6	-1	1
6	148	5	6	9	-3	9
7	158	14	2	1	1	1
8	163	12	1	2,5	-1,5	2,25
9	154	9	4	5	-1	1
10	155	7	3	7,5	-4,5	20,25
Sum	1483	89	x	x		86

$$R_{XY} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} =$$

$$= 1 - \frac{6 * 86}{10(10^2 - 1)} = 0.48$$

STATISTICA

TESTING THE SIGNIFICANCE OF PEARSON AND SPEARMAN CORRELATION COEFFICIENTS

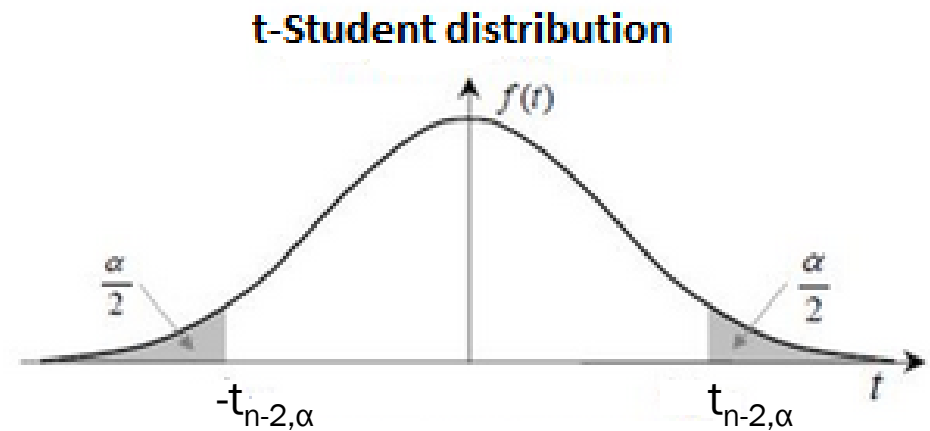
$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0 \quad (< \quad >)$$

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2} \sim t\text{-Student} \quad v = n-2$$

$$|t| \geq t_\alpha \text{ reject } H_0$$

$$|t| < t_\alpha \text{ not reject } H_0$$



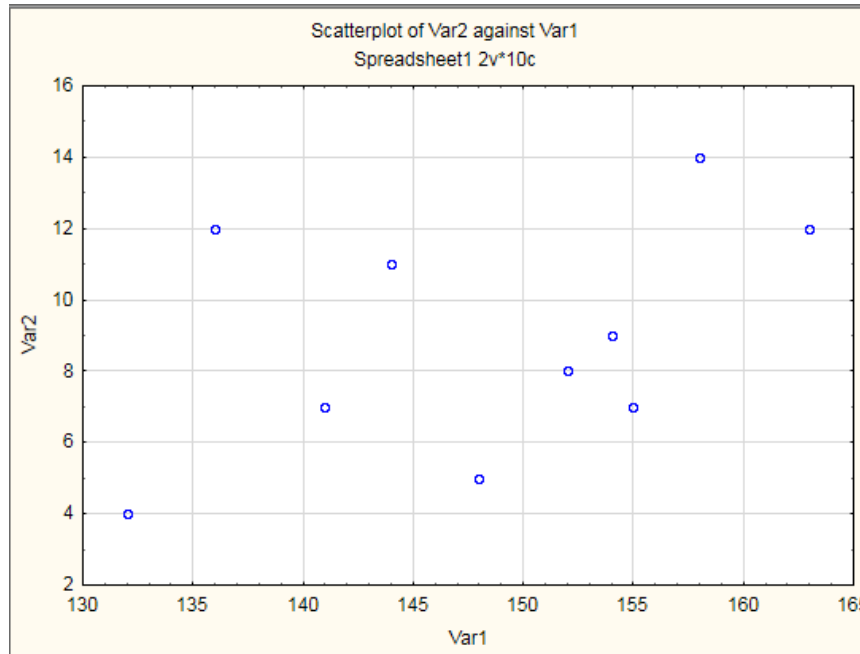
TASK 1. PEARSON'S CORRELATION COEFFICIENT

The following table shows the points scored on the exam in Statistics and the hours of studying which 10 students had spent before the exam. Create a scatter plot. Find and interpret the corresponding Pearson's correlation coefficient and coefficient of determination.

X	136	132	141	144	152	148	158	163	154	155
Y	12	4	7	11	8	5	14	12	9	7

Statistics>Basic statistics/tables>Correlation matrices>...
Graphs>Scatterplots...

HINT



Var. X & Var. Y	Correlations (Spreadsheet1)						
	Mean	Std.Dv.	r(X,Y)	r2	t	p	N
Var1	148,3000	9,944848					
Var2	8,9000	3,281260	0,433458	0,187886	1,360453	0,210777	11

TESTING THE SIGNIFICANCE OF PEARSON CORRELATION COEFFICIENT

TASK 2. SPEARMAN'S RANK CORRELATION COEFFICIENT

The following table shows the points scored on the exam in Statistics and the hours of studying which 10 students had spent before the exam.

Find and interpret the corresponding Spearman's correlation coefficient.

X	136	132	141	144	152	148	158	163	154	155
Y	12	4	7	11	8	5	14	12	9	7

Statistics>Nonparametric statistics>Correlations>Spearman rank...

HINT

Spearman Rank Order Correlations (Spreadsheet1)			
MD pairwise deleted			
Marked correlations are significant at $p < ,05000$			
Variable	Var1	Var2	
Var1	1,000000	0,475619	
Var2	0,475619	1,000000	

PREPARATION FOR THE NEXT CLASSES

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

**Thank you for your
attention**



**GDAŃSK UNIVERSITY
OF TECHNOLOGY**

FACULTY OF MANAGEMENT AND ECONOMICS