**EXERCISE 5 – CORRELATION**

1. Which of the following correlation coefficients expresses the strongest association?

   a) 0.55   b) 0.09   c) -0.77   d) 0.1   e) 1.05

2. For which of the coefficients from the previous question it applies that a person with above average scores X will probably also have above average scores Y?

3. We have five representative samples of people aged 15, 20, 30, 45 and 60 years who completed a questionnaire of political conservatism. In these 5 samples in the given order were the average scores of political conservatism as follows: 60, 85, 80, 70, 65. Correlation between age and political conservatism is:

   b) 1.0   b) -1.0   c) linear   d) nonlinear

4. For each scatterplot below choose corresponding association description:

   a) perfect positive linear relationship (r = 1.0)

   b) moderate positive linear relationship (r = 0.5)

   c) no linear relationship (r = 0)

   d) moderate negative relationship (r = -0.5)

   e) perfect negative linear relationship (r = -1.0)
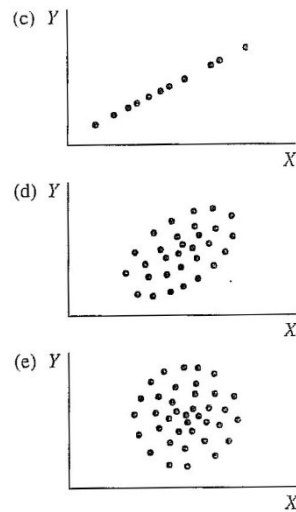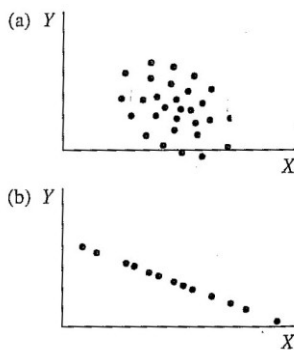
5. How is Pearson's coefficient influenced by:

   a) limited variability?

   b) differences in distribution of the correlated variables?

   c) outliers?

   d) using extreme groups as samples?

6. Estimate correlation between variable pairs listed below – is it positive, negative, or zero?

   a) height in cm, weight in kg

   b) age in months, time in run for 50 meters

   c) math grade, reading grade

d) math grade, number of missed school lessons in a year

e) IQ, personal identification number

f) interest in sports, interest in politics

g) mileage on car speedometer, year when was car produced

h) maximum daily temperature, household water consumption per day

7. Suppose the answer to the question 6.g was r = -0.8, how would the correlation coefficient change, if we instead variable "year when was car produced" use variable "car's age"?

9. If correlation between X and Y is 0.5, how does the correlation change if we transform X to T-scores?

10. If r = 1 and z_x = -0.5, what is z_y? If r= - 1. and z_x = 0.8, what is z_y?

11. Regarding interpretation, between correlations 0.2 and 0.4 and between correlations 0.5 and 0.7…:

a) is approximately the same difference

b) there is bigger difference between the first pair than between the second pair

c) there is smaller difference between the first pair than between the second pair

d) the difference can't be compared

Justify your answer.

12. IQ scores from test A are consistently higher about 10 points than IQ score from test B. What is theoretically the highest possible correlation between test A and test B?

13. Study 1 found correlation coefficient -0.2 and study 2 found correlation between the same variables 0.4. How did the proportion of shared variance change between study 1 and study 2?

a) The shared variance proportion increased twice.

b) The shared variance proportion increased four times.

c) The shared variance proportion increased six times.

d) The shared variance proportion decreased three times.

e) Can´t be determined, because one of the correlations is positive and the other is negative.

14. If standard deviations of two correlated variables are s_X = 3 a s_Y = 15, what is the highest possible covariance between them? ($r_{XY} = c_{XY}/s_X s_Y$)

16. One study about heart attacks states that people who regularly go to church are in lower risk of heart attack. Which of the following statements is correct?

a) If you start to go to church, your risk of getting heart attack will certainly get lower.

b) There certainly isn't any causal relationship between the two variables.

c) If you regularly go to church, your probability of getting heart attack is lower than in people who don't go to church.

d) This correlation certainly shows causal relationship.

17. One study showed relatively low correlation between IQ and creativity (r = 0.2). SD of IQ scores in their sample was only 5. How would the correlation change if the variability wasn't such limited?

18. A study on sample of 280 students of teaching states almost zero correlation (r = 0.1) between study results and the ability of teaching. Evaluation of these 280 students in the study was done by two independent experience teacher. Correlation between their evaluations of the 280 students was 0.21. How does this information influence your interpretation of the correlation between study results and the ability of teaching?

19. a) On the following sample of 10 pupils compute Pearson's correlation coefficient between IQ scores and arithmetic scores:

Pupil	IQ	Arithmetic	Pupil	IQ	Arithmetic
A	105	15	F	96	10
B	120	23	G	107	4
C	83	11	H	117	30
D	137	22	I	108	18
E	114	17	J	130	14

b) Transform IQ and arithmetic to rankings and draw scatterplot. Make the units in the scatterplot's axis X and Y the same. Does the relationship look linear? Does the correlation r you computed in the previous step seem reasonable?

c) How can the sample mean and sample standard deviation of IQ scores influence the correlation, if we know that population IQ mean is 100 and population IQ SD is 15?

20. What information can we gather based on scatterplot?

21. Compute correlation between the following set of z-scores:

z_x	z_y
$-.32$	$-.56$
$-.10$	0
$.42$	$-.12$
0	$.68$

22. Estimate r for each of the following graphs:

(a)   (b)

(c)   (d)
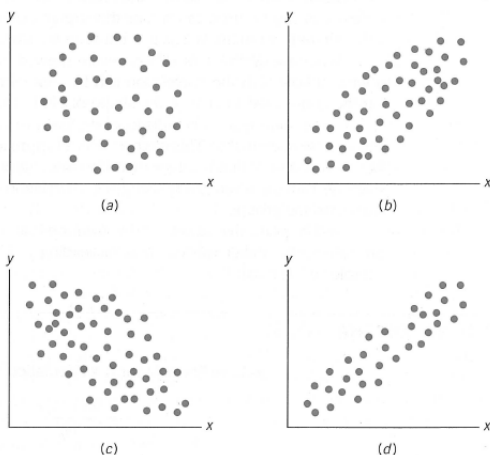
29. For a group of 75 subjects applies: $\sum(z_x z_y) = 64$. What is Pearson's correlation coefficient?

30. If you use Pearson's correlation coefficient for computing correlation between variables where the relationship between X and Y is not linear, how it influence the coefficient?

33. A psychologist is interested in strength of association between age and some tasks requiring motor abilities? Draw scatterplot for the following data. How does the relationship look like? Is Pearson's correlation coefficient suitable for describing these data?

Age in Years	Number of Errors
6	23
7	19
8	17
9	16
10	16
11	18
12	18
13	19
14	20
15	22

35. Carrie (1981) was measuring association between reporting symptoms during pregnancy and during menstruation and association of these reports with general tendency to report psychological and physiological symptoms. Among other results she discovered significant correlation between the number of symptoms experience during menstruation and the number of symptoms experienced during pregnancy. The following data are hypothetical, but consistent with the original study. Compute Pearson's correlation coefficient, don't forget to check assumptions.

Hypothetical questionnaire scores

Menstruation symptoms	Pregnancy symptoms
93	87
75	64
34	78
23	55
76	43
34	45
21	20
34	54
60	60
45	82
67	67
50	48
89	72
61	68
56	45
82	75
45	34
53	55
71	50
59	90
90	56
43	62
49	32

36. A researcher decided to explore relationship between height and self-esteem. He tested by subjects and already transformed all obtained score to z-scores:

ID	Height	Self-esteem
1	0,61	1,09
2	-0,77	-0,36
3	1,23	0,72
4	-1,23	1,45
5	0,15	0

   a) Compute Pearson's correlation coefficient

   b) What does the correlation coefficient says about the relationship between height and self-esteem?

   c) If we wanted to predict self-esteem form height, what percentage of self-esteem variability would height explain?

37. Look at the following scatterplot and from general factors that influence correlation choose one that in this case makes correlation higher compared to reality, and one that in this case makes correlation lower compared to reality.

Frequency Scatterplot (čápi 18v*971c)
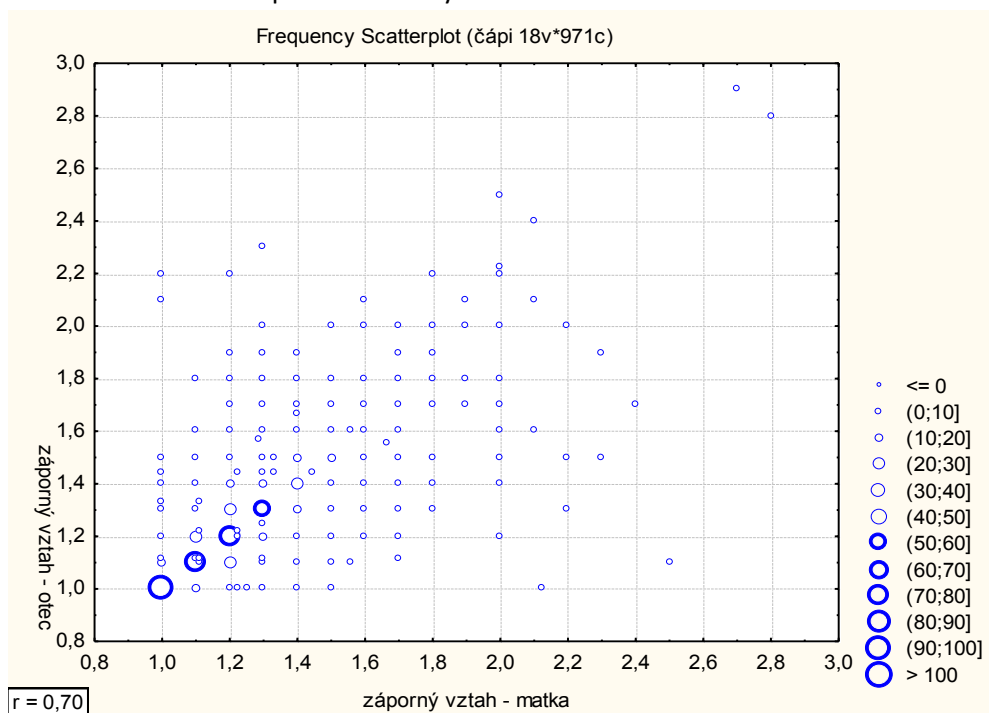
38. What is the full name for Kendall coefficient? Compute it on the following data:

| ID	Car's age	Person's age
a	16	25
b	8	27
c	2	30
d	1	45
e	25	89

39. What are marginal frequencies?

41. Estimate Pearson's correlation coefficient

   for the first scatterplot: a) -0.8, b) -0.3, c) 0.0, d) 0.3, e) 0.8

   for the second scatterplot: a) -0.8, b) -0.3, c) 0.0, d) 0.3, e) 0.8

   If we computed from the data in the first scatterplot Spearman correlation coefficient, would it be lower, higher or same as Pearson's correlation coefficient? Why?
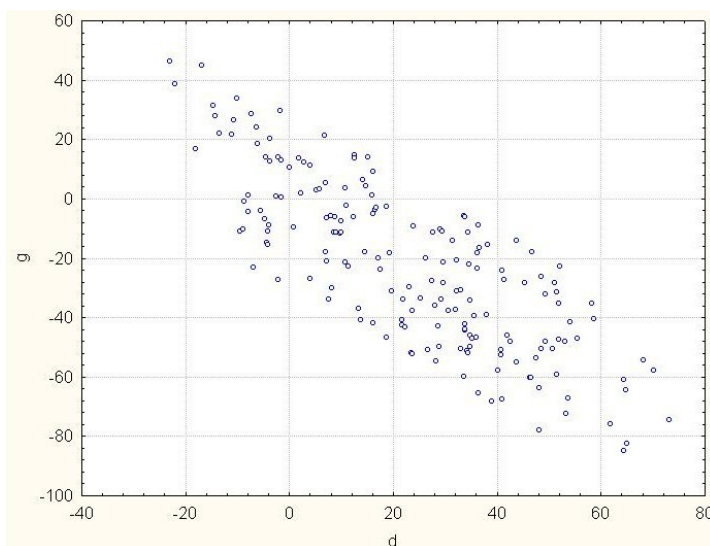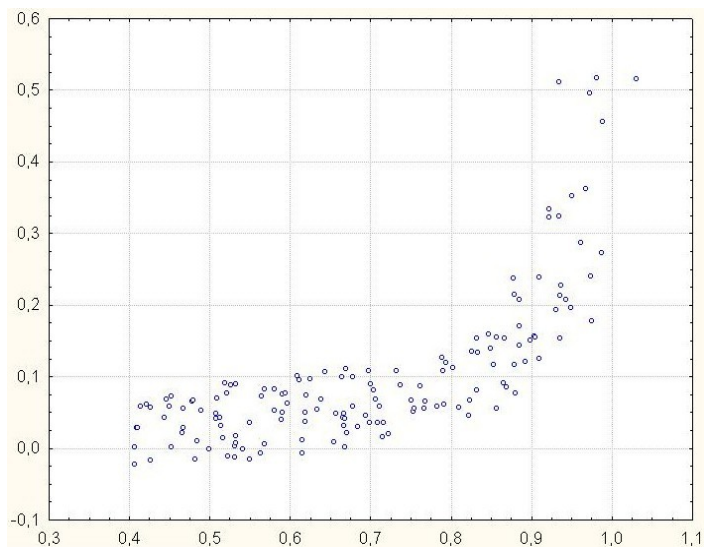
43. A researcher tested fear of death and religiosity in 15 people. His results are summarized in the following table:

ID	Fear of death	Religiosity
1	38	4
2	42	3
3	29	11
4	31	5
5	28	9
6	15	6
7	24	14
8	17	9
9	19	10
10	11	15
11	8	19
12	19	17
13	3	10
14	14	14
15	6	18

1. Compute mean, variance and standard deviations of both variables. What level of measurement is assumed for these computations?
2. Compute covariance and Pearson's correlation coefficient.
3. What does the Pearson's correlation coefficient say about the relationship between fear of death and religiosity?
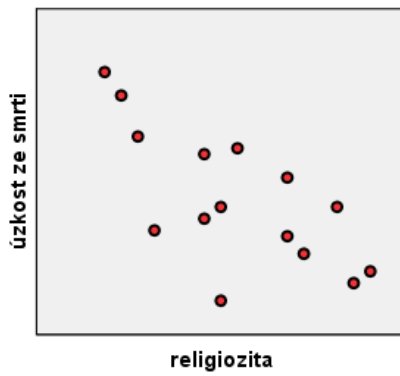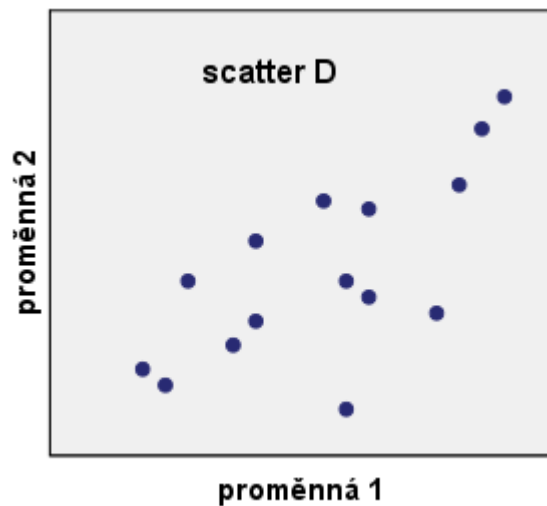4. Which of the following scatterplots corresponds with this relationship?
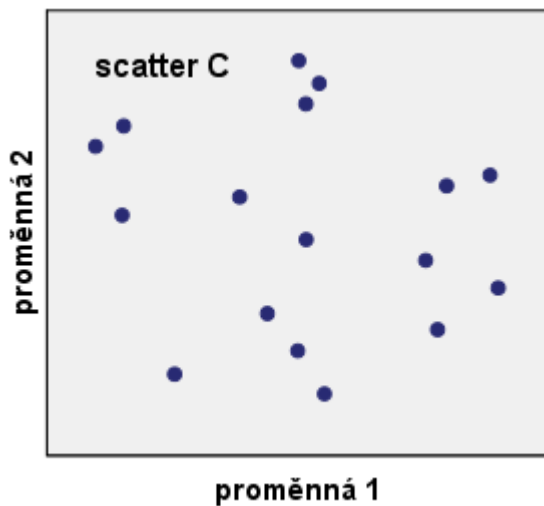
a)
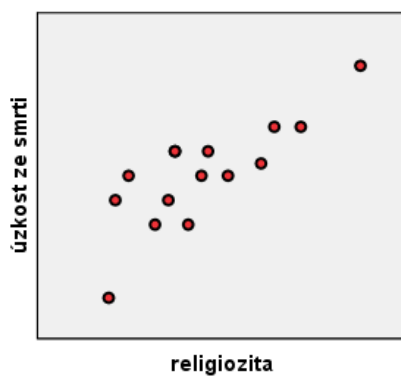
b)

c)

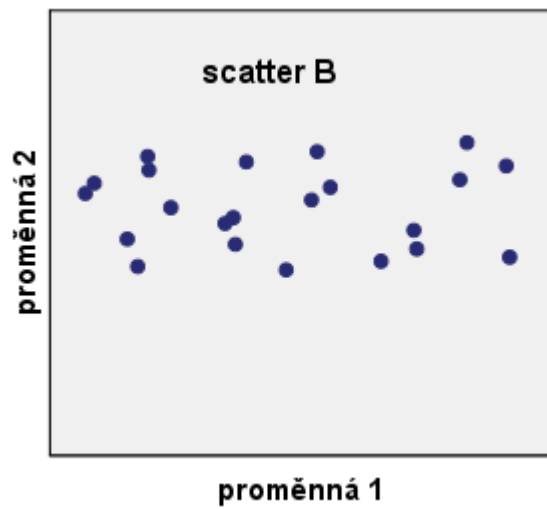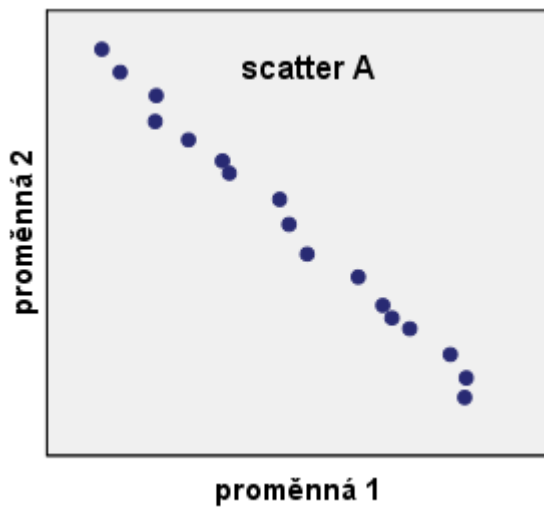**d)**

45. The following scatterplots display several associations:

1. Which of the scatterplots displays the strongest association? Try to estimate its correlation coefficient.

2. Which of the graphs display positive associations? Again try to estimate their correlation coefficients.

3. Which of the scatters displays no association? What will be their correlation?

46. A researcher explored relationship between height and self-efficacy. He tested 5 subjects, the results are in the following table:

ID	Height in inches	Self-efficacy
1	71	4,6
2	62	3,8
3	75	4,4
4	59	3,2
5	68	4

1. Compute means, variances, standard deviations, covariance and C. What do the results mean?

2. Another researcher decided to replicate the experiment. He tested the same five subjects, but this time he measured height in centimetres (height in cm = 2.5*height in inches) and he also used different measure of self-efficacy (new test scores = 5*old test scores). Compute again means, variances, standard deviations, covariance and Pearson's correlation coefficient.

3. How did the transformation influence the individual results?

47. A researcher decided to explore relationship between math grades and physics grade. The results are in the following table:

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
math	1	3	1	2	2	4	1	2	2	3	3	1	2	4	1	1	1	2	3	3
physics	2	3	1	1	2	3	1	1	2	2	3	3	2	3	1	2	2	1	3	2

1. Write contingency table from the data.

2. Covariance of the two grades is 0.474, math SD is 1.021 and physics SD is 0.795. What is Pearson's correlation coefficient?

3. Due to the nature of the data, which correlation coefficient would be the most suitable to use?

48. A researcher wanted to investigate the relationship between cultural literacy and number of cinema visit per month. He measured cultural literacy by a scale with possible scores from 0 (an absolute barbarian) to 20 (know everything about culture). In the following table, there are data from 12 examined students. However, student 9 made fun of the research, that's why his scores are so strange.

ID	1	2	3	4	5	6	7	8	9	10	11	12
Cultural literacy	14	19	12	15	14	17	19	11	1	17	13	17
Number of cinema visits	2	3	1	0	2	1	0	1	52	2	0	1

1. Compute Pearson's correlation coefficient between cultural literacy and number of cinema visit from all the data (the researcher didn't notice the weird scores from subject 9.

2. Can you estimate what would be the Pearson's correlation coefficient without the data from subject 9? How can you use this knowledge in your own research? Then compute the correlation coefficient.

3. Would Kendall coefficient be influence by the subject 9 scores in the same ways as Pearson's correlation coefficient?

4. And what about Spearman correlation?

5. How do we call resistance of some statistics against outliers?