

1. If we have a nominal variable, the most suitable frequency graph is:
 - a) bar chart
 - b) histogram

2. Which of the following distributions are probably negatively skewed?
 - a) family income in CZE per month
 - b) age at college graduation
 - c) number of inhabitants in Czech Republic towns
 - d) scores in an easy test

3. Imagine that 25 mathematicians and 25 people who are not good in math write a mathematical test. What distribution we'll probably get?
 - a) unimodal
 - b) bimodal
 - c) normal
 - d) skewed

5. Distribution of the following scores is:

1, 10, 6, 8, 7, 5, 5, 4, 9, 2, 9, 8, 6, 7, 8, 3, 4, 3, 5, 5, 7, 6, 4, 6, 6, 7

 - a) unimodal and approximately normal
 - b) bimodal and negatively skewed
 - c) normal and positively skewed
 - d) normal a negatively skewed

6. How do we call a symmetrical distribution where all the values have approximately the same frequency?

7. Lower you can see scores from a test. Create a frequency table including cumulative frequencies.
 - a) Create a frequency table for individual values
 - b) Create a frequency table for interval frequencies with interval width 10

40	98	63	90	70	60	45	43	78
67	56	54	78	87	43	90	81	81
77	80	79	80	81	66	75	88	84
49	63	78	79	80	92	89	84	77

8. Create a histogram for each frequency table you created in the previous exercise.

9. What is the difference between bar chart and histogram?

10. A primary school decided to evaluate its tuition quality. Every student was tested in four subjects. In each test it was possible to gain 0 to 100 points. The more points, the better result. Students' scores are displayed in the following table:

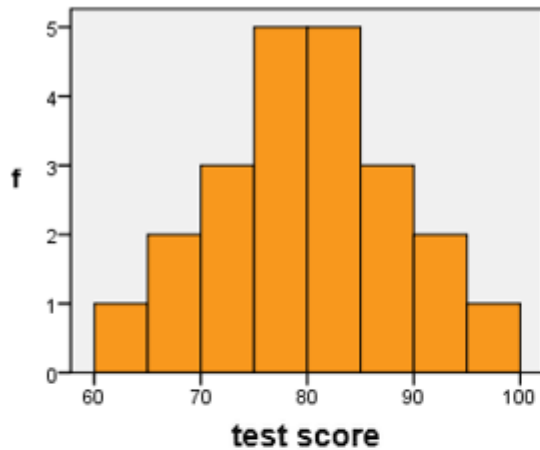
Student's ID	Math	Czech language	English language	Physical education
1	92	94	92	83
2	77	98	97	74
3	86	89	67	68
4	70	96	89	79
5	81	91	73	96
6	68	99	76	80
7	77	94	66	87
8	61	83	91	66
9	71	90	94	72
10	82	97	86	77
11	86	95	74	70
12	79	99	81	76
13	65	82	66	73

14	80	97	62	62
15	90	77	69	98
16	81	89	88	81
17	97	93	71	79
18	76	99	90	71
19	75	90	85	78
20	84	95	69	74
21	72	92	73	80
22	89	88	91	79

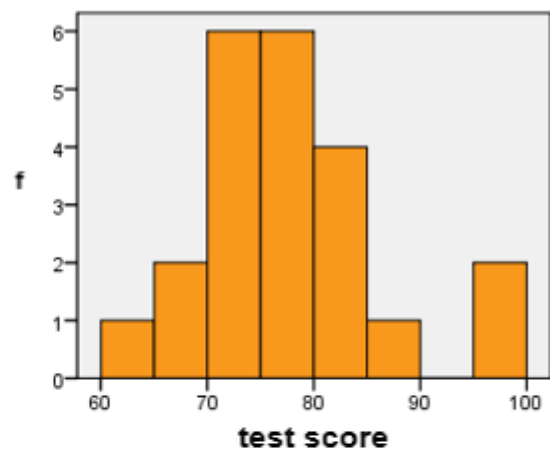
10.1 Create a frequency table including cumulative frequencies for scores from Physical education. Next, create the frequency table for Physical education for interval width 5.

10.2 Here you can see histograms for the individual school subjects. Which histogram belongs to which school subject?

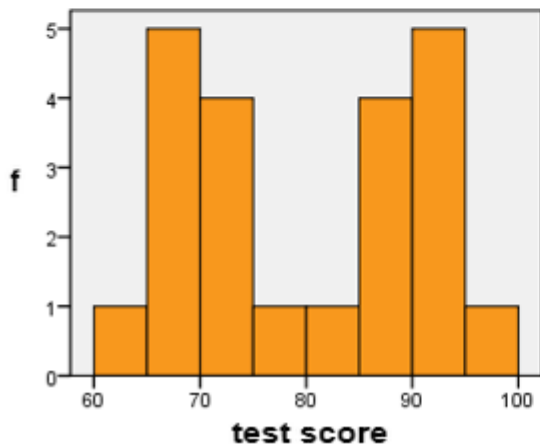
a)



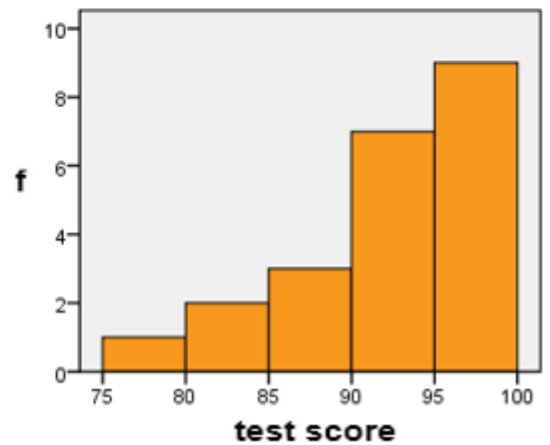
b)



c)



d)



10.3 Which of the histograms is the most similar to the normal distribution?

10.4 Which of the histograms doesn't have unimodal distribution? And what has probably caused it?

- a) Test from the give subject was too easy.
- b) Several students have extra tutoring and are better in this subject.
- c) Students are in the given subject divided into two groups: beginners and advanced.

10.5 Distribution of scores from Czech language (histogram d) is skewed. Which of the following statements about this histogram is correct?

- a) It's right-skewed, in other words positively skewed.
- b) It's right-skewed, in other words negatively skewed.
- c) It's left-skewed, in other words positively skewed.
- d) It's left-skewed, in other words negatively skewed.

10.6 The effect that caused it is called:

- a) The floor effect.
- b) The Gaussian effect.
- c) The ceiling effect.

10.7 As outlier we would consider:

- a) A score that nobody obtained.
- b) A student who had in the test the lowest score.
- c) The area between two modes (peaks).
- d) A score which is far from most other scores.

10.8 In which subject we can find outliers?

- a) In all the subjects.
- b) In physical education.
- c) In English language.
- d) In all the subjects except the Czech language.

10.9 The math teacher decided to use the test and give students the following grades:

- 100 – 91 ... grade 1
- 90 – 81 ... 2
- 80 – 71 ... 3
- 70 – 61 ... 4
- 60 and lower ... 5.

If he wanted to write the frequencies table and visualise it, what graphical form would he probably use?

- a) boxplot
- b) bar chart
- c) histogram
- d) pie chart

11. Ms. Johansson runs a small market in a village. She has lots of time, because not many customers visit her shop. So she started to write down, who and when came. The shop is open 6 days every day and every day come in average 12 customers in random times.

11.1 If we randomly took 50 hours from the total opening hours during a month and made during these hours observation how many customers came to the shop, the distribution would of the variable „number of customers per hour“ would probably look like:

- a) Poisson distribution
- b) Gaussian distribution
- d) normal distribution

11.2 What will be the value of λ in this distribution?

11.3 If we used „number of customers per day“ instead of „number of customers per hour“, we could quite precisely replace the Poisson distribution by:

- b) normal distribution
- c) bimodal distribution
- d) uniform distribution

12. Peter loves gambling and he already lost a lot of money in the dice game. He wants to play cleverer, so he decided to find out more about probabilities in the dice game. During the whole time, we will assume classical dices and that probabilities of tossing the individual numbers are even.

12.1 What is the distribution of one round of dice tossing?

- a) discrete uniform
- b) leptokurtic
- c) continuous normal
- d) Poisson

12.2 In the next round, Peter always tosses 6 dices and sums the tossed numbers. What distribution will be this distribution most similar to?

- a) continuous uniform
- b) Poisson
- c) right-skewed
- d) normal

13. If I ask random passers how old I am, what would be the distribution of their guesses?

- a) uniform
- b) bimodal
- c) normal
- d) Poisson

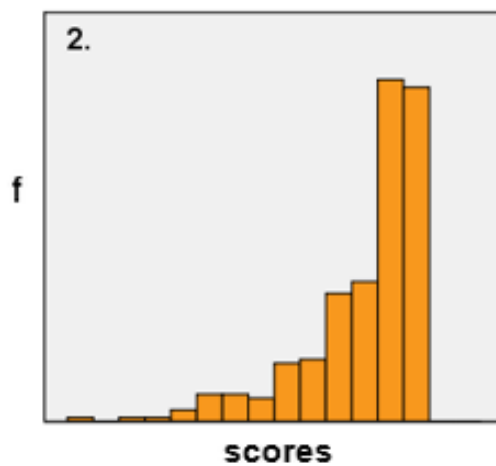
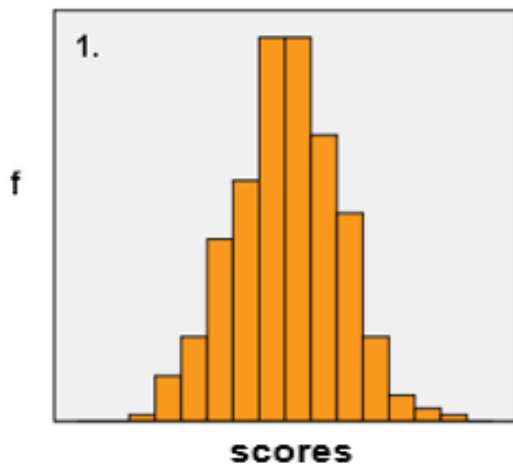
14. The following histograms (eventually bar charts) illustrate different distributions. Each of the distributions can be described by some of the following characteristics. Assign to each graph all suitable characteristics (one characteristic can be assign to more graphs):

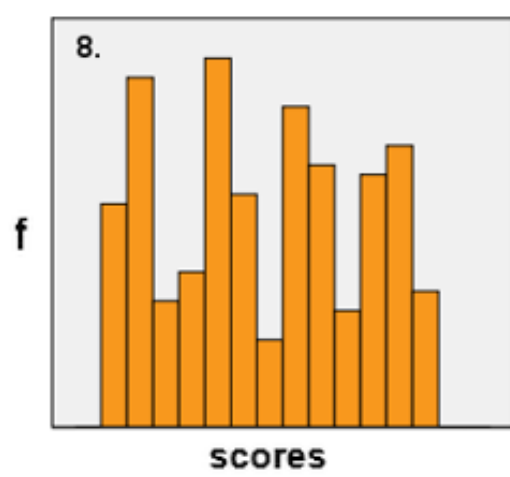
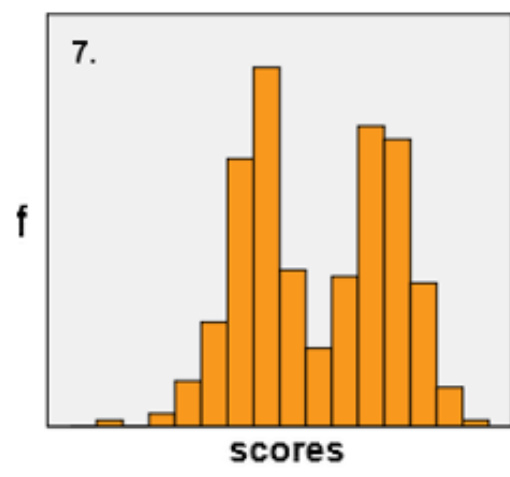
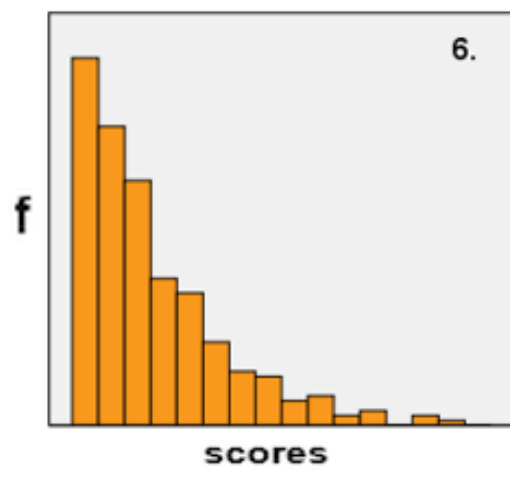
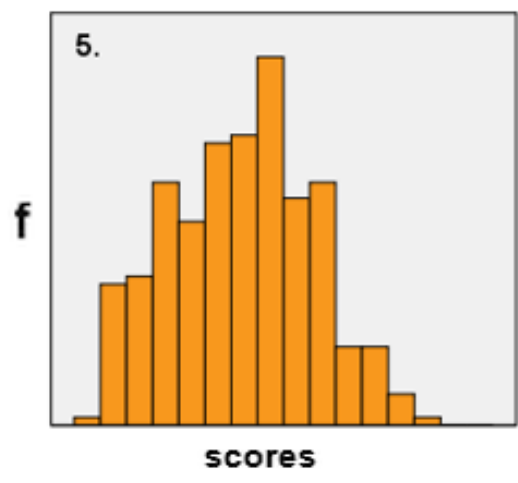
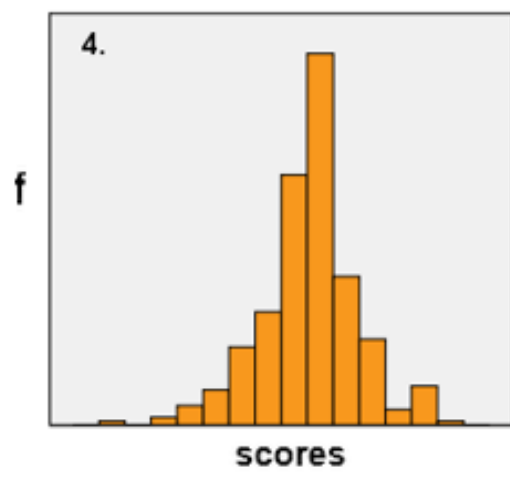
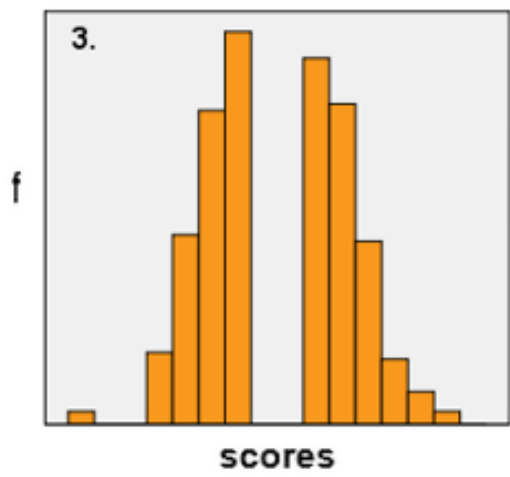
unimodal
right-skewed
uniform
binomial
continuous

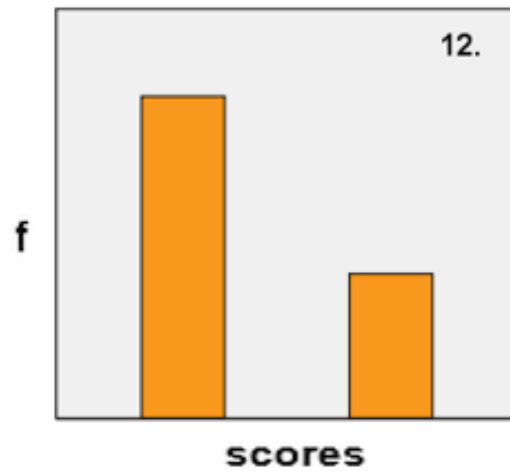
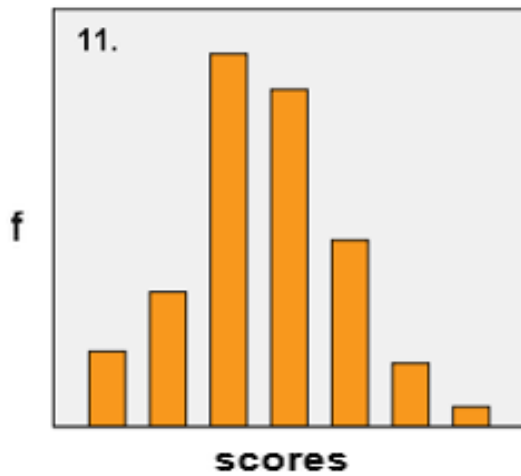
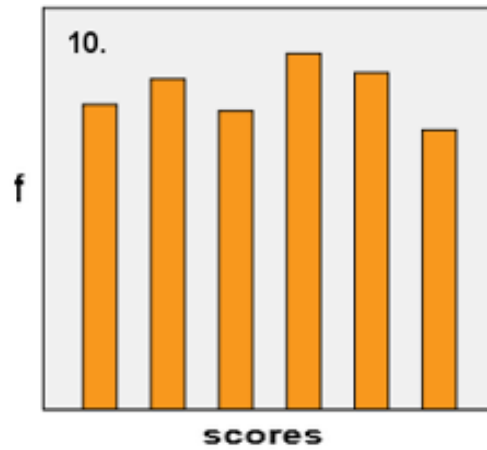
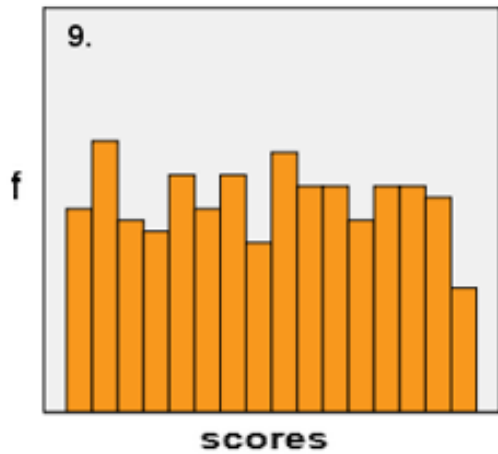
bimodal
left-skewed
normal
discrete

multimodal
positively skewed
platykurtic

negatively skewed
leptokurtic







15. The following sets of scores represent scores in different tests. Display the distributions graphically and describe the distributions.

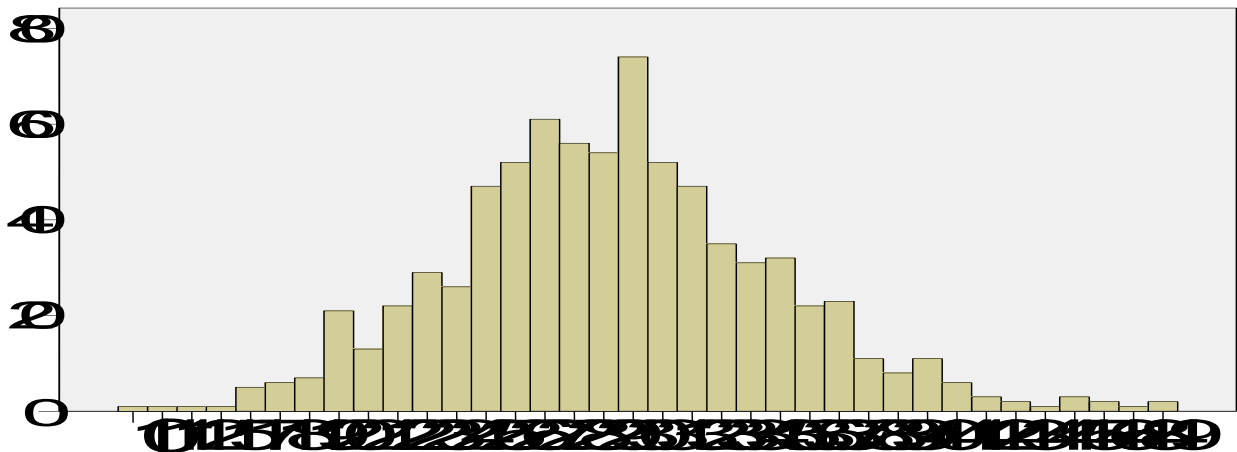
a. 10; 6; 8; 7; 5; 5; 4; 9; 2; 9; 8; 6; 7; 8

b. 2; 9; 5; 1; 2; 4; 2; 6; 7; 2; 8; 5; 3; 4; 7; 2; 3; 5; 4; 3; 6; 3

c. 10,0; 9,7; 9,0; 8,9; 8,7; 7,8; 7,5; 7,2; 6,9; 6,6; 6,0; 5,1; 4,8; 4,3; 3,0

d. 12; 12; 16; 19; 21; 23; 26; 36; 51; 56; 57; 60; 63; 68; 69; 71; 75; 75

16. Look carefully on the following graph or diagram. How is it called and why?



20. What distribution do usually infrequent events have?

21. What is percentile?

22. If is a distribution symmetrical, mode usually doesn't differ much from mean. In which condition this doesn't apply?

23. If a distribution substantially differs from normal distribution, informational value of descriptive statistics, which are not , decreases. (complete the sentence)

24. A researcher works with variable „the number of subjects during the whole college in which the student had to repeat the exam“. The following table with descriptive statistics represents the variable:

proměnná	N	rozpětí	M	SD
numb_of_subjects	897	20	2,1	2,5

Distribution of the variable is apparently:

- a) normal
- b) positively skewed
- c) bimodal
- d) uniform
- e) negatively skewed

25. For the following variable „age“ create stem-and-leaf diagram and describe the distribution.

10,2; 10,0; 9,7; 9,0; 8,9; 8,7; 7,8; 7,5; 7,2; 6,9; 6,9; 6,6; 6,1; 6,0; 5,6; 5,1; 4,8; 3,9; 3,1

26. A student writes her thesis about autobiographical memory of married couples. She wanted to know what's true about the statement, that women remember more about history of their relationship than their husbands. So she asked them, when and where they met, when and where they first kissed, when and where they first had sex etc. In sum, she asked everyone about 12 events and wrote down, in how much cases they remembered the datum and the place. She also asked about the length of the relationship in years. So far, we have data from 6 families:

Family 1

Relationship length = 4	
Wife	Husband
data=9	data=3
places=12	places=6

Family 2

Relationship length = 5	
Wife	Husband
data=7	data=2
places=9	places=9

Family 3

Relationship length = 8	
Wife	Husband
data=7	data=8
places=3	places=12

Family 4

Relationship length = 11	
Wife	Husband
data=6	data=7
places=6	places=11

Family 5

Relationship length = 14	
Wife	Husband
data=8	data=6
places=4	places=6

Family 6

Relationship length = 18	
Wife	Husband
data=11	data=4
places=8	places=10

We're interested whether are the distributions of variables „the number of remembered data“ and „the number of remembered places“ alike. Illustrate the variables distributions graphically so that it's possible to compare them visually.

27. In a histogram, what displays axis X and what axis Y?

28. A professor is interested in colours attractiveness. He made an experiment, in which 1000 respondents should fill in a test with a pencil. Every respondent got a test and he picked a pencil from a big box. The pencils were painted in different colours. The professor wasn't interested in the test results. Instead of that, he counted what pencil colour who picked and also whether he returned the pencil or he kept it. The results of the experiment are displayed in the following table:

		colour of the chosen pencil				
		red	black	blue	yellow	Total
women	returned	120	130	90	170	510
	kept	20	25	25	20	90
men	returned	85	45	90	100	320
	kept	20	25	20	15	80
Total		245	225	225	305	1000

Display visually the distribution of the variable „colour of the chosen pencil“, so that we can compare distribution of this variable between men and women.

29. 196 students answered the question how they most often spend their free time on Saturdays. The results are summarized in the following table:

Category	Reading	Sport	Theatre	Cinema	Dancing	Passive relaxation	Social activities	Studying	Working
f	10	25	3	32	29	37	34	15	11

Think of the most suitable graphical representation of the results. Justify your answer.