

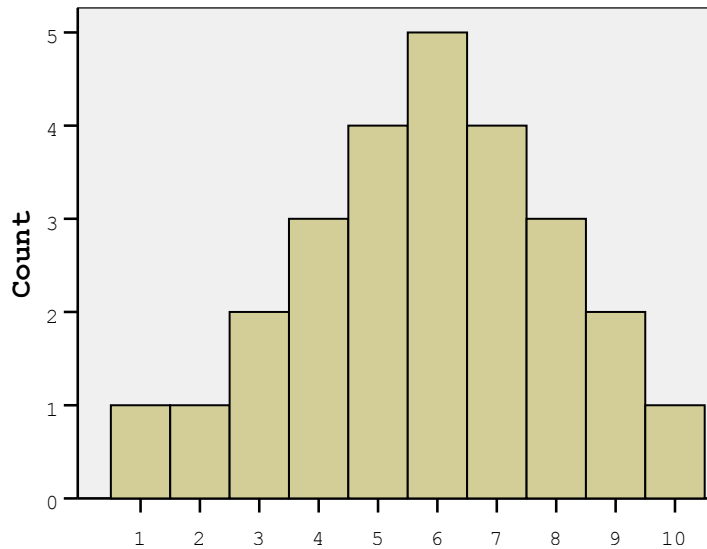
Answers

1. a

2. d

3. c

5. a



6. uniform

7. a) Test scores are measured on ratio level, so we could write to the table all possible values between 40 a 98 and write frequency 0 for the values that nobody obtained. Note that intervals have to be disjoint – each value has to be included in only one interval (40-49, 50-59 etc.).

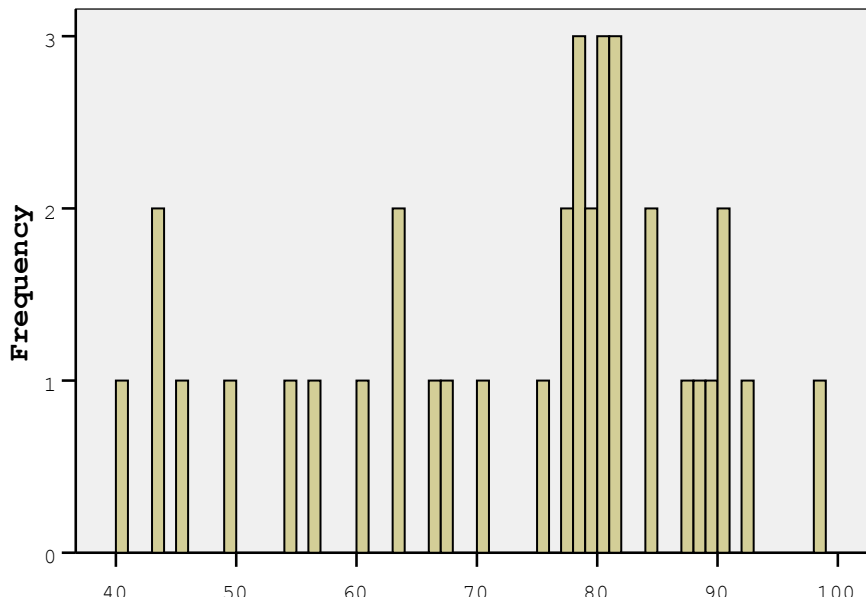
	Frequency	Relative frequency	Cumulative relative frequency
40	1	2,8	2,8
43	2	5,6	8,3
45	1	2,8	11,1
49	1	2,8	13,9
54	1	2,8	16,7
56	1	2,8	19,4
60	1	2,8	22,2
63	2	5,6	27,8
66	1	2,8	30,6
67	1	2,8	33,3
70	1	2,8	36,1
75	1	2,8	38,9
77	2	5,6	44,4
78	3	8,3	52,8
79	2	5,6	58,3
80	3	8,3	66,7
81	3	8,3	75,0
84	2	5,6	80,6
87	1	2,8	83,3
88	1	2,8	86,1
89	1	2,8	88,9
90	2	5,6	94,4

92	1	2,8	97,2
98	1	2,8	100,0
Celkem	36	100,0	

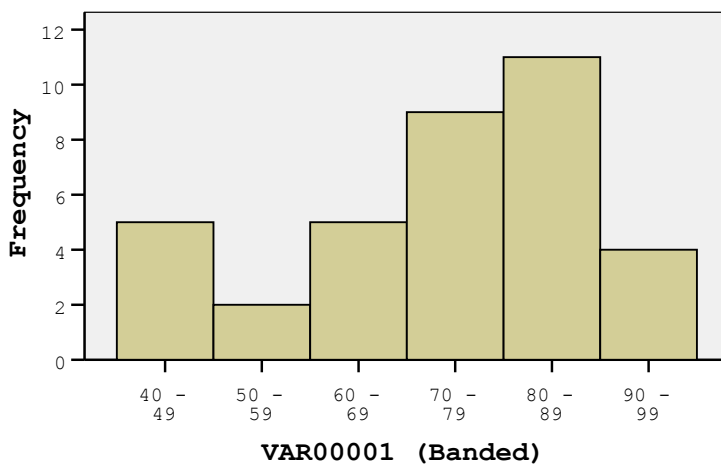
b)

	Frequency	Relative frequency	Cumulative relative frequency
40 - 49	5	13,9	13,9
50 - 59	2	5,6	19,4
60 - 69	5	13,9	33,3
70 - 79	9	25,0	58,3
80 - 89	11	30,6	88,9
90 - 99	4	11,1	100,0
Celkem	36	100,0	

8. a)



b)



9. Bar chart is graphical representation of frequency table (one value – one bar) and it is suitable for nominal or ordinal or interval data with a few values. Histogram is also graphical representation of frequency table,

but its axis X is real – it includes all values even if nobody obtained them. It is most suitable for ratio variables and interval variables with many values.

10.1 Frequency table of physical education scores

Score	Frequency	Cumulative frequency	Relative frequency (percent, %)	Cumulative percent
62	1	1	4,5	4,5
66	1	2	4,5	9,1
68	1	3	4,5	13,6
70	1	4	4,5	18,2
71	1	5	4,5	22,7
72	1	6	4,5	27,3
73	1	7	4,5	31,8
74	2	9	9,1	40,9
76	1	10	4,5	45,5
77	1	11	4,5	50
78	1	12	4,5	54,5
79	3	15	13,6	68,2
80	2	17	9,1	77,3
81	1	18	4,5	81,8
83	1	19	4,5	86,4
87	1	20	4,5	90,9
96	1	21	4,5	95,5
98	1	22	4,5	100
Total	22		100	

Frequency table for intervals:

Interval	Frequency	Cumulative frequency	Relative frequency (percent, %)	Cumulative percent
60 – 64	1	1	4,5	4,5
65 – 69	2	3	9,1	13,6
70 – 74	6	9	27,3	40,9
75 – 79	6	15	27,3	68,2
80 – 84	4	19	18,2	86,4
85 – 89	1	20	4,5	90,9
95 – 99	2	22	9,1	100
Total	22		100	

10.2 a) math, b) physical education, c) English language, d) Czech language

10.3 histogram a) – math

10.4 histogram c) – English language

The most probable cause is c), that is two different groups (populations) in one class, but also b) could cause it, if the tutoring is effective and there are many students with extra tutoring.

10.5 d)

10.6 c) Ceiling effect. In other words, the test was very easy, so most of the student will gain close to maximum points.

10.7 d).

10.8 b) – in physical education.

10.9 b) bar chart

Boxplot is not a graphical representation of frequency table. Histogram is used for interval and ratio variables with many values, pie chart isn't used almost at all.

11.1 a) Poisson distribution: describes distribution of infrequent events

11.2 $\lambda = 2$ Lambda is frequency per time unit

11.3 b) normal distribution. If the event occurs more often than 10 times per time unit, than the distribution can be approximated to normal distribution.

12.1 a) discrete uniform. Discrete mean that the distribution can't take all possible values, but only some values – here only the values on the dice. Uniform is distribution, which takes all values with the same probability.

12.2 d) normal. Sums of independent, equally distributed variables (here: two independent dices) approaching normal distribution.

13. c) normal. If is the variable determined by many independent influences, than its distribution is approximately normal. That's why e.g. intelligence or height are normally distributed.

14. It is not possible to say based just on histogram whether a variable is continuous or discrete. But the distributions represented by bar chart (the last three) are certainly discrete.

14.1 unimodal, normal – very common distribution, that's why it is called normal.

14.2 unimodal, left-skewed, negatively skewed – the last two attributes are the same thing (left-skewed and negatively skewed are synonyms).

14.3 bimodal – it has two modes (peaks)

14.4 unimodal, leptokurtic – leptokurtic distribution is similar to normal distribution, but it's „more spiked“, „sharper“.

14.5 unimodal, platykurtic – platykurtic distribution is on the other hand „less spiked“, „less sharp“, „flatter“.

14.6 unimodal, right-skewed, positively skewed – again, right -skewed and positively skewed are synonyms.

14.7 bimodal

14.8 multimodal – it has more than two modes (peaks).

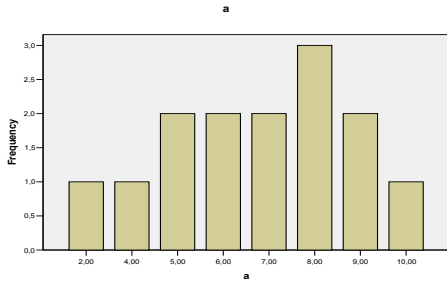
14.9 uniform – model uniform distribution would have rectangular shape, random data from uniform distribution can look like this. Uniform distribution takes every possible value with the same probability.

14.10 uniform, discrete – this distribution represents random data from throwing a dice. Each from the six values can occur with the same probability, other values are not possible.

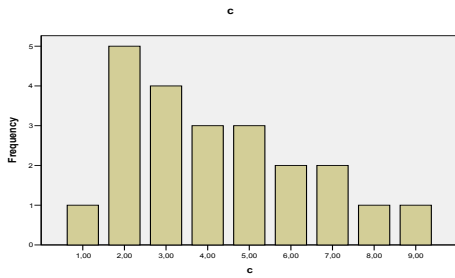
14.11 binomial, discrete – the binomial distribution is the discrete probability distribution of the number of successes in a sequence of n independent yes/no experiments, each of which yields success with probability p . Binomial distribution is distribution of the sum of binary distributions (yes/no options). Binomial distribution arises e.g. as the sum of several coin tosses (0=heads or 1=tails, that is the sum of binary distributions). In this case, the probability of tossing one coin side was probably twice higher.

14.12 binary, discrete – distribution which takes only two values – the first with probability p and the second with probability $1 - p$. It is a special case of binomial distribution, also called Bernoulli distribution. This distribution represents one coin toss, but the coin is unfair, it has NOT probability $p = 0,5$ of tossing one side (the probability of tossing one side is higher than the probability of tossing the other side).

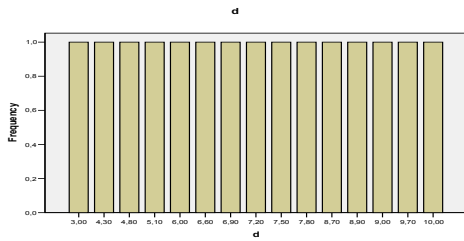
15. a. negatively skewed, unimodal



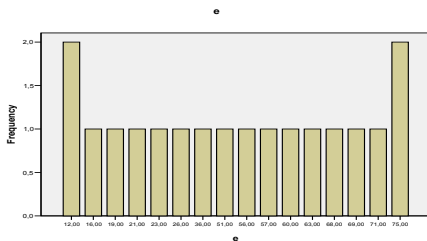
b. positively skewed, unimodal



c. uniform, symmetrical



d. bimodal, symmetrical



16. Bar chart – it is graphical representation of frequency table, it displays discrete data. The values distribution looks close to normal and the graph looks similar to histogram, but some values are missing on the X axis.

20. Poisson

21. X-th percentile is the value for which it holds that X % of people (events, phenomena) in the sample has/obtained the same or lower value

22. in the case of symmetrical bimodal distribution

23. robust (that's e.g. ordinal statistics)

24. b)

25.

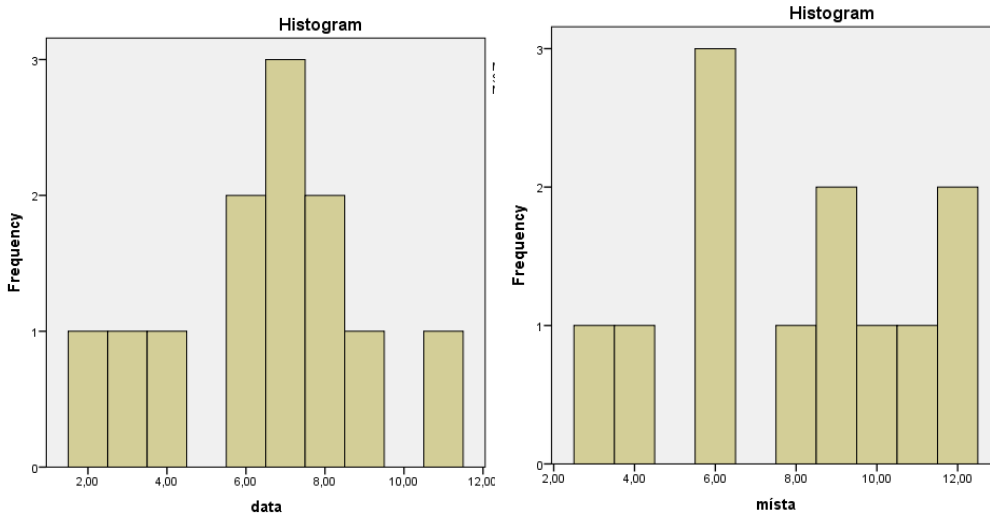
Frequency	Stem & Leaf
3,00	0 . 334
14,00	0 . 55666667778899
2,00	1 . 02
Stem width:	10,00
Each leaf	1 case(s)

nebo

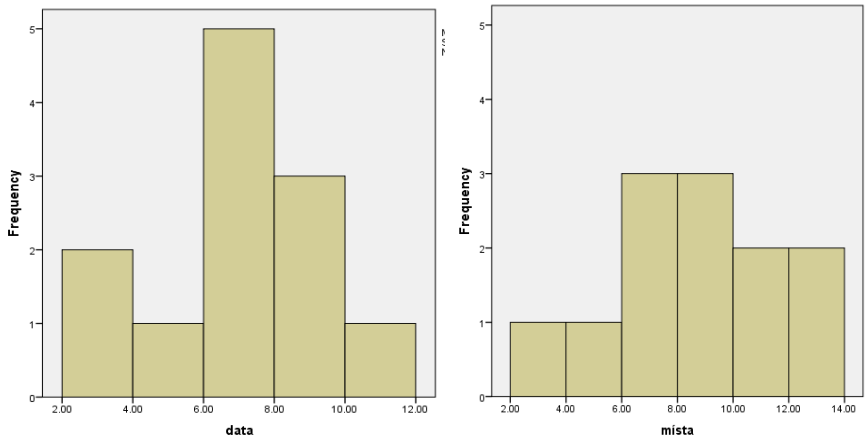
3		19
4		8
5		16
6		01699
7		258
8		7
9		07
10		02

The distribution is difficult to describe, because we have few values. However,, from the first diagram we can assume the distribution is unimodal and probably quite symmetrical.

26. In this case, it's the best to visualise the variables „number of remembered data“ and „number of remembered places“ by histogram, because the variables are ratio. In the histograms, we can see the “gaps” (zero frequencies of given values).



If we want to get rid of the gaps, we can group the values into intervals by 2 and make histograms again.



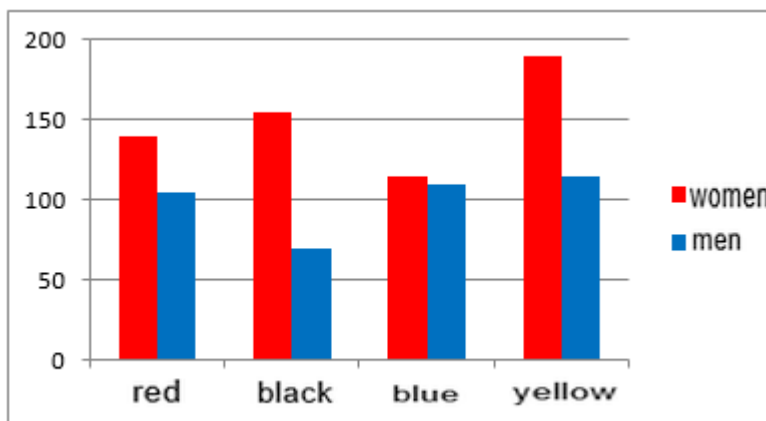
Here we can see that both variables are unimodal, histogram for places is quite symmetrical, the histogram for data is not. In the data histogram we can see moderate left skewness.

27. The axis X in histogram displays different values of the presented variable (discrete or continuous values or even intervals). The axis Y in histogram displays the frequencies of the individual values.

28. For visualisation is in this case the best a bar chart sorted by gender, because we want to compare the variable between men and women. Before making the graph, it is necessary to adapt the data by summing values of the variable „returned – kept“, because we are only interested in the pencil colour this time. Then we get this table:

	pencil colour				Total
	red	black	blue	yellow	
women	140	155	115	190	600
men	105	70	110	115	400
Total	245	225	225	305	1000

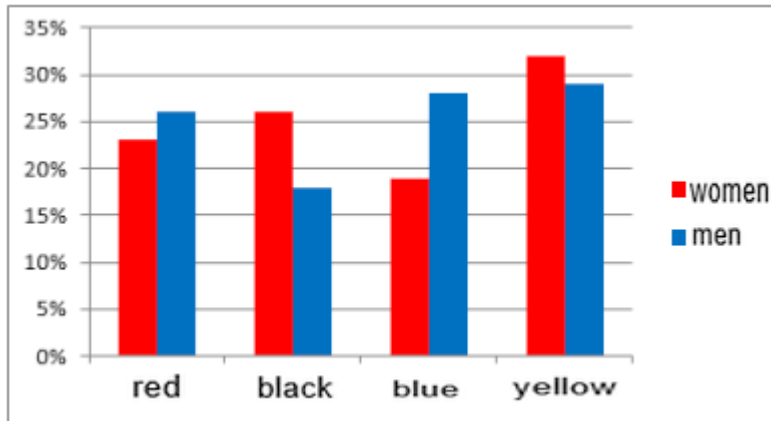
From the table we can make the following bar chart sorted by gender:



We can see that women have generally higher scores than men. This is cause simply by the fact that there are more women in the sample than men. How could we solve this “bias”? One of the possibilities is to display not absolute frequencies, but relative frequencies. First, we have to count the relative frequencies (percent) for every sub-sample (women and men):

	pencil colour				Total
	red	black	blue	yellow	
women	140	155	115	190	600
%	23%	26%	19%	32%	100%
men	105	70	110	115	400
%	26%	18%	28%	29%	100%
Total	245	225	225	305	1000

Bar chart for perceptual values will look like this:



Now we can see that men chose more often red and blue pencil, on the other hand women took more often black and yellow pencil.

29. The best option here is a bar chart, because we have a nominal variable. The graph can look like this. (we could also order the categories from the most (least) common to the least (most) common for easier interpretation).

