

1 DESCRIPTIVE STATISTICS

1.1 Means of expression in statistics

1.1.1 Exercises 1.1

1. According to the the Czech statistical office in 2006, there ware 29435.10³ tons of waste in the Czech Republic. Following table gives a proportional representation (%) of components.

<i>Kind of waste</i>	<i>Percentage</i>
agriculture and forestry	4.43 %
mining	1.56 %
water purification and water distribution	22.34 %
industry	1.40 %
building and demolition	29.50 %
energetics (apart from radioactivity)	6.95 %
treatment of a town	4.65 %
municipal waste	13.52 %
other waste	? %

What is the proportion of the other waste. Construct the „pareto“ graph.

2. Following table gives information about employees and their transportation to an office.

<i>Means of transport</i>	<i>Frequency</i>
Car (individual)	25
Car (sharing)	15
Bus	10
Train	5
Other means of transport	4

Compute relative frequencies and draw the pie chart (graph).

3. There ware 40 voluntary individuals (blood-transfusion) during a given period. Following table gives information about their blood-groups.

0	0	A	B	A	0	A	A	A	0
B	0	B	0	0	A	0	0	A	A
A	A	AB	A	B	A	A	0	0	A
0	0	A	A	A	0	A	0	0	AB

Construct the table with frequencies, find relative frequencies and draw the pie graph.

Solution.

1. 15.65;
2. 0.42; 0.25; 0.17; 0.08; 0.07;
3. 0: 16(0.40); A: 18(0.45); B: 4(0.10); AB: 2(0.05).

1.2 Basic data processing

1.2.1 Exercises 1.2

1. Due to a production inspection, there was carried out a distance measurement of 15 blockboards (cm):

49.8	50.2	50.3	49.5	50.0	49.3	50.0	50.9
50.4	50.0	49.7	50.6	50.2	49.9	50.1	

Sort the data according to size and find values $x_{(5)}$, $x_{(10)}$ and $x_{(11)}$ (i.e. the fifth, the tenth and the eleventh value of the sequenced data set). Construct the diagram of dispersion, interpret the graph regarding the concentration and the symmetry of the data.

2. A concentration of certain chemical is repeatedly verified by a special test. There are 50 chemical analyses with outcomes (%):

41	42	41	44	41	41	39	43	45	41
42	43	41	41	43	41	42	45	42	40
43	42	42	40	41	42	42	43	42	43
39	44	43	42	40	42	42	40	42	42
41	40	44	39	42	44	42	41	45	43

Construct the table with frequencies, find absolute frequencies, relative frequencies, cumulative frequencies and relative cumulative frequencies. Construct the frequency polygon and the cumulative curve. Interpret the outputs.

3. A test in mathematics consists of 10 questions with 5 possible answers and always only one is correct. Following table represents results of 60 students (number of correct answers):

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

Construct the table with frequencies, find absolute frequencies, relative frequencies, cumulative frequencies and relative cumulative frequencies. Construct the frequency polygon, the cumulative curve and the cumulative relative frequency distribution. Interpret the outputs.

4. Consider a data set which corresponds to earthquakes in California (Richter scale):

6.8	6.6	7.5	6.2	6.5	7.1	6.1	5.8	5.5	6.9
6.6	6.2	5.3	5.9	6.0	5.3	5.9	6.2	6.4	7.0
6.2	6.1	6.0	6.1	5.5	6.4	5.9	5.7	5.9	5.9
6.1	5.3	6.6	5.8	7.7	5.9	7.1	6.3	7.0	6.3
8.3	8.0	6.8	6.3	8.3	7.0	7.0	7.0	6.5	

Construct the interval distribution of frequencies, find absolute frequencies, relative frequencies, cumulative frequencies and relative cumulative frequencies. Construct the histogram and the cumulative histogram and the cumulative relative frequency distribution. Interpret the outputs.

5. Consider a given part of a highway. Following table represents speed values of motorcars (km/h):

147	125	127	134	126	143	134	129	151	143
138	116	122	142	115	132	125	128	129	163
120	134	131	126	142	104	120	141	114	130
155	130	145	133	121	137	116	146	140	118
139	122	122	128	136	159	139	167	158	132
157	142	117	134	126	150	129	115	146	153
152	157	133	158	143	111	136	120	164	118

Construct the interval distribution of frequencies, find absolute frequencies, relative frequencies, cumulative frequencies and relative cumulative frequencies. Construct the histogram and the cumulative histogram and the cumulative relative frequency distribution. Is the distribution symmetric? Interpret the outputs.

Solution.

1. 49.9; 50.2; 50.2;
2. 39: 3; 40: 5; 41: 11; 42: 16; 43: 8; 44: 4; 45: 3;
3. 2: 1; 3: 5; 4: 8; 5: 5; 6: 14; 7: 15; 8: 6; 9: 5; 10: 1;
4. e.g. for $k = 7, h = 0.5, a = 5.1$: (5.1; 5.6): 5; (5.6; 6.1): 15; (6.1; 6.6): 14; (6.6; 7.1): 10; (7.1; 7.6): 1; (7.6; 8.1): 2; (8.1; 8.6): 2;
5. e.g. for $k = 7, h = 10, a = 100$: (100, 110): 1; (110, 120): 12; (120, 130): 17; (130, 140): 16; (140, 150): 12; (150, 160): 9; (160, 170): 3.

1.3 Measure of location

1.3.1 Exercises 1.3

1. With reference to the properties of the average, simplify an expression

$$\frac{1}{n} \sum_{i=1}^n c_1(x_i + c_2),$$

where c_1 and c_2 are real constants.

2. Compute the arithmetic mean, the geometric mean and the harmonic mean in terms of a data set: 11, 13, 4, 8, 8, 7, 9, 15, 10, 14, 8, 10, 8, 12, 9. Furthermore, compute the median, the lower quartile and the upper quartile.
3. Compute the arithmetic mean, the median, the lower quartile and the upper quartile for the data set 1.2.1/1.
4. Compute the arithmetic mean, the mode, the median, the lower quartile, the upper quartile, the lower decile and the upper decile for the data set 1.2.1/3.
5. Compute the arithmetic mean, the modal interval, the median, the lower quartile, the upper quartile, the lower decile and the upper decile for the data set 1.2.1/5.

Solution.

1. $c_1(\bar{x} + c_2)$;
2. 9.733; 9.296; 8.801; 9(9); 8(8); 12(11.5);
3. 50.060; 50(50); 49.8(49.85); 50.3(50.25);
4. 6.1; 7; 6(6); 5(5); 7(7); 3.5(3.9); 8.5(8.1);
5. 134.786; (120; 130); 133.5(133.5); 125(125); 143(143); 116.5(116.9); 157(157).

1.4 Measure of dispersion

1.4.1 Exercises 1.4

1. Consider a data set of values x_i , n is the size, \bar{x} is the mean and s_n^2 is the variance of the set. Find the variance of a data set of values $y_i = c_1(x_i + c_2)$, where c_1 and c_2 are real constants.
2. Compute the range, the quartile range, the quartile deviation, the average deviation, the variance, the sample variance, the standard deviation, the sample standard deviation and the variation coefficient in terms of a data set: 11, 13, 4, 8, 8, 7, 9, 15, 10, 14, 8, 10, 8, 12, 9. (Advice: use results from 1.3.1/2).
3. Compute the range, the quartile range, the quartile deviation, the average deviation, the variance, the sample variance, the standard deviation, the sample standard deviation and the variation coefficient for the data set 1.3.1/3.

4. Compute the range, the quartile range, the decile range, the quartile deviation, the decile deviation, the average deviation, the variance, the sample variance, the standard deviation, the sample standard deviation and the variation coefficient for the data set 1.3.1/4.
5. Compute the range, the quartile range, the decile range, the quartile deviation, the decile deviation, the average deviation, the variance, the sample variance, the standard deviation, the sample standard deviation and the variation coefficient for the data set 1.3.1/5.

Solution.

1. $c_1^2 \cdot s_n^2$;
2. 11; 4(3.5); 2(1.75); 2.249; 7.796; 8.352; 2.792; 2.890; 0.287;
3. 1.6; 0,5(0.4); 0.25(0.2); 0.304; 0.156; 0.167; 0.395; 0.408; 0.008;
4. 8; 2(2); 5(4.2); 1(1); 0.625(0.525); 1.443; 3.290; 3.346; 1.814; 1.829; 0.297;
5. 63; 18(18); 40.5 (40.1); 9(9); 5.063(5.013); 11.733; 203.768; 206.722; 14.275; 14.378; 0.106.

1.5 Measure of concentration

1.5.1 Exercises 1.5

1. Compute the skewness and the kurtosis in terms of a data set: 11, 13, 4, 8, 8, 7, 9, 15, 10, 14, 8, 10, 8, 12, 9. (Advice: use results from 1.3.1/2 and 1.4.1/2).
2. Compute the skewness and the kurtosis for the data set 1.3.1/3. (Advice: use results from 1.3.1/3 and 1.4.1/3).
3. Compute the skewness and the kurtosis for the data set 1.3.1/4. (Advice: use results from 1.3.1/4 and 1.4.1/4).
4. Compute the skewness and the kurtosis for the data set 1.3.1/5. (Advice: use results from 1.3.1/5 and 1.4.1/5).

Solution.

1. 0.128; -0.349 ;
2. 0.136; -0.102 ;
3. -0.182 ; -0.567 ;
4. 0.286; -0.607 .