1 HYPOTHESIS TESTING

1.1 Statistical hypothesis, principle of hypothesis testing

1.2 Zero skewness and zero kurtosis test

1.2.1 Exercises 1.2

Use significance level $\alpha = 0.05$ in the following exercises.

1. Perform the zero skewness and the zero kurtosis test for the data set: 25.4 28.0 20.1 27.4 25.6 23.9 24.8 26.4 27.0 25.4.
2. Verify, if the data set

6.5	6.8	6.7	6.0	5.6	6.6	5.5	6.4	5.5	6.5
6.3	6.2	6.3	5.9	5.8	6.4	6.5	6.3	5.7	6.1

   could be considered as a random sample from a normal distribution.
3. Verify, if the data set

238	256	266	240	260	235	263	259
234	258	253	265	234	245	237	

   could be considered as a random sample from a normal distribution.
4. Verify, if the data set

914	1073	1073	1129	975	957	1100	1009	1007	973
709	782	1074	856	959	909	946	971	1051	992
1106	1060	995	909	1080	1173	1143	897	1235	968
1107	1091	962	1185	1084	1046	1038	883	1004	830
922	945	938	1070	961	1024	875	859	958	1047

   could be considered as a random sample from a normal distribution.

Solution.
1. reject both zero skewness and zero kurtosis;
2. not reject normal distribution;
3. not reject normal distribution;
4. not reject normal distribution.

1.3 Hypothesis testing for one population

1.3.1 Exercises 1.3

1. A random sample of 10 observations produced characteristics $\overline{x} = 32$ and $s^2 = 15$. Using the 5% significance level, test the hypothesis that the mean of the population is $\mu = 30$. Consider the population is normally distributed.

2. The required mean value of roasted coffee moisture is 4.2 % and the standard deviation is 0.4 %. Consider a random sample of 20 observations (%):

4.5	4.3	4.1	4.9	4.6	3.2	4.4	5.1	4.8	4.0
3.7	4.4	3.9	4.1	4.2	4.1	4.7	4.3	4.2	4.4

   It is known that the population of roasted coffee has a normal distribution (verify). Using the 5% significance level, find out, if the population follows required characteristics, i.e.
   a) the mean value of moisture,
   b) variance.

3. A random sample of 10 observations produced characteristic $s^2 = 2.0$. Using the 1% significance level, test the hypothesis that the variance of the population is $\sigma^2 = 0.2$. Consider the population is normally distributed.

4. Consider investigation of salaries of alumnae, especially the variance of this variable. Up to now, it is considered that it is the normally distributed variable with the standard deviation 995 Euro. Using the 5% significance level, verify assumption with respect to the actual variance, if a random sample of 25 observations produced characteristics $\overline{x} = 12494$ Euro, $s = 1152$ Euro.

5. The producer of non-alcoholic beer guarantees (for a capacity 2l) the standard deviation 0.05 l. Consider a random sample of 20 observations and corresponding capacities (l):

1.93	1.94	1.92	2.01	1.93	2.07	2.03	2.03	1.98	1.96
2.03	2.00	2.10	1.99	2.02	2.02	1.96	2.02	1.86	2.04

   Consider the population is normally distributed. Using the 5% significance level, verify declaration of the producer with reference to accuracy (i.e. the variance) of beer filling.

6. A travel agency organizes foreign trips regarding individual requests. On account of previous experience, it is known that 30 % of all organized trips have been to the same country X. Nowadays, there is a suspicion that the interest in the country X is lesser. A actual random sample of 150 clients produced 38 clients with demand on the country X. Using the 5% significance level, verify, if the suspicion comes true.

7. The producer of a certain measuring machine guarantees the accuracy by the standard deviation $\sigma = 0.9$. A random sample of 20 measurements produced the sample variance $s^2 = 1.44$. Using the 5% significance level, decide, if this reading is not in conflict with the assertion of the producer.

8. Environmental activists are against the construction of new industrial factory in given region, among others they consider that bad environmental conditions of this area are the main reason for the low birth weight of newborns in the area. Is it reasonable they use the low birth weight as a argument for their campaign? Consider the birth weight of healthy newborns has a normal distribution with the

mean 3500 g and a random sample of 50 newborns in the area produced the sample mean 3310 g and the sample standard deviation 500 g. Use the 1% significance level.

9. A producer is about to check the observance of the weight of products during their packing. Consider the producer does not wish the excessive filling of products. Consider a random sample of 15 observations (g)

238	256	266	240	260	235	263	259
234	258	253	265	234	245	237	

What is the result of the producer in case the population is normally distributed and the significance level is $\alpha = 0.05$?

10. A factory recently reduced the production and consequently there is a suspicion that the average number of passengers in one bus will decrease for some transport lines, because the factory employs mostly workers who commute from surroundings. On that account the transport company accomplishes a random sample of 40 observations (randomly chosen transport lines during traffic peaks and corresponding numbers of passengers):

number of passengers in one bus	25	28	29	34	35	38	40	42	45
number of observations	2	4	7	10	6	5	3	2	1

It is well known that the average number of passengers in one bus was 36 passengers (before reduction). It is clear that if the number of passengers is smaller the company will restrict business. Using the 5% significance level, what is the decision of the company?

11. According to long-term statistics, it is well known that the number of new-born boys is greater than the number of female newborns. The probability of birth for a boy is 0.5142. On the contrary, the proportion in certain month was only 0.492. Moreover, it was at the time of an environmental disaster and hence there is a suspicion (hypothesis) that it caused due to the disaster.
    a) Using $\alpha = 0.05$, would you reject the hypothesis on condition that the total number of newborns in given month is 250.
    b) Perform the test on condition that the number of newborns is 1500.

12. According to a recent survey for some company, 33 % of households prefer supermarkets as the main place for the shopping. The company assumes the further growth of popularity for this kind of shopping and that is why the company performs a random sample of 250 individuals and finds that 93 of them prefer supermarkets. Is the output of the survey in agreement with the assumption ($\alpha = 0.05$)?

Solution.

1. $H : \mu = 30,\ A : \mu \neq 30,\ t = 1.633$, not reject;
2. a) $H : \mu = 4.2,\ A : \mu \neq 4.2,\ t = 0.986$, not reject; b) $H : \sigma = 0.4,\ A : \sigma \neq 0.4$, $\chi^2 = 22.059$, not reject;
3. $H : \sigma^2 = 0.2,\ A : \sigma^2 \neq 0.2,\ \chi^2 = 90$, reject;

4. $H : \sigma = 995,\ A : \sigma \neq 995,\ \chi^2 = 32.171$, the standard deviation 995 Euro is still actual;
5. $H : \sigma = 0.05,\ A : \sigma \neq 0.05,\ \chi^2 = 24.928$, not reject the declaration;
6. $H : \pi = 0.3,\ A : \pi < 0.3,\ u = -1.247$, reject the suspicion;
7. $H : \sigma = 0.9,\ A : \sigma > 0.9,\ \chi^2 = 33.778$, the accuracy is lower;
8. $H : \mu = 3500,\ A : \mu < 3500,\ u = -2.687$, not reasonable;
9. $H : \mu = 250,\ A : \mu > 250,\ t = -0.147$, not excessive filling;
10. $H : \mu = 36,\ A : \mu < 36,\ u = -2.795$, the company must restrict business;
11. a) $H : \pi = 0.5142,\ A : \pi < 0.5142,\ u = -0.7023$, false suspicion; b)$u = -1.7203$, positive suspicion;
12. $H : \pi = 0.33,\ A : \pi < 0.33,\ u = 1.4123$, reject the growth.

1.4 Hypothesis testing for two populations

1.4.1 Exercises 1.4

1. Consider two laboratory technicians, the task is to compare the accuracy of their measuring of chemical concentration. The first technician performs 15 observations with sample standard deviation 0.32 and the second performs 20 observations with sample standard deviation 0.55. Using the 5% and 1% significance level, verify if both technicians measure with the same accuracy. Assume that the two populations have normal distributions.

2. There are weights of a wolf (separately for male and female) in the following table (kg):

female	28	38	41	32	32	35	31	33			
male	32	42	46	38	40	53	39	39	42	39	48

   At the 5% significance level, can you conclude that the weights are different for these two populations? Assume that the two populations have normal distributions.

3. We are about to compare two methods how to train workers for a given specialized work. Consider two independent random samples, 10 randomly chosen workers were trained via 1. method and similarly 10 different workers were trained via 2. method. Consequently, the aptitude for the work were checked (the speed of a given operation, minutes):

1. method	15	20	11	23	16	21	18	16	27	24
2. method	23	31	13	19	23	17	28	26	25	28

   At the 5% significance level, can you conclude that the workers trained by 1. method are (on average) quicker than workers trained by 2. method?

4. The task is to check feeders in two cement factories. Consider the sample of 13 randomly chosen bags from cement factory A ($\overline{x} = 25.03$ kg, $s = 0.11$ kg) and similarly 9 bags from cement factory B ($\overline{x} = 24.80$ kg, $s = 0.23$ kg). Decide at the 5% significance level, if feeders are comparable in the case of the filling of bags. Assume that the two populations have normal distributions.
5. Consider a stock feeding, we obtain different results for two different feeding diets A and B (for a given period). Using the 5% significance level, compare both diets by the help of the test for two populations means and decide, if diet A or B achieves better results. Assume that the two populations have normal distributions.

Diet A	26	29	24	29	20	22	16	23	20	21
Diet B	17	16	20	18	19	24	22	22	17	25

6. A consumer takes fluorescent lamps from two distributors. The quality of lamps is judged by the number of switches on the light, it is usually about 2000 switches. Test at the 5% significance level whether the lifetime is same for both distributors.

Distributor A	2139	2041	1968	1903	1952	1980	2089	1915	2389
	2163	2072	1712	2018	1792	1849			
Distributor B	1947	1602	1906	2031	2072	1812	1942	2074	2132

7. There was implemented the new technique of training (experimental population) in a sports school. Test at the 5% significance level whether this new technique yields better performance in 12 minutes run. Random samples produced (m):
   Experimental population: $n_1 = 57, \overline{x} = 2853.5$, $s_x = 159.5$.
   Reference population: $n_2 = 46, \overline{y} = 2963.1$, $s_y = 157.9$.
   Assume that the two populations have normal distributions.
8. The task is to investigate the maximum number of pull-ups for students of some university. Hence, for a random sample of students was determined this number and after two years was repeatedly determined this number. Test at the 5% significance level whether there is an improvement, assume that the populations have normal distributions.

Pull-ups before	9	11	9	8	6	10	10	8	12	7	10	6	14	7	8
Pull-ups after	11	12	10	10	11	12	12	9	14	8	10	8	15	10	10

9. In terms of a national survey of musical abilities, a test was completed, separately for students from countryside (90 children) and from cities (100 children). The outcome is:

	Countryside	City
Number	90	100
Sample mean	76.4	81.2
Sample st. deviation	8.2	7.6

   At the 5% and 1% significance level, can you conclude that the abilities are different for these two populations?

10. Before 2 years was performed a chemical analysis (quantity of chlorine) of water in some lake, a random sample of 85 water samples. For last two years, there has been the significant reduction of utilization of the salt during winter. Hence, nowadays is performed a comparative analysis, a random sample of 110 water samples.

	Before 2 years	Nowadays
Sample mean	18.3	17.8
Sample st. deviation	1.2	1.8

Test at the 5% and 1% significance level whether the reduction in utilization of the salt decreases the quantity of chlorine in the lake.

11. In terms of a public opinion poll, preferences of the left wing and the right wing were investigated separately in two regions A and B. According to the gathered data, is it correct that the left wing has significantly greater support in the region A than in the region B? Test at the 5% significance level.

	Number	Left wing	Right wing	Doesn't know
Region A	150	50	70	30
Region B	200	50	90	60

Solution.
1. $F = 0.339$, for $\alpha = 0.05$ significant difference;, for $\alpha = 0.01$ not significant difference;
2. homoscedasticity, $t = -3.344$. significant difference;
3. homoscedasticity, $t = -1.806$, significant difference;
4. heteroscedasticity, $t = 2.787$, significant difference;
5. homoscedasticity, $t = 1.830$, significant difference;
6. homoscedasticity, $t = 0.756$, significant difference;
7. homoscedasticity, $t = -3.482$, significant difference;
8. $t = -6.081$, significant difference;
9. $u = -4.170$, significant difference for both levels;
10. $u = 2.321$; for $\alpha = 0.05$ significant difference;, for $\alpha = 0.01$ not significant difference;
11. $u = 1.694$, significant difference.

1.5 Hypothesis testing of equality of probability distributions

1.5.1 Exercises 1.5

1. Consider a sample of 200 men (20 years old) and we measure their performance in the 12 minutes run. The sample produces characteristics $a_3 = -0.223$ and $a_4 = = 0.321$. Use the C-test (the modified C-test) and decide, if this random variable could has a normal distribution.

2. There is a suspicion that an opponent uses a weighted die. Hence, we roll the die 600 times. Use the goodness-of-fit test and decide, if the die is fair or not. Choose $\alpha = 0.05$.

Number	1	2	3	4	5	6
Quantity	78	93	102	95	106	126

3. Consider a shooter shoots at a target repeatedly three times, only the number of hits is registered. The shooter performs this random experiment in total 50 times.

Number of hits	0	1	2	3
Number of observations	1	4	22	23

Estimate the probability $\hat{\pi}$ that the shooter hits the target (one shot). Is it possible to describe this random variable (the number of hits for three shots) by a binomial distribution? At the 5% significance level test this hypothesis by the means of χ^2 goodness-of-fit test. (Advice: use the fact that the sample mean is an unbiased estimator of the mean which is in case of a binomial distribution $n\pi$, where n is the sample size and π is the probability of a success - hit in one shot.)

4. Consider a sample of 10 observations, where the random variable represents the fat content in milk (g/100 g milk):

1.59 1.57 1.47 1.65 1.59 1.39 1.49 1.48 1.69 1.32.

At the 5% significance level verify, if the random variable could has a normal distribution, use the Kolmogorov-Smirnov test.

Solution.
1. $C = 2.813$, normal distribution; $C' = ...$;
2. $\nu = 5, \chi^2 = 12.74$, unfair die;
3. $\hat{\pi} = 0.78$, $\nu = 1, \chi^2 = 0.437$, binomial distribution is appropriate;
4. $d_1 0 = 0.155$, $d_{10,0.95,} = 0.409$, normal distribution.