9. Parametric hypothesis testing

Statistical hypothesis is a statement about some parameter or probability distribution of one or more populations. When the hypothesis is about a parameter of distribution it is called **parametric**. The hypothesis testing is the decision-making procedure about the hypothesis.

The statement under test is called the **null hypothesis** H_0 and the contradictory statement is called the **alternative hypothesis** H_1. The null hypothesis always states that a parameter q of the population distribution is equal to some value q_0, e.g. $H_0(q = q_0)$, while the alternative states that it is not equal (*two-sided*: $H_1(q \neq q_0)$), less than (*left one-sided*: $H_1(q < q_0)$) or greater than (*right one-sided*: $H_1(q > q_0)$) that value.

A hypothesis testing procedure consists of eight steps:

1. From the problem context, identify the parameter of interest.
2. State the null hypothesis H_0.
3. Specify an appropriate alternative hypothesis H_1. It can be either two-sided or left/right one-sided.
4. Choose a significance level α. Usually, $\alpha = 0.05$.
5. Determine an appropriate test statistic.
6. State the rejection region S_C for the statistic.
7. Compute the test statistic value.
8. Decide whether or not H_0 should be rejected and report that in the problem context.

In hypothesis testing, regarding the actual validity of H_0, four different situations are possible:

	H_0 is true	H_0 is false
H_0 is rejected	Type 1 error, $P = \alpha$	Correct decision, $P = 1 - \beta$
H_0 is not rejected	Correct decision, $P = 1 - \alpha$	Type 2 error, $P = \beta$

The **type 1 error** is committed when a true null hypothesis H_0 is rejected. Probability α for committing the type 1 error is chosen before testing of the hypothesis.

The **type 2 error** is committed when a false null hypothesis H_0 is not rejected. Probability β for committing the type 2 error cannot be chosen in advance as it depends on the actual distribution of the population under test which is not known.

The **P-value** of the test (p) is the smallest level of significance α that would lead to rejection of the null hypothesis H_0 with the given data.

In the following, examples of the most frequently used test statistics and the corresponding rejection regions are listed. Only two-sided alternatives H_1 are listed. In a case of an one-sided alternative H_1, only the corresponding boundary of the rejection region is kept where $\alpha/2$ is replaced by α.

Mean m of a) normally distributed population with known variance and any value of sample size n or b) any population distribution with unknown variance and $n > 30$, where it is estimated $\sigma = \sqrt{S^2}$:

$$H_0(m = m_0), \qquad H_1(m \neq m_0), \qquad Z = \frac{\langle X \rangle_n - m_0}{\sigma/\sqrt{n}}, \qquad S_C = \{(z < -z_{\alpha/2}) \cup (z > z_{\alpha/2})\}$$

Mean m of normally distributed population with unknown variance and $n < 30$:

$$H_0(m = m_0), \qquad H_1(m \neq m_0), \qquad T = \frac{\langle X \rangle_n - m_0}{S/\sqrt{n}}, \qquad S_C = \{(t < -t_{n-1;\,\alpha/2}) \cup (t > t_{n-1;\,\alpha/2})\}$$

Variance σ^2 of normally distributed population:

$$H_0(\sigma^2 = \sigma_0^2),\ H_1(\sigma^2 \neq \sigma_0^2), \qquad \chi^2 = \frac{(n-1)S^2}{\sigma_0^2}, \qquad S_C = \{(\chi^2 < \chi^2_{n-1;\,1-\alpha/2}) \cup (\chi^2 > \chi^2_{n-1;\,\alpha/2})\}$$

Population proportion p, in case the binomial distribution can be approximated by a normal:

$$H_0(p = p_0), \qquad H_1(p \neq p_0), \qquad Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \qquad S_C = \{(z < -z_{\alpha/2}) \cup (z > z_{\alpha/2})\}$$

Sum/difference of means $m_1 \pm m_2$ a) normally distributed populations with known variances and any n_1 and n_2 or b) any population distributions with unknown variances and $n_1, n_2 > 30$, where it is estimated $\sigma_{1,2}^2 = S_{1,2}^2$:

$$H_0(m_1 \pm m_2 = \Delta_0),\ H_1(m_1 \pm m_2 \neq \Delta_0), \qquad Z = \frac{\langle X_1 \rangle_{n_1} \pm \langle X_2 \rangle_{n_2} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \qquad S_C = \{(z < -z_{\alpha/2}) \cup (z > z_{\alpha/2})\}$$

Sum/difference of means $m_1 \pm m_2$ of normally distributed populations with unknown, but similar, variances and any n_1 and n_2, $S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$:

$$H_0(m_1 \pm m_2 = \Delta_0),\ H_1(m_1 \pm m_2 \neq \Delta_0), \qquad T = \frac{\langle X_1 \rangle_{n_1} \pm \langle X_2 \rangle_{n_2} - \Delta_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, S_C = \{(t < -t_{n_1+n_2-2;\,\alpha/2}) \cup (t > t_{n_1+n_2-2;\,\alpha/2})\}$$

Sum/difference of population proportions $p_1 \pm p_2$, in case the binomial distributions can be approximated by normal distributions:

$$H_0(p_1 \pm p_2 = \Delta_0),\ H_1(p_1 \pm p_2 \neq \Delta_0), \qquad Z = \frac{\hat{p}_1 \pm \hat{p}_2 - \Delta_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}, S_C = \{(z < -z_{\alpha/2}) \cup (z > z_{\alpha/2})\}$$

Variance ratio σ_1^2/σ_2^2 of normally distributed populations:

$$H_0(\sigma_1^2/\sigma_2^2 = \Delta_0),\ H_1(\sigma_1^2/\sigma_2^2 \neq \Delta_0), \qquad F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}, \qquad S_C = \{(f < 1/f_{n_2-1,n_1-1;\,\alpha/2}) \cup (f > f_{n_1-1,n_2-1;\,\alpha/2})\}$$

Here, the following property of the Snedecor distribution was used: $f_{n_1-1,n_2-1;\,1-\alpha/2} = 1/f_{n_2-1,n_1-1;\,\alpha/2}$.

9. Parametric hypothesis testing – problems

1. Buyer agrees with manufacturer to buy a series of products if the average content of harmful substance in the product is less than 10 mg. It is assumed that the harmful substance content in the product is normally distributed. To verify the content of harmful substance, nine products in the series are randomly selected.

    a. The manufacturer offers a series of products that actually contains an average of 9 mg of harmful substance with the standard deviation of 2 mg. What is the risk of the manufacturer that the series will be rejected? R: $P = 0.067$

    b. If the manufacturer offers a series of products that actually contains 12 mg of harmful substance with the standard deviation of 2 mg, what is the risk of the buyer to accept such a series? R: $P = 0.001$

2. The maximum permissible average mass of CO_2 in the exhaust of a particular type of car is 200 mg/kg. In a test of a car 50 measurements are made and the exhaust average is determined at 204 mg/kg of CO_2 with the standard deviation of 10 mg/kg. It is assumed that the mass of CO_2 in the exhaust is normally distributed.

    a. Can we claim that the vehicle being tested complies with the regulation about the average mass of CO_2 in the exhaust? R: No. $z = 2.83$, $p = 0.002$

    b. Suppose that the true average mass of CO_2 in the exhaust of the tested vehicle is 204 mg/kg. What is the probability that the noncompliance with the regulation is not detected in the test? R: $\beta = 0.119$ at $\alpha = 0.05$

    c. Solve the problem using a sample of ten measurements. R: Yes. $t = 1.27$, $p = 0.118$, $\beta = 0.688$

3. In polishing of magnetic heads with diamond abrasive either monocrystalline or polycrystalline diamonds can be used. Using the monocrystalline diamonds on a sample of ten magnetic heads an average roughness of 42 nm with standard deviation of 4 nm was achieved, while using the polycrystalline diamonds on the same size of the sample the average roughness was 44 nm with standard deviation of 3 nm. It is assumed that the roughness of the magnetic heads is normally distributed.

    a. Can we claim that the same roughness results in both cases? R: Yes. $t = -1.27$, $p = 0.220$

    b. What should be the significance value in order to recognize the difference between both types of diamond abrasives based on given samples? R: $\alpha = p = 0.220$

4. In the past, 46 % of residents supported construction of a waste processing plant near their village. After an effort was made to persuade the residents to support the construction, a survey including 200 respondents was conducted where 101 of the respondents agreed with the construction.

    a. Can we claim, based on the survey, that the support for construction has increased significantly? R: No. $z = 1.28$, $p = 0.100$

    b. What should be the sample size to recognize the same resulting ratio as significant increase in support of the construction? R: $n \geq 332$

5. Usage of non-bank saving in Austria and Slovenia is compared by a telephone survey of a randomly selected sample from each country. Among the 300 respondents from Austria, 169 use non-bank saving while among the 300 respondents from Slovenia there are 143 using non-bank saving.

    a. Can we claim that the population proportions of non-bank savers in Austria and Slovenia are equal? R: No. $z = 2.13$, $p = 0.033$

    b. What is the minimum significance value where the difference of population proportions would be recognized as significant? R: $\alpha = p = 0.033$

6. In continuous welding using a laser beam, weld depth is not constant. At the average weld depth of 4 mm, the current process achieves a standard deviation of weld depth 650 µm. In testing of a slightly modified procedure of laser welding a deviation of 450 µm has been achieved on a sample of ten welds. It is assumed that the weld depth is normally distributed.

    a. Can we claim that the modified procedure significantly reduces the standard deviation of the weld depth? R: No. $\chi^2 = 4.31$, $p = 0.110$

    b. What is the minimum significance value where the standard deviation of the modified procedure would be recognized as significantly reduced? R: $\alpha = p = 0.110$

    c. What is the minimum sample size to recognize the above change in standard deviation as significant? R: $n = 15$

7. Effectiveness of a new sliding bearing lubricant is tested. In the first group of bearings the new lubricant is applied while on the other, so called control group, standard lubricant is applied. After a usage cycle, wear at the selected location is measured. Data in µm is gathered in the table below. It is estimated that wear is normally distributed in both cases. Can we claim that the new lubricant significantly reduces wear? Test means and standard deviations. R: Yes. $t = 2.53$, $p = 0.024$ and $f = 1.06$, $p = 0.941$

Standard lubricant	13	12	8	10	11	4	5	9
New lubricant	8	2	6	7	5	9	0	3

8. Machines A and B should produce tablets of the same weight. A few tablets from each machine are weighted and the results in mg are gathered in the table below. It is assumed that the weight of the tablets from each machine are normally distributed. Can we claim that the distributions of mass are the same on both machines? R: Yes. $t = -1.05$, $p = 0.314$ in $f = 0.792$, $p = 0.825$

A	49,0	48,5	50,3	49,5	50,7	50,3	/	/
B	50,6	48,8	51,4	50,6	51,1	49,4	49,2	50,8

9. Parametric hypothesis testing – additional problems

1. The buyer agrees to accept a series of synthetic fibers, if the average diameter of a sample of eight randomly selected fibers would fall between 10 and 12 µm. It is assumed that the diameter of the fibers is normally distributed.
    a. Suppose that the actual average diameter of the fibers in series is 11.8 µm, with the standard deviation of 0.4 µm. What is the probability that the buyer will reject the series based on a random sample? R: $P = 0.079$
    b. Suppose that the actual average diameter of the fibers in the series is 9.6 µm, with the standard deviation of 0.4 µm. What is the probability that the buyer will accept the series based on a random sample? R: $P = 0.002$
2. A new brake pad material is tested which is supposed to wear slower than the current material which reaches a critical wear at the average travel distance of 80 Mm. On a sample of 40 pieces of the new material the critical wear is identified at the average travel distance of 86 Mm with the standard deviation of 8 Mm. It is assumed that the travel distance length at the critical wear of brake pads is normally distributed.
    a. Can we claim that the new material wears slower? R: Yes. $z = 4.74$, $p = 10^{-6}$
    b. Suppose that the real average travel distance at the critical wear of the new material is 86 Mm. What is the probability that this is not recognized based on the given sample with the standard deviation of 8 Mm? R: $\beta = 9.7 \cdot 10^{-4}$ at $\alpha = 0.05$
    c. Solve the problem for a sample of 10 pieces. R: Yes. $t = 2.37$, $p = 0.021$, $\beta = 0.302$
3. The noise level of valves from two manufacturers is compared. Based on a sample of ten valves from the first manufacturer the average noise is 40 dB with the standard deviation of 3 dB, while a sample of nine valves from the second manufacturer gives the average noise level of 42 dB with the standard deviation of 2.5 dB. A normal distribution of the valve noise level is assumed for both manufacturers.
    a. Do the average noise levels of valves from the two manufacturers differ significantly? R: No. $t = -1.57$, $p = 0.135$
    b. Let us assume that the two manufacturers offer samples of the same size for the noise level testing. What should be the sample size in order to recognize the difference between the noise levels as significant with the above sample averages and standard deviations? R: $n \geq 16$
4. Quality of a series of joints between the seal and the housing is inspected. In a sample of 200 joints 18 are found inadequate. The customer allows 5 % of inadequate joints in the series.
    a. Can we claim that the series of joints meets the requirements of the customer? R: No. $z = 2.60$, $p = 0.005$
    b. Let the actual proportion of inadequate joints in a series be 0.08. What is the probability that based on a sample of 200 joints the customer will reject the series? R: $P = 0.603$

5. In a company the quality of production in the first and the second shift is checked. From each shift, 550 products were randomly selected and the products non-complying to regulations were counted. In the first shift there were 67 such products, while in the second shift there were 58.

    a. Can we claim that the average number of non-complying products in both shifts is the same? R: No. $z = 0.86$, $p = 0.395$

    b. What is the minimum significance level to identify the average numbers of non-complying products as significantly different? R: $\alpha = p = 0.395$

6. A product is manufactured by bending the wire. With the current process, the standard deviation of the critical dimension of the product is 1.0 mm. To reduce the deviation, a stringent check of the input material is applied. On a sample of twenty products, produced using a stringent check of the input material, the standard deviation of 0.7 mm was measured. It is assumed that the critical dimension of the product is normally distributed.

    a. Can we claim that the stringent checking of the input material reduced the standard deviation of the critical dimension of the product? R: Yes. $\chi^2 = 9.31$, $p = 0.032$

    b. Suppose that using the higher quality wire the standard deviation of the critical dimension of the product is 0.6 mm. What is the probability that the improvement is not identified based on a sample of 20 products? R: $\beta = 0{,}084$

7. Influence of fresh and sea water on the fatigue cracks growth rate in the material is compared. The table below gives the measured rates of the crack growth in fresh and sea water at the selected load frequency and amplitude. Can we claim that the fatigue cracks grow at the same rate in fresh and sea water? Test means and standard deviations. R: No. $t = 3.21$, $p = 0.005$ in $f = 0.45$, $p = 0.40$

Fresh water	2.06	2.05	2.23	2.03	2.11	2.08
Sea water	1.90	1.93	2.06	1.75	2.01	1.89