Statistical Inference 1 Winter Term 2024/2025

Lecture: Prof. Dr. Thomas Augustin, Prof. Dr. Christian Heumann
Exercises: Sergio Buttazzo, Jan Anders
Tutorial, Corrections: Sophie Hopp, Michael Speckbacher **Suggested Solution**

Exercise Sheet 8

Exercise 8.1 - T-Tests

We consider the Theatre dataset, in which the annual spending on culture in general (in SFR) of the inhabitants of a Swiss city is listed. In addition to their age, each of the respondents also indicated their gender and their income (in 1000 SFR). If you want to try out the code for this exercise yourself, it can be found in the corresponding Ex_8_1.R file and needs this dataset.

(a) In a previous study, it was found that on average, citizens of the Swiss city spent 225 SFR per year on culture. To test whether the mean annual expenditure has changed, we perform a t-test ($\alpha = 0.05$). The following R output shows the result of this test:
Note: n = 699; s = 51.802

```
> t.test(theater$kultur, mu = 225)
One Sample t-test
data:  theater$kultur
t = -2.6256, df = 698, p-value = 0.008838
alternative hypothesis: true mean is not equal to 225
95 percent confidence interval:
 216.0086 223.7024
sample estimates:
mean of x
 219.8555
```

Why do we need a t-test in this case? What is the hypothesis tested here? What is the decision of the test? Arrive at the test decision in three different ways and interpret the result. Also calculate the given test statistic by hand.

Since we are testing the expected value of a sample with unknown variance, we need to use a t-test here. The hypotheses are :

$$H_0\colon \mu = 225 \qquad \text{versus} \qquad H_1\colon \mu \neq 225.$$

The test decision is possible in three ways:

$$|t| = |-2.6256| = 2.6256 > 1.96 = z_{0.975} = t_{\infty;\,0.975} \quad \Rightarrow \quad H_0 \text{ is rejected}$$
$$p = 0.009 < 0.05 = \alpha \quad \Rightarrow \quad H_0 \text{ is rejected}$$
$$225 \notin [216.01;\ 223.70] \quad \Rightarrow \quad H_0 \text{ is rejected}$$

The expenditure for culture is significantly different from 225 ($\alpha = 0.05$).

Calculation of the test statistics:

$$t = \frac{\bar{x} - \mu_0}{s}\sqrt{n} = \frac{219.86 - 225}{51.802}\sqrt{699} = \frac{-5.144}{51.802} \cdot 26.44 \approx -2.623$$

(b) Have a look at the following output:
Note: $n_1 = 309; s_{X_1} = 51.101; n_2 = 390; s_{X_2} = 52.305$

```
> t.test(theater$kultur ~ theater$geschlecht)
Welch Two Sample t-test
data:  theater$kultur by theater$geschlecht
t = 1.3018, df = 697, p-value = 0.1934
alternative hypothesis:
true difference in means is not equal to 0
95 percent confidence interval:
 -2.602222 12.841554
sample estimates:
mean in group 0 mean in group 1
       222.7120        217.5923
```

What is the hypothesis tested here? What is the decision of the test at a significance level of $\alpha = 0.05$? Calculate the given test statistic by hand.

The following variables are relevant:

X_1: Expenditure for culture among men, $\mu_1 = E(X_1)$, $\sigma_1^2 = Var(X_1)$
X_2: Expenditure for culture among women, $\mu_2 = E(X_2)$, $\sigma_2^2 = Var(X_2)$

Two sample (Double) t-Test (Welch-Test) for testing equality of means with unknown variance:

$$H_0\colon \mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0 \qquad \text{vs.} \qquad H_1\colon \mu_1 \neq \mu_2 \text{ or } \mu_1 - \mu_2 \neq 0.$$

Again, the test decision is possible in three ways::

$$|t| = |1.302| = 1.302 < 1.96 = z_{0.975} = t_{\infty;\,0.975} \quad \Rightarrow \quad H_0 \text{ is not rejected}$$
$$p = 0.193 > 0.05 = \alpha \quad \Rightarrow \quad H_0 \text{ is not rejected}$$
$$0 \in [-2.622;\ 12.842] \quad \Rightarrow \quad H_0 \text{ is not rejected}$$

Test statistic:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_{X_1}^2}{n_1} + \frac{s_{X_2}^2}{n_2}}}$$

$$\begin{aligned} t &= \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_{X_1}^2}{n_1}+\frac{s_{X_2}^2}{n_2}}} = \frac{222.71-217.59}{\sqrt{\frac{51.101^2}{309}+\frac{52.305^2}{390}}} \\ &= \frac{5.12}{3.93} \approx 1.302 \end{aligned}$$

(c) To test whether spending on theaters has changed over the years, a paired t-test ($\alpha = 0.05$) is performed to compare the variables 'spending yesterday' and 'spending today'.

```
> t.test(theater$theaterheute, theater$theatergestern, paired = T)
Paired t-test
data:  theater$theaterheute and theater$theatergestern
t = 1.0925, df = 698, p-value = 0.275
alternative hypothesis:
true difference in means is not equal to 0
95 percent confidence interval:
 -2.481496  8.707533
sample estimates:
mean of the differences
               3.113019
```

What is the hypothesis tested here? Why is it not allowed to perform a regular two sample t-test here? What is the test decision? Interpret the results.

The following variables are considered:

X_{old}: Expenditure for theatre in the last year, $\mu_{old} = E(X_{old})$,

X_{new}: Expenditure for theatre this year, $\mu_{new} = E(X_{new})$.

Hypotheses:

$$H_0\colon \mu_{old} = \mu_{new} \text{ or. } \mu_{old} - \mu_{new} = 0 \quad H_1\colon \mu_{old} \neq \mu_{new} \text{ or. } \mu_{old} - \mu_{new} \neq 0.$$

A paired t-test is used because the sample is a connected sample: The same characteristic (theater expenditure) is observed at two different points in time on each test object (Swiss citizens). A double t-test would require the samples to be independent.

Note: Again, because of the central limit theorem and the large sample size, an approximate standard normal distribution can be assumed for the test statistic.

The test decision can be made in three ways:

$$|t| = |1.0925| = 1.0925 < 1.96 = z_{0.975} = t_{\infty;\,0.975} \quad \Rightarrow \quad H_0 \text{ is not rejected}$$
$$p = 0.275 > 0.05 = \alpha \quad \Rightarrow \quad H_0 \text{ is not rejected}$$
$$0 \in [-2.481;\ 8.708] \quad \Rightarrow \quad H_0 \text{ is not rejected}$$

Spending on theatres has not changed significantly this year compared to last year.

Exercise 8.2 - Duality of Tests and Confidence Intervals

The best seller of Edumm cheese dairy is their 1000g ball cheese. The actual weight of the cheese varies during production and can be regarded as normally distributed with known variance $\sigma^2 = 49[g^2]$. The cheesemaker suspects that in reality, the average cheese weight is less than 1000g, which would of course not be acceptable. Thus, the dairy wants to perform (based on a sample of size n) a statistical test at level $\alpha = 0.05$ on the true mean weight of their ball cheese.

(a) Formulate the hypotheses for this scenario. Which test would you use and why?

The research hypothesis of interest comes in H_1, since only H_1 can be significantly verified by a test:

$$H_0 : \mu \geq 1000 \quad \text{vs.} \quad H_1 : \mu < 1000$$

Since the weight X_i of the cheese is normally distributed with known $\sigma^2 = 49[g^2]$ and we onlz want to check whether the weight might be too low, a one-sided Gauß-test at significance level $\alpha = 0.05$ and sample size n is suitable.

(b) Which test decision do you arrive at for a measured mean cheese weight of 997g? Perform the test for $n = 10$ and $n = 30$. Interpret the results. Also calculate the p-values and rejection region of the tests.

From the specification, we take $n_1 = 10$ and $n_2 = 30$, $\bar{x} = 997$. The one-sided Gaussian test:

$$z = \frac{\bar{x} - \mu_0}{\sigma}\sqrt{n} = \frac{997 - 1000}{\sqrt{49}}\sqrt{n}$$

$$z_{n_1} = -\frac{3}{7}\sqrt{10} = -1.3552$$

$$z_{n_2} = -\frac{3}{7}\sqrt{30} = -2.3474$$

with the rejection region

$$Z < z_\alpha = z_{0.05} = -1.65$$

is used. Because of $z_{n_1} = -1.3552 > -1.65$ H_0, cannot be rejected for $n_1 = 10$. For $n_2 = 30$ $z_{n_2} = -2.3474 < -1.65$ is valid and H_0 can be rejected. This shows that the detectable difference, given equal variance, is dependent on the sample size. Here this means that with $n_2 = 30$ it can be assumed that the ball cheese sold is too light. With $n_1 = 10$ this statement cannot be made yet.

The p-value is calculated in the following way:

$$P_{n_1}(Z \leq z_{n_1}) = P_{n_1}(Z \leq -1.3552) = \Phi(-1.3552) = 0.0877.$$

$$P_{n_1}(Z \leq z_{n_2}) = P_{n_1}(Z \leq -2.3474) = \Phi(-2.3474) = 0.0095.$$

Or in R (same code for $n = 30$, accordingly):

```
z10 <- ((997 - 1000) / sqrt (49)) * sqrt(10)
1 - pnorm(z10) # 0.08766707
# Or:
1 - pnorm(997, mean = 1000, sd = sqrt(49/10))
```

The probability of observing the given realisation of the test statistic or a more extreme value under H_0 is $p_{n_1} = 0.0877$ and $p_{n_2} = 0.0095$.

With help of the p-value we can also derive a test decision at $\alpha = 0.05$:

- For $n_1 = 10$: Since $p = 0.0877 > 0.05 = \alpha$, the realisation of the statistic $z = -1.3552$ is not in the rejection region of the null hypothesis at level $\alpha = 0.05$. H_0 can not be rejected.
- For $n_2 = 30$: Since $p = 0.0095 < 0.05 = \alpha$, $z = -2.3474$ is in the rejection region of the null hypothesis for $\alpha = 0.05$. H_0 can be rejected in favour of $H1$.

(c) The diagram below shows the density of the sample mean under the null hypothesis at a sample size of $n = 30$. Fill in the values on the x-axis and calculate and mark the rejection region of the null hypothesis. Also mark the region of the density where the p-value is 'located'.

See solution below in (d) and in Ex_8_2.R.

(d) Calculate a 90% confidence interval for μ at $n = 30$ and also mark it in the diagram from (c).

The 90% confidence interval for the population mean μ is given by:

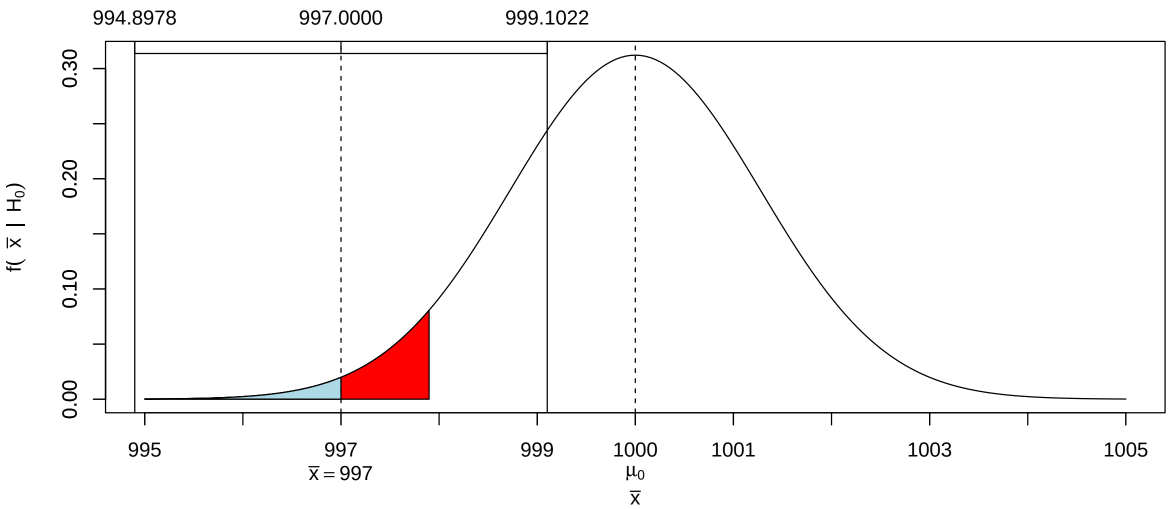
$$[\bar{x} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}].$$

Substituting the values:

$$[997 \pm 1.645 \cdot \sqrt{\frac{49}{30}}] \approx [997 \pm 2.102] = [994.898, 999.102].$$

Or in R:

```
> qnorm(c(0.05, 0.95), mean = 997, sd = sqrt(49/30))
[1] 994.8978 999.1022
```



Exercise 8.3 - Tests

As part of a broader study on traffic and car usage in large cities, the average monthly rent for permanent parking spots or garages is examined between Hamburg (H) and Berlin (B). It is assumed that on average, Hamburg residents pay more than 3€ more than Berliners for a parking spot. In Hamburg, 200 parking slots were randomly selected and in Berlin, 250 randomly selected parking spots were examined. The prices for parking can be considered approximately normal. The following estimates were collected:

$$\bar{x}_H = 175 \quad \bar{x}_B = 168.4$$

Let X_H be the random variable for the cost of parking spaces in Hamburg and X_B the variable for the cost in Berlin. We assume a normal distribution with

$$X_H \sim N(\mu_H, \sigma_H^2) \qquad X_B \sim N(\mu_B, \sigma_B^2)$$

The hypothesis can then be formulated as

$$H_0 : \mu_H - \mu_B \leq 3 \qquad H_1 : \mu_H - \mu_B > 3$$

with $n_H = 200$ and $n_B = 250$ and

$$\hat{\mu}_H = \bar{x}_H = 175.0 \qquad \hat{\mu}_B = \bar{x}_B = 168.4$$

(a) Can the scientists' assumption be maintained if from previous investigations $\sigma_H^2 = 600$ and $\sigma_B^2 = 250$ is known ($\alpha = 0.05$)?

Since σ_X^2, σ_Y^2 are known, we use a Gauss test and the test statistic is

$$Z = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}. \tag{1}$$

Note that in the case of variance homogeneity, i.e. $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, this would simplify to

$$Z = \frac{\bar{X} - \bar{Y} - \delta_0}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}, \tag{2}$$

which is the formula that is known from the lecture. In general, using (1), we can also conduct a Gauss test if the variances are unequal, but known.

Since $\sigma_H^2 = 600$ and $\sigma_B^2 = 250$ are clearly very different, we plug into (1) and get

$$z = \frac{\bar{x}_H - \bar{x}_B - 3}{\sqrt{\frac{\sigma_H^2}{n_H} + \frac{\sigma_B^2}{n_B}}} = \frac{175.0 - 168.4 - 3}{\sqrt{\frac{600}{200} + \frac{250}{250}}} = \frac{3.6}{2} = 1.80$$

as test statistic with the rejection criterion

$$Z > z_{1-\alpha} = z_{0.95} = 1.65$$

Because of $z = 1.80 > 1.65$, H_0 can be rejected, i.e. the assumption can be confirmed that on average, Hamburgers pay more than 3 euro more than Berliners.

(b) What changes compared to (a) if σ_H^2 and σ_B^2 are assumed to be unknown but the same, where from the sample we calculated $s_H^2 = 665$ and $s_B^2 = 288$?

If the variances are estimated by $s_H^2 = 665$ and $s_B^2 = 288$ but are theoretically assumed to be equal, then as test statistic we can use

$$t = \frac{\bar{x}_H - \bar{x}_B - 3}{\sqrt{\left(\frac{1}{n_H} + \frac{1}{n_B}\right) \frac{(n_H-1)s_H^2 + (n_B-1)s_B^2}{n_H+n_B-2}}} = \frac{175.0 - 168.4 - 3}{\sqrt{\left(\frac{1}{200} + \frac{1}{250}\right) \frac{199 \cdot 665 + 249 \cdot 288}{448}}} = \frac{3.6}{2.0246} = 1.78$$

with the rejection range

$$T > t_{n_H+n_B-2,\ 1-\alpha} = t_{448,\ 0.95} \approx z_{0.95} = 1.65$$

Because of $t = 1.78 > 1.65$, H_0 can be rejected as in (a).

Note: Assuming the variances to be theoretically equal with such a large difference in the actual sample is negligent!

(c) What changes compared to (b) if σ_H^2 and σ_B^2 are unknown and not equal?

If the variances are estimated by $s_H^2 = 665$ and $s_B^2 = 288$ and are theoretically not equal, then in variation to (a) we use

$$t = \frac{\bar{x}_H - \bar{x}_B - 3}{\sqrt{\frac{s_H^2}{n_H} + \frac{s_B^2}{n_B}}} = \frac{175.0 - 168.4 - 3}{\sqrt{\frac{665}{200} + \frac{288}{250}}} = \frac{3.6}{2.1159} = 1.70$$

as a test statistic with the rejection range

$$T > t_{k,1-\alpha}$$

and the Welch modification to the degrees of freedom of the distribution under H_0:

$$\begin{aligned} k &= (S_X^2/n + S_Y^2/m) / \left(\frac{1}{n-1} \left(\frac{S_X^2}{n} \right)^2 + \frac{1}{m-1} \left(\frac{S_Y^2}{m} \right)^2 \right) \\ &= 73.5313 \end{aligned}$$

We obtain the rejection range (here, the approximation to the normal is already a bit imprecise)

$$t_{74,0.95} = 1.666 \quad (\approx z_{0.95} = 1.65).$$

Because of $t = 1.70 > 1.67$, H_0 is still rejected, i.e. there is sufficient evidence that the assumption in H_1 holds.