Statistical Methods Fall 2021

Assignment 3: Estimation and testing

Deadline: see Canvas

Topics of this assignment

The exercises below concern topics that were covered in Lectures 5, 6, 7 and 8: estimation of population proportion and mean, tests for population proportion or mean based on one sample, tests for differences of two means or proportions based on two samples (see the respective sections in Chapters 6, 7, and 8 of the book and the slides of Lectures 5–8). Before solving the assignment, study these topics.

How to make the exercises? See Assignment 1.

If you are asked to perform a test, do not only give the conclusion of your test, but report:

– the hypotheses in terms of the population parameter of interest;
– the significance level;
– the test statistic and its distribution under the null hypothesis;
– the observed value of the test statistic (the observed score);
– the P-value or the critical values;
– whether or not the null hypothesis is rejected and why.
If applicable, also phrase your conclusion in terms of the context of the problem.

Theoretical exercises

For the three theoretical exercises below use Tables 2 and 3 from the Appendix in the book (also given on Canvas) to find critical values. Do not use R. If you need to use a t-distribution with the number of degrees of freedom not included in Table 3, report the number of degrees of freedom, and use the critical value based on a t-distribution with the next lower number of degrees of freedom found in the table.

Exercise 3.1 You have been employed by an airline and given the task of estimating the percentage of flights that arrive on time, i.e. no later than 20 minutes after the scheduled arrival time. How many flights should you survey in order to be 95% confident that your estimate is within two percentage points of the true population percentage? Solve this problem under each of the following assumptions:

a) nothing is known about the percentage of on-time flights of your airline;

b) for a previous year, 90% of all flights of your airline were on time (based on an official statistic from the government).

Also state the formulas that you are using.

Exercise 3.2 Consider a study about the brain volumes (in cm^3) of 30 pairs of twins. The aim of the study was to quantify the difference in the population mean volumes of first- and second-born twins. The average volume of the 30 first-born twins was $\bar{x}_1 = 1124.3$, and that of the 30 second-borns was $\bar{x}_2 = 1118.1$. Some more statistics that you may or may not use are: $s_1 = 130.5, s_2 = 124.7, s_d = 57.8$. Construct a 90% confidence interval for the difference of the population means.
Also state the formulas that you are using.

Exercise 3.3 Consider a study about the brain volumes (in cm^3) of 25 first-borns and 20 (unrelated) second-borns. The aim of the study was to test whether there is a difference in the population mean volumes of first- and second-born babies. The average volume of the 25 first-borns was $\bar{x}_1 = 1131.3$, and that of the 20 second-borns was $\bar{x}_2 = 1123.8$. Some more statistics that you may or may not use are: $s_1 = 129.0, s_2 = 127.2, s_p = 128.2$. Conduct the statistical test of interest and use $\alpha = 5\%$. Do we need to make specific assumptions in order to be allowed to use the test?
(See the first page of the assignment for detailed instructions about testing)

R-exercises

Hints concerning R:

- The R-functions `pnorm` and `qnorm` can be used for computing probabilities and quantiles of normally distributed random variables. Similarly, `pt(...,df=k)` and `qt(...,df=k)` can be used to compute probabilities and quantiles of a t-distributed random variable with k degrees of freedom.
- For computing a confidence interval based on a t-distribution and for performing a t-test the R-function `t.test` can also be used. For example, to perform a t-test with significance level 5% for testing the null hypothesis $\mu = 10$ against the alternative hypothesis $\mu > 10$, using the data set `example` use the command `t.test(example,mu=10,alt="greater")`. The arguments of `t.test` can be adjusted to test with other significance levels and other null and alternative hypotheses as well: study `help(t.test)`.
- Note that the function `t.test` reports also the P-value. Use `t.test(...)$p.value` to access it without printing the whole output of `t.test()`.
- In order to get a proportion of elements of the vector `x` that are, for instance, less than some value `M` you can use `mean(x<M)`.
- The R-function `t.test` can also be used for *two*-sample problems: put the values of the two samples into two vectors, `x` and `y`, say; the command `t.test(x,y, ...)` performs a two-sample test (based on the t-distribution) and computes a confidence interval as well, with default significance level $\alpha = 5\%$. Values of other arguments, like `"paired = TRUE"` to perform a paired t-test, can be specified by inserting them at the position of the dots.

Exercise 3.4 Alice and Bob work evening shifts in a supermarket. Alice has complained to the manager that she works, on average, much more than Bob. The manager claims that on average they both work the same amount of time, i.e. the competing claim is that the average working hours are different. After a short discussion between the manager and Alice, the manager randomly selected 50 evenings when Alice and Bob both worked. The datasets `Alice` and `Bob` in the file `Assign3.RData` contain the number of hours they have worked per evening.

a) Give an estimate and also a 90% confidence interval for the difference of mean working time per evening of Alice and Bob.
Also state the formulas that you are using.
(Remark: sometimes estimates, i.e. statistics which are single numbers, are also called *point estimates*, whereas confidence intervals are sometimes called *interval estimates*.)

b) Investigate the manager's claim with a suitable test. Take significance level 10%. Motivate your choice.
(See the first page of the assignment for detailed instructions about testing)

c) Now investigate Alice's claim with a suitable test. Take significance level 1%. Motivate your choice.
(See the first page of the assignment for detailed instructions about testing)

Exercise 3.5 Alice from the previous exercise has another concern. By contract the employees are supposed to work 3.8 hours per evening. Alice claims that the proportion of evenings on which she worked more than 3.8 hours is larger than the proportion of evenings during which Bob worked more than 3.8 hours.
In contrast to Exercise 3.4, we now suppose that the data given in `Assign3.RData` were all collected on different evenings for Suppose again that the data were collected on 100 different evenings. We will reuse that dataset in this exercise.

a) Based on the data find an estimate for the difference in proportion of evenings that Alice and Bob have worked more than 3.8 hours.

b) Investigate Alice's claim with a suitable test. Take significance level 1%. Motivate your choice.
(See the first page of the assignment for detailed instructions about testing)

Assignment 3

Statistical Methods 2021

Project Group 3:
Caspar M. S. Grevelhörster (2707848)
Zülfiye Tülin Manisa (2664758)
Sidharth R. Singh (2708756)

Vrije Universiteit Amsterdam,
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

Exercise 1

a) When nothing is known about the probability that flights are on time, one can use the following formula to determine the minimum sample size n: $n \geq (\frac{Z_{\alpha/2}}{2*E_{max}})^2$ where E_{max} is the maximum margin of error (0.02) and $Z_{\alpha/2}$ the z-score found in **Table 2** (1.96). The calculation results in $n \geq (\frac{1.96}{2*0.02})^2 = 2401$. At least 2401 flights should be checked to have a 95% confidence that flights are within 2% range of the real value.

b) The sample size n can be calculated using the following formula: $n \geq \lceil Z^2_{\alpha/2} * \frac{p(1-p)}{E_{max}} \rceil$ where E_{max} is the maximum margin of error (0.02) and p the probability of flights being on time (0.90). This results in the following calculation: $\lceil 1.96^2 * \frac{0.9(0.1)}{0.02^2} \rceil = 865$. This means, that at least 865 flights should be picked in order to have a 95% confidence that flights are within a 2% range of the real value given that 90% are on time.

Exercise 2

We know that the pairs are matched so the formula we can use to construct the confidence interval for $\mu_1 - \mu_2$ is: $\left[\bar{d} - t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}}, \bar{d} + t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}}\right]$. First we calculate the mean of the differences: $\bar{d} = \bar{x}_1 - \bar{x}_2 = 1124.3 - 1118.1 = 6.2$. Next we calculate the margin of error. It is given that $n = 30$, $s_d = 57.8$ and since we construct a 90% confidence interval, $\alpha = 0.1$. In **Table 3** of the book, we see that $t_{29,0.05} = 1.699$ ($\alpha/2$ because it is a two-tailed test). Thus: $E = t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}} = 1.699 * \frac{57.8}{\sqrt{30}} \approx 17.93$. It follows that the 90% confidence interval is $\left[\bar{d} - E, \bar{d} + E\right] = [6.2 - 17.93, 6.2 + 17.93] = [-11.73, 24.13]$.

Exercise 3

Step 0: Population parameter:
$\mu_1 - \mu_2$: Differenence in mean brain volumes of first-born (μ_1) and second-born (μ_2) babies.
Step 1: Hypotheses and significance level α:
The null hypothesis is that there is no difference in mean brain volumes between first and second born babies: H_0: $\mu_1 = \mu_2$ vs. The alternative hypothesis is that there is a difference: H_1: $\mu_1 \neq \mu_2$ with significance level $\alpha = 0.05$
Step 2: Data:
$n_1 = 25$ first born babies and $n_2 = 20$ second born babies, $\bar{x}_1 = 1131.3$, $s_1 = 129.0$; $\bar{x}_2 = 1123.8$, $s_2 = 127.2$. Since s_1 and s_2 are relatively close we can assume that $\sigma_1 = \sigma_2$.
Step 3: Test statistic and critical region:
Since the samples are independent with (assumed to be) equal population standard deviations, we use the test statistic: $T_2^{eq} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2/n_1 + S_p^2/n_2}}$. Under $H_0, T_2^{eq} \sim t_{n_1+n_2-2} = t_{43}$. The pooled sample variance is: $s_p^2 = 128.2^2 = 16435.24$. The observed value of the test statistic is: $t_2^{eq} = \frac{(1131.3-1123.8)}{\sqrt{16435.24/25+16435.24/20}} \approx 0.195$. Critical region using **Table 3** of the book, where we check for an area of 0.05 in two tails, because this is a two tailed test (using 40 degrees of freedom since the area for 43 degrees of freedom is not in the table): $t_2^{eq} \leq -2.021$ or $t_2^{eq} \geq 2.021$. It is clear that the observed value 0.195 is not in the critical region.
Step 4: Conclusion:
There is not sufficient evidence to reject the null hypothesis H_0, which claims that there is no difference in the mean brain volumes between first- and second-born babies.

In order to conduct the test as we did above, we must assume that both samples come from a normal population because the sample sizes $n1, n2$ are smaller than 30. Furthermore, as explained in **Step 2** we assume that $\sigma_1 = \sigma_2$.

Exercise 4

a) A point estimate for the difference of mean working time can be calculated by subtracting the two sample means of the datasets: $\bar{x}_{Alice} - \bar{x}_{Bob} = 3.957444 - 3.802945 = 0.1544989$. It is assumed that the datasets are paired, since the manager "randomly selected 50 evenings when Alice and Bob *both worked*". The confidence interval for the t-distribution can be computed using the following formula: $\left[\bar{d} - t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}}, \bar{d} + t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}}\right]$. The mean of the differences was already calculated and equals 0.1544989. The standard deviation of differences can was calculated using $s^2 = \frac{\sum(x-y)^2}{n-1} = 0.5696574$ Since we construct a 90% confidence interval, $\alpha = 0.1$.In **Table 3** of the book, we see that $t_{49,0.05} \approx 1.676$. Thus: $E = t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}} = 1.676 * \frac{0.5696574}{\sqrt{50}} \approx 0.1350$. It follows that the 90% confidence interval is $\left[\bar{d} - E, \bar{d} + E\right] = [0.1545 - 0.1350, 0.1545 + 0.1350] = [0.0194, 0.2895]$.

The difference of mean working time is t-distributed and the confidence interval of this distribution can thus also be computed using the t.test function in R. It needs to be specified that the two samples are dependent (R: "paired"), since Alice and Bob are working together. Also, the confidence level (R: "Conf.level") is set to 0.90, following the instructions. The appropriate t-test was performed using R and the confidence interval taken from the test:

```
load("Assign3.RData")
test = t.test(Alice, Bob, paired = TRUE, conf.level = 0.9)
conf_int = c(test$conf.int[1], test$conf.int[2])
conf_int

## [1] 0.0194331 0.2895647
```

For the specified two-sided test, R computed the confidence interval CI [0.0194331, 0.2895647] using the same formula as we used in the manual part above. It had the same results representing the interval estimate and thereby confirmed the manually calculated results.

b) The manager claims that the working times of Alice and her colleague are equal. Alice's claims the opposite by saying they are unequal. To test those claims, two hypotheses are used: The null hypothesis, claiming that they are equal and thus representing the manager $H_0 : \mu_{Alice} - \mu_{Bob} = 0$ as well as the alternative hypothesis, claiming that Alice works unequal amounts, so $H_a : \mu_{Alice} - \mu_{Bob} \neq 0$. The significance level α of this two-sided test equals 0.1. The t-test function is included in R and computes the observed T_d-value using the equation of $T_d = \frac{\bar{D}-d_0}{S_d/\sqrt{n}}$ where $\bar{D}$ is the average difference between all pairs, S_d the standard deviation between all pairs and d_0 the expected mean difference under the H_0-hypothesis. Whenever the *t.test()*-functionality of R is used in this assignment, the calculation of the observed value of the test statistic will be based on this exact formula.

The appropriate t-test for the t-distributed means can be performed using R:

```
load("Assign3.RData")
t.test(Alice, Bob, conf.level = 0.9, paired = TRUE)

##
##  Paired t-test
##
## data:  Alice and Bob
## t = 1.9178, df = 49, p-value = 0.06098
## alternative hypothesis: true difference in means is not equal to 0
## 90 percent confidence interval:
##  0.0194331 0.2895647
## sample estimates:
## mean of the differences
##               0.1544989
```

Fig. 1: Graphical representation of the two-tailed t-test.

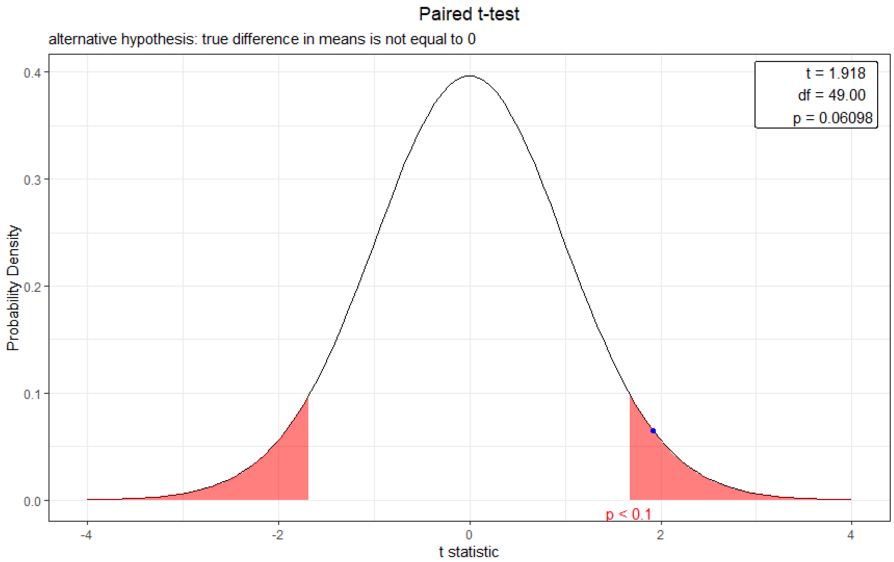

Fig. 1 describes the t-distribution of means of differences of working times. The critical values can be found in approximated form in **Table 3** or exactly computed using R:

```
c(qt(p = 0.1/2, df = 49, lower.tail = TRUE), qt(p = 0.1/2,
    df = 49, lower.tail = FALSE))

## [1] -1.676551  1.676551
```

The red area in fig. 1 is located at $x \leq -1.676551, x \geq 1.676551$ since those are the two critical t-values for $df = 49$ and $\alpha = 0.1$. R calculated, that the observed t-value of the test statistic lies at $t = 1.918$ (also represented by the blue dot in fig. 1). The observed value of the test statistic lies inside the critical region i.e. is larger than the critical value of 1.676551, so $t_{0.1/2} < t$. Because of the value lying inside of the critical region, H_0 is rejected. Since H_0 is rejected, there is enough evidence to reject the manager's claim that the working times are equal and Alice is thus indeed working different amount of hours than Bob.

c) Alice's claim is investigated in this task. She claims that she is working more than Bob, thereby opposing the manager's claim. So $H_0 : \mu_{Alice} - \mu_{Bob} = 0$ representing the manager and $H_a : \mu_{Alice} - \mu_{Bob} > 0$ representing Alice. The significance level α of this one-tailed test equals 0.01. The appropriate t-test can be performed using R:

```
load("Assign3.RData")
t.test(Alice, Bob, conf.level = 0.99, alternative = "greater", paired = TRUE)

##
##  Paired t-test
##
## data:  Alice and Bob
## t = 1.9178, df = 49, p-value = 0.03049
## alternative hypothesis: true difference in means is greater than 0
## 99 percent confidence interval:
##  -0.03924329         Inf
## sample estimates:
## mean of the differences
##               0.1544989
```

Fig. 2: Graphical representation of the one-tailed t-test.

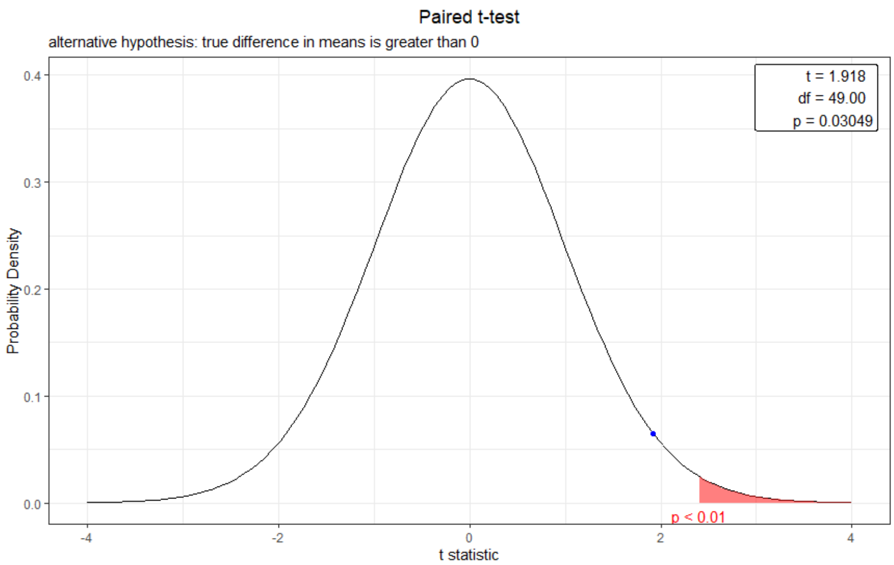

Fig. 2 is the same one as in task **b)**, except for the red area that is smaller and only on one side (since Alice's claim is that she's working *more* not *unequal* amounts). The critical value is again calculated in R:

```
qt(p = 0.01, df = 49, lower.tail = FALSE)

## [1] 2.404892
```

Therefore, $t_{\alpha=0.1} > t_{\alpha=0.01}$. The important difference to task **b)** is that now, the observed t-value $t = 1.918$ lies *outside* of the critical t-value $t_{0.01} = 2.404892$ (i.e. it is lower). With the significance level α of 1%, it is not possible to reject the H_0 hypothesis. This means that there is not sufficient evidence to support the claim of Alice that she is working more hours than Bob.

Exercise 5

a) The estimate for the difference in proportion can be computed using the number of evenings where the persons' worked hours are greater than 3.8 hours divided by the total number of considered evenings: $\frac{X_1}{n_1} - \frac{X_2}{n_2} = \frac{|\{x \in Alice: x > 3.8\}|}{50} - \frac{|\{y \in Bob: y > 3.8\}|}{50} = 0.7 - 0.46 = 0.24$. This can also be computed in R:

```
load("Assign3.RData")
mean(Alice > 3.8) - mean(Bob > 3.8)

## [1] 0.24
```

b) Just like in **a)**, The H_0 hypothesis represents the claim that Alice works equal amounts of days more than 3.8 hours as Bob: $H_0 : \frac{|\{x \in Alice: x > 3.8\}|}{50} - \frac{|\{y \in Bob: y > 3.8\}|}{50} = 0$. The alternative hypothesis $H_a : \frac{|\{x \in Alice: x > 3.8\}|}{50} - \frac{|\{y \in Bob: y > 3.8\}|}{50} > 0$ claims that Alice works more than 3.8 hours on more evenings than Bob does. To test this claim, the following test statistic can be used:

$$z_p = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{P}(1-\hat{P})}{n_1} + \frac{\hat{P}(1-\hat{P})}{n_2}}}$$

Where the two sample proportions are $\hat{p}_1 = \frac{X_1}{n_1}$, $\hat{p}_2 = \frac{X_2}{n_2}$ and the pooled sample proportion is $\hat{P} = \frac{X_1+X_2}{n_1+n_2}$. Following this calculation, $\hat{p}_{Alice} = \frac{35}{50} = 0.7$, $\hat{p}_{Bob} = \frac{23}{50} = 0.46$ and $\hat{P} = \frac{58}{100} = 0.58$. With these values on mind, the observed score of the normally distributed test statistic can be calculated:

$$z_p = \frac{0.7 - 0.46}{\sqrt{\frac{0.58(1-0.58)}{50} + \frac{0.58(1-0.58)}{50}}} \approx 2.4313$$

The next step is to find the p-value using **Table 2** for this result: $p(2.4313 \leq Z_p) \approx 0.0075$. Since $p \approx 0.0075 < \alpha = 0.01$, there is enough evidence to reject H_0 and to accept H_a; the proportion of evenings on which Alice worked more than 3.8 hours is larger than the proportion of evenings during which Bob worked more than 3.8 hours.

1 Appendix

1.1 Code of exercise 4b

```
# the following two packages need to be installed before plotting is possible:
# install.packages('devtools') devtools::install_github('cardiomoon/webr') the
# following two packages need to be imported: require(moonBook) require(webr)
load("Assign3.RData")
test = t.test(Alice, Bob, conf.level = 0.9, paired = TRUE)
plot(test)

## Error in xy.coords(x, y, xlabel, ylabel, log): 'x' is a list, but does not have components
'x' and 'y'
```

1.2 Code of exercise 4c

```
# the following two packages need to be installed before plotting is possible:
# install.packages('devtools') devtools::install_github('cardiomoon/webr') the
# following two packages need to be imported: require(moonBook) require(webr)
load("Assign3.RData")
t = t.test(Alice, Bob, conf.level = 0.99, alternative = "greater", paired = TRUE)
plot(t)

## Error in xy.coords(x, y, xlabel, ylabel, log): 'x' is a list, but does not have components
'x' and 'y'
```