CHAPTER 5. HYPOTHESIS TESTING (TWO SAMPLES)

Hypothesis testing: Two samples

- Comparison of means (independent samples)
- Paired data test (dependent samples)
- Selecting the right statistical test

Chapter 5: Assignments (Hypothesis testing: Two samples)

1. The average relative humidities (%) for two cities, are: 72.9 (city A) and 70.8 (city B), based on 25 measurements of relative humidity in each city. The standard deviations of these measurements are: 2.5 (city A) and 2.8 (city B). Based on the samples, can it be concluded with 95% confidence level that the relative humidity in the two cities is significantly different?

- First, we can see, that we have got two different averages of two different cities. In this case we can see that we have got 72,9 and 70,8 based on 25 measurements. Hereafter we have got the standard deviations of these measurements which are 2,5 and 2,8. Out of these levels and variable we would try and use the formula of two independent samples.

$$\boxed{\frac{(\overline{x_1}-\overline{x_2})-(\mu_1-\mu_2)}{s_{\overline{x_1}-\overline{x_2}}}} \approx \boxed{t_{N1+N2-2}}$$

- In this case we can see that the averages are 72,9 and 70,8. Because of that we can start by subtracting these two numbers.

$$72{,}9 - 70{,}8 \approx 2{,}1$$

- Now we will go forward, and then we will try and look at the difference between means of two populations. In our case, it is 0.
- Because we have the values of the upper part of the bracket, then we can afterwards use the pooled standard deviation, where we will use the sample length and then we will subtract with 1 and then multiply with the standard deviation.

$$s_{x_1-x_2} = \sqrt[2]{\frac{(N_1-1)\cdot s_1^2 + (N_2-1)\cdot s_2^2}{N_1+N_2-2}}$$

$$\sqrt[2]{\frac{(25-1)\cdot 2{,}5^2 + (25-1)\cdot 2{,}8^2}{25+25-2}} \cdot \sqrt[2]{\frac{1}{25}+\frac{1}{25}} = 0{,}75$$

- If we put the values together, then we can see that the following brackets will be like this.

$$\frac{(2{,}1)-0}{0{,}75} \approx 2{,}8$$

- Now, we can see that we have got the level of significance to be at 95%. In this case we can see that the a=0,05. In this case we will have to find the critical value and then use the table to find the correct value of student t distribution.

$$t_{N_1+N_2-2}$$
$$25 + 25 - 2 = 48$$

- Now we can go forward and use the student t-value calculator and see if what value we get.

$$t_{table} = 2{,}01$$

- Now we can go forward and create our interval.

$$-2{,}01 < 0{,}75 < 2{,}01$$

2. The dean of a university claims that the average scores in Software Engineering education, of those students that were educated in public high schools is higher than the average scores of those students that were educated in private high schools. A sample of the 50 students from each group is randomly selected. The average scores of the students from public schools are 8.6 (out of 10) and with a standard deviation equal to 3.3. The average scores of the students front private schools are 7.9 (out of 10) and with a standard deviation equal to 3.3. Are we 90°% confident that the statement of the dean is true?

- We will start by finding the main important values. In this case, we can start by taking looking after the means. In this case we can see that the means are mentioned as 8,6 and 7,9. Whereas if we look forward towards the amount of people, then we can see that we have got two samples which are 10. Then we can see that the standard deviation of both schools is the same, which are 3,3.

$$\boxed{\frac{(\overline{x_1}-\overline{x_2})-(\mu_1-\mu_2)}{s_{\overline{x_1}-\overline{x_2}}}} \approx \boxed{t_{N1+N2-2}}$$

$$(8{,}6-7{,}9)-0$$

- In this case we can see that we don't have the lower part of the bracket. Therefore, we will try and calculate it.

$$\sqrt[2]{\frac{(10-1)\cdot 3{,}3^2+(10-1)\cdot 3{,}3^2}{10+10-2}\cdot\sqrt[2]{\frac{1}{10}+\frac{1}{10}}}$$

$$\sqrt[2]{\frac{196{,}02}{18}*\sqrt[2]{\frac{1}{10}+\frac{1}{10}}}=2{,}206$$

Now we can see that we've got the value of 2,206. In this case, if we put the brackets together. It will result the following bracket.

$$\frac{(0{,}7)-0}{2{,}206}\approx 0{,}3173164$$

If we go forward, then we can see that we need to find the critical value. And in this case, we can just take the lower part of the bracket which we found when we calculated the difference between the means of standard deviation.

$$10+10-2=18$$

Now we need to look for the student t's distribution in the table. Because we have got the number eighteen, then we can look under 0,05 in the table with the row 18.

We will get the number 2,1009.

Therefore, we can conclude, that the following interval would become as follows.

$$-2{,}10 < 0{,}31 < 2{,}10$$

3. We would like to know if the concentration of a compound in two brands of yogurt is different. We select 50 bottles of each type. The average concentration in one of the brands is 88.42 wig/L and in the other one is 80.61 ing/L. The standard deviations of the populations are 5.62 nag/L and 4.83 wig/L, respectively. Can we be 95% confident that there is a significant difference among the two brands? What about 99°% confident?

First, we will start by looking after the values which are important in the task.

In this case we can see that we have got the 50 bottles of each type. This shows the sample length. Whereas if look closer towards the difference between the concentrations of the two bottles. Then we can see that the first bottle has 88,42 wig/L and the other one has 80,01 ing/L.

Then we can see that we have got the standard deviations. In this case it is 5,2 and 4,83.

We will follow the same $88{,}42 - 80{,}61 \approx 7{,}81$. We have 0 at the difference between the means of the two populations. Now we need to find the lower part of the bracket and in this case we can use the same method as we did before.

$$\sqrt[2]{\frac{(50-1)\cdot\ 5{,}62^2 + (50-1)\cdot\ 4{,}83^2}{50+50-2}} \cdot \sqrt[2]{\frac{1}{50}+\frac{1}{50}}$$

$$\sqrt[2]{\frac{2690{,}7517}{98}} \cdot \sqrt[2]{\frac{1}{50}+\frac{1}{50}}$$

$$5{,}239 * 0{,}2 = 1{,}0478$$

Now we will put it inside the brackets.

$$\frac{7{,}81 - 0}{1{,}0478} \approx 7{,}453713$$

Now we have to find the critical value.

$$50 + 50 - 2 = 98$$

No, we have to find the values with 0,05 and 98.

- We can see, that out of the two tailed theory - we've got 1,98.

$$1{,}98 < 7{,}45 < 1{,}98$$

- Rejected

And the same method has to be used with 0,01 and 98.

- We can see, that out of the two tailed theory - we've got 2,6269.

$$1{,}98 < 2{,}62 < 1{,}98$$

- Rejected

4. **We want to know whether or not a certain training program is able to increase the maximin long jump of athletes. We recruit a simple random sample of 20 long jump athletes and measure *each* of their maximum long jump. Then, we have each athlete rise the training program for one month and then measure their maximum long jump again at the end of the month. These are the results (below). Does the training program have any effect on the maximum long jump? (use level of significance = 0.05)**

- First, we will start by looking after the average. First, we need to find the average of the two groups, which are maximum long jump before training program. Then the group of maximum long jump after training program. We have done this in Excel.
- We found out, that the standard deviation was in this case 2,56 and 4,013.
- In this case we will use the following formula and insert the following values. But first we will subtract the two averages from each other. The following is this: 4,02-3,87=0,15.

$$\sqrt[2]{\frac{(20-1)\cdot\ 4{,}02^2+(20-1)\cdot\ 3{,}87^2}{20+20-2}}\cdot\sqrt[2]{\frac{1}{20}+\frac{1}{20}}$$

$$\sqrt[2]{\frac{19\cdot\ 4{,}02^2+19\cdot\ 3{,}87^2}{38}}\cdot\sqrt[2]{\frac{1}{20}+\frac{1}{20}}=1{,}24$$

- Now we will put it inside the brackets and get the following answer.

$$\frac{0{,}15-0}{1{,}24}\approx 0{,}1209677$$

- Now we will find the critical value and then we will find the interval.

$$\mathbf{20+20-2=38}$$

- And because we have found the critical value, then we need to use the student t-formula to find out the interval value.
- **The critical value would be as follows, according to the two tailed theory: 2,02.**

$$\mathbf{2,02<0,12<2,02}$$

- We can that the training program, does have an effect, as the max long jump of player is different before and after the participation to the training program.

Athlete	Maximum long jump before training program	Maximum long jump after training program
1	3.7	4.0
2	3.3	3.7
3	3.2	3.2
4	4.0	3.7
5	4.2	4.7
6	4.2	4.3
7	4.7	4.7

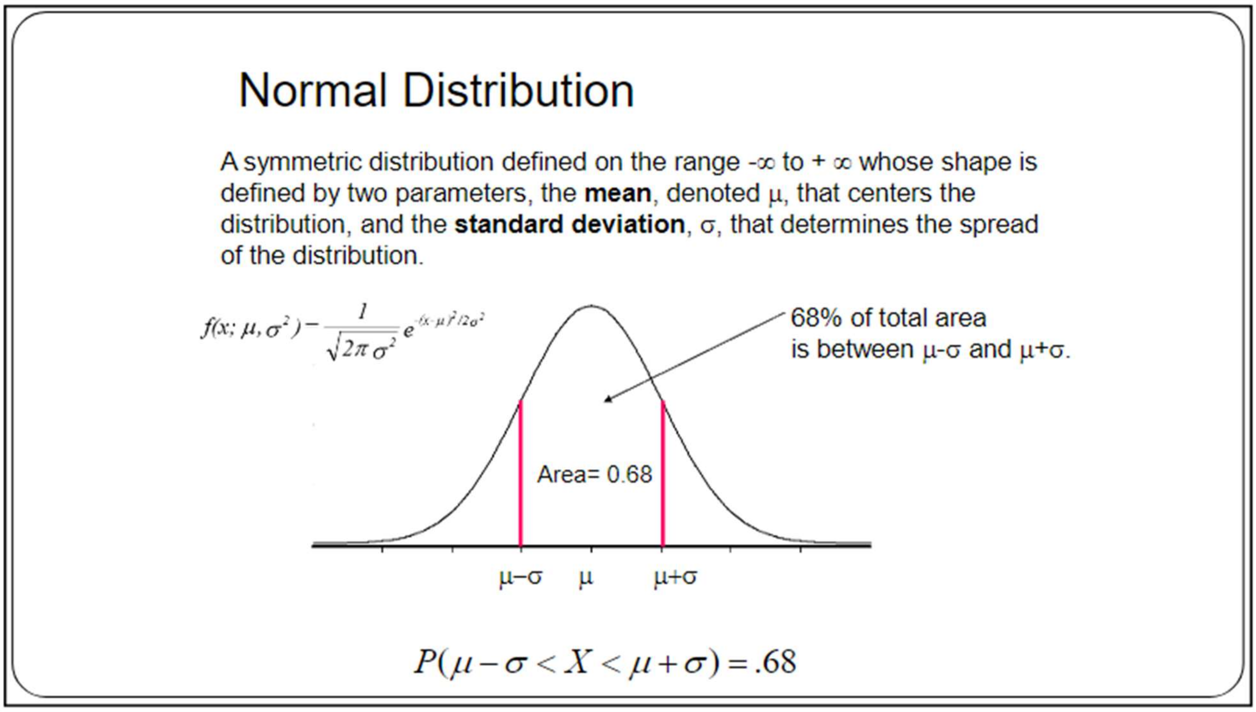
8	3.7	4.0
9	5.0	5.0
10	4.5	4.8
11	4.0	4.2
12	3.0	3.3
13	2.7	2.8
14	3.2	3.0
15	3.2	3.0
16	4.7	4.7
17	4.0	4.3
18	4.2	4.5
19	4.2	4.5
20	3.8	4.0

Hej @Vivek,

På slide 39 (side 18) i Lecture 3 kan du se, at arealet under kurven (area under curve (AUC)) mellem mean + standard deviation og mean - standard deviation fylder 68% af den totale AUC i en normal distribution. Vi får at vide i opgaven, at der er 1000 studerende. Dvs. at der 680 studerende mellem mean + standard deviation og mean - standard deviation. I slidet kan du se, at de sidste 32% må dække de resterende områder, som ligger uden for de røde streger, som er det hvide område i starten af kurven (16%) og enden af kurven (16%).

I opgaven bliver vi spurgt ca. hvor mange elever der har karakter højere end 80 (som kan tolkes 80 ud fra en 100-point skala, altså 80%). Vi ved nu, at mean + standard deviation og mean - standard deviation AUC er 68%, og at AUC for det hvide område i starten af kurven er 16%, som sammenlagt giver 84% AUC. Det vil sige, at der ca. er 840 studerende, som har under 80 i karakter, og at de resterende 160 studerende (16%) må have over 80 (som er det resterende AUC). Altså er svaret 160 studerende.

Spørg endelig hvis du er i tvivl. (edited)



39

$$\sqrt[2]{\frac{48\cdot\ 10{,}75^2 + 48\cdot\ 10{,}75^2}{49+49-2}\cdot\sqrt[2]{\frac{1}{49}+\frac{1}{49}}}$$

$$\sqrt[2]{\frac{48\cdot\ 10{,}75^2 + 48\cdot\ 10{,}75^2}{49+49-2}} = 10{,}75$$

$$\sqrt[2]{\frac{1}{49}+\frac{1}{49}} = \frac{\sqrt{2}}{7} \approx 0{,}2020305$$

$$10{,}75 \cdot 0{,}2020305 \approx 2{,}171828$$

We have to different means.

$$72 - 10{,}25 = 61{,}75$$

$$\frac{61{,}75 - 0}{2{,}17} \approx 28{,}45622$$

$$49 + 49 - 2 = 96$$

$$0{,}2020305$$

-0,8<0,2<0,8