Tutorial Business Analytics

Tutorial 2 - Solution

Exercise 2.1

The consumption per person is measured in index values, where a high index value represents a high consumption. The following table embodies index values for 10 individuals before and after a tax increase.

Individual no. i	Index value previous to the tax increase (a)	Index value after the tax increase (b)	Difference (d=a-b)
1	27	40	**-13**
2	31	36	**-5**
3	23	43	**-20**
4	35	34	**1**
5	26	25	**1**
6	27	41	**-14**
7	26	32	**-6**
8	18	29	**-11**
9	22	21	**1**
10	21	36	**-15**

a.) Determine if there is a significant difference between consumption prior to the tax increase and after, utilizing a hypothesis test (significance level $a = 0.05$). The difference is assumed to be normally distributed.

b.) Check your result applying t.test() in R.

Solution

a.) Use the "test manual" to solve the exercise.

1.) i) 2 samples ii) dependent

2.) $\mu_D = \mu_{before} - \mu_{after}$

$H_0 : m_{before} = m_{after} \quad \Leftrightarrow \quad H_0 : \mu_D = \mu_0 = 0$

$H_1 : \mu_{before} \neq \mu_{after} \quad \Leftrightarrow \quad H_1 : \mu_D \neq \mu_0 = 0$

3.) Paired t-Test

$$t_0 = \frac{\bar{d}-\mu_0}{s_d}\sqrt{n} \qquad \bar{d} = \frac{1}{n}\sum_i d_i \qquad s_d = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}}$$

Average difference $\bar{d}$	Standard deviation of differences (s_d)
-8.1	7.5931

Test statistic $t_0 = \frac{\bar{d}-\mu_0}{s_d}\sqrt{n} = \frac{-8.1-0}{7.5931}\sqrt{10} = -3.3734$

4.) $a = 0.05$

5.) $\rightarrow \; t^c_{1-\frac{\alpha}{2};\, n-1} = t^c_{0.975;\, 9} = 2.262$

6.) Check: $|t_0| = 3.3734 > 2.262 = t^c_{0.975;\, 9}$

$\Rightarrow$ H_0 is rejected

Interpretation: Regarding a significance level of $a = 0.05$ the tax increase does have an effect on consumption.

b.)

```
#use paired=T
a <- c(27,31,23,35,26,27,26,18,22,21)
b <- c(40,36,43,34,25,41,32,29,21,36)
t.test(a,b,alternative = "two.sided", paired=T)

#<=> test if difference is significantly different from zero
d <- c(-13,-5,-20,1,1,-14,-6,-11,1,-15)
t.test(d)
```

$\Rightarrow$ H_0 is rejected

Exercise 2.2

According to the information supplied by the manufacturer of a certain type of car, its gas consumption in city traffic is approximately normally distributed with expected value $\mu = 9.5l/100km$. The standard deviation $\sigma = 2.5l/100km$ is commonly known (to the general public and the manufacturer). In order to review the manufacturers prediction, a consumer organization has performed a test on 25 cars which yielded the following result:

Average gas consumption: $\bar{x} = \frac{10.5l}{100km}$

Check the manufacturers statement with a suitable test at significance level of $a = 0.05$ and a second time with $a = 0.01$.

Solution

Use the “test manual” to solve the exercise.

1.) i) 1 sample ii) σ_X known

2.)

$H_0 : \mu_x = \mu_0 = 9.5$ (information supplied by the manufacturer is correct)
$H_1 : \mu_x \neq \mu_0 = 9.5$ (information supplied by the manufacturer is not correct)

3.) Gauss Test

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma_X}\sqrt{n}$$

$$Z_0 = \frac{10.5 - 9.5}{2.5} \cdot \sqrt{25} = 2$$

4.) a) $a = 0.05$

b) $a = 0.01$

5.) Since it is two-sided test we use $\alpha/2$ to find the critical value :

a) $1 - \alpha/2 = 1 - 0.025 = 0.975;$ $z^c = z_{0.975} \approx 1.96$

b) $1 - \alpha/2 = 1 - 0.005 = 0.995;$ $z^c = z_{0.995} \approx 2.58$

6.) a) $Z_0 = 2 > Z^c = 1.96$ $\Rightarrow$ H_0 is rejected

b) $Z_0 = 2 < Z^c = 2.58$ $\Rightarrow$ H_0 is not rejected

Alternative solution, using the p-value criterion instead of the test statistics criterion:

5.) Calculating p value corresponding to the test statistic:
$\frac{p}{2} = 1 - \phi(z_0) \approx 1 - 0.97725 = 0.02275 \approx 0.023$
(Note: since it is two sided test what we get from the test statistic is p/2)

6.) We compare p/2 with $\alpha/2$:
a) $p/2 \approx 0.023 < 0.025 = \alpha/2$ $\Rightarrow$ $\mathrm{p} < \alpha$ $\Rightarrow$ H_0 is rejected
b) $p/2 \approx 0.023 > 0.005 = \alpha/2$ $\Rightarrow$ $\mathrm{p} > \alpha$ $\Rightarrow$ H_0 is not rejected

Exercise 2.3

During a recent study project, a friend of yours asked 8 men and 10 women how many hours per day they wear a mask during the ongoing COVID 19 pandemic. The following table shows their answers. Afterwards he/she set the hypothesis to "On average, women wear their mask longer per day".

Individual no. i	Hours per day	Gender
1	4	**female**
2	2	**female**
3	3	**female**
4	5	**female**
5	7	**female**
6	2	**female**
7	7	**female**
8	3	**female**
9	5	**female**
10	2	**female**
11	2	**male**
12	1	**male**
13	5	**male**
14	3	**male**
15	1	**male**
16	3	**male**
17	2	**male**
18	3	**male**

a) Test the hypothesis “by hand” with significance level $a = 0.05$ and 16 degrees of freedom.

b) Try to find out how to solve this exercise using R.

Solution

a)

1.) i) 2 samples ii) independent

2.) $H_1: \mu_f > \mu_m$ (on average, women wear their mask longer)

$H_0: \mu_f \le \mu_m$ (on average, women wear their mask shorter or equally long)

⇔ $H_1: \mu_D = \mu_f - \mu_m > \mu_0 = 0$ and $H_0: \mu_D = \mu_f - \mu_m \le \mu_0 = 0$

3.) Welch-Test:

$t_0 = \frac{\overline{x}_f - \overline{x}_m - \mu_0}{s_{\overline{f}-\overline{m}}}$ and $s^2_{\bar{x}-\bar{w}} = \frac{s_f^2}{n_f} + \frac{s_m^2}{n_m}$ (taken from test manual – 3rd step)

$$\overline{x}_f = \frac{4+2+3+5+7+2+7+3+5+2}{10} = 4 \text{ and } \overline{x}_m = \frac{2+1+5+3+1+3+2+3}{8} = 2.5$$

$$s_f^2 = \frac{(4-4)^2+(2-4)^2+(3-4)^2+(5-4)^2+(7-4)^2+(2-4)^2+(7-4)^2+(3-4)^2+(5-4)^2+(2-4)^2}{10-1}$$

$$= \frac{0^2+(-2)^2+(-1)^2+1^2+3^2+(-2)^2+3^2+(-1)^2+1^2+(-2)^2}{9} = 3.778$$

$$s_m^2 = \frac{(-0.5)^2+(-1.5)^2+2.5^2+0.5^2+(-1.5)^2+0.5^2+(-0.5)^2+0.5^2}{7} = 1.714$$

$s^2_{\bar{f}-\bar{m}} = \frac{3.778}{10} + \frac{1.714}{8} = 0.592$ → $s_{\bar{f}-\bar{m}} = 0.769$

$t_0 = \frac{1.5}{0.769} = 1.949$

4.) $a = 0.05$

5.) df=16 (from exercise, but can also be calculated: $df = \frac{\left(s^2_{\bar{f}-\overline{m}}\right)^2}{\frac{s_f^4}{n_f^2(n_f-1)} + \frac{s_m^4}{n_m^2(n_m-1)}}$)

➔ results in $t^c_{0.95;16} = 1.746$ (taken from t-table)

6.) $t_0 = 1.949 > 1.746 = t^c$

⇒ H_0 can be rejected. Regarding a significance level of $a = 0.05$ it can be concluded that on average, women wear their mask longer per day.

b) The same result can be achieved by using R as follows:

```
> female <- c(4,2,3,5,7,2,7,3,5,2)
> male <- c(2,1,5,3,1,3,2,3)
> t.test(female, male, alternative="greater", paired=F)

      Welch Two Sample t-test
data:  female and male
t = 1.9494, df = 15.637, p-value = 0.03471
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.1547037       Inf
```