Chair of Decision Sciences and Systems
TUM School of Computation, Information and Technology
Technical University of Munich

Business Analytics & Machine Learning
Homework sheet 1: Statistics – Solution

Prof. Dr. Martin Bichler, Prof. Dr. Jalal Etesami
Julius Durmann, Markus Ewert, Johannes Knörr, Yutong Chao

Exercise H1.1 *Gas consumption*

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 $\mu = 9.5\,\ell/100km$. The standard deviation $\sigma = 2.5\,\ell/100km$ 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: $\bar{x} = 10.5\,\ell/100km$.

Check the manufacturers statement with a suitable test for significance levels $\alpha = 0.05$ and $\alpha = 0.01$.

Solution

1) i) One sample, ii) σ_X known

2) The null hypothesis is $H_0 : \mu_x = \mu_0 = 9.5$ and states that information supplied by the manufacturer is correct, whereas the alternative hypothesis $H_1 : \mu_x \neq \mu_0 = 9.5$ states that the information supplied by the manufacturer is **not** correct.

3) Gauss test:
$$z_0 = \frac{\bar{x} - \mu_0}{\sigma_0}\sqrt{n} = \frac{10.5 - 9.5}{2.5}\sqrt{25} = 2$$

4) a) $\alpha = 0.05$
   b) $\alpha = 0.01$

5) Since it is two-sided test we use $\alpha/2$ to find the critical value:
   a) $1 - \frac{\alpha}{2} = 1 - 0.025 = 0.975, \quad z^c = z_{0.975} \approx 1.96$
   b) $1 - \frac{\alpha}{2} = 1 - 0.005 = 0.995, \quad z^c = z_{0.995} \approx 2.58$

6) Test decision:
   a) $z_0 = 2 > z^c = 1.96$. Thus, H_0 is rejected.
   b) $z_0 = 2 < z^c = 2.58$. Thus, 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 a two sided test, what we get from the test statistic is $p/2$.)

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

Exercise H1.2 *Caloric intake*

32 individuals take part in a study about nutritional behavior. One aspect of the study is comparing carnivore diets to non-carnivore diets in terms of daily caloric intake. The research hypothesis states, that the daily average caloric intake of individuals following a non-carnivore diet is lower, compared to individuals following a carnivore diet. Out of 32 participants, 12 adhere to a non-carnivore diet, yielding an average caloric intake of $\bar{x}_1 = 1780$ kcal. In contrast, the remaining 20 participants following a carnivore diet average to $\bar{x}_2 = 1900$ kcal per day. The respective estimated standard deviations result in $s_1 = 230$, and $s_2 = 250$. The daily caloric intake of an individual is assumed to be a normally distributed variable.

a) Give a 95% confidence interval of the average daily caloric intake for each of the groups.

b) Which conclusions can be drawn from the computed confidence intervals?

c) Identify and apply a suitable hypothesis test using a significance level of $\alpha = 0.05$.

Solution

a) Group 1:
   $\left[\bar{x}_1 \pm t_{1-0.5\alpha}(n_1 - 1)\frac{s_1}{\sqrt{n_1}}\right] = [1633.86, 1926.14]$

   Group 2:
   $\left[\bar{x}_2 \pm t_{1-0.5\alpha}(n_2 - 1)\frac{s_2}{\sqrt{n_2}}\right] = [1783.00, 2017.00]$

b) The confidence intervals overlap to a great extend, making a decisive inference about the hypothesis not possible. From the given data, it can not be concluded whether the daily caloric intake is depending on whether a carnivore or non-carnivore diet was followed.

c) Since the two groups produced the data indepentently, a suitable test is Welch's t-test. Let the hypothesis H_1 be $(\mu_1 < \mu_2)$. First, compute the suitable degrees of freedom,

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}} = 24.88 \approx 25.$$

   With $\alpha = 0.05$, this yields a critical value of $t^c_{\alpha,df} = -t^c_{1-\alpha,df} = -t^c_{0.95,25} = -1.708$ (see t-table, one-tail). Compute t_0 as

$$t_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \approx \frac{-120}{86.79} \approx -1.38.$$

   H_0 can be rejected, if $t_0 < t^c$. Since this does not hold, H_0 is failed to be rejected. This means, regarding the correctness of H_1, no conclusion can be drawn from the hypothesis test.

Exercise H1.3 *Population mean*

Determine (with $\alpha = 0.05$) if the following sample was obtained from a population with zero mean:

$$2, 3, 2, 4, 2, 4, 5, 2, 1, 4, 3, 0, 3, 2, 4, 5, 3, 3, 0, 1.$$

Solution

1) Single sample with unknown σ_X
2) $H_0 : \mu_x = \mu_0 = 0$
3) t-Test:

$$\bar{X} = 2.65, S_X^2 = 2.134$$
$$\implies \qquad t_{=} \frac{\bar{X} - \mu_0}{S_X}\sqrt{n} = \frac{2.65}{1.461}\sqrt{20} \approx 8.112$$

4) $\alpha = 0.05$
5) $t^c_{1-\frac{\alpha}{2},n-1} = t^c_{0.975,19} = 2.093$ (see t-table)
6) $t^c_{0.975,19} < t_0 \quad \Rightarrow \quad H_0$ is rejected.