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

Business Analytics & Machine Learning
Tutorial sheet 1: Statistics

Prof. Dr. Martin Bichler
Julius Durmann, Markus Ewert, Yutong Chao, Dr. Mete Ahunbay

Exercise T1.1 *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:

$$\text{Average gas consumption: } \bar{x} = 10.5\,\ell/100km.$$

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

Exercise T1.2 *Effect of tax on consumption*

The following table contains data of 10 individuals' consumption levels before and after a tax increase, measured by an index value. High index values correspond to high consumption levels. The rows represent individuals' identifiers i, their index values prior to the tax increase a_i, and after the tax increase b_i.

i	1	2	3	4	5	6	7	8	9	10
a_i	27	31	23	35	26	27	26	18	22	21
b_i	40	36	43	34	25	41	32	29	21	36
$d_i = a_i - b_i$	-13	-5	-20	1	1	-14	-6	-11	1	-15

a) Perform a hypothesis test in order to find out whether there is a significant ($\alpha = 0.05$) difference between consumption levels prior to the tax increase and consumption levels after the tax increase. Assume, that the difference is normally distributed.

b) Verify your result by applying `stats.ttest_rel()` in Python using the *SciPy* package.

Exercise T1.3 *Masks during Covid19*

In the context of the COVID-19 pandemic, 8 children and 10 adults were asked how many hours per day they wear a mask. The following table shows their answers. The hypothesis is "On average, adults wear their mask longer per day than children". It can be assumed, that the average time people wear their mask is normally distributed.

Individual no. (i)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Hours per day	4	2	3	5	7	2	7	3	5	2	2	1	5	3	1	3	2	3
Adult	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N

a) Test the hypothesis "by hand" with a significance level of $\alpha = 0.05$ and 16 degrees of freedom.

b) Search for the corresponding functions in Python and use them to verify your result.