Statistics.
Comments to exercises 2. lecture

Calculations are aided by python.

Exercise [EL] 7

Case of T-distribution:

$$P(T > 2) = 1 - \texttt{tcdf}(2, 2) = 0.0918; \quad P(T > 3) = 1 - \texttt{tcdf}(3, 2) = 0.0477,$$

while normal distribution yields

$$P(U > 2) = 1 - \texttt{normcdf}(2) = 0.0228; \quad P(U > 3) = 1 - \texttt{normcdf}(3) = 0.0013.$$

In conclusion: The T-distribution has thikker tail than the normal distribution.

The T-distribution pops up, when considering normalized residuals in connection with least squares, where the apriory variance is unknown. Apparently, in case of only 2 redundants, a residual exceding 4 times the standard deviation has a probability, which is not negligible.

Since the residual may be both positive and negative, the probability is $2 \times (1 - \texttt{tcdf}(4, 2)) = 2 \times 0.0286 = 0.0572$, i.e. almost 6%, which cannot be neglected - as the normal distribution would suggest.

Increasing the number of degrees of freeedom changes the picture. In case of 10 redundants:

$$P(|T| > 4) = 2 * (1 - \texttt{tcdf}(4, 10)) = 0.0025.$$

Exercise 13

Three cases:

(a)

$$\left[2.7\sqrt{\frac{10}{\texttt{chi2.ppf}(0.975, 10)}}, 2.7\sqrt{\frac{10}{\texttt{chi2.ppf}(0.025, 10)}}\right] = [1.8865, 4.7383],$$

(b)

$$\left[2.7\sqrt{\frac{20}{\texttt{chi2.ppf}(0.975, 20)}}, 2.7\sqrt{\frac{20}{\texttt{chi2.ppf}(0.025, 20)}}\right] = [2.0657, 3.8990],$$

(c)

$$\left[2.7\sqrt{\frac{100}{\texttt{chi2.ppf}(0.975, 100)}}, 2.7\sqrt{\frac{100}{\texttt{chi2.ppf}(0.025, 100)}}\right] = [2.3721, 3.1340].$$

Note the length of the interval - even when the number of redundants is 100.

Exercise 14

Solution in python script file `exe14.py`.

Exercise 15

We obtain $\bar{x} = 1412.818$ og $s = 0.1362$. The calculations associated with (a) and (b) can be found in the python scriptfile `exe15.py`.

(a) 95% interval: $[1412.743, 1412.893]$.
   99% interval: $[1412.713, 1412.923]$.
   Remark, that we use the t-distibution with 14 degrees of freedom, as there are 14 redundants.

(b) 50% interval: $[0.1232, 0.1599]$.
   95% interval: $[0.0997, 0.2149]$.
   The χ^2-distribution with 14 degrees of freedom is needed.

(c) $H_0 : \mu_0 = 1413.00$mm is tested by a T-test, where the aposteriori deviance is $s = 0.1362$. The observered test statistic:

$$t_{\text{obs}} = \frac{1412.818 - 1413}{0.1362/\sqrt{15}} = -5.175.$$

With 14 degrees of freedom and level of significance $\alpha = 5\%$ we obtain the acceptance region:

$$A_{0.05} = [\texttt{t.ppf}(0.025, 14), \texttt{t.ppf}(0.975, 14)] = [-2.1448, 2.1448]. \qquad (1)$$

As t_{obs} is outside A_α, we *reject* the hypothesis at level 5%.

Alternatively, you might consider the 95% confidence interval above: $\mu_0 = 1413$ is outside and we reject.

(d) Similarly:

$$t_{\text{obs}} = \frac{1412.818 - 1412.75}{0.1362\sqrt{1/15}} = 1.934,$$

which is within the acceptance region.

I.e we *accept* the hypothesis for $\alpha = 5\%$, and correspondingly we find that 1412.75 is within the 95% confidence interval for μ.

(e) Use the χ^2-test as $Y = ds^2/\sigma_0^2 = \hat{r}^\top C\hat{r}/\sigma_0^2$ is χ^2-distributed with $d = 15 - 1$ degrees of freedom. We obtain

$$y_{\text{obs}} = 14 * (0.1362)^2/0.08^2 = 40.58.$$

As the alternative is twosided, the acceptance region is:

$$A_{0.05} = [\texttt{chi2.ppf}(0.025, 14), \texttt{chi2.ppf}(0.975, 14)] = [5.63, 26.12], \qquad (2)$$

The hypothesis is rejected for $\alpha = 5\%$, as y_{obs} is not within the acceptance region.

(f) Similar to (e) except that we need a onesided test. The test statistic is

$$y_{\text{obs}} = 14 * (0.1362)^2/0.11^2 = 21.47$$

The onesided acceptance region with significance level 5% corresponds to the 95% quantile of the $\chi^2(14)$-distribution:

$$A_{0.05} = [0, \texttt{chi2.ppf(0.95, 14)}] = [0, 23.68]. \tag{3}$$

i.e. we accept the manufacturers claim.

(g) Use the z-test as $Z = \sqrt{n}(\bar{X} - \mu_0)/\sigma_0$ is standard normal. We obtain

$$z_{\text{obs}} = \sqrt{15} * (1412.818 - 1412.75)/0.11 = 2.39.$$

With level of significance $\alpha = 5\%$ we obtain the acceptance region:

$$A_{0.05} = [\texttt{norm.ppf(0.025)}, \texttt{norm.ppf(0.975)}] = [-1.96, 1.96]. \tag{4}$$

As z_{obs} is outside A_α, we *reject* the hypothesis at level 5%.