Statistics.
Comments to exercises 3. lecture

Calculations are aided by python.

Exercise 16

(a) Data are available for python in `exe16.py`.

Denoting the samples by x_1 and x_2 yields

$$\bar{x}_1 = \texttt{mean(x1)} = 0.28, \quad s_1 = \texttt{std(x1)} = 7.2544$$

and

$$\bar{x}_2 = \texttt{mean(x2)} = -0.46, \quad s_2 = \texttt{std(x2)} = 7.8805.$$

(b) For each sample we test the hypothesis $H_0 : \mu = 0$. As we dont know the variance we refer to the t-test statistics:

$$t_1 = \frac{\bar{x}_1 - 0}{s_1/\sqrt{25}} = \frac{5 \times 0.28}{7.2544} = 0.1930,$$

and

$$t_2 = \frac{\bar{x}_2 - 0}{s_2/\sqrt{25}} = \frac{5 \times 0.46}{7.8805} = -0.2919,$$

In both cases we get an observation within the area of acceptance, when $\alpha = 5\%$:

$$A_{0.05} = [\texttt{t.ppf}(0.025, 24), \texttt{t.ppf}(0.975, 24)] = [-2.0639, 2.0639].$$

In conclusion, there is no evidence of systematic errors.

(c) We test the hypothesis $H_0 : \sigma_1 = \sigma_2$ by an F-test yielding

$$f_{\text{obs}} = \texttt{var(x1)}/\texttt{var(x2)} = 0.8474,$$

and area of acceptance

$$A_{0.05} = [\texttt{f.ppf}(0.025, 24, 24), \texttt{f.ppf}(0.975, 24, 24)] = [0.4407, 2.2693],$$

i.e. we accept that there is no significant difference between the measurement errors.

(d) In case of common variances, we calculate the estimate of the common standard error with 48 degrees of freedom:

$$s_0 = \sqrt{(24s_1^2 + 24s_2^2)/48} = 7.5739.$$

This now enters the test statistic for comparing the means:

$$t_{\text{obs}} = \frac{\bar{x}_1 - \bar{x}_2}{s_0\sqrt{1/25 + 1/25}} = 0.3454,$$

which is also within the acceptance area with $\alpha = 5\%$:

$$A_{0.05} = [\texttt{t.ppf}(0.025, 48), \texttt{t.ppf}(0.975, 48)] = [-2.0106, 2.0106],$$

We conclude that there is no significant difference between the means.

(e) This leaves us with a single sample of size 50 and the test statistic analogous to (b):

$$\bar{x} = \texttt{mean}(\texttt{x}) = -0.09, \quad s = \texttt{std}(\texttt{x}) = 7.5056.$$

with an observed value

$$t = \frac{\bar{x} - 0}{s/\sqrt{50}} = \frac{7.0711 \times -0.09}{7.5056} = -0.0848,$$

which is within the acceptance area:

$$A_{0.05} = [\texttt{t.ppf}(0.025, 49), \texttt{t.ppf}(0.975, 49)] = [-2.0096, 2.0096],$$

(f) 90% confidence interval:

$$\bar{x} \pm s * \texttt{t.ppf}(0.95, 49)/\sqrt{50} = [-1.87, 1.69].$$

Note that this interval actually leaves room for a systematic error of size fex 1 gon.

Exercise 17

(a) As we know the variance, this leaves us with the Z-test as given by

$$Z = \bar{x}/(7.5/\sqrt{50}) = \texttt{sqrt}(50) * \texttt{mean}(\texttt{x})/7.5 = -0.0849$$

and

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

We accept that there is no evidence of a systematic error. But maybe that is due to lack of power?

(b) The power function when the systematic error is of size δ:

$$\begin{aligned}
\beta(\delta) &= 1 - P(-1.96 \leq Z \leq 1.96 \,|\, \mu = \delta) \\
&= 1 - P(-1.96 * 7.5/\sqrt{50} \leq \bar{X} \leq 1.96 * 7.5/\sqrt{50} \,|\, \mu = \delta) \\
&= 1 - \Phi\left(1.96 - \frac{\delta}{7.5/\sqrt{50}}\right) + \Phi\left(-1.96 - \frac{\delta}{7.5/\sqrt{50}}\right) \\
&= 1 - \texttt{norm.cdf}(1.96 - \texttt{sqrt}(50) * \delta/7.5) + \texttt{norm.cdf}(-1.96 - \texttt{sqrt}(50) * \delta/7.5),
\end{aligned}$$

which can be plotted in python (see fx. `exe17.py`). In case of 50 observations, the plot illustrates that the systematic error has to be of size $3 - 4$ gon, before we have a fair chance of detection, as:

$$\beta(3) = 0.8074, \quad \beta(4) = 0.9649.$$

(b) We fix the systematic error to $\delta = 1$and consider the power as a function of sample size:

$$\beta(n) = 1 - \texttt{norm.cdf}(1.96 - \texttt{sqrt}(n) * 1/7.5) + \texttt{norm.cdf}(-1.96 - \texttt{sqrt}(n) * 1/7.5).$$

A plot reveals that (see `exe17.py`) you need around 600 observations in order to detect an error of size 1 gon. More accurately, when $n = 591$ we obtain $\beta(n, 1) = 0.90$.

Exercise 19

Level of significance is chosen to be $\alpha = 5\%$

We need to compare variances, i.e. the F-distribution is into play.

At first, we look at the traditional method and do pairwise comparisons between groups witin the same period.

Groups 1 and 2:

$$F = \frac{11^2}{10^2} = 1.21.$$

Acceptance area with $(145, 142)$ degrees of freedom:

$$A_{0.05} = [\texttt{f.ppf}(0,025, 145, 142), \texttt{f.ppf}(0.975, 145, 142)] = [0.7202, 1.3894].$$

No significant difference, i.e. we have a common variance, which is estimated by

$$\hat{\sigma}^2_{\text{old}} = (145 * 11^2 + 142 * 10^2)/287 = 110.6098$$

such that

$$\hat{\sigma}_{\text{old}} = \sqrt{110.6098} = 10.5171.$$

The estimate has $145 + 142 = 287$ degrees of freedom.

Analogously for Groups 3 and 4:

$$F = \frac{6.1^2}{4.7^2} = 1.6845.$$

and

$$A_{0.05} = [\texttt{f.ppf}(0.025, 71, 69), \texttt{f.ppf}(0.975, 71, 69)] = [0.6240, 1.6051].$$

which means that we reject the hypothesis and we end up with 3 different groups of measurements.

These groups are in turn compared to the photogrammetric method:

$$\begin{aligned} F_1 &= 110.6098/4.7^2 = 5.0072 \\ F_2 &= 6.1^2/4.7^2 = 1.6845 \\ F_2 &= 4.7^2/4.7^2 = 1. \end{aligned}$$

with areas of acceptance:

$$\begin{aligned} A^1_{0.05} &= [\texttt{f.ppf}(0.025, 287, 36), \texttt{f.ppf}(0.975, 287, 36)] = [0.6386, 1.7236] \\ A^2_{0.05} &= [\texttt{f.ppf}(0.025, 71, 36), \texttt{f.ppf}(0.975, 71, 36)] = [0.5785, 1.8252] \\ A^3_{0.05} &= [\texttt{f.ppf}(0.025, 69, 36), \texttt{f.ppf}(0.975, 69, 36)] = [0.5764, 1.8289]. \end{aligned}$$

In conclusion the photogrammetric method is significantly better that the conventional before 1966, whereas there is no significant difference after 1966.