MATH1712 Probability and Statistics II
Homework 7

http://www1.maths.leeds.ac.uk/~voss/2017/MATH1712/
Jochen Voss, J.Voss@leeds.ac.uk
2017/18, semester 2

This exercise sheet will be discussed in the tutorials of the week beginning 12th March.

The main topic of this exercise sheet are hypothesis tests.

During the tutorial.

(1) Discuss the difference between type I and type II errors. Why were we only bounding the type I error when we derived tests in lectures?

(2) The test statistic for the z-test is $|z|$, where

$$z = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{x_i - \mu_0}{\sigma} = \sqrt{n} \frac{\bar{x} - \mu_0}{\sigma}$$

Discussed how z can be found, if data is given in binned form, *e.g.* if we have

range	0–1	1–2	2–3	3–4	4–5
count	17	83	127	55	20.

(3) In lectures we have considered one-sided z-tests. Assume that we have observed samples $x_1, \ldots, x_{n_x}$ and $y_1, \ldots, y_{n_y}$, which we can describe using the model $X_1, \ldots, X_{n_x} \sim \mathcal{N}(\mu_x, \sigma_x^2)$ i.i.d. and $Y_1, \ldots, Y_{n_y} \sim \mathcal{N}(\mu_y, \sigma_y^2)$ i.i.d., where the variances σ_x^2 and σ_y^2 are known. Following the same steps we used for other variants of the z-test, derive tests for the hypotheses $H_0 \colon \mu_x \leq \mu_y$ and $H_0 \colon \mu_x \geq \mu_y$, respectively.

Homework questions. Your solutions to these questions contribute towards your final mark for the module. Please hand in your solutions **to your tutor** via the silver pigeon holes (down the stairs from the maths reception) by **Tuesday, 17th April, 5pm** (after the Easter break).

a) Clearly mark your solution with your name, your student ID, and your tutor's name.
b) Staple the sheets of your solution together. Do not use plastic sleeves *etc.*
c) Write clearly and legibly, and leave margins for the marker to write comments in.
d) Write complete sentences, including correct punctuation.
e) Explain how you obtained your solution. Just giving the final answer is not enough, all intermediate steps are also required.

Exercise 21. For the following four cases, perform statistical tests with significance level $\alpha = 5\%$, testing the hypothesis $H_0 \colon \mu = 0$ against the alternative $H_1 \colon \mu \neq 0$. Show how you perform each test and state the outcome.

a) We have observed 100 independent, normally distributed values with known variance $\sigma^2 = 1$. The average of the observed values is -0.18.

b) We have observed 10 independent samples of a normal distribution with known variance $\sigma^2 = 4$: The observed values are 0.595, 3.923, 0.12, 1.078, 0.578, 1.575, 2.26, -0.031, 1.733, and 1.135.

c) We have observed 10 independent samples from a normal distribution. The sample mean is $\bar{x} = -1.405$, the exact variance is unknown but the sample variance is $\mathrm{s}_x^2 = 1.456$.

d) We have observed the independent samples given in the file

`http://www1.maths.leeds.ac.uk/~voss/2017/MATH1712/ex07-q21d.csv`

Hint: be careful when importing this csv file into R.

Exercise 22. The density of the $\chi^2(\nu)$ distribution is

$$\varphi_\nu(x) = \begin{cases} \dfrac{1}{\Gamma(\frac{\nu}{2})2^{\nu/2}}\, x^{\nu/2-1}\mathrm{e}^{-x/2}, & \text{if } x > 0, \text{ and} \\ 0 & \text{otherwise,} \end{cases}$$

where

$$\Gamma(t) = \int_0^\infty x^{t-1}\mathrm{e}^{-x}\, dx$$

is the gamma function. For every $\nu \in \mathbb{N}$, find the point at which φ_ν has its maximum. Hints: The algebra can be simplified by appropriate use of logarithms. Also, it may be best to treat the cases $\nu = 1$ and $\nu = 2$ separately.

Exercise 23. For the data set about fishing vessels from the practical, perform statistical tests for the following hypotheses at significance level $\alpha = 5\%$:

a) Vessels built from 2008 on have the same engine power as vessels built before 2008.

b) Vessels with a "scallop licence" have the same engine power as vessels without a "scallop licence".

c) Vessels with a "scallop licence" are newer than vessels without a "scallop licence".

d) Vessels with home port Kilkeel (Northern Ireland) have larger engines than vessels with home port Newlyn (Cornwall).

Exercise 24. Let $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ be i.i.d. where μ is known, but σ^2 is unknown. Consider the estimate

$$\tilde\sigma^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$$

for the variance.

a) Show that

$$C := \sum_{i=1}^n \Big(\frac{X_i - \mu}{\sigma}\Big)^2 \sim \chi^2(n).$$

b) For $\alpha \in (0,1)$, denote the α-quantile of $\chi^2(n)$ by $q_n(\alpha)$. Using this notation, find numbers $a_n, b_n \in \mathbb{R}$ such that $P(\tilde\sigma^2 < a_n\sigma^2) = 2.5\%$ and $P(\tilde\sigma^2 > b_n\sigma^2) = 2.5\%$.

c) Using these results, develop a test which can be used to test the hypothesis $H_0\colon \sigma^2 = \sigma_0^2$ against the alternative $H_1\colon \sigma^2 \neq \sigma_0^2$.

d) Assume we have $n = 100$, $\mu = 0$ and we have observed $\tilde\sigma^2 = 4.81$. Test the hypothesis $H_0\colon \sigma^2 = 4$ *vs.* $H_1\colon \sigma^2 \neq 4$ at significance level 5%.