1st Semester, 2021-2022

3rd Year, Math & CompSci

Mathematical Statistics

Seminar Exercises: Week 11

Recap. *Throughout this class,* $X_1, X_2, \ldots, X_n, \ldots$ *will be i.i.d. random variables that follow the distribution of a given characteristic* X*.*

How to do hypothesis testing (finding the rejection region) and significance testing (finding the p-value):

Step 1: Choose a null hypothesis: $H_0 : \theta = \theta_0$*;*

Step 2: Choose the alternative hypothesis, according to whether the test is left-tailed, right tailed, or two-tailed:

$$\begin{aligned} H_1 : \; & \theta < \theta_0 \text{ (left-tailed)} \\ & \theta > \theta_0 \text{ (right-tailed)} \\ & \theta \neq \theta_0 \text{ (two-tailed)} \end{aligned}$$

The decision for whether the test is left-, right- or two-sided is made based on the question and not based on the data;

Step 3: Choose a significance level α*;*

Step 4: Select the proper test statistic TS (this depends on the application or context, see below). We usually denote the test statistic by Z *(particularly if it follows a normal law);*

Parameter θ to be tested	*Known*	*TS (test statistic)*
μ	σ	$Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \in \mathcal{N}(0,1)$
μ		$Z = \frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} \in T(n-1)$
σ		$Z = \frac{(n-1)s^2}{\sigma^2} \in \chi^2(n-1)$
p		$Z = \frac{\overline{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \in \mathcal{N}(0,1)$
$p_1 - p_2$		$Z = \frac{\overline{p}_1-\overline{p}_2-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}} \in \mathcal{N}(0,1)$

Step 5: Calculate $Z_0 = TS(\theta = \theta_0)$*;*

Step 6: (a) For ***hypothesis testing****: Find the rejection region (RR):*

$$RR = \begin{cases} (-\infty, z_\alpha], \text{ for a left-tailed test} \\ [z_{1-\alpha}, \infty), \text{ for a right-tailed test} \\ (-\infty, z_{\frac{\alpha}{2}}] \cup [z_{1-\frac{\alpha}{2}}, \infty), \text{ for a two-tailed test} \end{cases}$$

If $Z_0 \in RR$*, then the hypothesis* H_0 *is rejected. If not, then it is accepted.*

(b) For ***significance testing****: Let* F_Z *be the cdf of the test statistic* Z*. Find the* P*-value:*

$$P = \begin{cases} F_Z(Z_0), \text{ for a left-tailed test} \\ 1 - F_Z(Z_0), \text{ for a right-tailed test} \\ 2 \cdot \min\{F_Z(Z_0), 1 - F_Z(Z_0)\}, \text{ for a two-tailed test} \end{cases}$$

If $P \leq \alpha$*, then the hypothesis* H_0 *is rejected, otherwise it is accepted.*

Exercise 1. Consider the following sample data for the weight (in kg) of the people in a certain city:

$$71.6, 88.7, 92.1, 72.0, 68.2, 79.9, 73.2, 75.3, 86.4, 82.6$$

Assume that the weight is a characteristic that follows the normal distribution. Using two-tailed tests accept or reject with the $\alpha = 5\%$ significance level (risk probability) the following hypotheses:

(a) the mean value of the weight is 75 kg, given that the standard deviation of the weight is 10 kg.

(b) the mean value of the weight is 75 kg, given that the standard deviation of the weight is unknown.

(c) the standard deviation of the weight is 5 kg.

Exercise 2. In a pre-election poll, we are interested in the proportion p_A of people who plan to vote for candidate A against candidate B.

(a) Given that 530 persons out of a random sample of 1000 persons support A, is, at 5% significance level, candidate A favourite to win the elections? Find the corresponding p-value of the test.

(b) Estimate the minimum number n_A of persons that must support A out of a random sample of 900 persons, in order to conclude that candidate A is favourite to win the elections, at 5% significance level.

Exercise 3. A vending machine contains the following numbers of bills:

1 RON	5 RON	10 RON
25	15	10

Assume that each client pays the vending machine with only one bill. At a 5% significance level, test the following hypotheses:

(a) the mean value of the amount of money paid by a client is less than 5 RON.

(b) the standard deviation of the amount of money paid by a client is larger than 5 RON.

(c) the proportion of clients that pay the vending machine with a 5 RON bill is 40%.

Exercise 4. In an orange juice factory, cans are filled by a machine according to the normal distribution.

(a) Consider a random sample of 100 cans that contain a total amount of 24.8 liters of orange juice. Can we conclude, at 5% significance level, that each can is filled in mean with less than 250 ml of juice, given that the standard deviation for the filling machine is 5 ml?

(b) Estimate the minimum number of cans in a sample in order to conclude, at 5% significance level, that each can is filled in mean with at least 250 ml of juice, given that the sample mean is 249 ml and the standard deviation for the filling machine is 5 ml.

(c) Test, at 5% significance level, if the proportion of cans that contain between 249 ml and 251 ml is 70%, given the following sample:

251.2 250.2 249.6 247.2 250.4 250.2 251.4 251.0 250.3 250.4

251.3 250.5 249.1 248.4 251.3 250.5 250.2 251.7 250.0 250.6

250.0 250.1 247.7 249.7 249.2 249.8 249.3 249.5 249.5 249.7