1st Semester, 2021-2022

3rd Year, Math & CompSci

Mathematical Statistics

Seminar Exercises: Week 12

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:

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

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		$Z = \frac{\overline{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \in \mathcal{N}(0,1)$
$p_1 - p_2$		$Z = \frac{\overline{p}_1-\overline{p}_2-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}} \in \mathcal{N}(0,1)$
$\mu_1 - \mu_2$	$\sigma_1,\ \sigma_2$	$Z = \frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \in \mathcal{N}(0,1)$
$\mu_1 - \mu_2$	$\sigma_1 = \sigma_2$ *(unknown)*	$Z = \frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \in T(n_1+n_2-2)$
$\mu_1 - \mu_2$		$Z = \frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \in T(n)$
$\mu_1 - \mu_2$	*paired data*	$Z = \frac{\overline{X}_d-(\mu_1-\mu_2)}{\frac{s_d}{\sqrt{n}}} \in T(n-1)$
$\frac{\sigma_1^2}{\sigma_2^2}$		$Z = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \in F(n_1-1, n_2-1)$

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:

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

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

Additional notation for the TS table: *If we have paired data, that is, two random samples that are not independent $X_{11}, \ldots, X_{1n}$ and $X_{21}, \ldots, X_{2n}$, then with the notation $D_i := X_{1i} - X_{2i}$, $i = \overline{1, n}$, we have the following statistics:*

$$\overline{X}_d = \frac{1}{n} \sum_{i=1}^{n} D_i$$

$$s_d^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \overline{X}_d)^2$$

Exercise 1. A new type of battery for a certain brand of laptop is tested in order to replace the old one. 40 laptops were tested with the old type of battery and a sample mean of 3.9 hours and a sample standard deviation of 0.1 hours were recorded. 30 laptops were tested with the new type of battery and a sample mean of 4 hours and a sample standard deviation of 0.2 hours were recorded. Accept or reject with the 5% significance level the hypothesis that the new type of battery lasts longer (in mean) than the old one, if the corresponding standard deviations are unknown and:

(a) equal;

(b) unequal.

Exercise 2. A new medication for isolated systolic hypertension was tested on a sample of 10 patients. The following are the systolic blood pressures (in mmHg) before and after administration of the drug:

145 147 152 145 156 141 151 148 144 151

140 149 142 144 152 143 145 140 142 150

Assume that the systolic blood pressure is a characteristic that follows the normal distribution. Accept or reject with the 5% significance level the hypothesis that (in mean) the new medication reduces the systolic blood pressure.

Exercise 3. Consider the following random sample data for the height (in cm) of the 10-year-old boys of a city:

$$141, 138, 136, 142, 145, 140, 139, 137, 132, 140$$

Next, consider another independent random sample data for the height (in cm) of the 10-year-old girls of the same city:

$$137, 138, 133, 130, 131, 139, 140, 138, 135, 129$$

Assume the height is a characteristic that follows the normal distribution. Accept or reject with the 5% significance level the following hypotheses:

(a) the mean values of the height of a 10-year-old boy is equal to the mean value of the height of a 10-year-old girl, given that the standard deviation of the height of a boy is 2 cm and of a girl is 6 cm.

(b) the mean values of the height of a 10-year-old boy is larger than the mean value of the height of a 10-year-old girl, given that the standard deviations of the heights of a boy and a girl are equal but unknown.

(c) the mean values of the height of a 10-year-old boy is larger than the mean value of the height of a 10-year-old girl, given that the standard deviations of the heights of a boy and a girl are unknown.

(d) the variance of the height of a 10-year-old boy is smaller than the variance of the height of a 10-year-old girl.