

Statistical Techniques for CS/BIT 2021-1B

Practice test #2

Time: 2hrs 15min

Instructions. This test consists of 7 exercises. The formula sheet and the probability tables are provided. An ordinary calculator is allowed, not a programmable one (GR).

If a question asks “is there a statistical test that ...”, this should be interpreted as “a statistical test *covered in this course* that ...”.

1. Suppose that the profitability (in %) of a random sample of 51 companies in a specific economic sector was determined and the following numerical summaries are given:

- (1) The 5-number-summary is -0.3, 2.1, 2.8, 4.1, 7.5
- (2) The classical numerical summary is (the Standard Error of a statistic is indicated with “ \pm SE”)

Size	Mean	Stand. Dev.	Variance	Skewness	Kurtosis
36	3.0	1.40	1.96	0.51 ± 0.34	2.81 ± 0.48

- a. Determine whether there are outliers according to the $1.5 \times$ IQR-rule.
 - b. Does the classical numerical summary support the assumption of a normal distribution for the profitability of an arbitrary company in the sector? Why (not)?
 - c. Determine a 95%-confidence interval for the expected profitability of a company in the sector.
 - d. Is for the interval in c. the normality assumption strictly necessary? Why (not)?
2. Let T_1, T_2, \dots, T_n be a random sample of size n from a population. Which of the following is a better estimator for the population variance σ^2 ? Motivate your answer with your arguments.

(a) $1/n \sum_{i=1}^n (T_i - \bar{T})^2$

(b) $1/(n-1) \sum_{i=1}^n (T_i - \bar{T})^2$

3. Classify each of the following statements about statistical power as either true or false.
- (a) Power is the ability of a test to identify an effect given that an effect of a certain size exists in a population.
 - (b) Power is associated to the probability of making a Type II error.
 - (c) The power of a test is the probability that a given test is reliable and valid.
 - (d) Power can be used as a tool to determine how large a sample is required.
 - (e) The power of a statistical test depends on the sample size and whether the test is a one- or two-tailed test.
4. A test of $H_0 : \mu = 0$ versus $H_1 : \mu < 0$ is conducted on the same population independently by two separate researchers. They use the same sample size $n=100$ and also the same value of $\alpha = 0.01$ for different samples from the same population. Which of the following is true? (Check all that apply).
- (a) Both researchers have the same observed test statistic.
 - (b) Both researchers have the same decision whether or not to reject the null hypothesis.
 - (c) If the true $\mu = -10$ then both researches will have the same power.
 - (d) Both researchers will have the same p value.
 - (e) Both researchers will test the same assumptions.
5. Do the corona measures affect the student achievement at university level? The programme management of a study programme compared the results (number of completed EC's in a year) of two random samples of first year students, in 2018-2019 (pre-corona) and 2019-2020 (corona). Researchers are interested if the variances for the EC-numbers in the mentioned two years do not differ significantly (with a significance level of 1 %). Below are the relevant results:

	n	Mean	Std. Dev.
2018-2019	13	51.2	7.1
2019-2020	15	48.1	11.4

	Shapiro Wilk	
	Statistic	df
2018-2019	0.795	12
2019-2020	0.927	14

- (a) Which test can be applied to test the equality of variances? Write down the assumptions needed to check the equality of variances.
- (b) Check whether the assumptions you have stated hold. Justify your answers.

6. To test the new design, a group of 8 top athletes were invited for two races: one with the old design shoes and one with the new design shoes. A 3-day-rest between the 2 races was taken into account. 4 racers used the old design and the other 4 first used the new design shoes. The results (times in seconds) are as follows:

Athlete	1	2	3	4	5	6	7	8
Old design shoe	45.38	44.98	46.12	45.52	46.03	44.87	45.66	46.25
New design shoe	45.12	45.05	45.87	45.25	45.91	44.61	45.16	45.98

- (a) Assume there is enough statistical evidence that the above data follows a normal distribution. Are the racers on average faster on the newly designed shoes (according to the above sample)? Conduct an appropriate parametric test with a significance level $\alpha = 0.05$ to answer this question.
- (b) Assume now there is not enough statistical evidence to support the assumption of a normal distribution. Which test can serve as a non-parametric alternative for the test in part a? To answer this question :
- Define for this test: the test statistic and its observed value,
 - Give the hypotheses,
 - Determine the p-value,
 - Draw your conclusion, in words, as compared with $\alpha = 0.05$

$\text{Grade} = 1 + \frac{\# \text{ points}}{33} \times 9$ <p>Rounded to 1 decimal</p>
--

question	1	2	3	4	5	6	Total
points possible	10	3	3	3	7	7	33