Statistical Analysis

Master in Statistics and Information Management

Professor *Ana Cristina Costa* (ccosta@novaims.unl.pt)

HOMEWORK

LU6: HYPOTHESIS TESTING

***Suggestion**: Use the form to select the appropriate formula, and solve the exercises using statistical tables. Then, use Excel to find the exact solutions provided here.*

I) In the following multiple-choice questions, choose only one option.

1. A type I error is

a) Always the same as (1–type II error)

b) The error you make when rejecting the null hypothesis when it is true

c) The error you make when rejecting the alternative hypothesis when it is true

d) The error you make when not rejecting the null hypothesis when it is false

2. A type II error

a) Is typically smaller than the type I error

b) Is the error you make when rejecting the null hypothesis when it is true

c) Is the error you make when not rejecting the null hypothesis when it is false

d) Cannot be calculated when the alternative hypothesis contains an "="

3. The power of the test is

a) Dependent on whether you calculate a t or a Z statistic

b) One minus the probability of committing a type I error

c) One minus the probability of committing a type II error

d) None of the above

4. When studying the performance of the master students, it is expected, on average, that students study less than 2 hours per day. What type of test would you use to validate this claim?

a) A left-sided test for the mean

b) A two-sided test for the mean

c) A right-sided test for the mean

d) None of the above because it would depend on the sample size

5. Consider a hypothesis test with α=1% for the proportion of a given population, which resulted in the decision to reject the null hypothesis. If the significance level is increased to 5%, would you change your decision?

a) Yes, because when the significance level increases, the rejection region decreases.

b) Yes, because when the significance level increases, the acceptance region increases.

c) No, because when the significance level increases, the acceptance region increases.

d) No, because when the significance level increases, the acceptance region decreases.

***Hint**: Graph the rejection region for alpha = 1% and for alpha = 5% considering a one-sided test.*

6. The null hypothesis is strongly rejected when

a) The sample standard deviation is approximately equal to 1.

b) The observed value of the test statistic is approximately equal to 0.

c) The p-value is approximately equal to 1.

d) The p-value is approximately equal to 0.

7. If the p-value is greater than alpha (α) in a two-sided test, what conclusion should be drawn?

a) Do not reject the null hypothesis

b) A higher value of alpha (α) should be used to draw a conclusion.

c) Reject the null hypothesis

d) A one-sided test should be used

8. For a significance level of 10%, what critical value(s) would you use to test whether the mean population value is less than 100, assuming that the sample size is large, and the population standard deviation and the sample mean are known?

a) $z = 1.28$

b) $z = -1.28$

c) $z = 1.645$

d) $z = -1.645$

e) $z_1 = -1.645$ and $z_2 = 1.645$

9. The price of a business class trip follows a Normal distribution with a mean value of 3200 euros. In a sample of 40 trips, the standard deviation of the prices in business class is 10 euros. It is intended to test whether the variance of prices in business class is higher than 50 euros, considering a significance level of 1%. Which of the following statements is correct?

a) The observed statistic of the test is 78 and the critical value is 62.428, so there is evidence that the variance of prices in business class is higher than 50 euros.

b) The observed test statistic is 78 and the critical value is 62.428, so there is no evidence that the variance of prices in business class is higher than 50 euros.

c) The observed test statistic is 78 and the critical value is 21.426, so there is evidence that the variance of prices in business class is higher than 50 euros.

d) The observed test statistic is 78 and the critical value is 21.426, so there is no evidence that the variance of prices in business class is higher than 50 euros.

10. In an experiment involving 7 pairs of paired continuous values, a Pearson correlation coefficient of –0.7 was obtained. For a significance level of 5%, what is the most correct conclusion of the respective two-sided test?

a) Accept the null hypothesis that the two variables are independent

b) Accept the alternative hypothesis that the two variables are linearly correlated

c) Do not reject the null hypothesis that the two variables are independent

d) Reject the null hypothesis in favour of the alternative hypothesis

e) Sufficient data were not provided to calculate the test statistic.

II) Solve the following problems. If it is not stated, consider α=0.05.

1. The telephone operator *Pessimus* wants to control the mean time of telephone conversations during the period from 22h to 24h. To this end, a random sample of 130 customers was selected and a sample mean of 1.1538 minutes was obtained. Suppose that the time of such conversations follows a Normal distribution with a standard deviation of 1 minute. Test, for a significance level of 1%, the hypothesis of mean conversation time to be different from 1 minute.

2. The following data represents the gains in weight, in kilograms, in the first 6 months of life of a group of randomly chosen male children

4,1 4,5 3,6 2,8 3,6 3,2 4,1

Admit that it can be considered that the weight gains follow a Normal distribution.

a) It can be stated that the mean weight gain of male children is significantly (α = 0,05) greater than 3.1 kg?

b) What is the lowest significance level for which you reject the null hypothesis previously tested?

Hints*: This video explains the interpretation of the p-value (not its calculation):* https://www.youtube.com/watch?v=XyVDXRuA9Oc

This video explains its calculation with examples: https://www.youtube.com/watch?v=m6sGjWz2CPg

3. In the first 6 months of life, two random groups of children followed different feeding schemes: the group 1 followed the scheme *A* and the group 2 the scheme *B*. The following table presents the gains in weight, in kg, of those children.

Group 1	2,7	3,2	3,6	4,1	2,7	3,2	4,5	3,6	2,7
Group 2	4,1	4,5	3,6	2,7	3,6	3,2	4,1		

It is known that the children of the two groups had approximately equal weights at birth. Admit that the distribution of the weight gain of each group follows a Normal distribution with variance 0.36 and 0.32, respectively.

a) For a 1% significance level, it can be stated that the average weight gain of children fed according to scheme A is different to that of the children fed according to the scheme B?

b) What is the lowest significance level for which you reject the null hypothesis previously tested?

Hint*: Compute the p-value of the two-sided test of paragraph a)*: p-value = $2 \times P(Z > |Z_{obs}|)$.

4. Having selected some randomly chosen days, the maximum temperature (in °C) was observed in the Arctic and in the Antarctic. The values obtained are presented in the following table:

Arctic	2,7	3,2	3,6	4,1	2,7	3,2	4,5	3,6	2,7
Antarctic	4,1	4,5	3,6	2,7	3,6	3,2	4,1		

Admit that the distribution of maximum temperatures follows a Normal behaviour with equal variances.

a) For a 5% significance level, it could be said that, on average, the maximum temperature in the Arctic is different to that of the Antarctic?

b) What is the lowest significance level for which you reject the null hypothesis previously tested?

5. The *Community of Nations* is interested in studying the difference between the income, in monetary units (m.u.), in two communities of the Group of the Nations. To this end, two random samples were collected, having obtained the following results:

Community	**Size**	**Mean**	**Standard deviation**
Nativa	100	20	2
Vatina	110	18	8

The distributions of the income can be regarded as normal, but with different variances.

a) The Community of Nations suspected that the mean income of the Nativa community is not equal to the Vatina community. Test this hypothesis at a 5% significance level.

b) Calculate the p-value associated to the test of the previous paragraph.

c) Test the hypothesis that the mean income of the Nativa community to be less than that of the Vatina community, for a 1% significance level.

d) Determine the p-value associated with the test of the previous paragraph.

e) At the 2.5% significance level, can it be said that the mean income of the Nativa community is greater than that of the Vatina community?

f) Which is the lowest significance level that allows rejecting the null hypothesis of the previous test?

6. An agricultural enterprise has two types of mowing machines. The productivity of the machines of type A and type B are random variables with normal distributions and variances of 1.5 and 1, respectively. To compare the two types of machines, a wheat-field was divided into 20 longitudinal sections and each of two adjacent sections was treated by each one of the two types of machines. The observed productivities were the following:

Type A machine:	8,0	8,4	8,0	6,4	8,6	7,7	7,7	5,6	5,6	6,2
Type B machine:	5,6	7,4	7,3	6,4	7,5	6,1	6,6	6,0	5,5	5,5

a) Do the two machines have equal productivity? Calculate the p-value of the test. What can you conclude for $\alpha = 0.05$?

b) Suppose now that the variances are unknown, but equal. What would be your answer?

c) Suppose now that the variances are unknown, but different. What would be your answer?

7. A consultant company for human resources training claims that, due to their innovative pedagogical techniques, their trainees increase their qualifications measured by a *standard* exam at least 10 points, on average. A certifying entity has carried out a study to investigate on the reasonableness of the advertising of the company. For this purpose, a random sample of 15 trainees who attended the consultant's course was selected, and their skills were measured before and after attending the course. The scores obtained were the following:

Before	80	82	71	77	75	90	92	81	70	70	80	90	70	60	98
After	82	82	81	79	77	95	95	90	80	82	85	85	83	77	95

Is the consultant company claim true? Calculate the p-value of the test. What can you conclude for α=0.05?

***Hint**: Assume the population of differences has a Normal distribution (this is preferable than using an asymptotic test with a small sample). The application of a non-parametric test would be more appropriate*.

8. According to one of the directors of a large company, the variability of the salaries of the employees without managerial positions has a standard deviation below $\sqrt{10}$ euros. It is known that the salaries of these employees are normally distributed. A random sample of 10 wages under those conditions showed a standard deviation of $\sqrt{15}$ euros. Test the director's statement at a 5% significance level.

9. In the first 6 months of life, two random groups of children followed different feeding schemes: the group 1 followed the scheme *A* and the group 2 the scheme *B*. The following table presents the gains in weight, in kilograms, of those children.

Group 1	2,7	3,2	3,6	4,1	2,7	3,2	4,5	3,6	2,7
Group 2	4,1	4,5	3,6	2,7	3,6	3,2	4,1		

It is known that the children of the two groups had approximately equal weights at birth. Admit that the distribution of the weight gain of each group follows a Normal distribution.

a) Based on a hypotheses test, can you conclude that the variability is different in the two groups at the 5% significance level?

***Hint**: Use Excel to find the critical values.*

b) What would be your decision regarding the previous test if you consider a 1% significance level? Justify your answer without carrying out calculations.

***Hint**: Graph the rejection region for alpha = 5% and for alpha = 1%.*

10. On a poll in a given electoral circle, 1000 people were surveyed and 430 of them were in favor of a particular political party.

a) Is it reasonable to assume that this party has less than 50% of the preferences in that circle?

b) Suppose that sometime later a second survey, independent of the first one, was carried out in the same circle. 1200 people were surveyed and 540 of them said to be favorable to that political party. Does the collected data provide enough evidence of an actual increase of the favorable intentions of the electorate? Draw your conclusion based on the p-value of a statistical test.

11. In a certain city, a random sample of 150 men was collected and 54 of them stated that they watched the news every day.

a) Test if the proportion of men of that city who watch the evening news every day is different from 0.4 at a 10% significance level.

b) Calculate the p-value of the previous test and interpret it.

12. A farmer intends to estimate the total production of his orange grove from his "hunch" about the weight of the fruit of each one of the orange trees. In order to use his estimate, the correlation between the "hunch" *(X)* and the actual weight *(Y)* must be positive. Oranges of a random sample of 25 orange trees in this orange grove were harvested and weighted, and the sample correlation coefficient was 0.85. Assume that the pair (*X*, *Y*) follows a bivariated Normal distribution.

a) Is the correlation between the "hunch" (*X*) and the actual weight (*Y*) significant?

b) Since the estimate of the total production will be better the higher the correlation between X and Y, test if the correlation coefficient is less than 0.9.

13. A sample of 10 university students was collected to compare their grades on a math test in the beginning of the first year at the university (*X*), and their final grades on the *Calculus* course at the end of the first semester (*Y*). The results were the following:

Student	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
X	39	43	21	64	57	47	28	75	34	52
Y	65	78	52	82	92	89	73	98	56	75

a) Is there enough evidence to conclude that the math grades (X) and the Calculus grades (Y) are related?

b) Is there enough evidence to conclude that the correlation coefficient between the math grades (X) and the Calculus grades (Y) is greater than 0.5? Draw your conclusion based on the p-value of a statistical test.

SOLUTIONS

Group I) Multiple choice

1. b
2. c
3. c
4. a
5. d
6. d
7. a
8. b
9. a
10. c

Group II) Problems

Question 1) Observed statistic = 1.75359; Not reject H0

Question 2)

a) Observed statistic = 2.72246; Reject H0

b) 0.0173

Question 3)

a) Observed statistic = –1.08976; Not reject H0

b) 0.2758

Question 4)

a) Observed statistic = −1.006; Not reject H0

b) 0.3317

Question 5)

a) Observed statistic = 2.5363; Reject H0

b) 0.0125

c) Not reject H0

d) 0.9938

e) Reject H0

f) 0.0062

Question 6)

a) Observed statistic = 1.66; p-value = 0.097

b) Observed statistic = 1.881994; p-value = 0.0761

c) Observed statistic = 1.881994; p-value = 0.0794

Question 7) Observed statistic = −2.82773; p-value = 0.9933

Question 8) Observed statistic = 13.5; Not reject H0

Question 9)

a) Observed statistic = 1.1173; Not reject H0

Question 10)

a) Observed statistic = −4.4272; Reject H0

b) Observed statistic = −0.94079; p-value = 0.1734

Question 11)

a) Observed statistic = –1; Not reject H0

b) p-value = 0.3173

Question 12)

a) Observed statistic = 7.7384; Reject H0

b) Observed statistic = −1.0134; Not reject H0

Question 13)

a) r = 0.8398; Observed statistic = 4.375; Reject H0

b) Observed statistic = 1.775671; p-value = 0.0379