HOMEWORK

LU8: NONPARAMETRIC TESTING

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

1. The store Great Sales has verified that in the last few years 15% of its customers pay for their purchases by check, 38% with credit card, 32% with a debit card, and 15% in cash. A sample of 160 sales in the week preceding Christmas revealed the following results:

Type of payment	Check	Credit card	Debit card	Cash
Number of sales	27	65	48	20

Does the type of payment used during the Christmas season agree with the information that the store has? Formulate the hypothesis of the test and provide the conclusion based on the Chi-square test at the 10% significance level.

Hint: use Excel to compute the critical value for the 10% significance level.

2. Certain insurance company bases its premium's system for a certain risk on the premise that the number of casualties per insurance policy has Poisson distribution of parameter λ=0.2. A sample of 1000 policies was collected and the following data were obtained:

Number of casualties	0	1	2	≥ 3
Number of insurance policies	800	175	21	4

a) Does the sample data agree with the premise of the insurance company? Formulate the null hypothesis of the test.

b) Calculate the *p-value* of the Chi-square test and conclude.

Hint: solve using Excel, and group the last two classes to ensure that $n\pi_i \geq 5$.

3. The following results refer to 250 measurements of total airborne particles ($\mu g/m^3$) in a public garden:

Particles ($\mu g/m^3$)	[0, 12[	[12, 14[	[14, 16[	[16, 18[	[18, 20[	[20, 22[	[22, ∞[
Observed Frequency	7	22	55	81	53	24	8

Using the Chi-square test at the 10% significance level, test the hypothesis of the sample data to come from a population with

a) Normal distribution (μ=18, σ=3.5).

b) Normal distribution.

Hint: Use the following formulas to compute the sample mean and sample variance for grouped data:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{k} n_i C_i \; ; \; S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k} n_i (C_i - \bar{X})^2}$$

where n_i is the absolute frequency of interval i, and C_i is the midpoint of interval i. Consider 23 as the midpoint of the last interval.

4. In a real-estate market, pay rates at 12 months of 20 randomly chosen shares were the following:

4	4.3	4.7	5	5.2	5.5	5.8	6.1	6.4	6.7
7	7.3	7.4	8	8.3	8.5	8.6	9.1	9.2	9.5

Test the hypothesis that the data come from a normal distribution using the Shapiro-Wilk test.

5. Fifteen students received the following scores for enrolling at the university.

395	481	515	525	540	562	580	584
596	598	615	618	620	642	740	

Test the hypothesis that the data come from a normal distribution using the Shapiro-Wilk test.

6. At the beginning of the year, an analyst of a financial market has established a list of shares to buy and a list of shares to sell. Independent random samples of shares of each one of the lists had annual returns (%) that are listed in the following table. Test the hypothesis that the shares to buy had higher annual returns than the shares to sell.

Shares to buy	9.6	5.8	13.8	17.2	11.6	4.2	3.1	11.7	13.9	12.3
Shares to sell	–2.7	6.2	8.9	11.3	2.1	3.9	–2.4	1.3	7.9	10.2

7. The Director of the Human Resources department of a company believes that the operators of a Call-Centre with training in social skills leave a more favourable impression on customers than operators without this type of training. In a group of 22 operators, the impression of friendliness registered by 22

customers after a control call was evaluated. The degree of friendliness, assessed on an ordinal scale with 5 points (1 – not friendly; …; 5 – very friendly) for each operator is registered in the following table. Is the Director's believe correct?

Control group	Group with training in social skills
1	2
2	4
3	3
3	3
2	3
4	5
3	4
3	3
3	3
2	4
3	
2	

8. Six students took a certain diet with the perspective of losing weight, and the results are listed in the following table. Is this diet an effective way to reduce weight?

Student	Weight before	Weight after
A	174	165
B	191	186
C	188	183
D	182	178
E	201	203
F	188	181

9. A consultant for training human resources claims that, due to her innovative pedagogical techniques, their trainees increase on average their qualifications by at least 10 points measured by a standard test. A certifying entity has carried out a study in order 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 courses was chosen, and the trainees' skills were measured before and after attending the course. The obtained scores are in the following table. Is the consultant right?

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

10. It was requested to a market research company to assess the reason for the success of some mobile phone models. To this end, some university students were asked to give their opinion on 7 models on a scale of 1 to 5 (1-very poor, 2-poor, 3-average, 4-good, 5-very good). The number of mobile phones from these models that were sold was also recorded. The following table presents the results. Verify if there is a significant relationship between the opinion on the mobile phone models and the quantity sold.

Model	Opinion	Sales
A	1	475
B	2	500
C	2	420
D	3	650
E	4	920
F	5	1100
G	5	1050

11. It was requested to a market research company to assess the reason for the success of some laundry detergents. To this end, some housewives were asked their opinion about 7 laundry detergents (1-very poor, 2-poor, 3-average, 4-good, 5-very good). The quantity sold, in tons per year, of these detergents was also collected. The following table presents the results. Verify if there is a significant relationship between the opinion on the laundry detergents and the quantity sold.

Detergent	Opinion	Sales
A	1	9
B	2	9.2
C	2	9.1
D	3	9.5
E	4	9.6
F	5	9.8
G	5	9.7

SOLUTIONS

Question 1) Let A1= check; A2= credit card; A3= debit card; A4=cash

H_0: P(A1)=p1=0.15 , P(A2)=p2=0.38 , P(A3)=p3=0.32 , P(A4)=p4=0.15

H_1: At least one P(Ai)≠pi , i=1,2,3,4

Q_{obs}= 1.532; p-value = 0.6750

Question 2)

a) H_0: X~P(λ=0.2)

b) 0.1112

Question 3)

a) Q_{obs}= 44.76; p-value = 0.0000

b) Average = 16.9; standard deviation = 3.04; Q_{obs}= 10.6; p-value = 0.0314

Question 4) W_{obs}= 0.9555 > $W_{(20, 0.05)}$= 0.905 $\Rightarrow$ do not reject H_0 at the de 5% significance level

Question 5) W_{obs}= 0.9574 > $W_{(15, 0.05)}$= 0.881 $\Rightarrow$ do not reject H_0 at the 5% significance level

Question 6) W^*_{obs}= 137 > $W_{(10, 10; 0.95)}$= 127 $\Rightarrow$ reject H_0 at the 5% significance level

Question 7)

If X – Group with training: H_0: $F_X(x)$ =$F_Y(x)$ $\forall$x; H_1: $\exists$x $F_X(x)$ > $F_Y(x)$
W'^*_{obs}= 2.099 > Z(0.95) = 1.645 $\Rightarrow$ reject H_0 at the 5% significance level

If X – Control Group: H_0: $F_X(x)$ =$F_Y(x)$ $\forall$x; H_1: $\exists$x $F_X(x)$ < $F_Y(x)$
W'^*_{obs}= −2.099 < Z(0.05) = −1.645 $\Rightarrow$ reject H_0 at the 5% significance level

Question 8) W_{obs}= 20 > W(6;0.95)=18 $\Rightarrow$ reject H_0 at the 5% significance level

Question 9) Solution computed for H_1: E[D] < 0 with D=X_{before}−X_{after}:
Considering few ties: $W+_{obs}$ = 11.5 < W(14; 0.05) = 26 $\Rightarrow$ reject H_0 at the 5% significance level
Considering many ties: Z_{obs} = −2.58 < Z(0.05) = −1.645 $\Rightarrow$ reject H_0 at the 5% significance level

Alternatively, we could also consider H_1: E[D] > 0 with D=X_{after}r−X_{before}

Question 10) ρ_{obs} = 0.9286 > W(7; 0.025) = 0.786 $\Rightarrow$ reject H_0 (p-value=0.0229)

Question 11) ρ_{obs} = 0.9821 > W(7; 0.025) = 0.786 $\Rightarrow$ reject H_0 (p-value=0.0161)