
Group:**Name:****Problem 1 (25 pts)**

The table below shows data on the average percentage occupancy of hotels with respect to rating and distance from the city center.

	2 stars	3 stars	4 stars	5 stars
<3 km	92	98	94	89
3-5 km	96	89	87	87
>5 km	98	74	78	73

- (a) **(10 pts)** Fulfill the ANOVA table.
- (b) **(3 pts)** Is there a significant difference in occupancy between the hotels of different ratings? What about hotels with different distances from the center?
- (c) **(5 pts)** Construct a 90% confidence interval for the difference of mean occupancies between the three star and four star hotel.
- (d) **(7 pts)** Construct 95% simultaneous confidence intervals for the difference between mean occupancies of two star and three star and the difference between mean occupancies of three star and four star hotels.

Group:**Name:****Problem 2 (20 pts)**

Below is the data on the desire for career growth for 283 random people of different professions.

	desire for career growth		
	Yes	No	Total
White-collar workers	47	7	54
Skilled workers	49	30	79
Unskilled workers	65	85	150
Total	161	122	283

- (a) **(3 pts)** Based on the data in the table, and without conducting a significance test, describe the differences of career growth desire for different occupations.
- (b) **(10 pts)** Calculate the χ^2 statistic and use it to test for independence, using a 5% significance level. What do you conclude?
- (c) **(7 pts)** Two shooters fired 50 shots at the target. The first shooter hit the target 41 times, and the second – 36. Is it possible to say at the significance level of 0.1 that the first shooter is more accurate?

Group:**Name:****Problem 3 (30 pts)**

11 runners participated in a competition. Below is the data on the heights of the runners and their results in the race.

Runner	1	2	3	4	5	6	7	8	9	10	11
Height (cm)	192	186	182	179	178	176	175	173	169	168	165
Result (sec)	10.7	10.4	10.2	10.3	10.1	11.0	10.6	10.8	10.9	10.5	11.3

- (a) **(2 pts)** Graph the scatter plot and comment the type of the dependence.
- (b) **(4 pts)** Calculate the sample correlation coefficient.
- (c) **(3 pts)** Test the hypothesis of zero correlation at $\alpha = 0.05$.
- (d) **(8 pts)** Find Spearman's rank correlation and test its significance $\alpha = 0.05$.
- (e) **(6 pts)** Using the OLS method estimate a linear regression coefficients and graph the regression line.
- (f) **(7 pts)** Find the determination coefficient and comment on your findings.

Group:**Name:****Problem 4 (20 pts)**(a) **(9 pts)**

Below are the data on the yield (in hundredweight per hectare) when using two methods of pre-sowing seed treatment. Is it possible to assume at the significance level of 0.05 that the first method gives better productivity? Solve the problem under the assumption that the yield has a normal distribution.

Method I	20.0	17.9	20.6	22.0	21.4	23.8	21.4	18.4	
Method II	22.1	18.5	19.4	22.1	21.7	24.9	21.6	20.3	18.3

(b) **(9 pts)**

The psychologist has the hypothesis that the time for solving equivalent "game of 5" problems (i.e. having the same solution algorithm) will significantly decrease as a result of learning. To test the hypothesis the solution time (in minutes) of the first and third problems were measured for eight individuals. At 5% significance level can we claim that the observations are in favour of the psychologist's hypothesis?

Individual	1	2	3	4	5	6	7	8
1st problem	4.0	3.5	4.1	5.5	4.6	6.0	5.1	4.3
2nd problem	3.0	3.0	3.8	2.1	4.9	5.3	3.1	2.7

(c) **(2 pts)**

Explain the main difference between the problems in (a) and (b).

(d) **(BONUS 1 pt)**

Education helps us to develop the mind. Well, you get a bonus if you write a piece of any poem. Could you remember at least one?