STATISTICS PART
Instructor: Dr. Samir Safi

Name:_____ ID Number:_____

Question #1: (20 Points)

For each of the situations described below, state the sample(s) type the statistical technique that you believe is the most applicable.
For example: Two independent samples - t test.

1. As part of an attitude survey, a sample of men and women are asked to rate a number of statements on a scale of 1 to 5, according to whether they agree or disagree. We wish to determine whether there is a significant difference between the answers of men and women.
Answer: Two independent samples === Mann Whitney test.

2. Investors use many "indicators" in their attempts to predict the behavior of the stock market. One of these is the "January indicator." Some investors believe that if the market is up in January, then it will be up for the rest of the year. We wish to determine if there is a relationship between the market's direction in January and the market's direction the rest of the year.
Answer: Chi Square.

3. Bastien, Inc. has been manufacturing small automobiles that have averaged 50 miles per gallon of gasoline in highway driving. The company has developed a more efficient engine for its small cars and now advertises that its new small cars average more than 50 miles per gallon in highway driving. An independent testing service road-tested 25 of the automobiles. We wish to determine whether or not the manufacturer's advertising campaign is legitimate. *Assume the data is normally distributed*
Answer: One sample__t - test.

4. A large corporation wants to determine whether or not the “typing efficiency” course given at a local college can increase the typing speeds of its word processing personnel. A sample of 6 typists is selected, and are sent to take the course. We wish to test to see if it can be concluded that taking the course will actually increase the average typing speeds of the typists. *Assume the data is normally distributed.*
Answer: Paired Samples T test

5. One company hires employees for its management staff from three local colleges. The company's personnel has been collecting and reviewing annual performance ratings in an attempt to determine if there are differences in performance among the managers hired from these colleges. Performance-rating data are available from independent samples seven employees from college A, six employees from college B, and seven employees from college C. We wish to determine whether the three populations are identical with respect to performance evaluations.
Answer: Independent Samples – Kruskal - Wallis

Question #2: (10 Points)

The management of a chain electronic store would like to develop a model for predicting the weekly sales (in thousand of dollars) for individual stores based on the number of customers who made purchases. A random sample of 12 stores yields was used . **Using $\alpha = 0.05$ and the SPSS output, answer each of the following:**

(a) (1 Points) Write the estimated regression equation of to predict the weekly sales by the number of customers.

(b) (3 Points) Interpret the values of the estimated intercept and slope

(c) (3 Points) What is the value of the coefficient of determination? State your full interpretation.

(d) (3 Points) Test whether the number of customers is statistically significant at 0.05 level

SPSS Output for question #2

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.972[a]	.945	.940	.41905

a. Predictors: (Constant), Number of Customers

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.446	.558		2.591	.027
	Number of Customers	.010	.001	.972	13.146	.000

a. Dependent Variable: Sales (Thousands of Dollars)

Question #3: (14 Points)

A realtor wants to compare the average sales-to-appraisal ratios of residential properties sold in four neighborhoods (A, B, C, and D). Four properties are randomly selected from each neighborhood and the ratios recorded for each. **Using $\alpha = 0.05$ and the SPSS output, answer each of the following:**

a. (2 Points) Using Kolmogorov-Smirnov test, determine whether the normality assumption is satisfied for each neighborhood .

b. (2 Points) Using Levene test, determine whether the Homogeneity of Variances assumption is satisfied.

c. (4 Points) Conduct an ANOVA test of the null hypothesis that the average sales-to-appraisal ratios of residential properties sold in four neighborhoods are the same.

d. (6 Points) Discuss all the multiple comparisons.

SPSS Output for question #3

One-Sample Kolmogorov-Smirnov Test

neighborhoods		Sales
Neighborhoods A	Kolmogorov-Smirnov Z	.500
	Asymp. Sig. (2-tailed)	.964
Neighborhoods B	Kolmogorov-Smirnov Z	.343
	Asymp. Sig. (2-tailed)	1.000
Neighborhoods C	Kolmogorov-Smirnov Z	.427
	Asymp. Sig. (2-tailed)	.993
Neighborhoods D	Kolmogorov-Smirnov Z	.475
	Asymp. Sig. (2-tailed)	.978

Test of Homogeneity of Variances

Sales

Levene Statistic	df1	df2	Sig.
.287	3	12	.834

ANOVA

Sales

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.182	3	1.061	10.763	.001
Within Groups	1.183	12	.099		
Total	4.364	15			

Multiple Comparisons

Dependent Variable: Sales

Bonferroni

(I) neighborhoods	(J) neighborhoods	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Neighborhoods A	Neighborhoods B	-1.12500*	.22197	.002	-1.8248	-.4252
	Neighborhoods C	-.32500	.22197	1.000	-1.0248	.3748
	Neighborhoods D	-.07500	.22197	1.000	-.7748	.6248
Neighborhoods B	Neighborhoods A	1.12500*	.22197	.002	.4252	1.8248
	Neighborhoods C	.80000*	.22197	.022	.1002	1.4998
	Neighborhoods D	1.05000*	.22197	.003	.3502	1.7498
Neighborhoods C	Neighborhoods A	.32500	.22197	1.000	-.3748	1.0248
	Neighborhoods B	-.80000*	.22197	.022	-1.4998	-.1002
	Neighborhoods D	.25000	.22197	1.000	-.4498	.9498
Neighborhoods D	Neighborhoods A	.07500	.22197	1.000	-.6248	.7748
	Neighborhoods B	-1.05000*	.22197	.003	-1.7498	-.3502
	Neighborhoods C	-.25000	.22197	1.000	-.9498	.4498

*. The mean difference is significant at the .05 level.

Question #4: (6 Points)

A personnel director for large, research- oriented firm categorizes colleges and graduates. The director collects data on 156 recent graduates, and has each rated supervisor.

	Rating		
School	**Outstanding**	**Average**	**Poor**
Most desirable	21	25	2
Good	20	35	10
Adequate	4	14	7
Undesirable	3	8	6

Using $\alpha = 0.05$ and the SPSS output, answer each of the following:

a. (2 Points) Is there any problem in using the χ^2 approximation?

b. (4 Points) Can the director conclude that there is a relation between school type and rating? Explain.

SPSS Output for question #4

SCHOOL * RATING Crosstabulation

			RATING			
			Outstanding	Average	Poor	Total
SCHOOL	Most Desirable	Count	21	25	2	48
		Expected Count	14.8	25.5	7.7	48.0
	Good	Count	20	36	10	66
		Expected Count	20.3	35.1	10.6	66.0
	Adequate	Count	4	14	7	25
		Expected Count	7.7	13.3	4.0	25.0
	Undesirable	Count	3	8	6	17
		Expected Count	5.2	9.0	2.7	17.0
Total		Count	48	83	25	156
		Expected Count	48.0	83.0	25.0	156.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	15.967[a]	6	.014
Likelihood Ratio	16.577	6	.011
Linear-by-Linear Association	13.934	1	.000
N of Valid Cases	156		

a. 2 cells (16.7%) have expected count less than 5. The minimum expected count is 2.72.