Eastern University, Sri Lanka

Faculty of Commerce and Management

Third Year-Second Semester Examination in Bachelor of Business Administration Honours/ Bachelor of Business Administration Honours in HRM/ Bachelor of Business Administration Honours in MKT MGT/Bachelor of Commerce Honours/ Bachelor of Commerce Honours in ACC & FIN/ Bachelor of Commerce Honours in Business Economics - 2021/2022 (September 2024) (Proper/Repeat)

COM 3063 Statistical Software Applications in Business

Answer All Questions. **Time: 03 Hours**

01. **Write the letter that corresponds to the most correct answer in the given answer script book.**

(i) In SPSS for Windows, which menu option would you use to create a new data set?

a) File > Open
b) File > Data
c) File > New > Data
d) File > Open > Data

(ii) What is the process of selecting cases based on certain conditions in SPSS?

a) Data cleaning
b) Data visualization
c) Data Filtering
d) Data Transformation

(iii) What does the '.sav' file extension in SPSS represent?

a) A syntex file with command instructions
b) A data set containing data and variable definitions
c) An output file with results of analysis
d) A saved graphical representation

(iv) Which of the following is false about the use of the 'COMPUTE' command in SPSS?

a) Sorting cases based on the values of a new variable
b) Creating a new variable that is the average of two existing variables
c) Generating a new variable based on conditional statements
d) Transforming values from one variable into a new variable

(v) What does happen to the data set in the Data View after 'Split File' command has been applied?

a) The data set remains unchanged
b) The data set is sorted in ascending order based on the grouping variable
c) The data set is sorted in descending order based on the grouping variable
d) The data set is merged with other data sets

(vi) Which option is best for computing Cronbach's Alpha in SPSS?

a) Analyze > Descriptive Statistics > Frequencies
b) Analyze > Dimension Reduction > Factor
c) Analyze > Correlate > Bivariate
d) Analyze > Scale > Reliability Analysis

(vii) Which SPSS menu path is used to access the Chart Builder, a tool for creating various types of graphs?

a) Analyze > Graphs > Chart Builder
b) Data > Graphs > Chart Builder
c) Graphs > Chart Builder
d) Graphs > Legacy Dialogs > Chart Builder

(viii) Which option in the SPSS 'Graphs' menu allows you to create a bar chart for categorical data?

a) Graphs > Chart Builder > Bar
b) Graphs > Legacy Dialogs > Bar
c) Graphs > Chart Builder > Histogram
d) Graphs > Legacy Dialogs > Histogram

(ix) Which of the following allows you to customize the appearance of your graphs, including adding titles, changing colours, and modifying axes?

a) Data Editor
b) Output Viewer
c) Syntax Editor
d) Chart Editor

(x) Which of the following menus in SPSS contains the submenu "Options" that allows check and modify, where necessary, the options that IBM SPSS Statistics uses to display the data and the output that is produced.

a) Edit
b) File
c) Analyze
d) Transform

(xi) In SPSS, which test is appropriate for comparing more than two independent groups when the assumption of normality is not met?

a) One-way ANOVA
b) Friedman Test
c) Kruskal – wallis Test
d) Chi-Square test

(xii) In SPSS, what is the purpose of performing a Direct Oblimin Rotation in Factor analysis?

a) To simplify the interpretation of factors by making the loadings more distinct
b) To increase the reliability of the factors
c) To select the initial factor extraction method
d) To determine the correct number of factors to extract

(xiii) In SPSS, which of the following is used to diagnose multicollinearity in a multiple regressior analysis?

a) Durbin-Watson Test
b) Shapiro-Wilk Test
c) Variance Inflation Factor
d) Levene's Test

(xiv) Which statistic in SPSS output of a multiple regression analysis indicates how well the overall model fits the data?

a) t-statistic
b) Standard error of the estimate
c) Beta Coefficients
d) Adjusted R-squared

(xv) Which SPSS add-on is commonly used to perform simple mediation analysis?

a) PROCESS Macro
b) AMOS
c) Smart PLS
d) LISREL

(xvi) Which of the following steps is not required when setting up a mediation analysis in SPSS using the PROCESS macro

a) Specify the independent variable (IV), mediator (M), and dependent variable (DV)
b) Choose the number of bootstrap samples for the analysis
c) Calculate the mean of each variable
d) Select the model number for mediation analysis

(xvii) In SPSS, which of the following methods is commonly performed to assess interaction effect between two predictors?

a) Simple Linear Regression analysis
b) Hierarchical Regression analysis
c) Stepwise Regression analysis
d) Principle Component analysis

(xviii) In a simple moderation analysis, how is the interaction term typically created in SPSS?

a) By multiplying the independent variable by the moderator variable
b) By multiplying the independent variable by the dependent variable
c) By adding the independent variable and moderator variable
d) By mean centring the dependent variable

(xix) Which SPSS procedure is commonly used to assess the normality of a distribution of a continuous variable?

a) Descriptives
b) Frequencies
c) Linear
d) Explore

(xx) What does the Durbin-Watson test in SPSS assess in the context of regression analysis?

a) Normality of residuals
b) Multicollinearity among predictors
c) Independence of residuals
d) Linearity of the relationship between independent and dependent variables

(Total Marks 25)

02. **Write the appropriate terms in your answer script that fill the blanks correctly in the sentences given under the table using the terms given in the following table:**

Staked-bar chart	Chi-square test	Heteroscedastic	Effect size
Kolmogorov-Smirnov test	Beta coefficient	Post-hoc test	Regression coefficient
Reliability analysis	Shapiro-Wilk test	Part correlation	Eta-squared
Histogram	Homoscedastic	Kurtosis	Skewness
Residual	Scatterplot	Communality	Eigen value
Partial correlation coefficient	Correlation coefficient	Multiple correlation	Partial regression plot

a) ..... is used to make a visual comparison to the normal distribution.
b) In multiple regression analysis, ... is a graphical representation that depicts the relationship between the dependent variable and a single independent variable when the effects of other independent variables in the model are held constant.
c) ..... measures the peakedness or flatness of a distribution when compared with a normal distribution.
d) When the variance of error terms appears constant over a range of predictor variables, the data are said to be .....
e) In regression analysis, ... represent the unexplained portion of the dependent variable.
f) ..... represents the relationship between two metric variables portraying the joint values of each observation in a two-dimensional graph.
g) In exploratory factor analysis, ... represents the amount of variance accounted by a factor.
h) In multiple regression analysis, ... is a standardized regression coefficient that allows for a direct comparison between coefficients as to their relative explanatory power of the dependent variable.
i) ..... is a numerical value that measures the strength of the relationship between a dependent variable and a single independent variable when the predictive effects of the other independent variables in the model are removed.
j) The statistical significance test used to assess the normality of the data when the sample size is above 50 is .....

(Total Marks 10)

03. A study was conducted to evaluate whether a new training program has a significant effect on employee productivity. Productivity scores were collected before and after the training for 10 employees. The data were analyzed using paired samples t-test with IBM SPSS Statistics. The outputs of SPSS are given below.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Productivity score before training	77.00	10	7.513	2.376
	Productivity score after training	81.70	10	6.567	2.077

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Productivity score before training & Productivity score after training	10	.926	<.001

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Productivity score before training - Productivity score after training	-4.700	2.869	.907	-6.753	-2.647	-5.180	9	<.001

Paired Samples Effect Sizes

			Standardizer[a]	Point Estimate	95% Confidence Interval Lower	Upper
Pair 1	Productivity score before training - Productivity score after training	Cohen's d	2.869	-1.638	-2.589	-.652
		Hedges' correction	2.996	-1.569	-2.479	-.624

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

a) Formulate the null and alternative hypotheses for the above scenario.
b) State the assumptions of this test.
c) What is the average productivity score before training?
d) What is the average change in productivity? Interpret it.
e) Whether does new training program have a significant effect on employee's productivity? Justify your answer.
f) Interpret the results in the "Paired Samples Correlations".
g) Whether does the new training program practically improves productivity? Justify your answer
h) Write-up the conclusion of the test.

(Total Marks 20)

04. A company wants to understand whether the relationship between the amount of **job training** employees receive, and their subsequent **performance** is moderated by their **gender**. The goal is to see if the effect of job training on performance differs between male and female employees.

The data were collected on the above variables and were analyzed using PROCESS Macro (Hayes) with IBM SPSS Statistics. The results are summarized in the following tables:

Model Summary

R	R - sq	MSE	F	df1	df2	p
0.9822	0.9647	2.3680	893.8038	3.0000	98.0000	.0000

Model

	coeff	se	t	p	LICI	ULCI
Constant	59.313	1.162	51.030	.0000	56.848	61.777
Training (X)	0.096	.0123	7.599	.0000	.069	.122
Gender (M)	-24.412	1.644	- 14.852	.0000	- 27.897	- 20.927
Training × Gender (X ×M)	-.079	.018	- 4.485	.0004	- .117	- .042

R – square increase due to interaction

	R2 - change	F	df1	df2	p
X×M	.0381	105.9647	1.0000	98.0000	.0000

Conditional effect of X on Y at values of the Moderator

Gender	Effect	se	t	p	LICI	ULCI
Male	.0955	.0126	7.5989	.0000	.0689	.1222
Female	.0159	.0125	1.2665	.2235	- .0107	.0425

Use the above information to answer the following questions.

a) Identify the independent, dependent, and moderating variables in this study.
b) Does job training significantly affect employees' performance? Justify your answer.
c) Does gender significantly affect employees' performance? Justify your answer.
d) Assess the significance of the interaction term to determine if gender significantly moderates the relationship between job training and performance.
e) Write the estimated statistical model that describes the relationship among job training, gender, and employees' performance.
f) Show graphically the relationship identified among job training, gender, and employees' performance.
g) What are the key findings you derived from this study?

(Total Marks 20)

05. The owner of a food delivery company is concerned about setting up appropriate costs for each food delivery. The goal is to identify which factors significantly affect food delivery cost. To investigate this, the owner gathered information on a random sample of 50 recent food deliveries on the following variables:

Cost: The cost in rupees incurred for each delivery
Distance: Distance in kilometre between the food delivery company and the delivery location
Delivery time: The actual travel time in minutes from the company to the customer location
Preparation time: The time in minutes between when the customized order is phoned into the company and when it is ready for delivery

The collected data were analysed using multiple linear regression with IBM SPSS Statistics and the results are provided below.

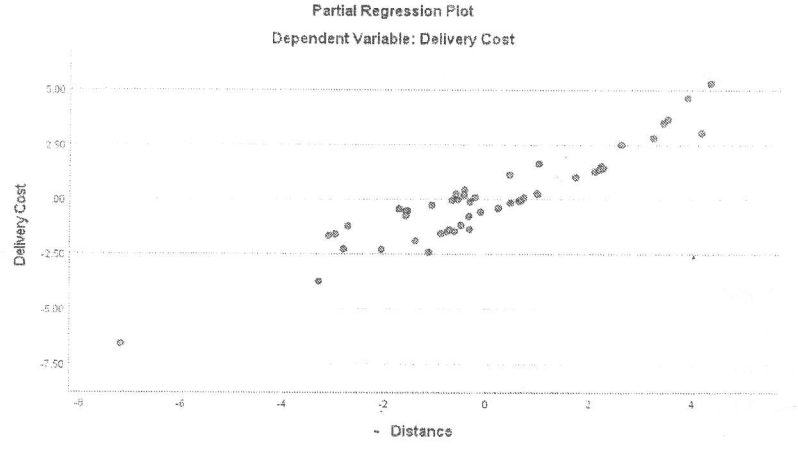
R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
.978	.957	.954	.772	2.336

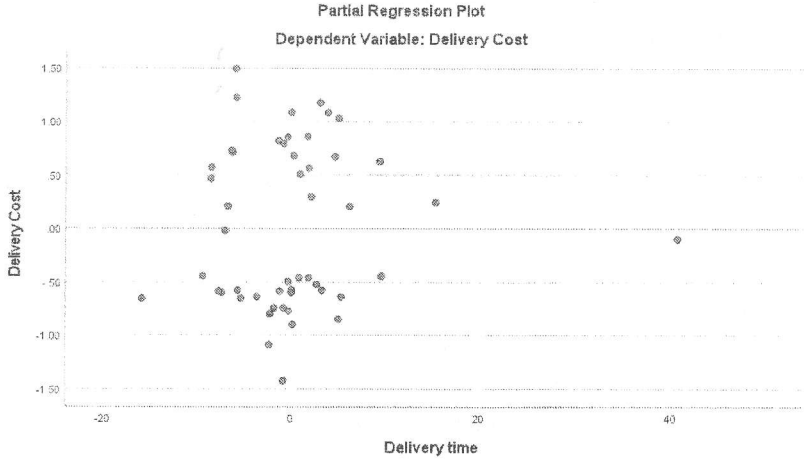
ANOVA

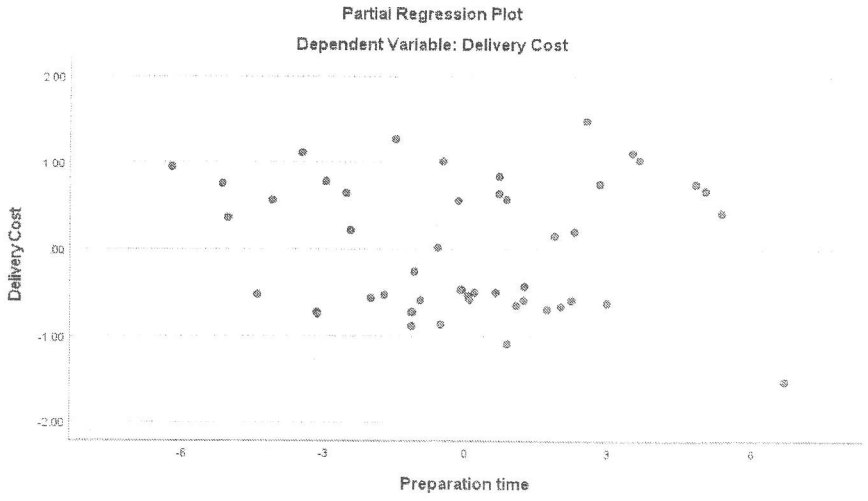
Model		Sum of Squares	df	Mean Square	F	Sig.
	Regression	604.614	4	201.538	337.882	.0000
	Residual	27.438	46	.596		
	Total	632.052	49			

Coefficients

	Unstandardized Coefficients		Standardized Coefficients			Correlations			Collinearity Statistics	
	B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
(Constant)	13.400	.470		28.528	<.001					
Distance	.880	.050	.963	17.602	<.001	.978	.933	.541	.315	3.172
Delivery time	.004	.014	.017	.314	.755	.812	.046	.010	.316	3.160
Preparation time	-.015	.039	-.012	-.382	.704	-.072	-.056	-.012	.988	1.012







Answer the following questions using the above information.

a) What is the objective of this study?
b) Identify the dependent and independent variables for this study?
c) Formulate three research questions that you can answer from the above results?
d) Interpret the results in the "Model Summary" table.
e) Which independent variables have strong or weak correlations with the dependent variable. Justif your answer.
f) Write down the multiple linear regression model that predicts delivery cost.
g) Do the variables in the model explain a reasonable amount of the variation in the dependent variable Justify your answer.
h) To what extent does the preparation time impact on delivery cost?
i) How much of variance in delivery time is uniquely explained by distance?
j) Determine whether the predicting model you wrote in part (f) is significant. Justify your answer.
k) Estimate the food delivery cost that takes 10 minutes for preparation, takes 30 minutes to deliver, and must cover a distance of 14Km.
l) Which predictor has the largest unique impact on delivery costs? Justify your answer.
m) Are there any predictors that should be dropped from the predicting model you wrote in part (f)? I so, what are the predictors should be dropped from the predicting model? Justify your answer.
n) Is there multicollinearity problem among the independent variables? Justify your answer.
o) Is each independent variable linearly related to the dependent variable? Justify your answer.
p) Write-up the results of this study
q) What do you recommend to the food delivery company?

(Total Marks 2