

56



(Knowledge for Development)

KIBABII UNIVERSITY
UNIVERSITY EXAMINATIONS
2020/2021 ACADEMIC YEAR
THIRD YEAR SECOND SEMESTER
MAIN EXAMINATION
FOR THE DEGREE OF BACHELOR OF SCIENCE

COURSE CODE: STA 324

COURSE TITLE: STATISTICAL COMPUTING

DATE: 14/10/2021

TIME: 9:00 AM - 11:00 AM

INSTRUCTIONS TO CANDIDATES

Answer Question One and Any other TWO Questions

TIME: 2 Hours

This Paper Consists of 6 Printed Pages. Please Turn Over.

QUESTION ONE

- a) Describe the different between data view and variable view in the SPSS data editor
- b) Explain the use of the following in statistical computing
 - (i) Cross tabulation
 - (ii) Boxplot
 - (iii) Correlation coefficient
 - (iv) Bar chart
- (c) When talking about statistical inference, what is alpha? What are the typical sizes of alpha? What does it mean?
- (d). Define the following terms as used in R. (i) Assign, (ii) Comment
 - (iii) Operator, (iv) Function
- (e) Discuss the four panes of the R Studio environment
- (f) Why do we use standard deviation, not variance?
- (g) Statistical methods that are sensitive to their underlying assumptions are commonly referred to as Parametric Methods while those that do not are known as Non-parametric Methods. Fill the corresponding tests

Type of test	Parametric test	Nonparametric test
2-sample		Mann-Whitney U-test
Paired sample	paired t-test	
Distribution		Kolmogorov-Smirnov
> 2 samples	1-way Anova	
Correlation		Spearman's r
Crossed comparisons		Friedman's;

QUESTION TWO

- (a) Briefly describe how you can carry out the following procedure using SPSS
 - (i) Descriptive
 - (ii) Cross-tabulation
 - (v) Kruskal Wallis H test
 - (ii) Explore
 - (iii) Mann – whitney U test
- (b) Explain 5 ways in which SPSS can transform data

QUESTION THREE

- (a). State five classical assumptions of linear correlation
- (b). Interpret the following the regression SPSS output

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.807 ^a	.651	.650	4.622

a. Predictors: (Constant), Vehicle Weight (lbs.)

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	15794.632	1	15794.632	739.503	.000 ^a
	Residual	8457.943	396	21.358		
	Total	24252.575	397			

a. Predictors: (Constant), Vehicle Weight (lbs.)

b. Dependent Variable: Miles per Gallon

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	45.492	.841		54.110	.000
	Vehicle Weight (lbs.)	-.007	.000	-.807	-27.194	.000

a. Dependent Variable: Miles per Gallon

QUESTION FOUR

A market researcher is interested in the coffee drinking habits of males and females. He asks a sample of male and female office workers to record the number of cups of coffee they consume during a week.

- a) Which parametric statistical technique could the researcher use to determine if males and females differ in terms of the number of cups of coffee consumed in a week?

Justify your answer and

- (b) describe how you would obtain this statistic using SPSS.
- c) What are the key values you would look for in the output?
- d) What assumptions should you check for when using the technique that you chose in question (a), above.
- e) What non-parametric technique could be used to address this research question?

QUESTION FIVE (20 MKS)

Consider the following SPSS output

Tests of Between-Subjects Effects

Dependent Variable: TEP: Average Text Editing Performance (Seconds)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4371 ^a	2	2185.5	16.2	.000
Intercept	114650	1	114650.5	849.3	.000
abilitycluster	4371	2	2185.5	16.2	.000
Error	9720	72	135.0		
Total	128741	75			
Corrected Total	14091	74			

a. R Squared = .310 (Adjusted R Squared = .291)

- i. Which three rows correspond to our typical: between, within and total sums of squares?
- ii. If you did not know the sample size or number of groups, how could you work this out from this table?
- iii. What is the overall text editing standard deviation for the entire sample?
- iv. What two numbers can be added to give us corrected total?
- v. What two numbers are used to obtain between treatments mean square?
- vi. What two numbers are used to obtain within mean square?
- vii. What two numbers are used to obtain the F value?
- viii. What is critical F at .05? What about .01? What effect does decreasing alpha have on critical F?
- ix. What effect does decreasing alpha have on statistical power?

- x. If the sample size was 203, what would critical F .05 be? What about a sample size of 18?
- xi. What effect does increasing sample size have on critical F? What effect does it have on power
- xii. What if there were 4 groups? What if there were 5 groups? What would critical F be? What effect does increasing the number of groups have on critical F?
- xiii. Is the obtained F statistically significant at .05? What about .01?
- xiv. Can we conclude that all three groups differ? If not, what could we do to find out?

(b) POST HOC ANALYSIS

TEP: Average Text Editing Performance (Seconds)

Tukey HSD^{a,b}

Ability Cluster	N	Subset	
		1	2
2 Higher Ability Group	25	31.1865	
3 Lower General Ability / Higher Psychomotor Ability Group	25	36.6923	
1 Lower Ability Group	25		49.4159
Sig.		.222	1.000

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = 134.997.

a. Uses Harmonic Mean Sample Size = 25.000.

b. Alpha = .05.

Multiple Comparisons

Dependent Variable: TEP: Average Text Editing Performance (Seconds)

Tukey HSD

(I) Ability Cluster	(J) Ability Cluster	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1 Lower Ability Group	2 Higher Ability Group	18.2295*	3.28629	.000	10.3650	26.0940
	3 Lower General Ability / Higher Psychomotor Ability Group	12.7236*	3.28629	.001	4.8591	20.5881
2 Higher Ability Group	1 Lower Ability Group	-18.2295*	3.28629	.000	-26.0940	-10.3650
	3 Lower General Ability / Higher Psychomotor Ability Group	-5.5059	3.28629	.222	-13.3704	2.3586
3 Lower General Ability / Higher Psychomotor	1 Lower Ability Group	-12.7236*	3.28629	.001	-20.5881	-4.8591
	2 Higher Ability Group	5.5059	3.28629	.222	-2.3586	13.3704

Based on observed means

- (i) What would Tukey's HSD .05 be using the exam formula sheet?
- (ii) Which pairs of groups differ by more than this amount?
- (iii) Looking at the homogenous subsets output, which groups are meant to be homogenous?