



(Knowledge for Development)

KIBABII UNIVERSITY COLLEGE

**A CONSTITUENT COLLEGE OF MASINDE MULIRO UNIVERSITY OF
SCIENCE AND TECHNOLOGY**

**UNIVERSITY EXAMINATIONS
2014/2015 ACADEMIC YEAR
THIRD YEAR SECOND SEMESTER**

**MAIN EXAMINATION
FOR THE DEGREE OF BACHELOR OF SCIENCE
AND BACHELOR OF EDUCATION**

COURSE CODE: STA 348

COURSE TITLE: STATISTICAL COMPUTING

DATE: 29/4/15 **TIME:** 3.00PM-5.00PM

INSTRUCTIONS TO CANDIDATES

Answer all question in section A and any other two in section B

TIME: 2 Hours

SECTION A

QUESTION ONE

Sample questionnaire

1. Sex ____ Male ____ Female

2. Age in years _____

3. Education level (please indicate the highest level of schooling that you completed)

_____ Year 10 _____ Year 12 _____ University or College

4. Are you currently on a diet to lose weight? ____ Yes ____ No

Please indicate how much you either agree or disagree with each of the following statements.

Write a number from 1 to 4 on the line next to each statement.

strongly disagree 1 2 3 4 strongly agree

1. _____ I have little control over the things that happen to me

2. _____ I can do just about anything I really set my mind to

3. _____ There is really no way I can solve some of the problems I have

4. _____ There is little I can do to change many of the important things in my life

5. _____ What happens to me in the future mostly depends on me

6. _____ I often feel helpless in dealing with the problems of life

7. _____ Sometimes I feel that I'm being pushed around in life

- a) Prepare a codebook for this questionnaire, detailing each of the variable names and codes to be used to prepare the data for entry into SPSS.
- b) Using the codebook you developed, describe how you would create value labels for Question 3 (Education level). Describe the steps you would use in SPSS.

- c) Using SPSS, create a new data file for this questionnaire. Enter some pretend data in this data file (for at least five cases).
- (d) Briefly describe how you can carry out the following procedure
- Explore
 - Crosstabulation

QUESTION TWO

- a) The regression SPSS output of the above question was as below. Use it to answer the following questions

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

a. Dependent Variable: Miles per Gallon

- i) How much variance in Miles Per Gallon is explained by Car Weight?
- ii) Is this variance explained significantly different to 0?
- iii) What is the constant
- iv) What is the slope?
- v) Is the slope statistically significant?
- vi) Write out the model regression equation
- vii) What is the standardised regression coefficient for vehicle weight?
- viii) If a car weighed 1000 pounds, what would be the predicted miles per gallon?
- ix) What is the standard error of the estimate?
- x) What would be the approximate 95% confidence interval of our prediction

SECTION B(answer any two)

QUESTION THREE

ANOVA

Age of Respondent

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2784.189	2	1392.095	4.409	.012
Within Groups	477048.215	1511	315.717		
Total	479832.404	1513			

Multiple Comparisons

Pendent Variable: Age of Respondent

LSD

(I) Race of Respondent	(J) Race of Respondent	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
White	Black	2.930*	1.344	.029	.29	5.57
	Other	5.609*	2.587	.030	.53	10.68
Black	White	-2.930*	1.344	.029	-5.57	-.29
	Other	2.679	2.828	.344	-2.87	8.23
Other	White	-5.609*	2.587	.030	-10.68	-.53
	Black	-2.679	2.828	.344	-8.23	2.87

*. The mean difference is significant at the 0.05 level.

Consider the SPSS output above

- (a) State the Procedure for one-way between-groups ANOVA with post-hoc tests
- (b) Explain the interpretation of output from one-way between-groups ANOVA with post-hoc tests

QUESTION FOUR

- (a) When talking about statistical inference, what is alpha? What are the typical sizes of alpha? What does it mean?
- (b) You want to test the following hypotheses. What type of statistical test do you run? Why do you run this test (what can you tell from it?). Describe the procedure for the above test
 - (i) Research hypothesis #1: Males have a higher mean level of income than females. Income is measured in thousands of dollars earned per year.
 - (ii) Research hypothesis #2: Males are more likely to be alcoholics than females. “Alcoholic” is coded “yes/no.”
- (c) In statistical analysis what is the difference between correlation and regression

QUESTION FIVE

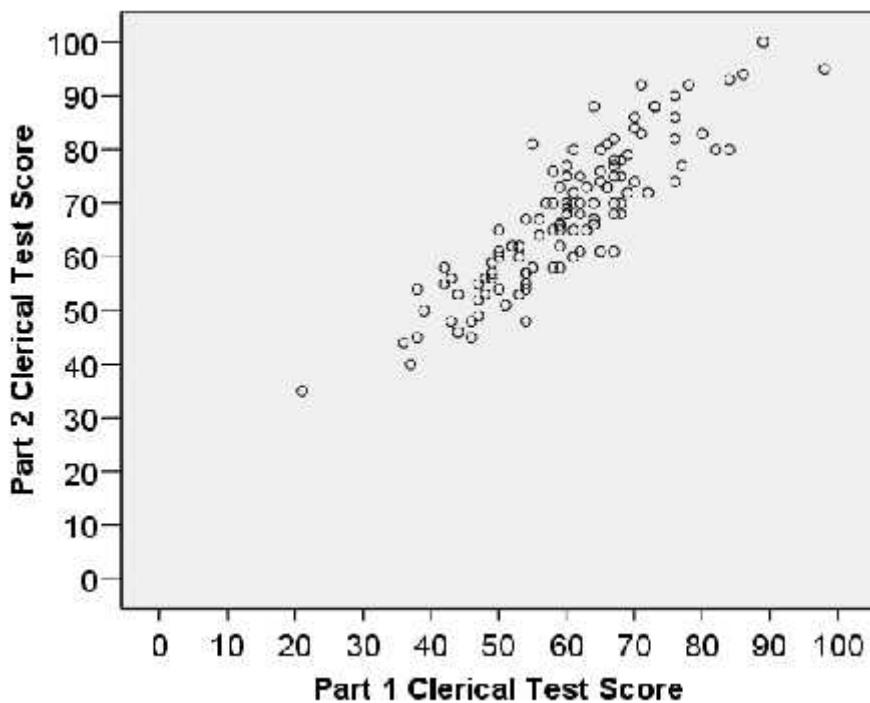
A market researcher is interested in the coffee drinking habits of males and females. He asks a sample of male and female office workers to record the number of cups of coffee they consume during a week.

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

QUESTION SIX

- (a) In SPSS, how can one combine values of a variable into smaller number of categories?
- (b) Describe how you can create new variables out of old ones using RECODE

(c) Consider the following SPSS output



Correlations

		PSA: Clerical Speed Adjusted Total Part 2	PSA: Clerical Speed Adjusted Total Part 1
PSA: Clerical Speed Adjusted Total Part 2	Pearson Correlation	1	.882
	Sig. (2-tailed)		.000
	Sum of Squares and Cross-products	19759.948	16207.026
	Covariance	173.333	142.167
	N	115	115
PSA: Clerical Speed Adjusted Total Part 1	Pearson Correlation	.882	1
	Sig. (2-tailed)	.000	
	Sum of Squares and Cross-products	16207.026	17072.487
	Covariance	142.167	149.759
	N	115	115

- (i) Looking at the scatterplot, describe the relationship in terms of direction, form, and degree.
- (ii) What is the largest absolute value a correlation can take?
- (iii) Can we justifiably say that part 1 test score caused their part 2 test score?
- (iv) The sample was made up of roughly 25% first year psychology students and 50% friends of one of the researchers typically aged around 22 and 25% parents of friends of the researchers, do you think restricted range might be an issue here?
- (v) Are there any extreme scores?
- (vi) If instead of using the raw scores, we used the rank of each participant, and performed a correlation on these ranks, what type of correlation would we have?
- (vii) If we split part 1 scores into high and low groups and correlated it with the raw version of part 2, what type of correlation would we have?
- (viii) Is the correlation statistically significant? What is our df, what is critical r (two-tailed) in G&W?