

## **Statistical Exploration and Reasoning 2017-18**

### **DOING THE SUMMATIVE ASSIGNMENT**

#### **The objectives of each element of the assignment:**

Question 1 is designed to test:

- Your understanding of levels of measurement
- Your ability to select appropriate graphs to describe variables at different levels of measurement.
- Your ability to produce appropriate descriptive statistics for variables at each level of measurement.
- Your ability to interpret your graphs and summary statistics. This should be 2-3 sentences for each variable.

Question 2 (a and b) is designed to test your ability to:

- Construct and discuss a table relating two categorical variables. Remember to generate row percentages so that it is easier to compare the conditional distribution of the column variable at different values of the row variable.
- Use Chi Squared to test the null hypothesis of no association between the variables in the population.
- Use the gamma statistic to measure the strength and direction of relationship between the variables.
- Interpret the results of these operations. What do they tell us substantively?

Question 3 is designed to test your ability to:

- Recognise which types of explanatory and response variables are appropriate for two-group difference-in-means analysis.
- Briefly explain how you expect the variables you have used to be associated in the populations of interest and why you expect this
- Carry out appropriate two-group difference-in-means analyses on these variables
- Interpret the results of this analysis. Is there evidence for the association you expected? What is the magnitude of this association? What are the implications of this?

Question 4 is designed to test your ability to:

- Recognise which types of explanatory and response variables are appropriate for ANOVA analysis.

- Briefly explain how you expect the variables you have used to be associated in the populations of interest and why you expect this
- Carry out the appropriate ANOVA significance test on these variables and produce multiple comparisons confidence intervals.
- Interpret the results of this analysis. Is there evidence for the association you expected? What is the magnitude of the association? What are the substantive implications of this?

Question 5(a) is designed to test your ability to:

- Briefly explain how you expect the variables you have used to be associated in the populations of interest and why you expect this
- Use scatter plots and correlation coefficients as descriptive tools to explore the association between the two variables within your data
- Carry out simple linear regression
- Interpret the simple linear regression output, making inferences about the nature and magnitude of the association between these two variables in the population of interest. Interpret regression coefficients and corresponding significance tests and confidence intervals carefully.
- Based on your results, briefly discuss whether there is evidence for the association you expected to find.

Question 5(b) is designed to test your ability to:

- Briefly explain how you expect the variables you have used to be associated in the population of interest and why you expect this
- Carry out multiple linear regression, using dummy variable methods to include a categorical variable.
- Write out your estimated regression equation.
- Interpret the multiple linear regression output, making inferences about the nature and magnitude of the partial association between each of your explanatory variables and your response variable in the population of interest. Interpret regression coefficients and corresponding significance tests carefully, especially for the dummy variable(s).
- Based on your results, briefly discuss whether there is evidence for the association you expected to find.

## Handy hints

1. BE CAREFUL when you are looking at SPSS output. Make sure you know exactly what each element of output (e.g. a P-value or a coefficient) is for before you use it.
2. BE CAREFUL in the terminology you use when interpreting your results. For example, don't confuse samples with populations. Think about the meanings of the words you are using.
3. Always make inferences about the population of interest in questions 2 through to 5. Be careful to explicitly state your population of interest and to clearly state any inferences with reference to that population.
4. Make sure you are clear on levels of measurement and the differences between quantitative and categorical variables. This is crucial in order for you to choose appropriate variables for the different questions.
5. If you are unsure about anything, refer to textbooks, together with the lecture slides and computer class instructions.

## The Data Set

You can also get more details on the questions used to create each variable in the WVS survey (F00001101-WV6\_Official\_Questionnaire\_v4\_June2012.pdf), from the DUO assignments folder.

In the exercises below, you will choose the variables you want to analyse, so **make sure you select variables that you are comfortable interpreting**. Also, you may need to recode some variables or re-define missing values in some instances. If you recode variables, do this carefully using the SPSS re-coding techniques we covered in session one. If you re-define missing values for some variables, make sure you look at the handout on "Missing Values in SPSS", available in the DUO Teaching Materials/Practicals folder. If you re-code variables look at the computer class instructions from week 1 (general instructions) and from week 10 (specific instructions for creating dummy variables).

The data has plenty of nominal and ordinal variables but fewer quantitative interval/ratio variables. Think carefully about the variables you use for each question. You might decide to treat an ordinal variable as a quantitative variable, but if you do then **make sure that you note this and justify it briefly in your write-up**.