

STA10003 FOUNDATIONS OF STATISTICS WORK INTEGRATED LEARNING (WIL) ASSIGNMENT PART 1



This **Assignment Part 1** is worth **20%** of your final mark for STA10003.

Scenario

You have been employed as a new graduate researcher at Social Interactions Australia. Social Interactions Australia conducts research about relationships and other social issues. You have been given a dataset which contains the results of a survey given to Melbourne adults who visited a psychologist in 2022. You have been asked to analyse the data and answer several questions of interest that are presented on the following pages.

Data Preparation

For your assignment you should use the data set **STA10003 Sem 2 2024 Assignment Data.sav** provided in **Week 06: Assignment Part 1 Instructions & Data File** which can be accessed within the Week 06: Assignment Part 1 page. You must use SPSS to draw a random sample of 5000 cases from the 6000 cases in the data file. You will conduct your analysis on this sample of 5000 cases. Instructions on how to generate your random sample are on pages 5 to 8 of this document.

Submission Instructions

- Your submission must be a single Word file or PDF file.
- Although a cover page is not required, you should include your name and student number within the document [e.g., in footer].
- You must submit your file via the **SUBMIT ASSIGNMENT** button on the Week 06: Assignment Part 1 page in Canvas by **Sunday September 8 by 11:59pm**.
- Only the last document you submit will be marked.
- Once submitted, please review your submission to ensure the correct file has been submitted.
- This is an individual assignment. Do not share your work with other students. They will have a different sample of data, so any copying will be detected.

Plagiarism and Maintaining Academic Integrity

This is an individual assignment, so you are expected to complete it by yourself. It is important that you demonstrate good Academic Integrity by ensuring that your assignment work is entirely your own, because this shows that you have understood what you have learnt. You should not use the work of anyone else, including that of another students, for this assignment. Whilst you may seek help if there is anything about this assignment you do not understand, the assignment must be your own work.

Use of generative AI such as ChatGPT is not permitted. Any form of copying or submitting work that is not entirely done by you is a breach of academic integrity and could attract academic penalties. Your work will be checked for breaches of academic integrity.

For your Assignment Part 1, you are required to complete the first three (3) questions by producing the appropriate analyses using SPSS and writing the relevant report for each question. You are also required to complete questions 4 and 5, which contain short answer questions.

For each of the first three questions requiring SPSS, you should include the relevant output immediately following your report. Graphs which are part of the report should be included within the report as shown in the report writing examples used in the course materials – as shown in the lectures and in tutorials. The document *Reading B: Reporting Information about Single Variables* also has more report writing examples.

Question 1: Satisfaction with life

The variable Life measures the level of satisfaction with life. Using SPSS, produce the relevant graph and table to summarise the Life variable and write a paragraph explaining the key features of the data observed in the output in the style presented in the course materials.

Question 2: Household size

The variable Household measures household size. Using SPSS, produce the relevant graph and tables to summarise the Household variable and write a paragraph explaining the key features of the data observed in the output in the style presented in the course materials.

Question 3: Income per week (\$)

The variable Income measures income per week in \$. Using SPSS, produce the relevant graph and tables to summarise the Income variable and write a paragraph explaining the key features of the data observed in the output in the style presented in the course materials.

Question 4: [does not require SPSS]

Jasmine likes to take part in competitions where participants assemble mechanical puzzles. Participants use an assortment of metal and plastic parts to make a specific object. Participants are timed to see how long it takes them to assemble a mechanical puzzle. Jasmine participated in four competitions last year. Jasmine completed the car puzzle in 420 seconds, she completed the bus puzzle in 340 seconds, she completed the train puzzle in 670 seconds, and she completed the plane puzzle in 550 seconds. Completion times for people who did the car puzzle competition are normally distributed with a mean of $\mu = 469$ seconds and a standard deviation $\sigma = 54$ seconds. Completion times for people who did the bus puzzle competition are normally distributed with a mean of $\mu = 300$ seconds and a standard deviation $\sigma = 29$ seconds. Completion times for people who did the train puzzle competition are normally distributed with a mean of $\mu = 724$ seconds and a standard deviation $\sigma = 42$ seconds. Completion times for people who did the plane puzzle competition are normally distributed with a mean of $\mu = 505$ seconds and a standard deviation $\sigma = 36$ seconds. In which puzzle competition was Jasmine's performance best, relative to others who did the puzzle competitions? Justify your answer, quoting relevant statistics as part of your explanation.

Question 5: [does not require SPSS]

Umami Papi is a crispy aromatic chilli oil that is made in Melbourne. The company makes two versions of the chilli oil, the Original and Extra Spicy. Umami Papi comes in two sizes, a 225g jar and a 750g jar. The production manager at the production plant wants to make sure that the machine that fills the jars is working correctly. The production manager takes a random sample of 1200 225g jars of Umami Papi Original chilli oil produced in March 2024. This sample is then used to check if bottles are filled with the correct amount of chilli oil, or if perhaps the machine needs to be adjusted.

- a. What is the population we can draw conclusions about in this study?

To answer questions (b) to (d), only consider the sampling distribution shown in *Figure 1*.

No calculations are needed.

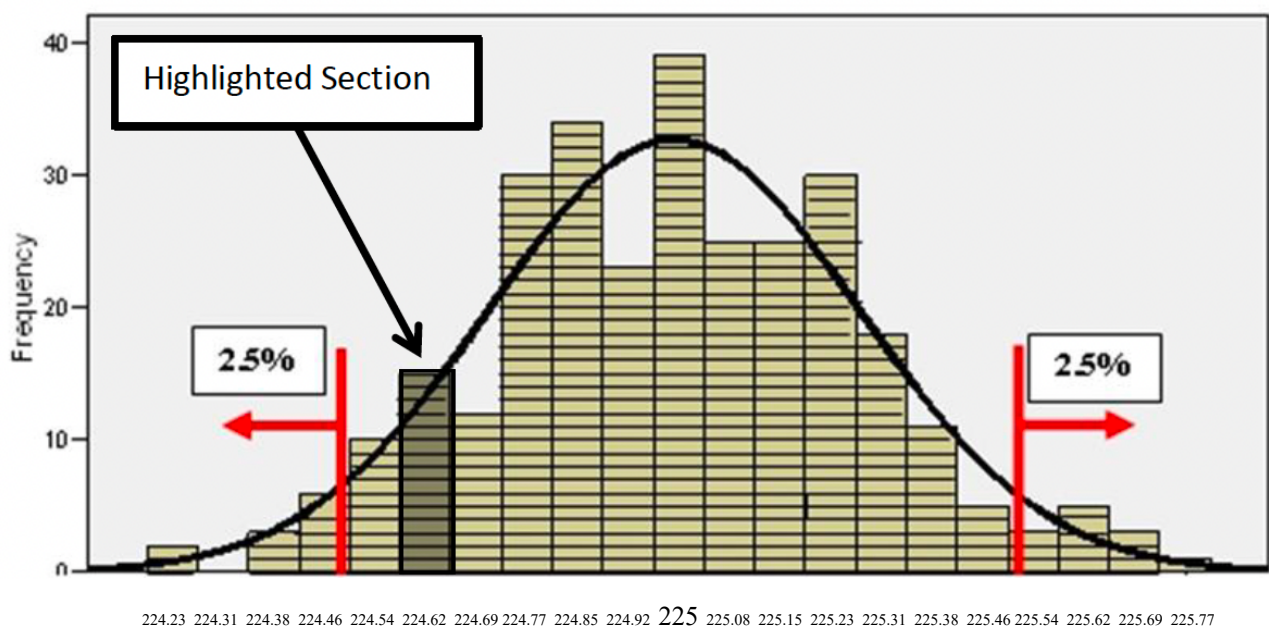
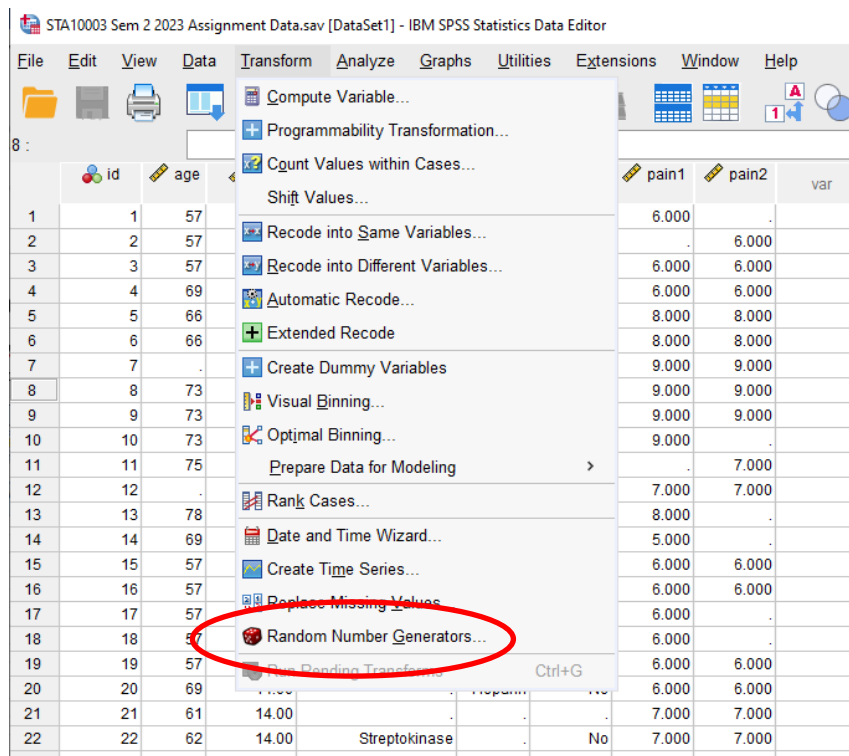


Figure 1: Distribution of sample means in 300 samples of size 1200, taken from a population where the mean is 225 and the standard deviation is 8

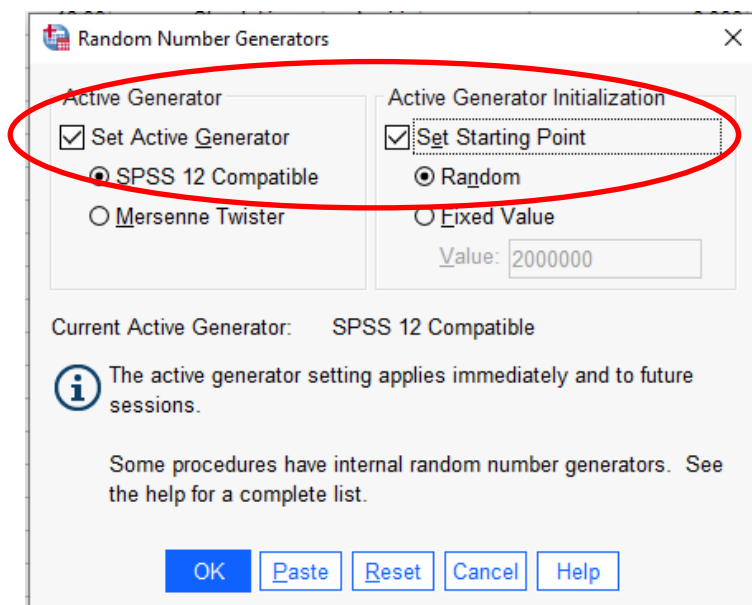
- b. What does the highlighted section of the sampling distribution in *Figure 1* represent?
- c. The random sample of 1200 jars taken by the production manager had a mean weight of 225.62g. Does this sample look like it belongs to the sampling distribution displayed in *Figure 1*? Justify your answer.
- d. The production manager wants to know if bottles are filled with the correct amount of chilli oil. What specific conclusion can be made based on part (c)?

How to generate your random sample of 5000 observations.

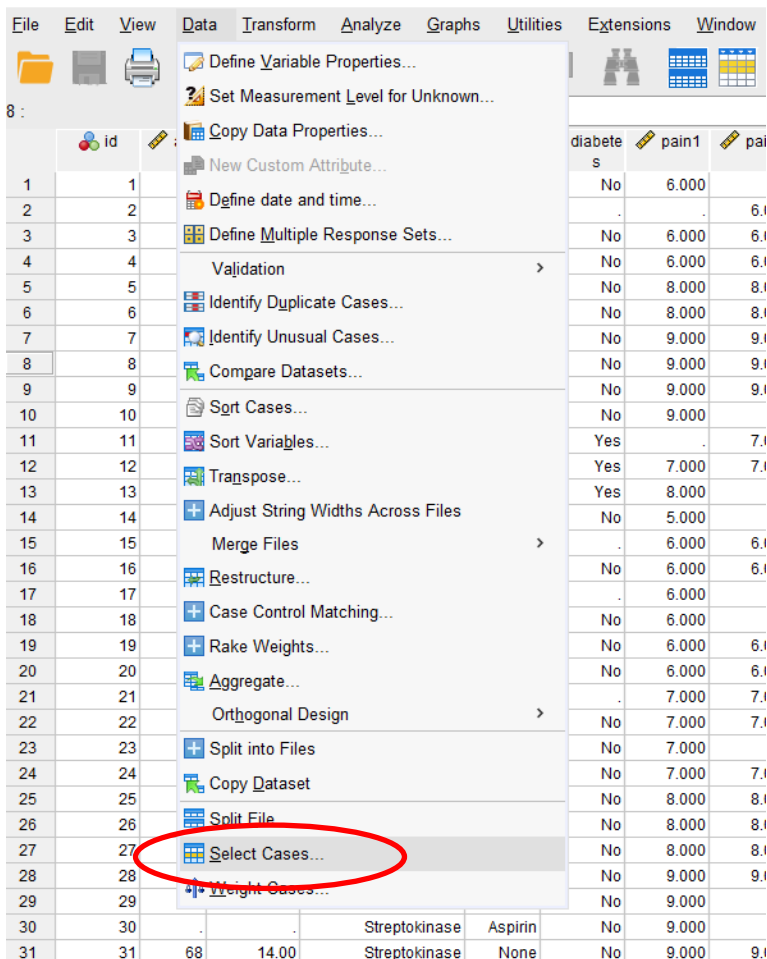
1. Open the **STA10003 Sem 2 2024 Assignment Data.sav** data file. From the **Transform** drop-down menu, select **Random Number Generators** from the menu



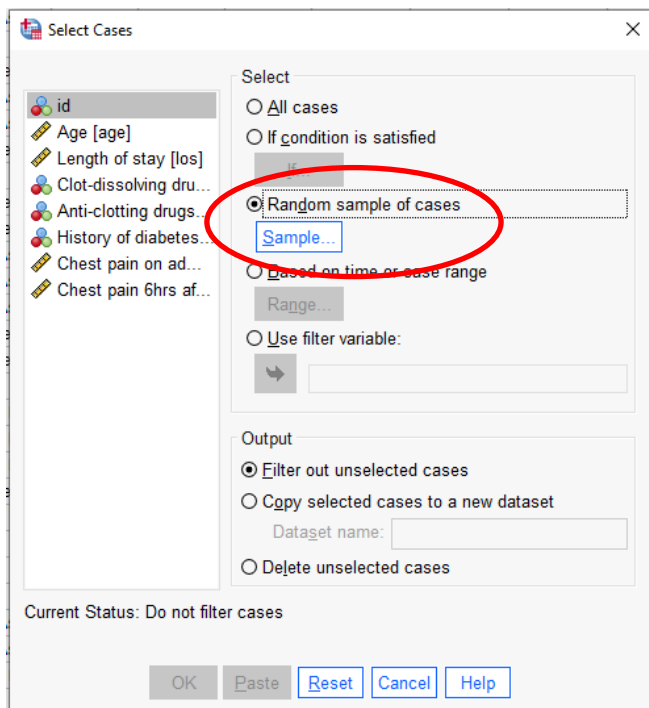
2. From the **Random Number Generators** dialogue box, click the boxes beside **Set Active Generator** and **Set Starting Point** as shown below. Then click **OK**.



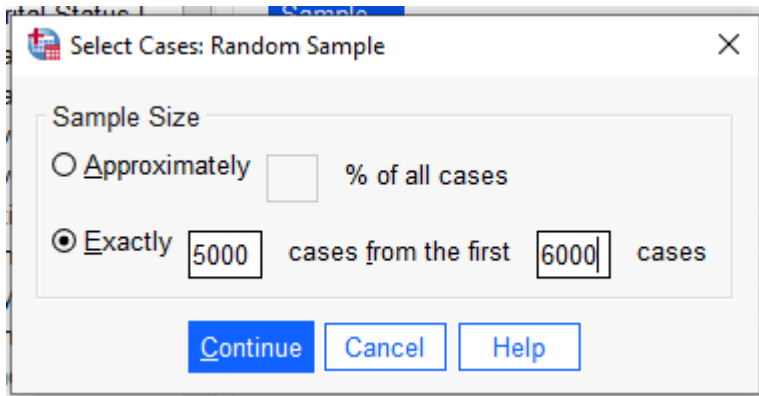
3. From the **Data** drop-down menu, choose **Select Cases**



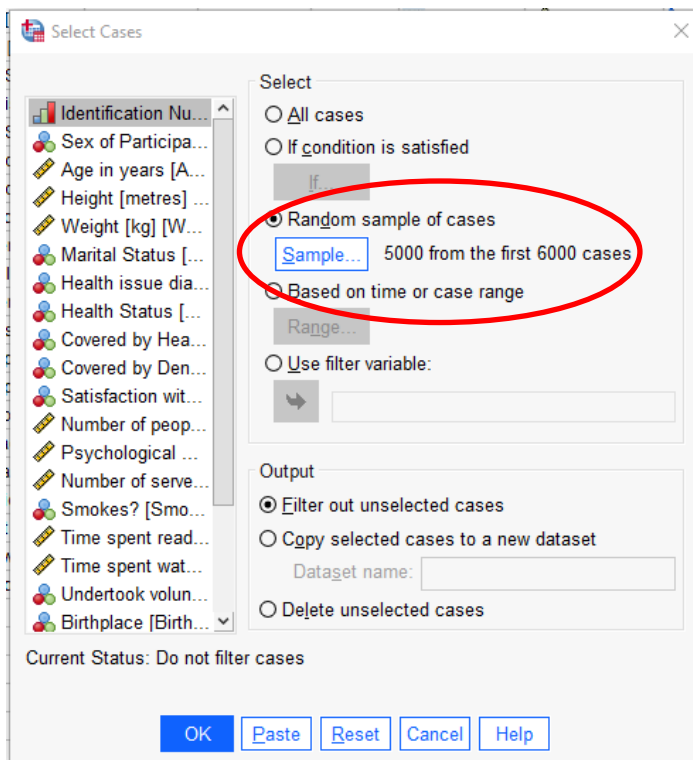
- From the **Select Cases** dialogue box, choose **Random Sample of Cases** and then click the **Sample** button.



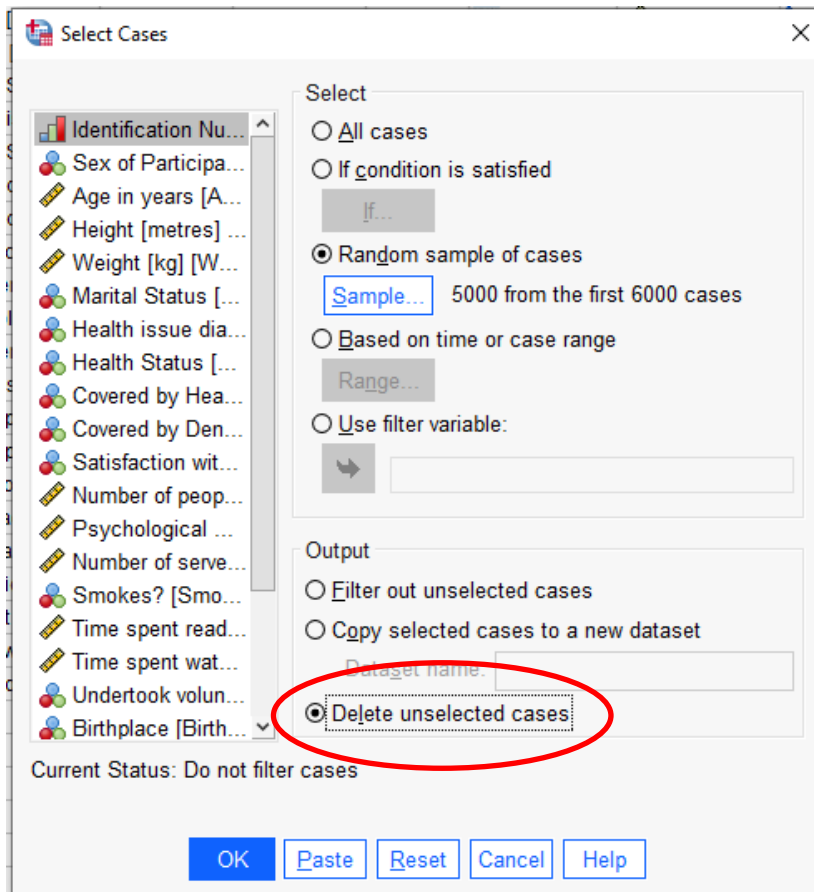
5. From the **Select Cases: Random sample** dialogue box, click **Exactly** and type **5000** cases from the first **6000**.



6. After entering the information above, click **Continue** [this returns you to the **Select Cases** Dialogue Box where you will see, next to the **Sample** button, confirmation of the 5000 cases selected].



7. We can remove the unselected cases by clicking the **Delete unselected cases** button under the **Output** heading.






8. After clicking **OK**, your data set will now only show the 5000 cases selected.

9. You should now **save the data file with a new name**.

The data file is ready to use for your Assignment!

Note: This data file will also be used in the Assignment Part 2.

STA10003 Assignment Part 1 Marking Rubric [out of 42]

STA10003 Assignment Part 1							  
Criteria	Ratings					Pts	
Question 1 Summary of Categorical Variable	10 to >9.0 Pts Full marks allocated Report has no errors.	9 to >0.0 Pts Partial marks allocated Report has 1 to 9 errors	0 Pts No mark allocated Question not answered, no report, report covers no relevant/correct information.			10 pts	
Question 2 Summary of Metric Variable	10 to >9.0 Pts Full marks allocated Report has no errors.	9 to >0.0 Pts Partial marks allocated Report has 1 to 9 errors.	0 Pts No mark allocated Question not answered, no report, report covers no relevant/correct information.			10 pts	
Question 3 Summary of Metric Variable	10 to >9.0 Pts Full marks allocated Report has no errors.	9 to >0.0 Pts Partial marks allocated Report has 1 to 9 errors.	0 Pts No mark allocated Question not answered, no report, report covers no relevant/correct information.			10 pts	
Question 4 Short Answer Question	3 Pts Full marks allocated Correct Conclusion. All relevant/correct statistics quoted.		2 Pts Partial marks allocated Correct Conclusion. Partial relevant/correct statistics quoted.		1 Pts Partial marks allocated Correct Conclusion. No relevant/correct statistics quoted.	0 Pts No mark allocated No attempt. Incorrect conclusion.	3 pts
Question 5a	2 Pts Full marks allocated Correct answer, no errors.		1 Pts Partial marks allocated. Answer has 1 or more errors.		0 Pts No mark allocated No attempt. Incorrect answer.		2 pts
Question 5b	2 Pts Full marks allocated Correct answer, no errors.		1 Pts Partial marks allocated Answer has 1 or more errors.		0 Pts No mark allocated No attempt. Incorrect answer.		2 pts
Question 5c	2 Pts Full marks allocated Correct answer, no errors.		1 Pts Partial marks allocated Answer has 1 or more errors.		0 Pts No mark allocated No attempt. Incorrect answer.		2 pts
Question 5d	3 Pts Full marks allocated Correct answer, no errors.	2 Pts Partial marks allocated. Answer has 1 error.	1 Pts Partial marks allocated. Answer has 2 errors.	0 Pts No mark allocated No attempt. Incorrect answer. Answer has 3 or more errors.			3 pts
Total points: 42							

Marking Details

Prior to submitting your Assignment, use the following checklist as a guide to ensure that you have provided all of the relevant information.

Q1 – Should include [*as appropriate*]:

A graph and Frequency output appropriate for a categorical variable.

A paragraph that includes mention of largest group, and all the other groups with correct percentages, and any other relevant patterns based on *Reading B: Reporting information about single variables*.

Q2 – Should include [*as appropriate*]:

A graph and Explore output appropriate for a metric variable.

A paragraph that includes a description of the shape of distribution, centre, spread, outliers if present based on *Reading B: Reporting information about single variables*.

Q3 – Should include [*as appropriate*]:

A graph and Explore output appropriate for a metric variable.

A paragraph that includes a description of the shape of distribution, centre, spread, outliers if present based on *Reading B: Reporting information about single variables*.

Q4 – The answer should be presented in a short paragraph, quoting relevant statistics in support of the response.

Q5 – The answers should be presented with sections (a) to (d) clearly identified.

Checklist:

- Correct variable used to produce output [note that many of the variables have similar names so it is important to double-check that the correct variable has been used]
- Correct procedure performed
- Graphs appropriately edited and labelled [eg edited variable names; “*Figure 1. The distribution of ...*”]
- All figures quoted in report correct according to your own output
- Correctly referring to the sample or population when appropriate
- Proof reading of reports for errors