STA10003 FOUNDATIONS OF STATISTICS ASSIGNMENT PART 1

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

Scenario

You are a new graduate researcher at a social and psychological sciences research institute. You have been given a dataset based on a survey of Australian public service sector employees, and asked to analyse the data to answer several questions of interest that are presented on the following pages.

For your assignment you should use the data set provided for you in the **Assignments > Assignment Part 1** section of the STA10003 Canvas site **STA10003 Sem 2 2020 Assignment Data.sav**.

Data Preparation

For the purpose of your report, you must use **SPSS** to draw a **random sample of 1500 cases** of the 4352 observations. **You will conduct your analysis on this sample of 1500**. Instructions on how to generate your random sample are on pages 5 to 8 of this document. Note, however, that some variables contain missing values, so each of your analyses might not contain the entire 1500 cases.

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 Canvas link by Monday 21 September at 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 random sample of data, so any copying will be detected.

ASSIGNMENT PART 1

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 all the relevant output either directly before your report, or after the report. Graphs, which are part of the report should be included within the report as shown in the report writing examples (see *Supplement B: Reporting Information about Single Variables* and *Supplement C: Choosing the Correct Analysis*)

Question 1

The variable **WorkEnjoyment** indicates the extent to which each employee agrees with the statement 'I enjoy my work'. Produce the relevant graph and table to summarise the **WorkEnjoyment** 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

The variable **JobEngagement** is a scale measurement that indicates how engaged an employee is with the job they work in. This variable was measured on a scale that can take values from 0 to 20, with higher values representing greater employee engagement with their job. Produce the relevant graph and tables to summarise the **JobEngagement** 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

The variable **AgencyEngagement** is a scale measurement that indicates how engaged an employee is with their Agency/Department. This variable was measured on a scale that can take values from 0 to 20, with higher values representing greater employee engagement with their Agency/Department. Produce the relevant graph and tables to summarise the **AgencyEngagement** 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]

James the postie is accused by his manager of being too slow in completing his mail deliveries. James feels he is completing his designated routes within a reasonable amount of time and calls for an investigation. He makes a point of recording the time it takes him to complete each of his four routes and it turns out they each take 110 minutes.

On which of the following routes would James be considered unusually ***slow*** compared to the other posties who deliver on the same routes?

Route 1 where the mean time is $\mu = 60$ minutes and the standard deviation $\sigma = 20$ minutes

Route 2 where the mean time is $\mu = 120$ minutes and the standard deviation $\sigma = 10$ minutes

Route 3 where the mean time is $\mu = 100$ minutes and the standard deviation $\sigma = 10$ minutes

Route 4 where the mean time is $\mu = 140$ minutes and the standard deviation $\sigma = 15$ minutes

Justify your answer, quoting the relevant statistics as part of your explanation. Assume the route times are normally distributed.

Question 5: [does not require SPSS]

The Tinned Fruit Company introduced a new range of mid-sized tins of fruit earlier this year. The company claims that the mean weight of these tins of fruit are on average 200g, with a standard deviation of 12g. Recently however, a consumer organisation has received a number of complaints from consumers who believe that the mean weight of their tins containing nectarines is actually less than 200g. To investigate this discrepancy, the consumer organisation randomly selects 1500 tins of nectarines produced in July and records the weight of these tins. This sample is then used to assess whether the company is marketing tins of nectarines that weigh on average less than 200g.

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

We have produced a sampling distribution for the mean weight using 200 samples of size 1500, taken from a population where the mean is 200g, with a standard deviation of 12g. The sampling distribution is displayed in *Figure 1*.

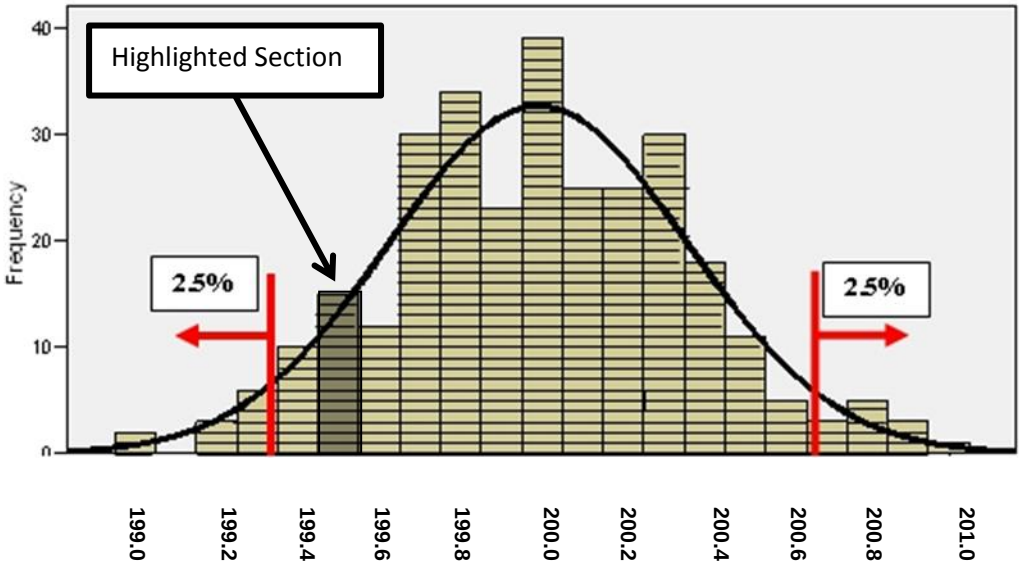


Figure 1: Distribution of sample means in 200 samples of size 1500, taken from a population where the mean is 200g and the standard deviation is 12g

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

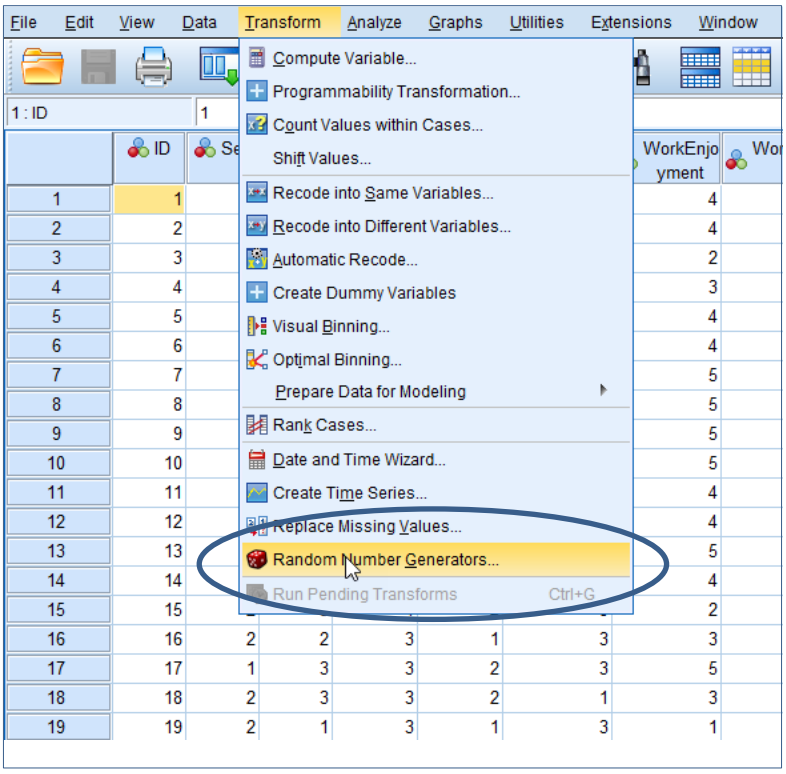
b. What does the highlighted section of the sampling distribution in *Figure 1* represent?

c. The random sample of 1500 tins of nectarines taken by the consumer organization turned out to have a mean weight of 200.8g. Does it look like this sample belongs to the sampling distribution displayed in *Figure 1*? Justify your answer.

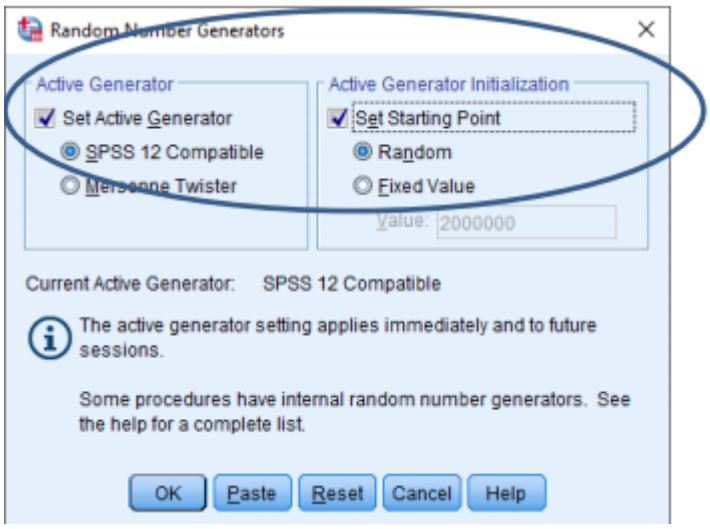
d. Given that the sample was randomly selected from all tins of nectarines produced by the Tinned Fruit Company during July, and given that the weight was measured accurately, what conclusion can we reach from part (c)?

How to generate your random sample of 1500 observations.

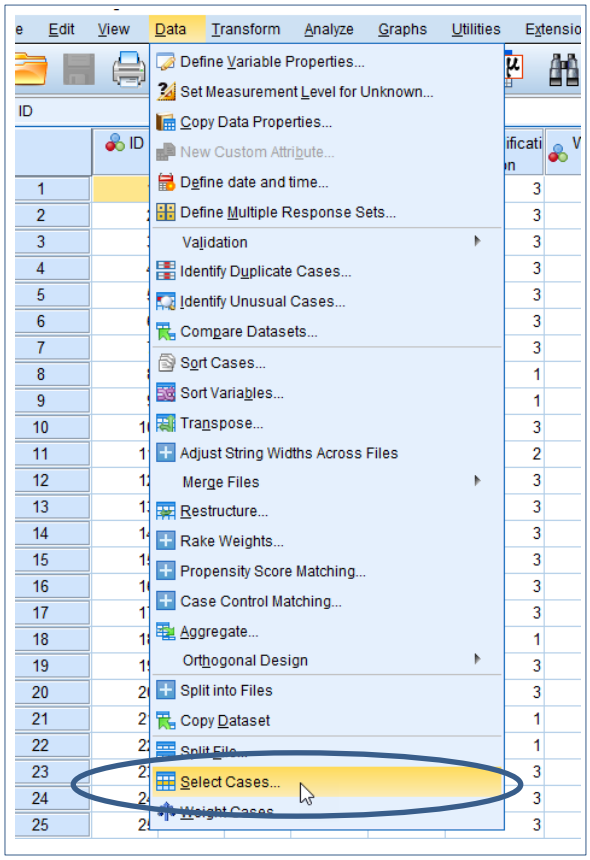
1. Open the **STA10003 Sem 2 2020 Assignment Data.sav** data file. From the **Transform** drop-down menu, select **Random Number Generators** from the menu



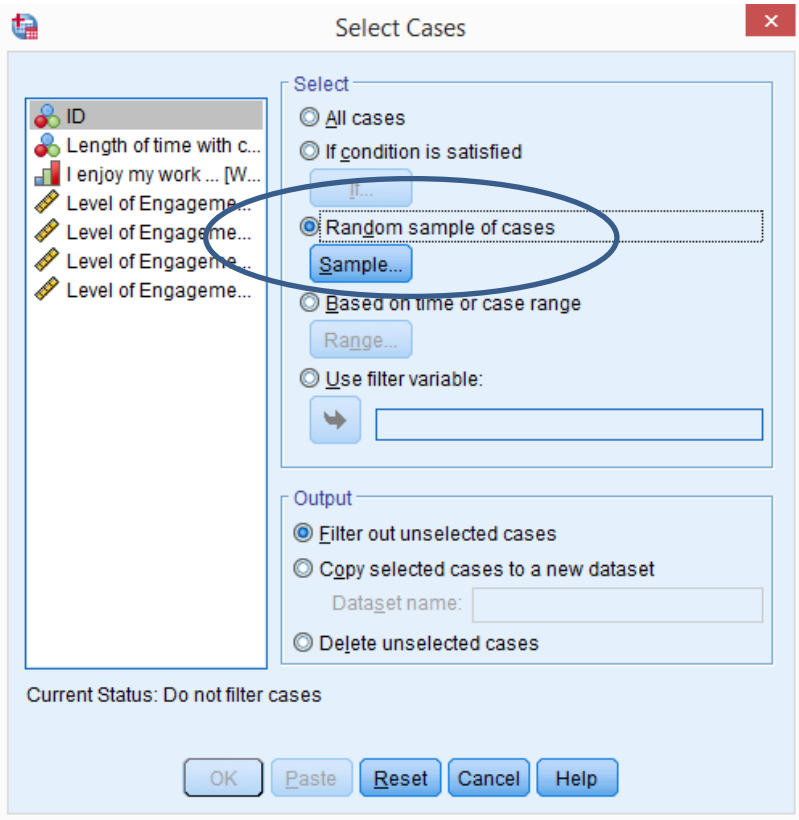
2. From the **Random Number Generators** dialogue box, click the boxes beside **Set Active Generator** and **Set Starting Point** as shown below. Then click **OK.**



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

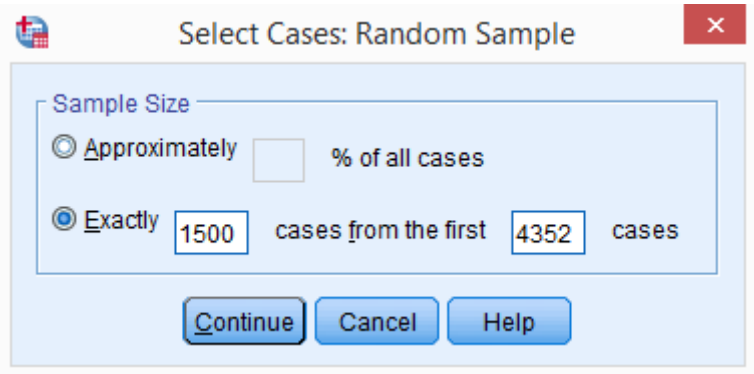


4. From the **Select Cases** dialogue box, choose **Random Sample of Cases** and then click the **Sample** button [the **Sample** button is pale blue until you select the **Random sample of cases** choice]:

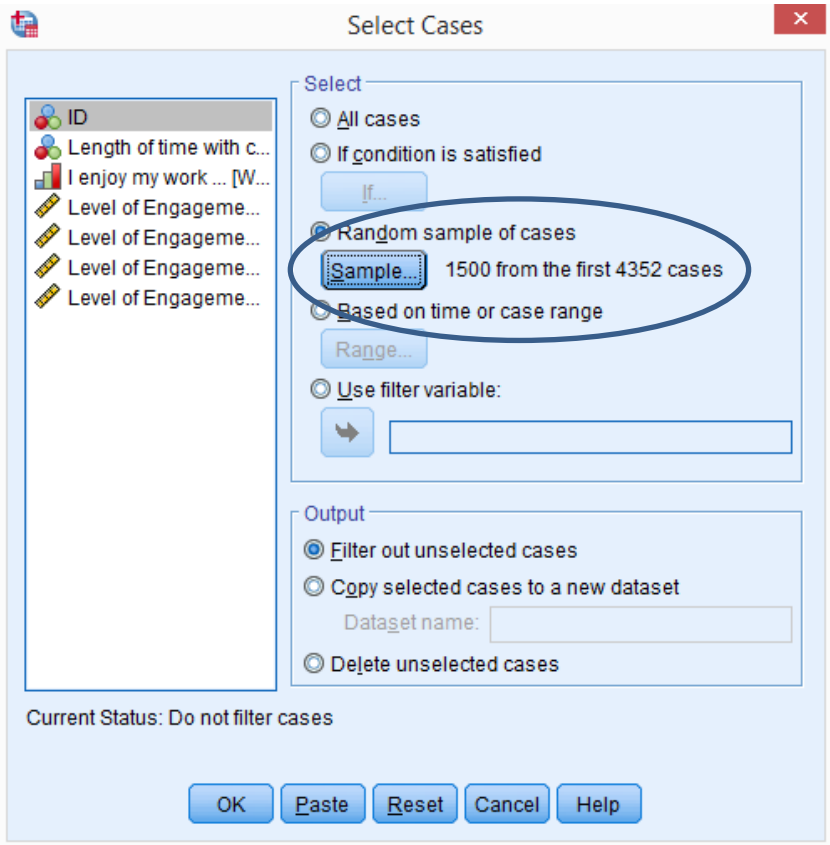


5. From the **Select Cases: Random sample** dialogue box, click **Exactly** and type **1500** cases from the first **4352**.

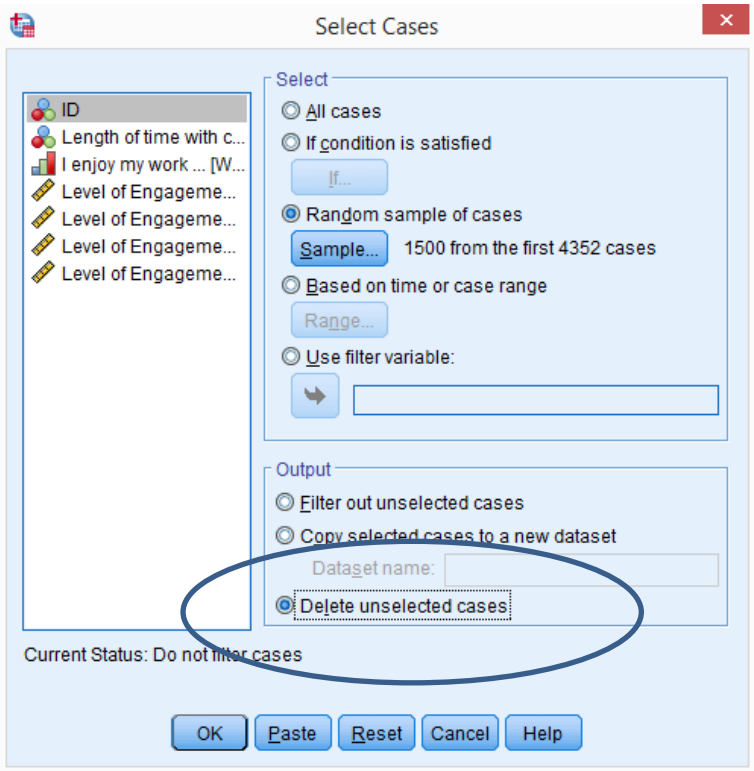
   [again the information is pale blue until you select '**Exactly**'. We want to generate a random sample of 1500 from the entire data set, so enter **1500** cases from the first **4352** cases]



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 1500 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 1500 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], due later in the Study Period.

STA10003 Assignment – *Part 1* Marking Rubric [out of 20]:

	0	1	2	3	4
Question 1 Summary of Categorical Variable [4 marks]	Inappropriate graphs or tables used	Appropriate graphs and tables for categorical data No report	Appropriate graphs and tables for categorical data Reasonable attempt to summarise the information shown in the output	Appropriate graphs and tables for categorical data Graphs suitably edited for presentation (e.g. figure title) Written summary covers all relevant features of output	Appropriate graphs and tables for categorical data Graphs suitably edited for presentation (e.g. figure title) Written summary covers all relevant features of output and is clearly and concisely written
Question 2 Summary of Metric Variable [4 marks]	Inappropriate graphs or tables used	Appropriate graphs and tables for metric data No report	Appropriate graphs and tables for metric data Reasonable attempt to summarise the information shown in the output	Appropriate graphs and tables for metric data Graphs suitably edited for presentation (e.g. figure title) Written summary covers all relevant features of output	Appropriate graphs and tables for metric data Graphs suitably edited for presentation (e.g. figure title) Written summary covers all relevant features of output and is clearly and concisely written
Question 3 Summary of Metric Variable [4 marks]	Inappropriate graphs or tables used	Appropriate graphs and tables for metric data No report	Appropriate graphs and tables for metric data Reasonable attempt to summarise the information shown in the output	Appropriate graphs and tables for metric data Graphs suitably edited for presentation (e.g. figure title) Written summary covers all relevant features of output	Appropriate graphs and tables for metric data Graphs suitably edited for presentation (e.g. figure title) Written summary covers all relevant features of output and is clearly and concisely written
Question 4 Short Answer Question [3 marks]	No attempt / incorrect	Partial information provided	Correct Conclusion Appropriate statistics quoted	Correct Conclusion Appropriate statistics quoted Statistics correctly interpreted to give full justification for conclusion.	Not applicable
Q5 a [1 mark]	Incorrect answer and / or Incomplete answer	Correct answer- all required specific details included.	Not applicable	Not applicable	Not applicable
Q5 b [2 marks]	Incorrect answer or no answer	Correct answer Partial information provided	Correct answer Full information provided	Not applicable	Not applicable
Q5 c [1 mark]	Incorrect answer / no justification provided	Accurate answer, with correct justification provided	Not applicable	Not applicable	Not applicable
Q5 d [1 mark]	Incorrect answer and / or Incomplete answer	Correct answer - all required specific details included	Not applicable	Not applicable	Not applicable

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, smallest group, spread of percentages [as appropriate], and any other relevant patterns based on the report writing examples in the report writing supplements in Canvas.

Q2 – Should include [*as appropriate*]:

A graph and Explore output appropriate for a metric variable.

A paragraph that includes a description of the centre, spread, outliers if present, and shape of distribution based on the report writing examples in the report writing supplements in Canvas.

Q3 – Should include [*as appropriate*]:

A graph and Explore output appropriate for a metric variable.

A paragraph that includes a description of the centre, spread, outliers if present, and shape of distribution based on the report writing examples in the report writing supplements in Canvas.

Q4 and Q5 – The answers should be presented in written form, with sections (a) to (d) in Question 5 clearly identified

Checklist:

- Correct variable used to produce output
- Correct procedure performed
- Graphs appropriately edited and labelled [e.g. 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