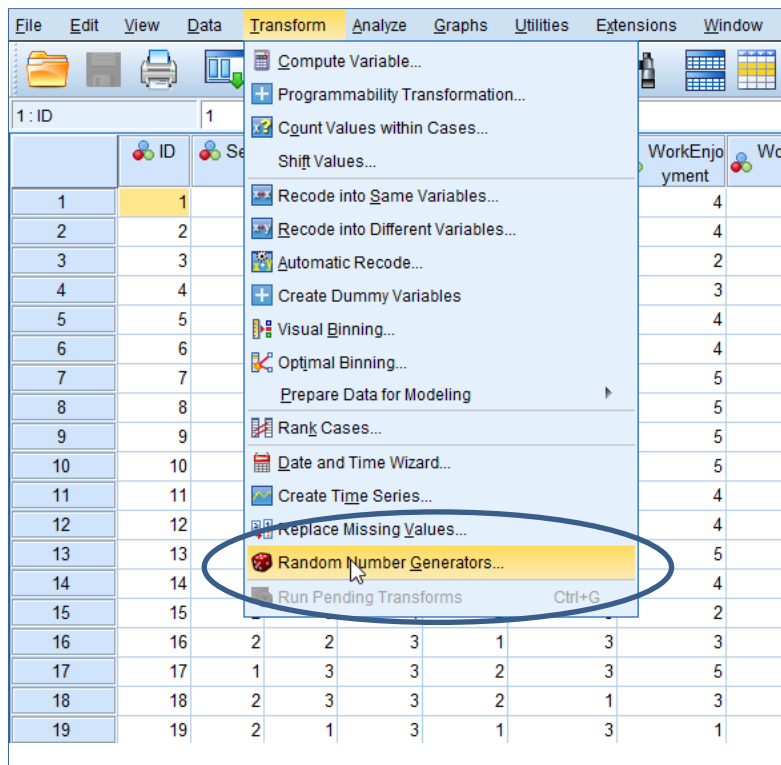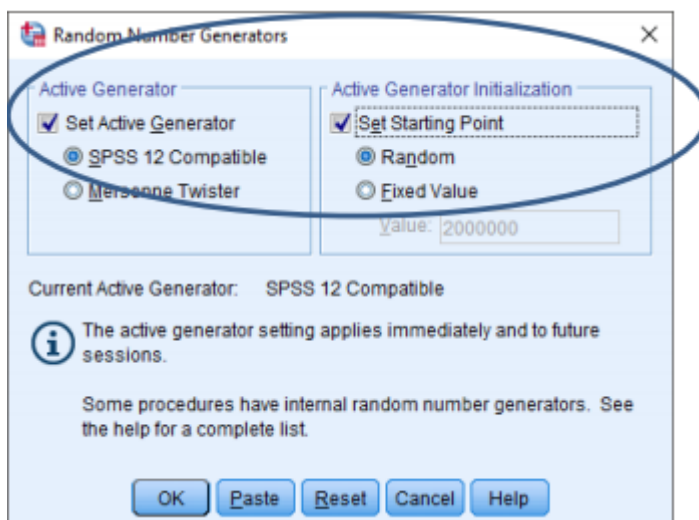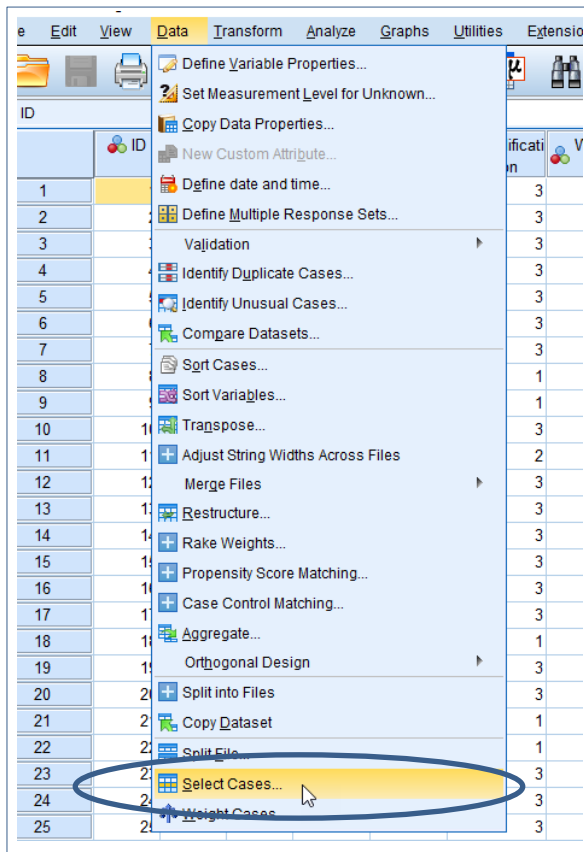# How to generate your random sample of 1500 observations.

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

File	Edit	View	Data	Transform	Analyze	Graphs	Utilities	Extensions	Window

Transform menu items:
- Compute Variable...
- Programmability Transformation...
- Count Values within Cases...
- Shift Values...
- Recode into Same Variables...
- Recode into Different Variables...
- Automatic Recode...
- Create Dummy Variables
- Visual Binning...
- Optimal Binning...
- Prepare Data for Modeling ▸
- Rank Cases...
- Date and Time Wizard...
- Create Time Series...
- Replace Missing Values...
- Random Number Generators...
- Run Pending Transforms   Ctrl+G

1 : ID   1

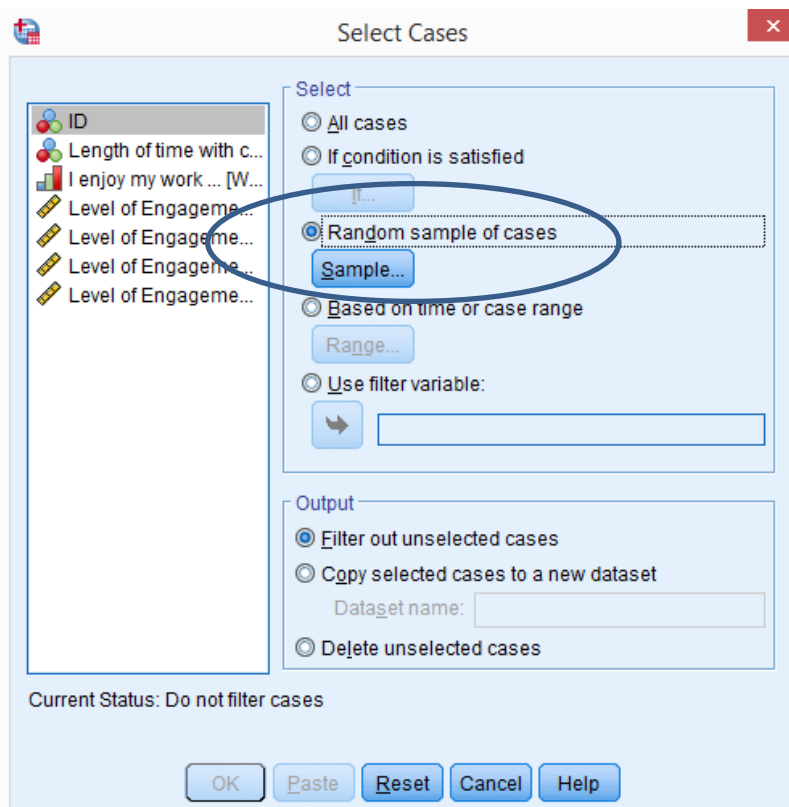|	ID	Se					WorkEnjo yment	Wor
1	1						4	
2	2						4	
3	3						2	
4	4						3	
5	5						4	
6	6						4	
7	7						5	
8	8						5	
9	9						5	
10	10						5	
11	11						4	
12	12						4	
13	13						5	
14	14						4	
15	15						2	
16	16	2	2	3	1	3	3	
17	17	1	3	3	2	3	5	
18	18	2	3	3	2	1	3	
19	19	2	1	3	1	3	1	

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.**

Random Number Generators

Active Generator
- ☑ Set Active Generator
  - ⦿ SPSS 12 Compatible
  - ◯ Mersenne Twister

Active Generator Initialization
- ☑ Set Starting Point
  - ⦿ Random
  - ◯ Fixed Value
    - Value: 2000000

Current Active Generator: SPSS 12 Compatible

ⓘ The active generator setting applies immediately and to future sessions.

Some procedures have internal random number generators. See the help for a complete list.

[OK] [Paste] [Reset] [Cancel] [Help]

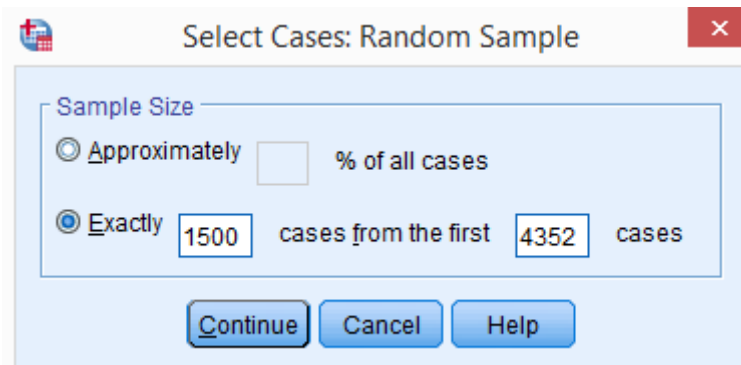3. From the **Data** drop-down menu, choose **Select Cases**

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]:
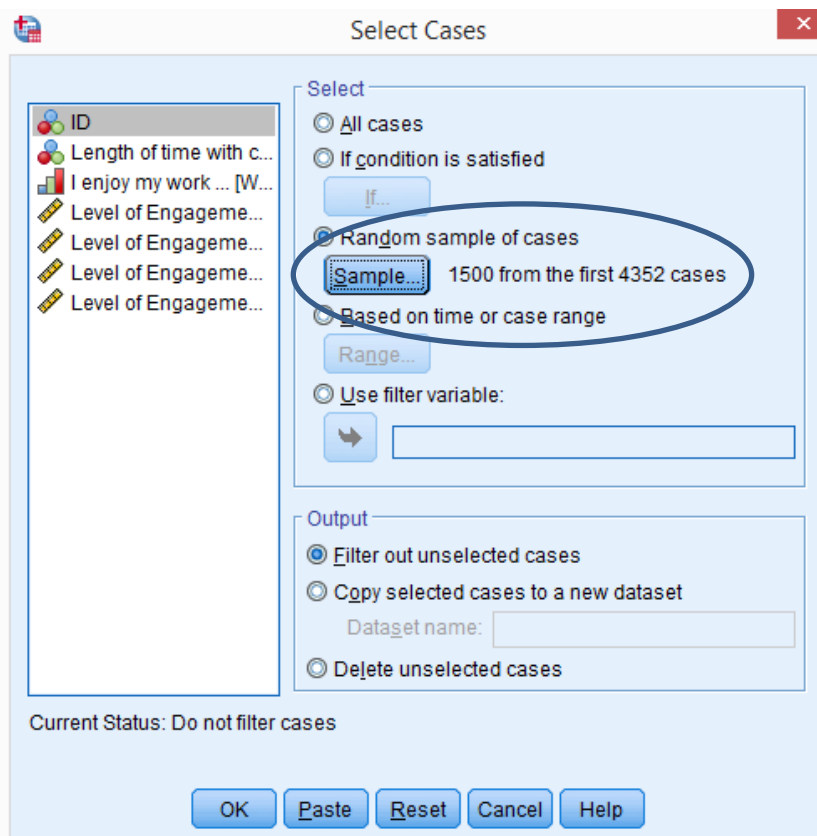
5. From the **Select Cases: Random sample** dialogue box, click **Exactly** and type **1500** cases from the first **4352**.
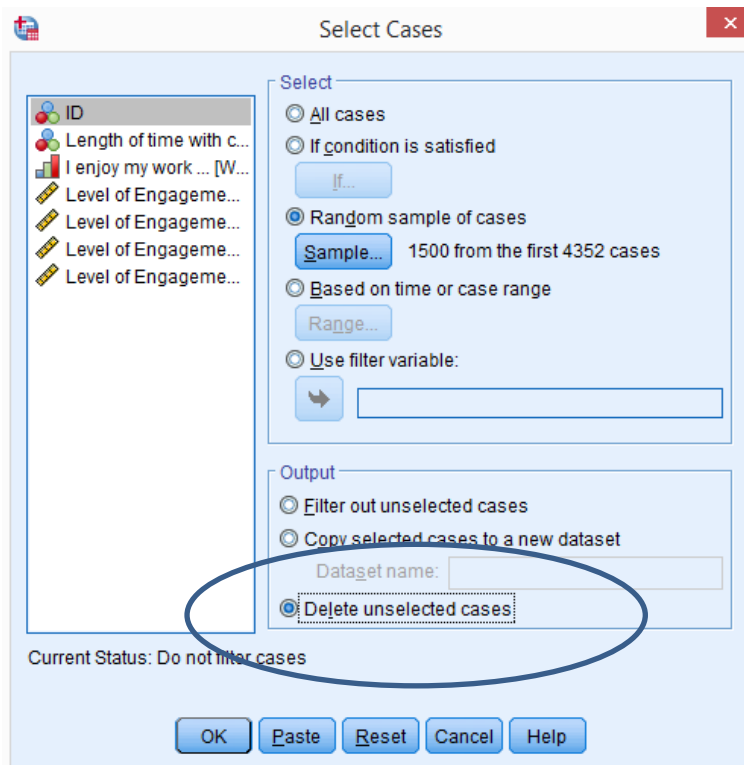
[again the information is pale blue until you select '**Exactly**'. We want to generate a random sample of 1500 from the entire data set, so enter **1500** cases from the first **4352** cases]

Select Cases: Random Sample

Sample Size
○ Approximately [ ] % of all cases
● Exactly [1500] cases from the first [4352] cases

Continue   Cancel   Help

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].

Select Cases

ID
Length of time with c...
I enjoy my work ... [W...
Level of Engageme...
Level of Engageme...
Level of Engageme...
Level of Engageme...

Select
○ All cases
○ If condition is satisfied
  If...
● Random sample of cases
  Sample...   1500 from the first 4352 cases
○ Based on time or case range
  Range...
○ Use filter variable:

Output
● Filter out unselected cases
○ Copy selected cases to a new dataset
  Dataset name: [ ]
○ Delete unselected cases

Current Status: Do not filter cases

OK   Paste   Reset   Cancel   Help

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!