



# CEMASTEA Training on Data Analysis

# GETTING STARTED: Planning for a study

## Several decisions and actions to be made:

1. What is the purpose of the study?
2. What will be the title of the study?
3. Why is the overall objective of the study?
4. What are the specific objectives of the study?
5. What methodology will be employed in carrying out the study?
6. How will the findings be presented?

## Key sections of study report

1. Introduction
2. Literature Review
3. Design and Methodology
4. Findings
5. Conclusion and Recommendations
6. References
7. Annexes

**IMPACT OF SMASE IN-SERVICE EDUCATION AND TRAINING (INSET) PROGRAMME IN TEACHING AND LEARNING: A TRACER STUDY OF PRIMARY AND SECONDARY SCHOOLS IN KENYA**

---

**Main research question**

What is the impact of SMASE INSET on teaching and learning in primary and secondary schools in Kenya?

**Specific research questions**

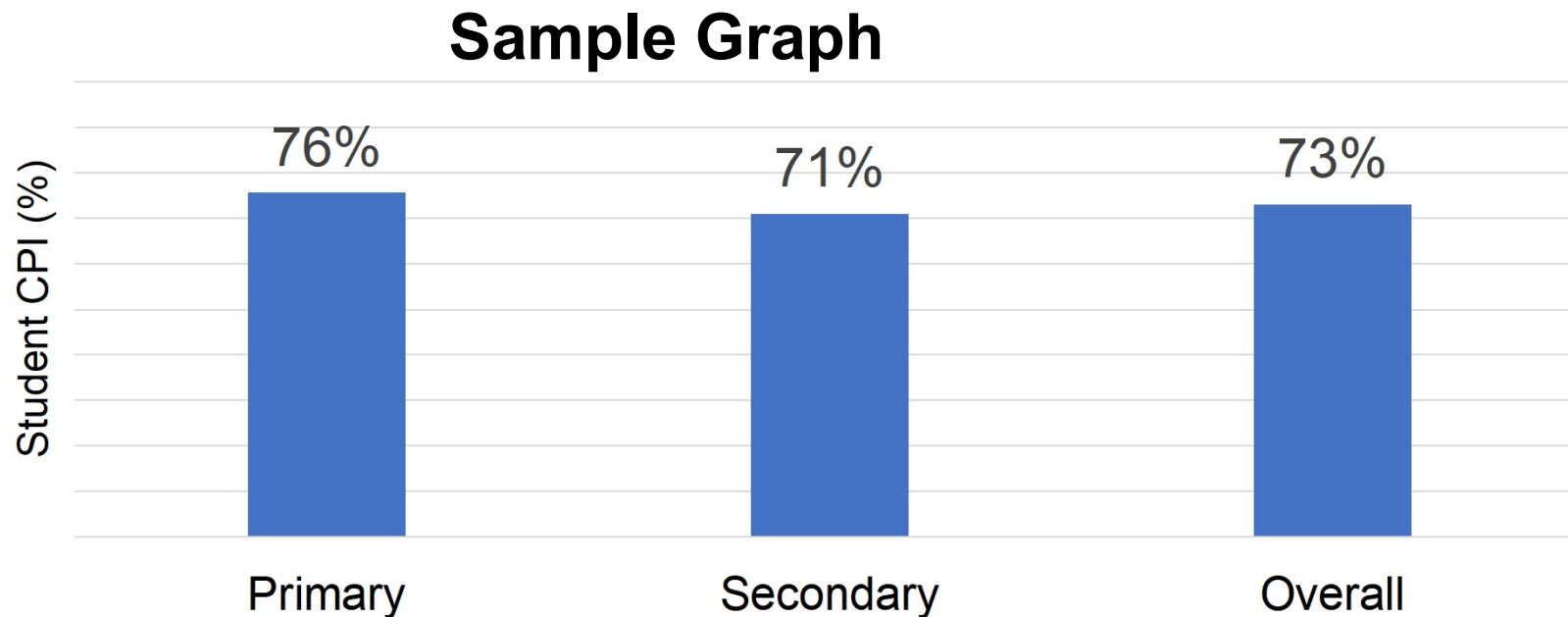
1. How has the SMASE INSET affected learner achievement?
2. To what extent are teachers who have attended SMASE INSET using learner-centred pedagogies?
3. What are the principals' perceptions of the SMASE INSET influence on the teacher's classroom practice?
4. What are the teacher's perceptions of the influence of the SMASE INSET on their classroom practice?
5. What are the learners' perceptions of their classroom experience in learning mathematics and science?
6. What kind of support exists in schools for effectively implementing SMASE INSET teaching and learning strategies?

# Tools Development Matrix

| Specific Questions  | Variables | Type of variables | Measurement Scale | Name of tool | Specific Items |
|---|-----------|-------------------|-------------------|--------------|----------------|
| How has the SMASE INSET affected learner achievement?   |           |                   |                   |              |                |
| To what extent are teachers who have attended SMASE INSET using learner-centred pedagogies?                       |           |                   |                   |              |                |
| What are the principals' perceptions of the SMASE INSET influence on the teacher's classroom practice?            |           |                   |                   |              |                |
| What are the teacher's perceptions of the influence of the SMASE INSET on their classroom practice?               |           |                   |                   |              |                |
| What are the learners' perceptions of their classroom experience in learning mathematics and science?             |           |                   |                   |              |                |
| What kind of support exists in schools for effectively implementing SMASE INSET teaching and learning strategies? |           |                   |                   |              |                |
| Background Information  |           |                   |                   |              |                |

# Dummy Output Tables and Graphs

**For each research question, how will the data be presented?**  
e.g., How has the SMASE INSET affected learner achievement?

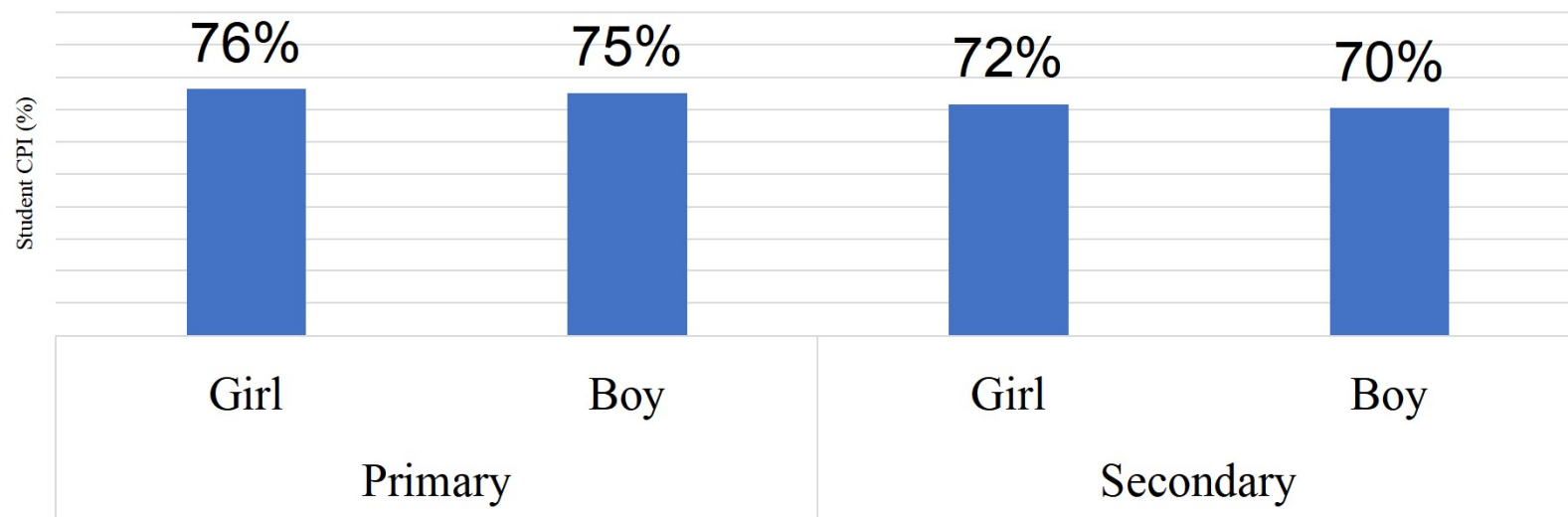


**Figure 4: Learner perception of classroom practice**

# Dummy Output Tables and Graphs

**For each research question, how will the data be presented?**  
e.g., How has the SMASE INSET affected learner achievement?

## Sample Graph



**Figure 5: Student CPI by School Level and Gender and School Level**

# Statistics

## Types of Variables

- **Nominal:** lowest scale, rank only (equality or difference, same or not-the-same).
- **Ordinal:** scale with rank and order only. Measures equality, difference, greater than, less than. However it does not measure the actual distances between the ranks
- **Interval:** scale measures rank, order and differences between ranks: sameness, differences between ranks, greater than, less than, equality or intervals, no true zero.
- **Ratio:** this scale measures rank, order, differences between ranks and there is an absolute zero or true zero.



# Statistics

## 1. Descriptive statistics

- a. Percentages, means, ratios and proportions
- b. Reduce massive amount of data into manageable sizes, summarization, organisation and description of data
- c. Enable visualisation of trends

## 2. Inferential statistics

- a. Correlation, Regression, Chi-Square, T-Tests, ANOVA
- b. Allow generalizations, conclusions, estimates, decisions, predictions and inferring characteristics of the population from the sample characteristics



# Statistics

## 1. Descriptive Analysis

- a. Means, Frequency, Percentages

## 2. Data Quality Checks: Tests of Normality, Reliability of scales, Factor Analysis

## 3. Assess Relationships: Correlation, Regression, Chi-Square

## 4. Group Comparisons

- a. T-Tests: Independent, Paired samples, One Sample
- b. ANOVA: One Way ANOVA, ANCOVA, MANOVA

## 4. Trend Analysis

- a. Repeated Measures ANOVA
- b. Times Series



# Overview of SPSS

- a. **SPSS:** Statistical Package for Social Sciences
- b. Statistical analyses, manipulating data, generating tables & graphs
- c. Analysis of descriptive & advanced inferential statistics
- d. Tools for manipulating & recoding data, computing new variables, merging and aggregating datasets
- e. Display data in the form of tables and graphs



# Overview of SPSS

SPSS has 6 different windows, but 3 are commonly used: *Data Editor*, *Output Viewer*, *Syntax Editor*, *Pivot Table Editor*, *Chart Editor*, and *Text Output Editor*.

- a. **Data Editor** : open at start-up and used to enter and store data
- b. **Output Viewer**: opens automatically when you execute an analysis
- c. **Syntax Window**: used to execute commands and is an alternative to the dialog box



# Overview of SPSS

- Starting SPSS
- Opening an existing data file
- Working with data files
- SPSS windows
- Menus
- Dialog boxes
- Closing SPSS
- Getting help



# ENTERING DATA IN SPSS

- **Data Editor:** displays the contents of the working dataset
- Spreadsheet format with variables in columns and cases in rows
- Two sheets in data editor: Data View & Variable View

# Data Editor Window

\*Tracer\_Secondary\_Teacher\_Dataset.sav [DataSet2] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

18 : Age3 3 Visible: 113 of 113 Variables

|    | Timestamp   | Index_No | Sex1 | County2 | Age3 | Prof_Qual4 | Teaching_Experience5 | SMASE_Attendance6   | Before2012 | Year2012 | Year2013 | Year2014 | Year2015 | Year2016 | Year2017 | Year2018 | Y |
|----|-------------|----------|------|---------|------|------------|----------------------|---------------------|------------|----------|----------|----------|----------|----------|----------|----------|---|
| 1  | 31-Oct-2022 | 1        | 2    | 5       | 4    | 3          | 4                    |                     | 0          | 0        | 0        | 0        | 0        | 1        | 0        | 1        |   |
| 2  | 31-Oct-2022 | 2        | 1    | 5       | 4    | 4          | 4                    | 2016, 2018          | 0          | 0        | 0        | 0        | 0        | 1        | 0        | 1        |   |
| 3  | 31-Oct-2022 | 3        | 1    | 5       | 3    | 3          | 2                    | 2018, 2019          | 0          | 0        | 0        | 0        | 0        | 0        | 0        | 1        |   |
| 4  | 31-Oct-2022 | 4        | 2    | 5       | 3    | 3          | 3                    | Earlier than 201... | 1          | 1        | 1        | 1        | 1        | 1        | 1        | 1        |   |
| 5  | 31-Oct-2022 | 5        | 1    | 5       | 5    | 2          | 4                    | Earlier than 2012   | 1          | 0        | 0        | 0        | 0        | 0        | 0        | 0        |   |
| 6  | 31-Oct-2022 | 6        | 1    | 5       | 2    | 3          | 2                    | 2019, 2020, 20...   | 0          | 0        | 0        | 0        | 0        | 0        | 0        | 0        |   |
| 7  | 31-Oct-2022 | 7        | 1    | 1       | 3    | 3          | 2                    | 2016                | 0          | 0        | 0        | 0        | 0        | 1        | 0        | 0        |   |
| 8  | 31-Oct-2022 | 8        | 1    | 1       | 4    | 4          | 4                    | Earlier than 201... | 1          | 0        | 0        | 0        | 0        | 0        | 0        | 1        |   |
| 9  | 31-Oct-2022 | 9        | 2    | 1       | 4    | 3          | 3                    | 2017                | 0          | 0        | 0        | 0        | 0        | 0        | 1        | 0        |   |
| 10 | 31-Oct-2022 | 10       | 2    | 1       | 4    | 3          | 3                    | 2022                | 0          | 0        | 0        | 0        | 0        | 0        | 0        | 0        |   |
| 11 | 01-Nov-2022 | 11       | 2    | 1       | 3    | 3          | 2                    |                     | 0          | 0        | 0        | 0        | 0        | 0        | 0        | 0        |   |
| 12 | 01-Nov-2022 | 12       | 1    | 1       | 1    | 3          | 1                    |                     | 0          | 0        | 0        | 0        | 0        | 0        | 0        | 0        |   |
| 13 | 01-Nov-2022 | 13       | 1    | 1       | 1    | 3          | 1                    |                     | 0          | 0        | 0        | 0        | 0        | 0        | 0        | 0        |   |
| 14 | 01-Nov-2022 | 14       | 1    | 3       | 3    | 3          | 2                    | 2016, 2017          | 0          | 0        | 0        | 0        | 0        | 1        | 1        | 0        |   |
| 15 | 01-Nov-2022 | 15       | 1    | 3       | 3    | 3          | 3                    | 2013, 2014, 2015    | 0          | 0        | 1        | 1        | 1        | 0        | 0        | 0        |   |
| 16 | 01-Nov-2022 | 16       | 1    | 3       | 4    | 3          | 4                    | Earlier than 201... | 1          | 0        | 0        | 0        | 0        | 0        | 0        | 0        |   |
| 17 | 01-Nov-2022 | 17       | 2    | 3       | 3    | 3          | 3                    | 2020, 2021          | 0          | 0        | 0        | 0        | 0        | 0        | 0        | 0        |   |
| 18 | 01-Nov-2022 | 18       | 1    | 3       | 3    | 3          | 2                    | 2018                | 0          | 0        | 0        | 0        | 0        | 0        | 0        | 1        |   |
| 19 | 01-Nov-2022 | 19       | 1    | 3       | 3    | 3          | 2                    | 2019                | 0          | 0        | 0        | 0        | 0        | 0        | 0        | 0        |   |
| 20 | 01-Nov-2022 | 20       | 1    | 5       | 3    | 2          | 2                    | 2018                | 0          | 0        | 0        | 0        | 0        | 0        | 0        | 1        |   |
| 21 | 01-Nov-2022 | 21       | 1    | 5       | 4    | 3          | 2                    | 2020                | 0          | 0        | 0        | 0        | 0        | 0        | 0        | 0        |   |
| 22 | 01-Nov-2022 | 22       | 1    | 5       | 4    | 2          | 4                    | Earlier than 201... | 1          | 0        | 0        | 0        | 0        | 0        | 1        | 0        |   |
| 23 | 01-Nov-2022 | 23       | 2    | 5       | 4    | 999        | 3                    | 2019, 2021          | 0          | 0        | 0        | 0        | 0        | 0        | 0        | 0        |   |
| 24 | 01-Nov-2022 | 24       | 1    | 5       | 5    | 2          | 5                    | Earlier than 2012   | 1          | 0        | 0        | 0        | 0        | 0        | 0        | 0        |   |
| 25 | 01-Nov-2022 | 25       | 1    | 1       | 3    | 3          | 2                    | 2019                | 0          | 0        | 0        | 0        | 0        | 0        | 0        | 0        |   |
| 26 | 01-Nov-2022 | 26       | 2    | 1       | 3    | 3          | 1                    |                     | 0          | 0        | 0        | 0        | 0        | 0        | 0        | 0        |   |

Data View Variable View

# Variable View Window

Tracer\_Secondary\_Teacher\_Dataset.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

|    | Name                               | Type    | Width | Decimals | Label                              | Values | Missing | Columns | Align | Measure | Role  |
|----|------------------------------------|---------|-------|----------|------------------------------------|--------|---------|---------|-------|---------|-------|
| 59 | Classroom_practice1                | String  | 159   | 0        | Classroom_practice1                | None   | None    | 50      | Left  | Nominal | Input |
| 60 | Classroom_practice2                | String  | 99    | 0        | Classroom_practice2                | None   | None    | 50      | Left  | Nominal | Input |
| 61 | Classroom_practice3                | Numeric | 12    | 0        | Classroom_practice3                | None   | None    | 12      | Right | Nominal | Input |
| 62 | Classroom_practice4                | Numeric | 12    | 0        | Classroom_practice4                | None   | None    | 12      | Right | Nominal | Input |
| 63 | Classroom_practice5                | Numeric | 12    | 0        | Classroom_practice5                | None   | None    | 12      | Right | Nominal | Input |
| 64 | Classroom_practice6                | Numeric | 12    | 0        | Classroom_practice6                | None   | None    | 12      | Right | Nominal | Input |
| 65 | Classroom_practice7                | Numeric | 12    | 0        | Classroom_practice7                | None   | None    | 12      | Right | Nominal | Input |
| 66 | Inyourownopinionwhatistheinflue... | String  | 167   | 0        | Inyourownopinionwhatistheinflue... | None   | None    | 50      | Left  | Nominal | Input |
| 67 | Profession1                        | Numeric | 12    | 0        | Profession1                        | None   | None    | 12      | Right | Nominal | Input |
| 68 | Profession2                        | Numeric | 12    | 0        | Profession2                        | None   | None    | 12      | Right | Nominal | Input |
| 69 | Profession3                        | Numeric | 12    | 0        | Profession3                        | None   | None    | 12      | Right | Nominal | Input |
| 70 | Profession4                        | Numeric | 12    | 0        | Profession4                        | None   | None    | 12      | Right | Nominal | Input |
| 71 | Profession5                        | Numeric | 12    | 0        | Profession5                        | None   | None    | 12      | Right | Nominal | Input |
| 72 | Profession6                        | Numeric | 12    | 0        | Profession6                        | None   | None    | 12      | Right | Nominal | Input |
| 73 | Profession7                        | Numeric | 12    | 0        | Profession7                        | None   | None    | 12      | Right | Nominal | Input |
| 74 | Whatareaswouldyouliketobetrain...  | String  | 150   | 0        | Whatareaswouldyouliketobetrain...  | None   | None    | 50      | Left  | Nominal | Input |
| 75 | Subject_content_mastery1           | Numeric | 12    | 0        | Subject_content_mastery1           | None   | None    | 12      | Right | Nominal | Input |
| 76 | Subject_content_mastery2           | Numeric | 12    | 0        | Subject_content_mastery2           | None   | None    | 12      | Right | Nominal | Input |
| 77 | Subject_content_mastery3           | Numeric | 12    | 0        | Subject_content_mastery3           | None   | None    | 12      | Right | Nominal | Input |
| 78 | Subject_content_mastery4           | Numeric | 12    | 0        | Subject_content_mastery4           | None   | None    | 12      | Right | Nominal | Input |
| 79 | Subject_content_mastery5           | Numeric | 12    | 0        | Subject_content_mastery5           | None   | None    | 12      | Right | Nominal | Input |
| 80 | Subject_content_mastery6           | Numeric | 12    | 0        | Subject_content_mastery6           | None   | None    | 12      | Right | Nominal | Input |
| 81 | Subject_content_mastery7           | Numeric | 12    | 0        | Subject_content_mastery7           | None   | None    | 12      | Right | Nominal | Input |
| 82 | d.Pedagogicalcontentmastery        | String  | 109   | 0        | d.Pedagogicalcontentmastery        | None   | None    | 50      | Left  | Nominal | Input |
| 83 | Pedagogical_content_mastery1       | Numeric | 12    | 0        | Pedagogical_content_mastery1       | None   | None    | 12      | Right | Nominal | Input |
| 84 | Pedagogical_content_mastery2       | Numeric | 12    | 0        | Pedagogical_content_mastery2       | None   | None    | 12      | Right | Nominal | Input |
| 85 | Pedagogical_content_mastery3       | Numeric | 12    | 0        | Pedagogical_content_mastery3       | None   | None    | 12      | Right | Nominal | Input |
| 86 | Pedagogical_content_mastery4       | Numeric | 12    | 0        | Pedagogical_content_mastery4       | None   | None    | 12      | Right | Nominal | Input |

# Output View Window

\*Tracer\_Secondary\_Teacher\_Output.spv [Document1] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Output

- Log
- Descriptives
  - Title
  - Notes
  - Descriptive Statist
- Log
- Crosstabs
  - Title
  - Notes
  - Case Processing
  - Sex1 \* County2 Cr

|                    |     |   |   |   |
|--------------------|-----|---|---|---|
| School_support7    | 177 | 4 | 1 | 5 |
| School_support8    | 177 | 4 | 1 | 5 |
| School_support9    | 174 | 4 | 1 | 5 |
| School_support10   | 175 | 4 | 1 | 5 |
| Valid N (listwise) | 137 |   |   |   |

CROSSTABS  
 /TABLES=Sex1 BY County2  
 /FORMAT=AVALUE TABLES  
 /CELLS=COUNT EXPECTED  
 /COUNT ROUND CELL.

**Crosstabs**

Case Processing

|                |  |       |         |
|----------------|--|-------|---------|
|                |  | Valid |         |
|                |  | N     | Percent |
| Sex1 * County2 |  | 174   | 98.3%   |

Sex1

|       |        |                |      |      |      |      |      |      |       |
|-------|--------|----------------|------|------|------|------|------|------|-------|
|       |        | Bungom         |      |      |      |      |      |      |       |
| Sex1  | Male   | Count          | 20   | 18   | 25   | 14   | 20   | 20   | 117   |
|       |        | Expected Count | 20.2 | 19.5 | 26.9 | 13.4 | 16.1 | 20.8 | 117.0 |
|       | Female | Count          | 10   | 11   | 15   | 6    | 4    | 11   | 57    |
|       |        | Expected Count | 9.8  | 9.5  | 13.1 | 6.6  | 7.9  | 10.2 | 57.0  |
| Total |        | Count          | 30   | 29   | 40   | 20   | 24   | 31   | 174   |
|       |        | Expected Count | 30.0 | 29.0 | 40.0 | 20.0 | 24.0 | 31.0 | 174.0 |

Crosstabs

Row(s): Sex1 [Sex1]

Column(s): County2 [County2]

Layer 1 of 1

Previous Next

Display layer variables in table layers

Display clustered bar charts

Suppress tables

OK Paste Reset Cancel Help

Double-click to activate

Double click to edit Log

IBM SPSS Statistics Processor is ready Unicode:ON H: 104, W: 1249 pt.



# Output View Window

\*Tracer\_Secondary\_Teacher\_Output.spv [Document1] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Output

- Log
- Descriptives
  - Title
  - Notes
  - Descriptive Statist
- Log
- Crosstabs
  - Title
  - Notes
  - Case Processing
  - Sex1 \* County2 Cr

|                    |     |   |   |   |
|--------------------|-----|---|---|---|
| School_support7    | 177 | 4 | 1 | 5 |
| School_support8    | 177 | 4 | 1 | 5 |
| School_support9    | 174 | 4 | 1 | 5 |
| School_support10   | 175 | 4 | 1 | 5 |
| Valid N (listwise) | 137 |   |   |   |

```

CROSSTABS
  /TABLES=Sex1 BY County2
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT EXPECTED
  /COUNT ROUND CELL.
    
```

→ **Crosstabs**

**Case Processing Summary**

|                | Cases |         |         |         |       |         |
|----------------|-------|---------|---------|---------|-------|---------|
|                | Valid |         | Missing |         | Total |         |
|                | N     | Percent | N       | Percent | N     | Percent |
| Sex1 * County2 | 174   | 98.3%   | 3       | 1.7%    | 177   | 100.0%  |

**Sex1 \* County2 Crosstabulation**

|       |        |                | County2 |           |       |        |       |        | Total |
|-------|--------|----------------|---------|-----------|-------|--------|-------|--------|-------|
|       |        |                | Bungoma | Kirinyaga | Kitui | Nakury | Nyeri | Vihiga |       |
| Sex1  | Male   | Count          | 20      | 18        | 25    | 14     | 20    | 20     | 117   |
|       |        | Expected Count | 20.2    | 19.5      | 26.9  | 13.4   | 16.1  | 20.8   | 117.0 |
|       | Female | Count          | 10      | 11        | 15    | 6      | 4     | 11     | 57    |
|       |        | Expected Count | 9.8     | 9.5       | 13.1  | 6.6    | 7.9   | 10.2   | 57.0  |
| Total |        | Count          | 30      | 29        | 40    | 20     | 24    | 31     | 174   |
|       |        | Expected Count | 30.0    | 29.0      | 40.0  | 20.0   | 24.0  | 31.0   | 174.0 |

IBM SPSS Statistics Processor is ready | Unicode:ON



# ENTERING DATA IN SPSS

- Data can be entered directly into SPSS
- Data imported from other applications

# IMPORTING DATA INTO SPSS

Opening Excel Data Source

C:\Users\daniel.muraya\OneDrive - Aga Khan Development Network\Documents\19\_CEMASTE A Training\Tracer Secondary Teachers\_ Questionnaire .xlsx

Read variable names from the first row of data

Worksheet: Teacher\_dataset [A1:DJ178]

Range:

Maximum width for string columns: 32767

OK Cancel Help



# IMPORTING DATA INTO SPSS

- **Importing data from other applications e.g. MS excel**
- To open an Excel file: **File, Open...**
- Select the desired location on disk using the *Look in* option
- Next, select Excel from the *Files of type* drop-down menu.



# IMPORTING DATA INTO SPSS

- Select a spreadsheet from within the Excel Workbook
- Option of reading variable names from the Excel directly into SPSS



# MODIFYING & ORGANIZING DATA

## Creating and Defining Variables

- Variable & value
- Defining variables: for interpretation of outputs and enable SPSS to understand the data
- Define variables in Variable view tab



# MODIFYING & ORGANIZING DATA

- Data Type
- Width
- Decimal
- Label
- Values
- Missing Values
- Measure



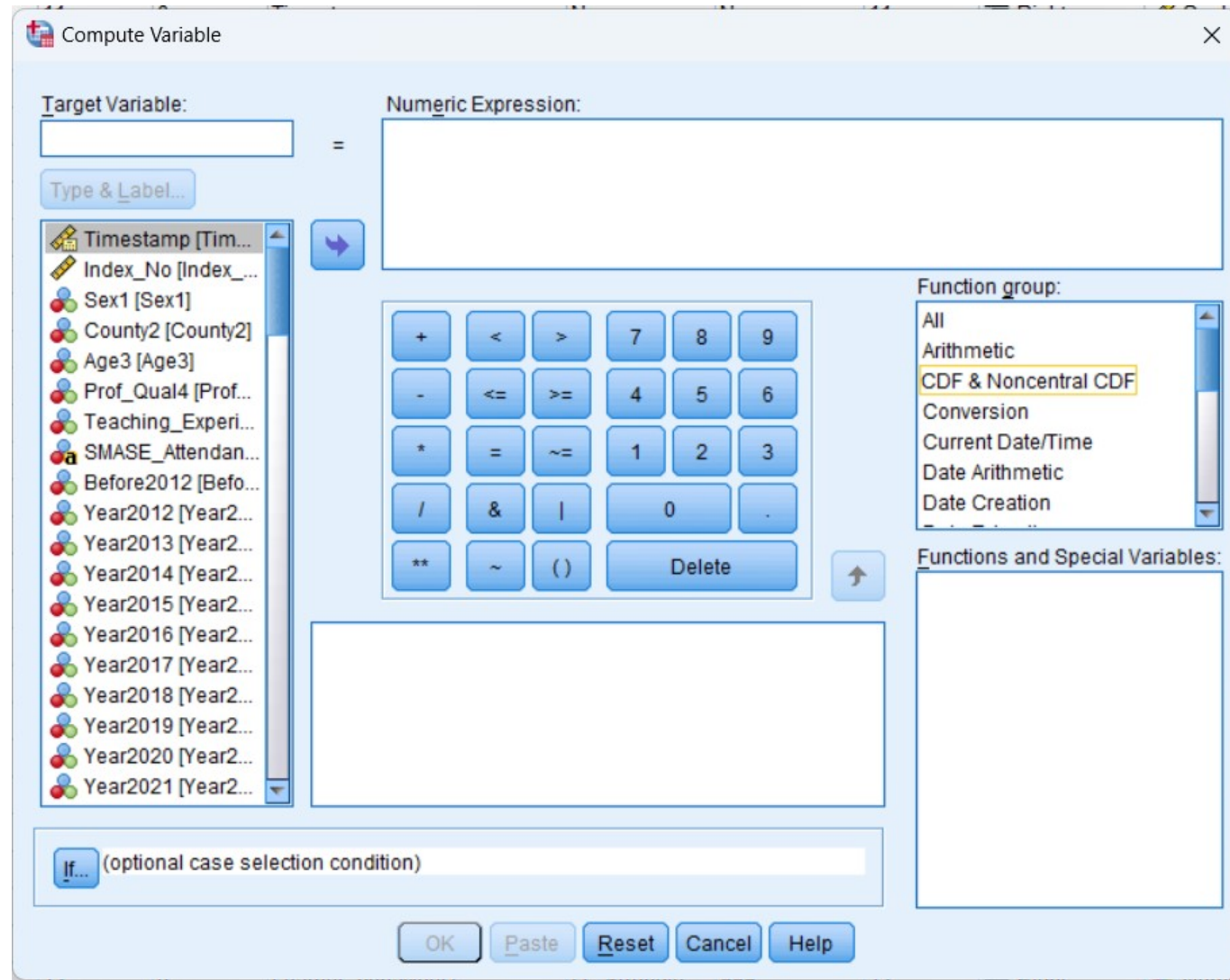
# MODIFYING AND ORGANIZING DATA

## Manipulating data

- **Inserting and Deleting Cases and Variables**
- **Transforming variables:** Computing New Variables, Recoding Variables, Replacing Missing Values



# Computing a New Variable



# Computing a New Variable – Teacher Perception Score

Compute Variable

Target Variable: **Teacher\_Perception\_Score**

Numeric Expression:  $(\text{Teacher\_Perception1} + \text{Teacher\_Perception2} + \text{Teacher\_Perception3} + \text{Teacher\_Perception4} + \text{Teacher\_Perception5} + \text{Teacher\_Perception6} + \text{Teacher\_Perception7} + \text{Teacher\_Perception8} + \text{Teacher\_Perception9} + \text{Teacher\_Perception10} + \text{Teacher\_Perception11} + \text{Teacher\_Perception12})/60*100$

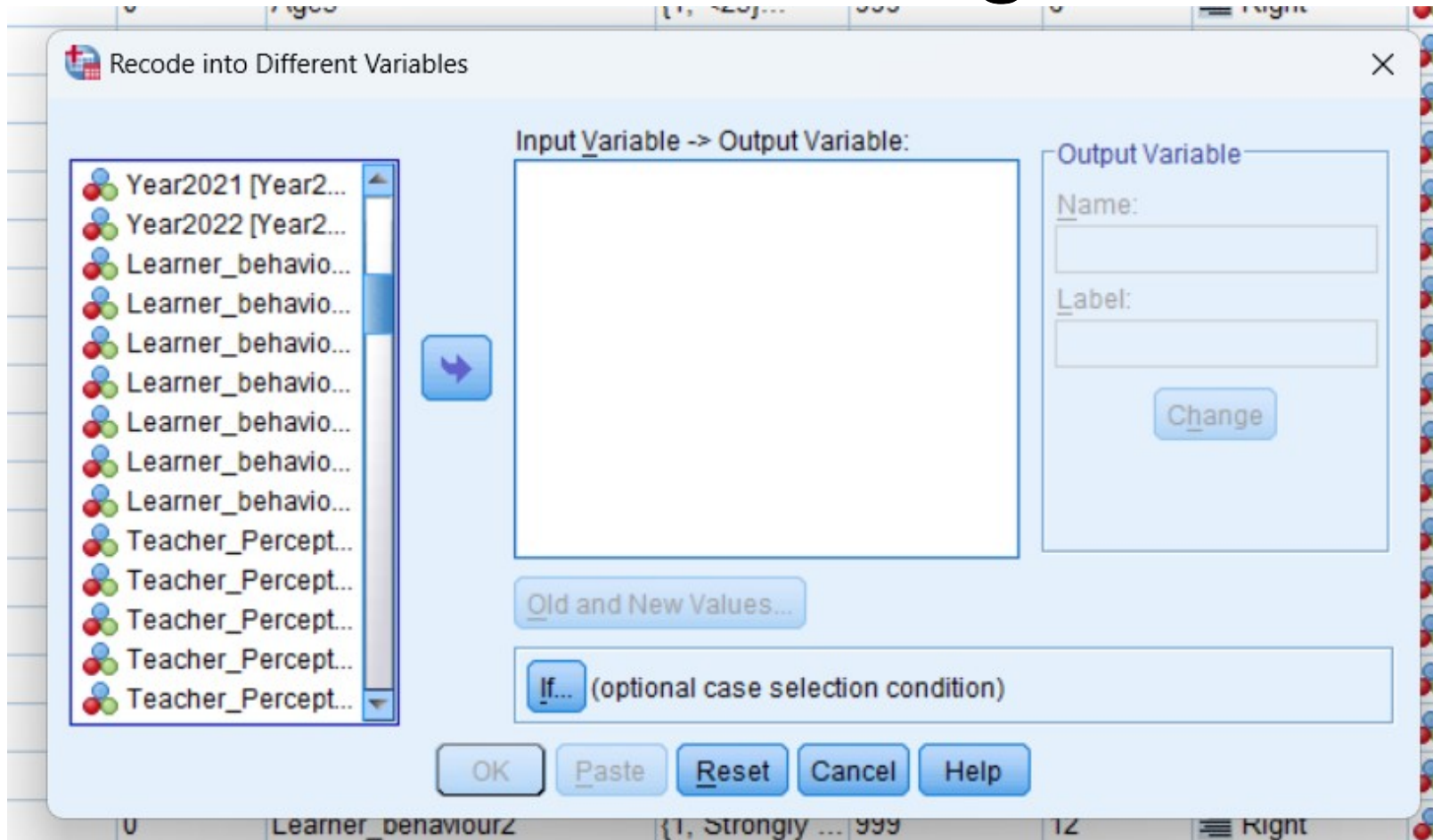
Function group: All

Functions and Special Variables:

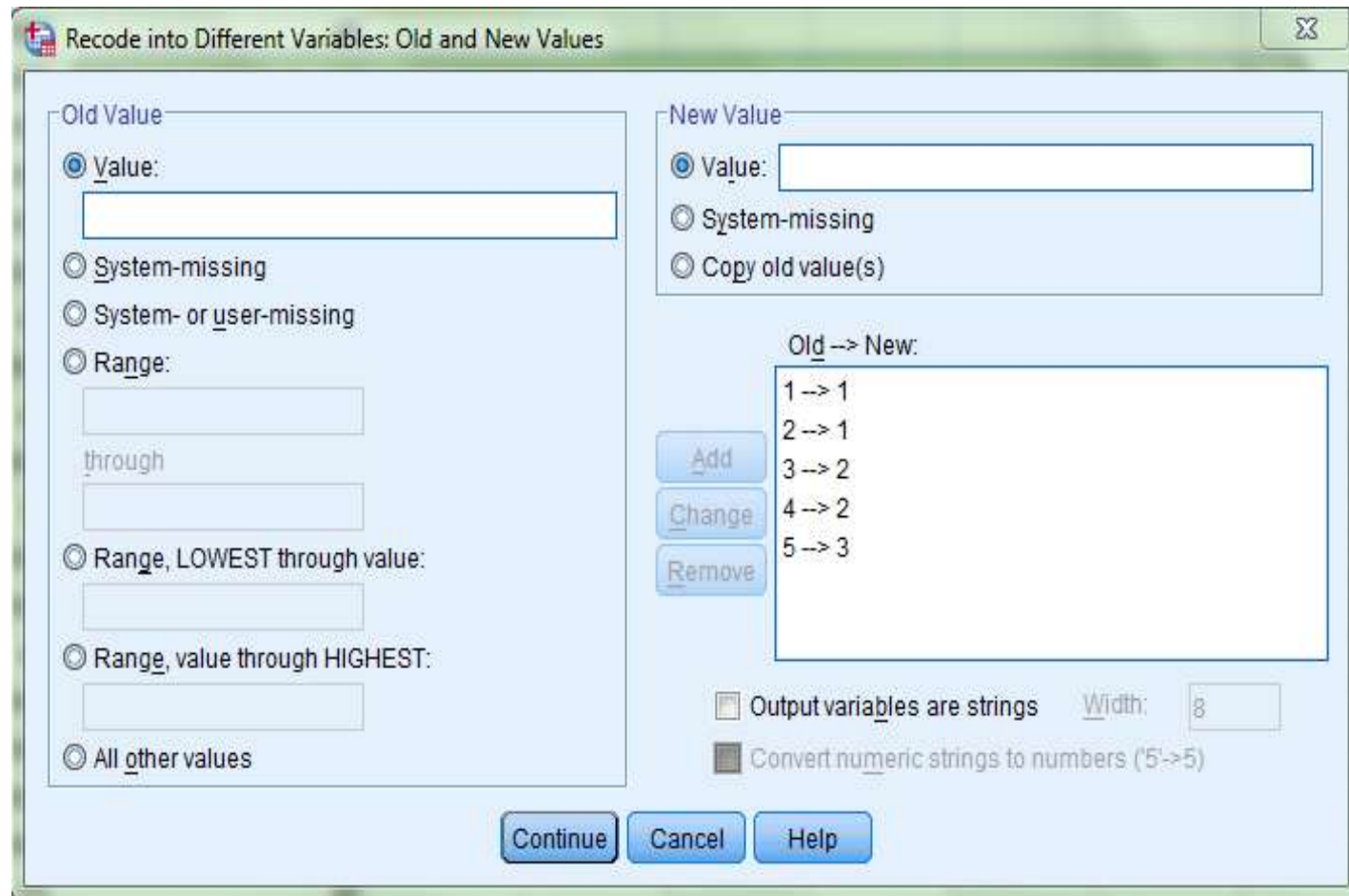
If... (optional case selection condition)

OK Paste Reset Cancel Help

# Recoding Variables



# Recoding Variables



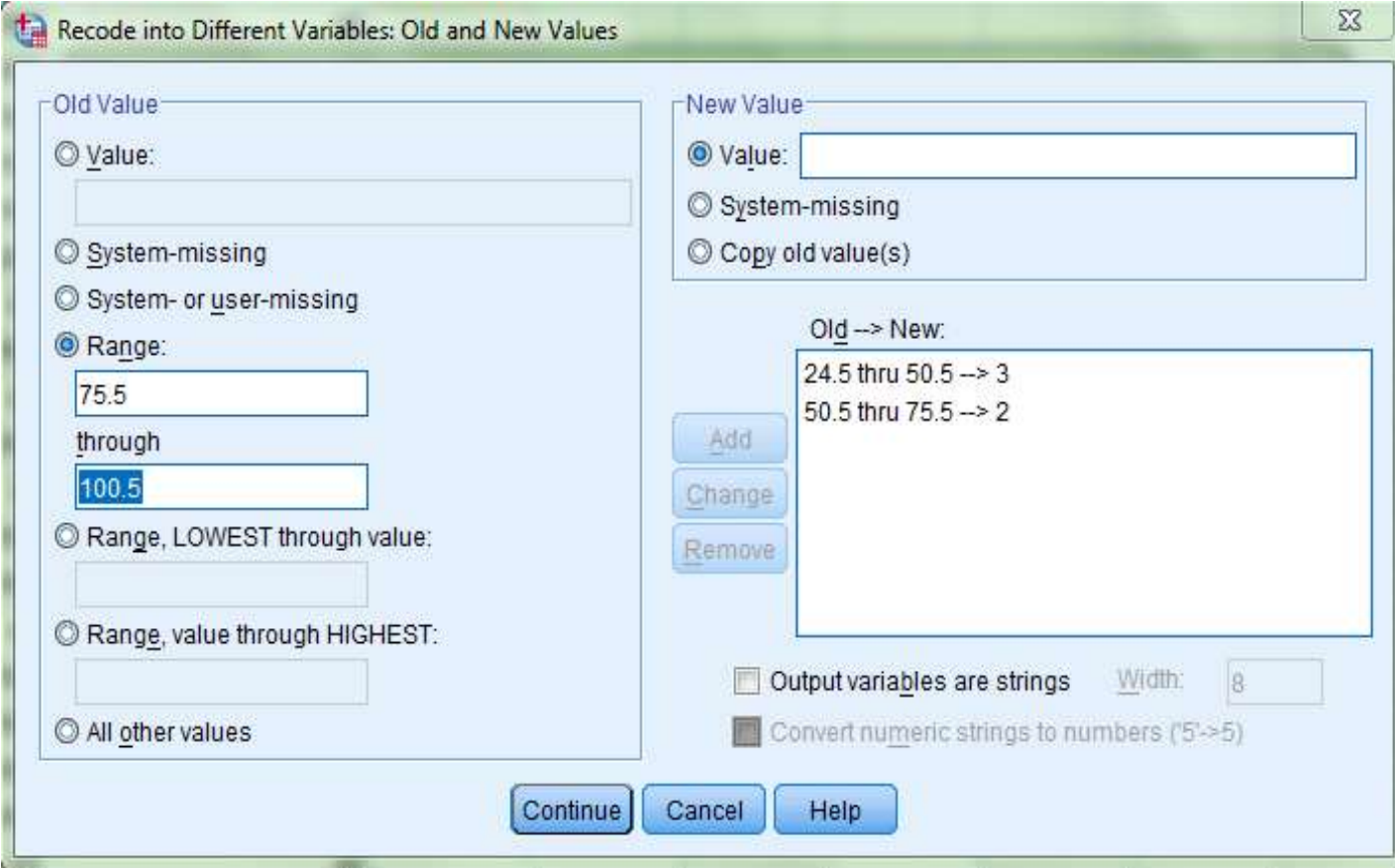
The image shows the 'Recode into Different Variables: Old and New Values' dialog box in SPSS. The window title is 'Recode into Different Variables: Old and New Values'. The dialog is divided into two main sections: 'Old Value' and 'New Value'. In the 'Old Value' section, the 'Value:' radio button is selected, with an empty text box below it. Other options include 'System-missing', 'System- or user-missing', 'Range:' (with two empty boxes for start and end values), 'Range, LOWEST through value:', 'Range, value through HIGHEST:', and 'All other values'. In the 'New Value' section, the 'Value:' radio button is selected, with an empty text box below it. Other options are 'System-missing' and 'Copy old value(s)'. Below these sections is a list box titled 'Old --> New:' containing the following entries: '1 --> 1', '2 --> 1', '3 --> 2', '4 --> 2', and '5 --> 3'. To the left of this list are three buttons: 'Add', 'Change', and 'Remove'. At the bottom right, there are two checkboxes: 'Output variables are strings' (unchecked) with a 'Width:' field containing the number '8', and 'Convert numeric strings to numbers ('5' -> 5)' (checked). At the bottom center are three buttons: 'Continue', 'Cancel', and 'Help'.



# Recoding a Continuous Variable into Groups

- In some cases, you may establish benchmarks for teacher perception score and then compute the proportion of cases that follow under each category
- For example, you may want to classify self-efficacy into three categories: (a) high; (b) medium; (c) low.

# Recoding a Continuous Variable into Groups



The image shows the 'Recode into Different Variables: Old and New Values' dialog box in SPSS. The 'Old Value' section has the 'Range' option selected, with '75.5' in the first input field and '100.5' in the second. The 'New Value' section has the 'Value' option selected. The 'Old --> New' list contains two entries: '24.5 thru 50.5 --> 3' and '50.5 thru 75.5 --> 2'. The 'Output variables are strings' checkbox is unchecked, and the 'Width' is set to 8. The 'Convert numeric strings to numbers' checkbox is checked. At the bottom are 'Continue', 'Cancel', and 'Help' buttons.

Recode into Different Variables: Old and New Values

Old Value

Value:

System-missing

System- or user-missing

Range:

75.5

through

100.5

Range, LOWEST through value:

Range, value through HIGHEST:

All other values

New Value

Value:

System-missing

Copy old value(s)

Old --> New:

24.5 thru 50.5 --> 3

50.5 thru 75.5 --> 2

Add

Change

Remove

Output variables are strings Width: 8

Convert numeric strings to numbers ('5' -> 5)

Continue Cancel Help



# Work To Do

- Recode the **teacher perception score** into a new variable with three categories:
  - High
  - Medium
  - Low
- **Hint!:** Remember to use the lower and upper class limits

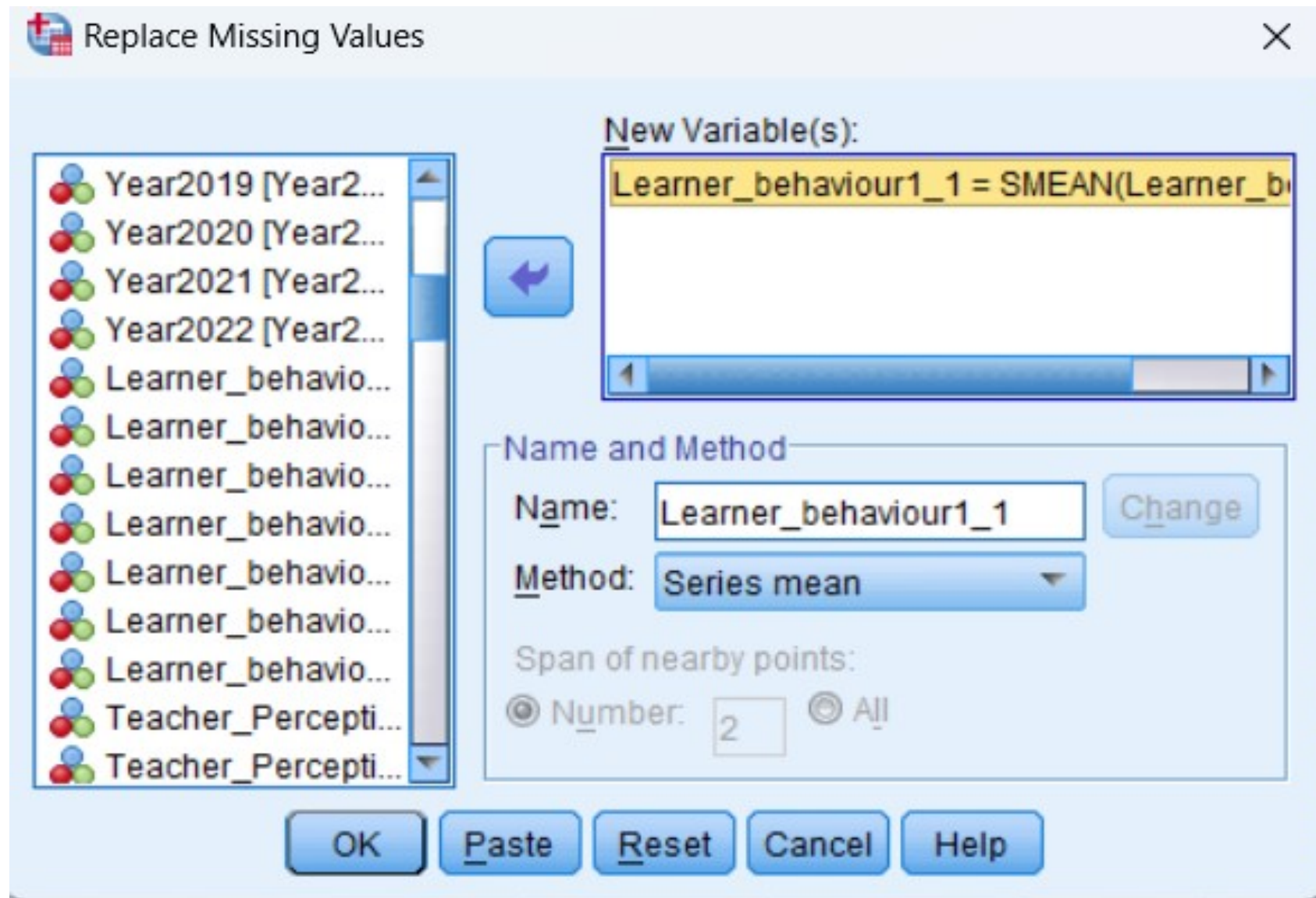


# Replacing Missing Values

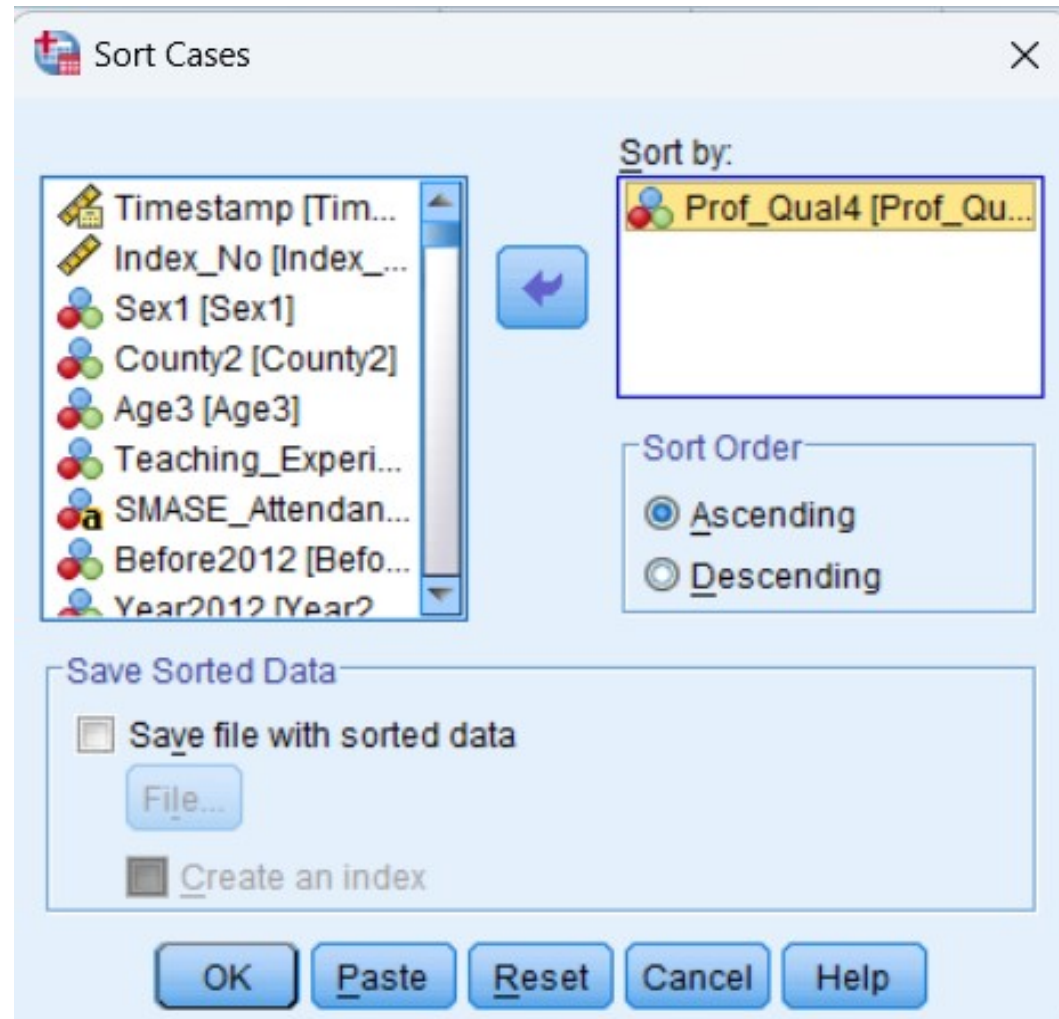
- ***Replacing Missing Categorical Values:*** category of 'unknown'
- ***Replacing Missing Continuous Values***
- SPSS have 5 methods
  - a. Series mean
  - b. Mean of nearby points
  - c. Median of nearby points
  - d. Linear interpolation;
  - e. Linear trend at point



# Replacing Missing Values



# Sorting





# Merging Datasets

- Merging Cases
- Merging Variables

# Selecting cases

The screenshot displays the 'Select Cases: If' dialog box in SPSS. On the left, a list of variables is shown, with 'Sex1 [Sex1]' selected. The central expression field contains the text 'Sex1 = 2'. Below this is a numeric keypad with buttons for arithmetic operators (+, -, \*, /, \*\*, <, >, <=, >=, ~), logical operators (&, |), and a 'Delete' button. To the right, there are two lists: 'Function group:' and 'Functions and Special Variables:'. The 'Function group:' list includes categories like 'All', 'Arithmetic', 'CDF & Noncentral CDF', 'Conversion', 'Current Date/Time', 'Date Arithmetic', 'Date Creation', and 'Date Extraction'. The 'Functions and Special Variables:' list is currently empty. At the bottom of the dialog are three buttons: 'Continue', 'Cancel', and 'Help'.

# Selecting cases

\*Tracer\_Secondary\_Teacher\_Dataset.sav [DataSet2] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

18 : Age3 3 Visible: 114 of 114 Variables

|    | Timestamp   | Index_No | Sex1 | County2 | Age3 | Prof_Qual4 | Teaching_Experien<br>ce5 | SMASE_Attendan<br>ce6 | Before2012 | I2 | Year |
|----|-------------|----------|------|---------|------|------------|--------------------------|-----------------------|------------|----|------|
| 1  | 31-Oct-2022 | 1        | 2    | 5       | 4    | 3          | 4                        |                       | 0          | 0  |      |
| 2  | 31-Oct-2022 | 2        | 1    | 5       | 4    | 4          | 4                        | 2016, 2018            | 0          | 0  |      |
| 3  | 31-Oct-2022 | 3        | 1    | 5       | 3    | 3          | 2                        | 2018, 2019            | 0          | 0  |      |
| 4  | 31-Oct-2022 | 4        | 2    | 5       | 3    | 3          | 3                        | Earlier than 201...   | 1          | 1  |      |
| 5  | 31-Oct-2022 | 5        | 1    | 5       | 5    | 2          | 4                        | Earlier than 2012     | 1          | 1  |      |
| 6  | 31-Oct-2022 | 6        | 1    | 5       | 2    | 3          | 2                        | 2019, 2020, 20...     | 0          | 0  |      |
| 7  | 31-Oct-2022 | 7        | 1    | 1       | 3    | 3          | 2                        | 2016                  | 0          | 0  |      |
| 8  | 31-Oct-2022 | 8        | 1    | 1       | 4    | 4          | 4                        | Earlier than 201...   | 1          | 1  |      |
| 9  | 31-Oct-2022 | 9        | 2    | 1       | 4    | 3          | 3                        | 2017                  | 0          | 0  |      |
| 10 | 31-Oct-2022 | 10       | 2    | 1       | 4    | 3          | 3                        | 2022                  | 0          | 0  |      |
| 11 | 01-Nov-2022 | 11       | 2    | 1       | 3    | 3          | 2                        |                       | 0          | 0  |      |
| 12 | 01-Nov-2022 | 12       | 1    | 1       | 1    | 3          | 1                        |                       | 0          | 0  |      |
| 13 | 01-Nov-2022 | 13       | 1    | 1       | 1    | 3          | 1                        |                       | 0          | 0  |      |
| 14 | 01-Nov-2022 | 14       | 1    | 3       | 3    | 3          | 2                        | 2016, 2017            | 0          | 0  |      |
| 15 | 01-Nov-2022 | 15       | 1    | 3       | 3    | 3          | 3                        | 2013, 2014, 2015      | 0          | 0  |      |
| 16 | 01-Nov-2022 | 16       | 1    | 3       | 4    | 3          | 4                        | Earlier than 201...   | 1          | 1  |      |
| 17 | 01-Nov-2022 | 17       | 2    | 3       | 3    | 3          | 3                        | 2020, 2021            | 0          | 0  |      |

Data View Variable View

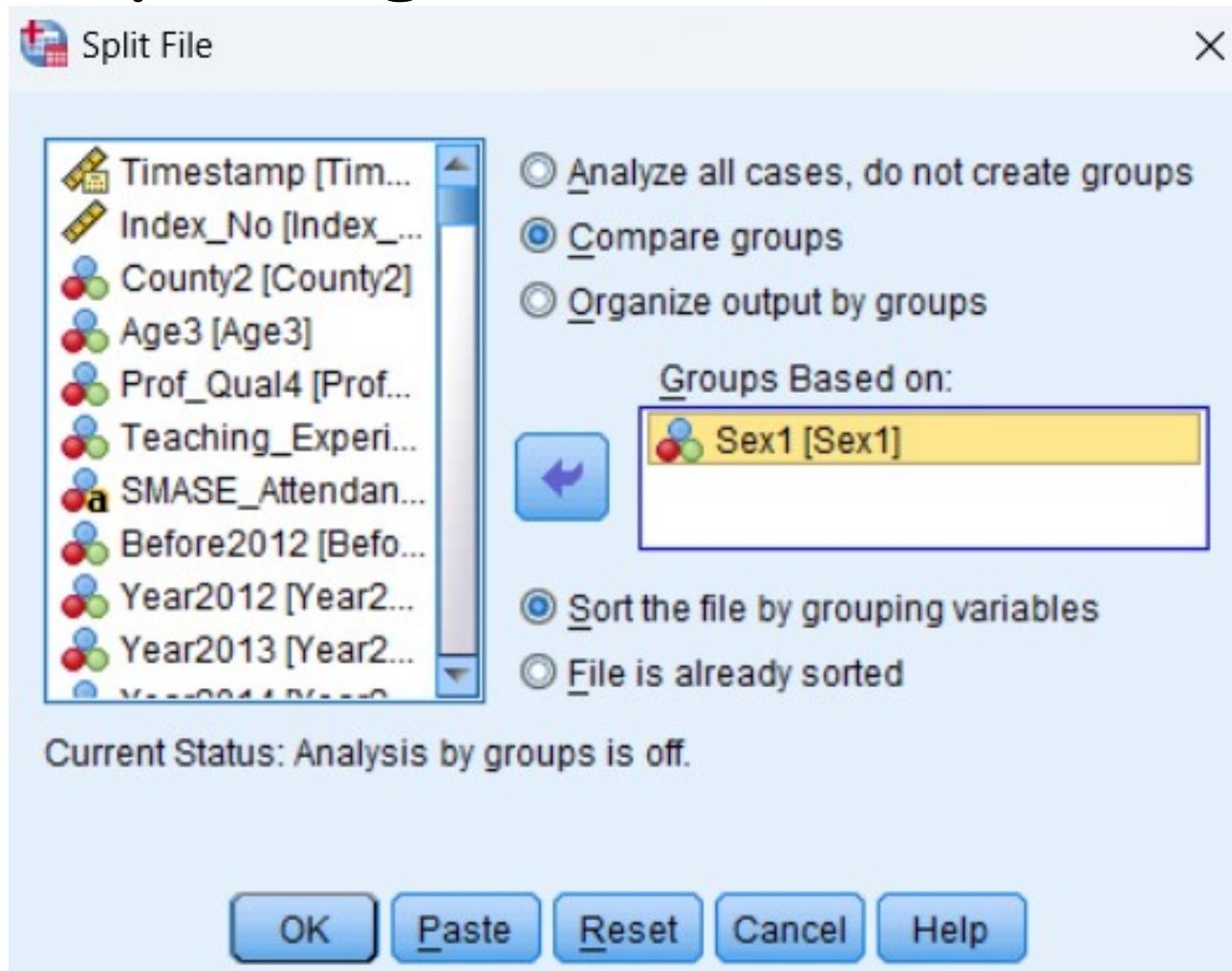
IBM SPSS Statistics Processor is ready Unicode:ON Filter On



# Splitting Files

- You may want to perform analysis on different groups within the same dataset, for example, comparing gender differences

# Splitting Files





# Summarizing Data

## Descriptive statistics

- **Categorical variables**

*Proportion of Teachers by County*

Frequencies and Percentages

## Analyze

### Descriptive Statistics

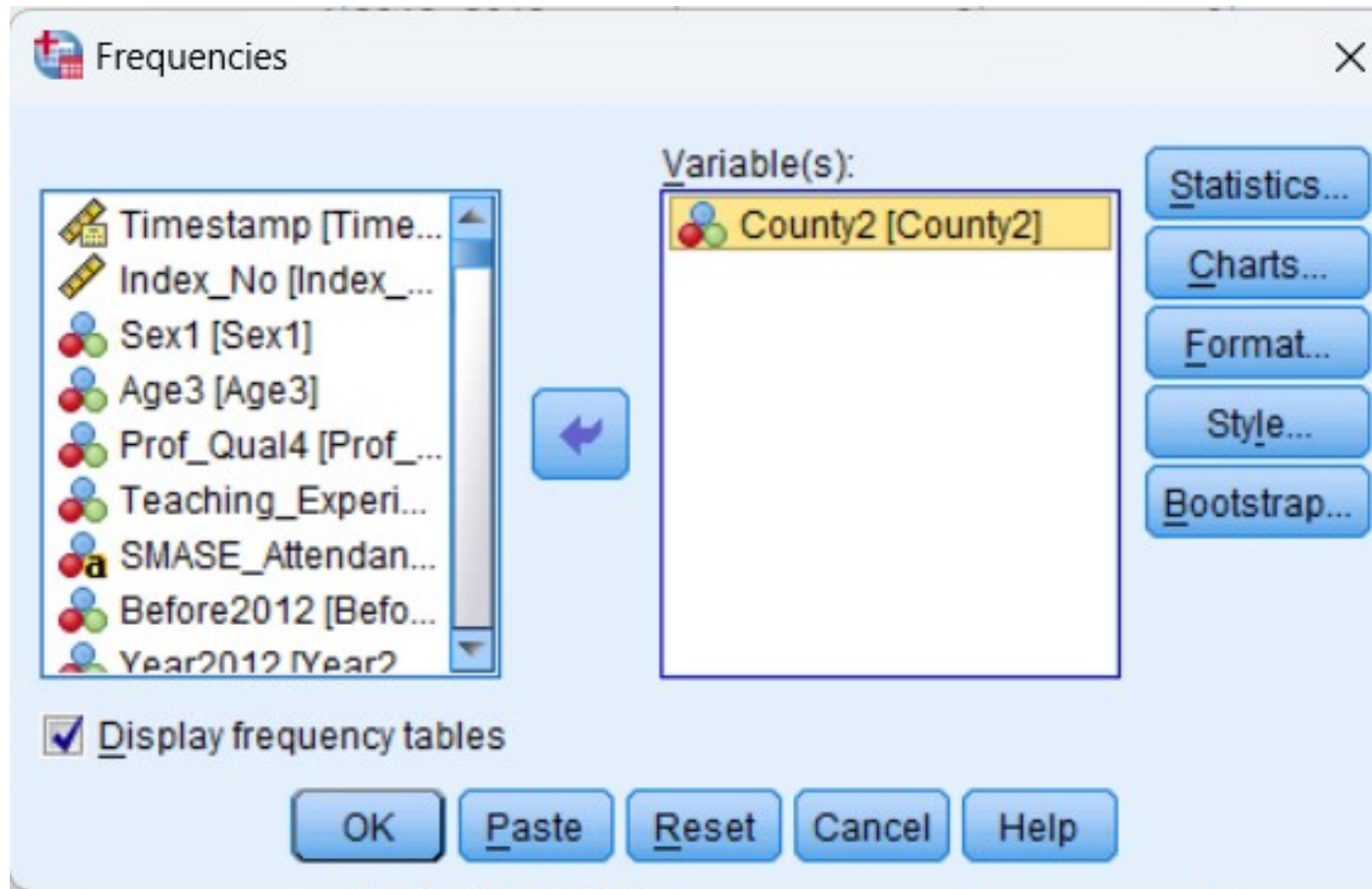
#### Frequency

Select '**County**' and transfer to '**Variables Window**'

- Click '**OK**'



# Proportion of Teachers by County



# Proportion of Teachers by County

\*Tracer\_Secondary\_Teacher\_Output.spv [Document2] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help

VARIABLE LABELS filter\_\$ 'Sex1 = 2 (FILTER)'.  
 VALUE LABELS filter\_\$ 0 'Not Selected' 1 'Selected'.  
 FORMATS filter\_\$ (f1.0).  
 FILTER BY filter\_\$.  
 EXECUTE.  
 FILTER OFF.  
 USE ALL.  
 EXECUTE.  
 SPLIT FILE OFF.

FREQUENCIES VARIABLES=County2  
 /ORDER=ANALYSIS.

→ **Frequencies**

[DataSet1] C:\Users\daniel.muraya\OneDrive - Aga Khan Development Network\Documents\19\_CEMASTE A Training\Tracer\_Secondary\_Teacher\_Dataset.sav

**Statistics**

| County2 |         |     |
|---------|---------|-----|
| N       | Valid   | 177 |
|         | Missing | 0   |

**County2**

|           | Frequency | Percent | Valid Percent | Cumulative Percent |
|-----------|-----------|---------|---------------|--------------------|
| Valid     |           |         |               |                    |
| Bungoma   | 30        | 16.9    | 16.9          | 16.9               |
| Kirinyaga | 30        | 16.9    | 16.9          | 33.9               |
| Kitui     | 40        | 22.6    | 22.6          | 56.5               |
| Nakury    | 20        | 11.3    | 11.3          | 67.8               |
| Nyeri     | 25        | 14.1    | 14.1          | 81.9               |
| Vihiga    | 32        | 18.1    | 18.1          | 100.0              |
| Total     | 177       | 100.0   | 100.0         |                    |

Double-click activate

Tuesday, July 07, 2025 11:55 IBM SPSS Statistics Processor is ready | Unicode ON | H: 682 W: 1249 nt



# 'Work To Do'

1. Proportion of Teachers by Gender
2. Proportion of Teachers by Professional Qualification
3. Proportion of Teachers by Gender
4. Proportion of Teacher by Subject



# Summarizing Data

## ***Cross-tabulations***

- What is the proportion of Teachers by Gender and County?

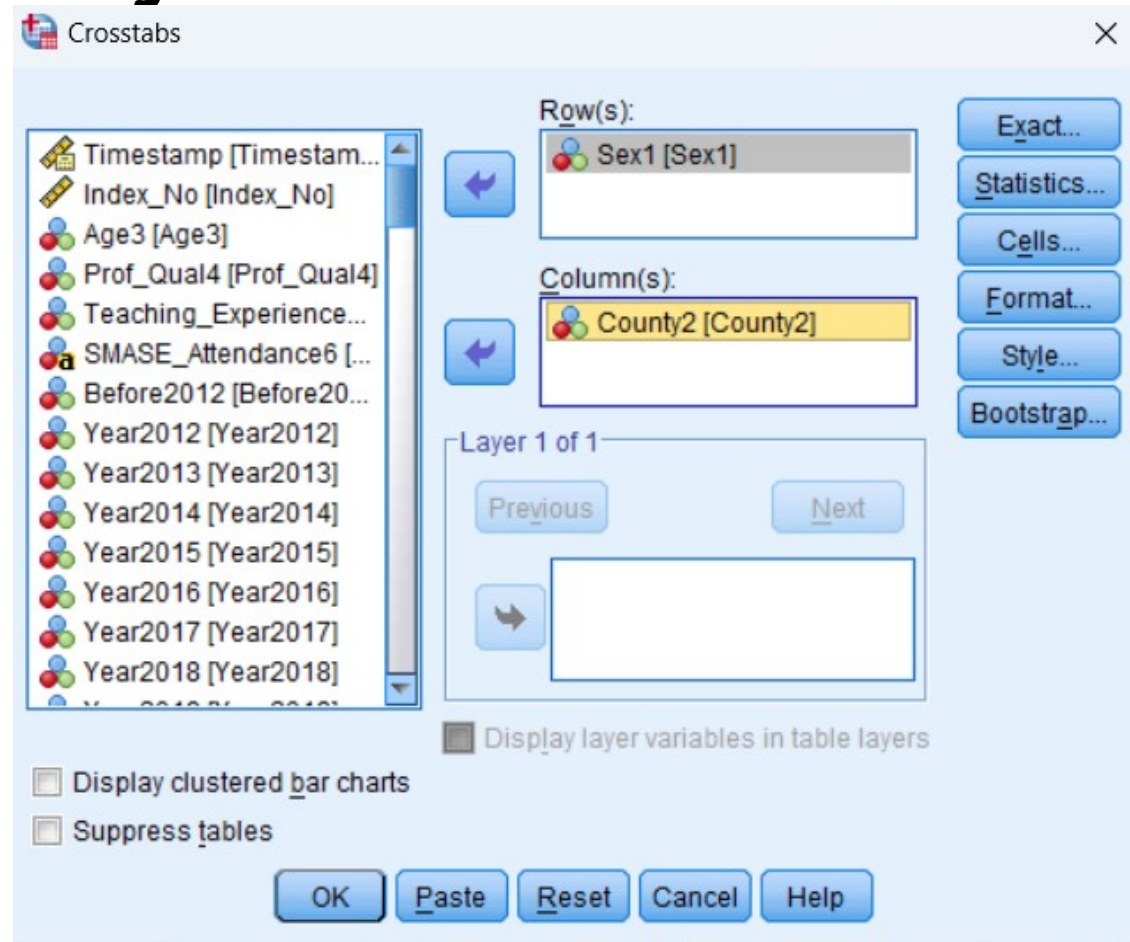
### **Analyze**

### **Descriptive Statistics**

### **Crosstab**

- Select '**Gender**' and transfer to '**Row(s)**'
- Select '**County**' and transfer to '**Column (s)**'

# Cross-Tabulation: Gender & County



# Cross-Tabulation: Gender & County

\*Tracer\_Secondary\_Teacher\_Output.spv [Document2] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Case Processing Summary

|                | Cases |         |         |         |       |         |
|----------------|-------|---------|---------|---------|-------|---------|
|                | Valid |         | Missing |         | Total |         |
|                | N     | Percent | N       | Percent | N     | Percent |
| Sex1 * County2 | 174   | 98.3%   | 3       | 1.7%    | 177   | 100.0%  |

Sex1 \* County2 Crosstabulation

Count

|       |        | County2 |           |       |        |       |        | Total |
|-------|--------|---------|-----------|-------|--------|-------|--------|-------|
|       |        | Bungoma | Kirinyaga | Kitui | Nakuru | Nyeri | Vihiga |       |
| Sex1  | Male   | 20      | 18        | 25    | 14     | 20    | 20     | 117   |
|       | Female | 10      | 11        | 15    | 6      | 4     | 11     | 57    |
| Total |        | 30      | 29        | 40    | 20     | 24    | 31     | 174   |

```

CROSSTABS
  /TABLES=Sex1 BY County2
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ROUND CELL.
    
```

IBM SPSS Statistics Processor is ready | Unicode:ON | H: 682. W: 1249 pt



# **‘Work to Do’**

Conduct Cross tabulations to determine the proportion of teachers by:

1. Gender and professional qualification
2. Identify other relevant tabulations

# Summarizing Data

## Descriptive Statistics

- **Continuous variables**
- [Means, Variance, Standard Deviations, Standard Error of the Mean, Range]

*Teacher Mean Perception Score*

## **Data**

### **Split File**

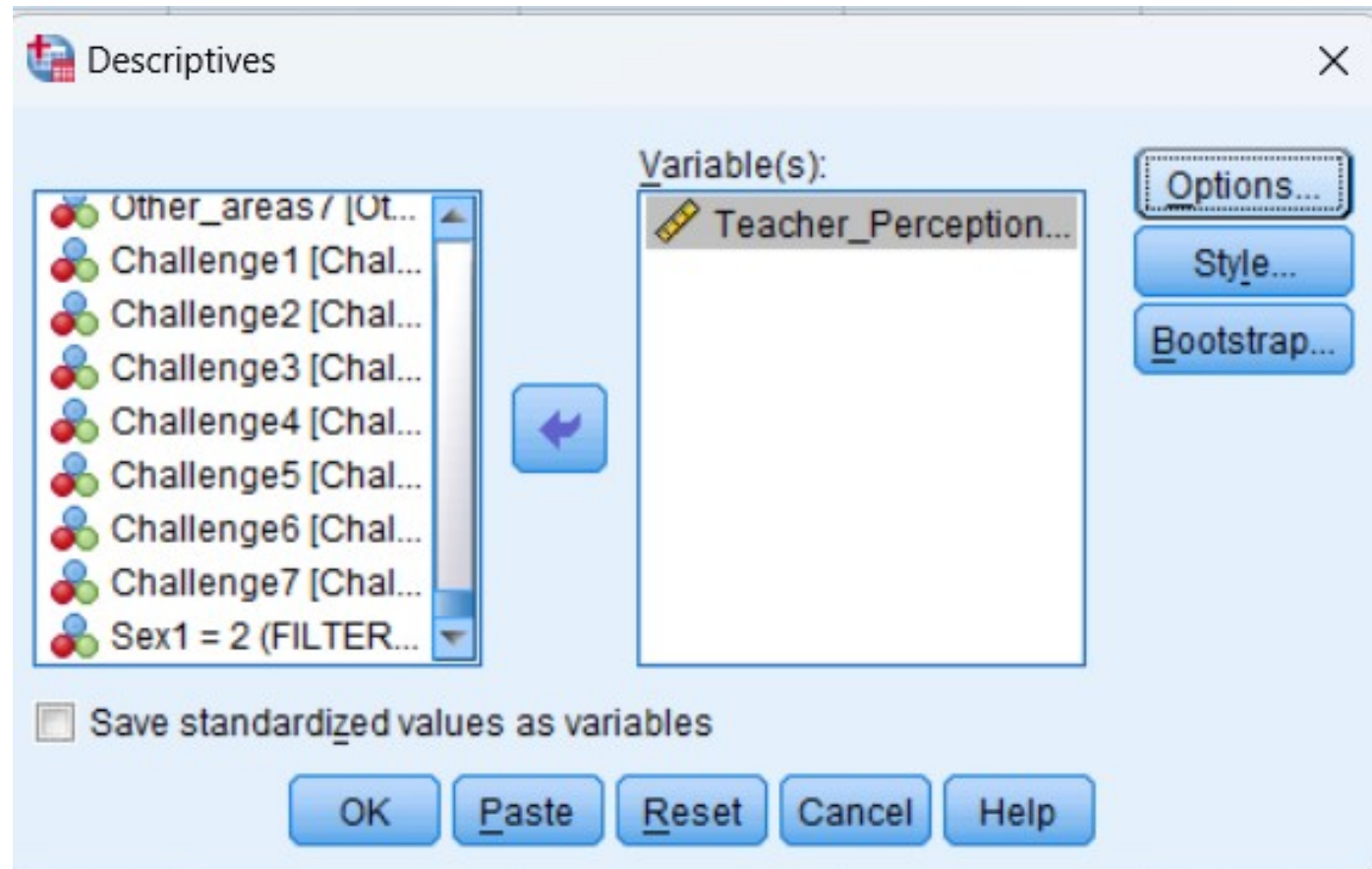
Select '**Compare Groups**'

Select '**Gender**'

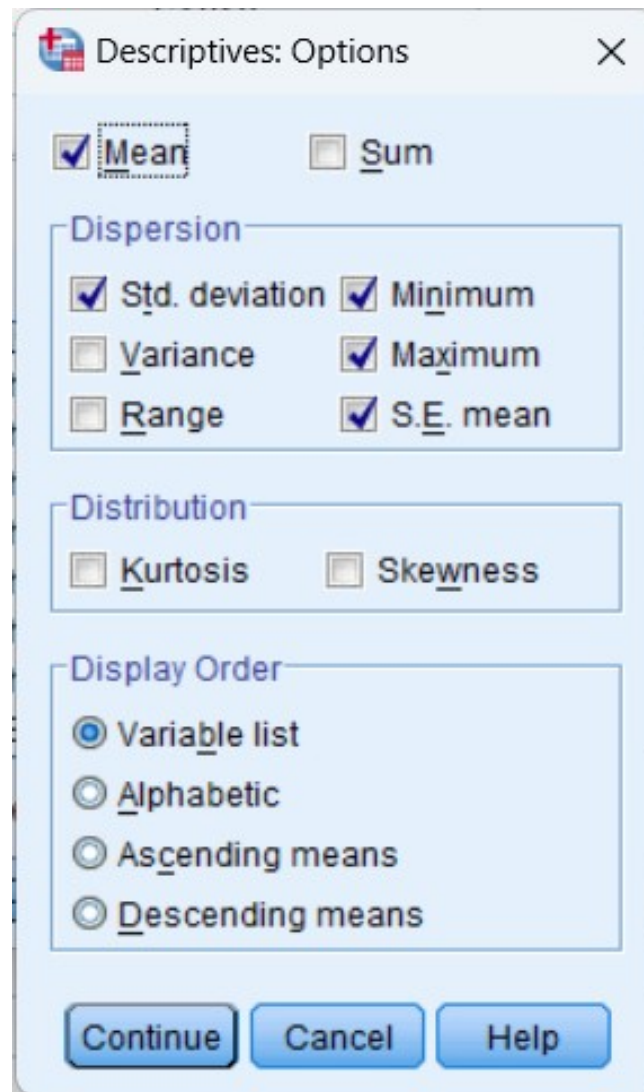
Click '**OK**'



# Summarizing Data



# Summarizing Data



SPSS

# Summarizing Data

IBM SPSS Statistics Viewer

```

SAVE OUTFILE='C:\Users\daniel.muraya\OneDrive - Aga Khan Development '+'
'Network\Documents\19_CEMASTEA Training\Tracer_Secondary_Teacher_Dataset.sav'
/COMPRESSED.
SORT CASES BY Sex1.
SPLIT FILE LAYERED BY Sex1.
DESCRIPTIVES VARIABLES=Teacher_Perception_Score
/STATISTICS=MEAN STDDEV MIN MAX SEMEAN.
    
```

→ Descriptives

| Sex1   |                          | N         | Minimum   | Maximum   | Mean      |            | Std. Deviation |
|--------|--------------------------|-----------|-----------|-----------|-----------|------------|----------------|
|        |                          | Statistic | Statistic | Statistic | Statistic | Std. Error | Statistic      |
| Male   | Teacher_Perception_Score | 106       | 37        | 100       | 86.82     | .927       | 9.541          |
|        | Valid N (listwise)       | 106       |           |           |           |            |                |
| Female | Teacher_Perception_Score | 56        | 50        | 100       | 87.05     | 1.122      | 8.397          |
|        | Valid N (listwise)       | 56        |           |           |           |            |                |
| 999    | Teacher_Perception_Score | 3         | 78        | 98        | 90.56     | 6.186      | 10.715         |
|        | Valid N (listwise)       | 3         |           |           |           |            |                |

IBM SPSS Statistics Processor is ready | Unicode:ON | H: 682, W: 1249 pt.

# Work to Do

Use the Split File Data option to compute descriptive statistics for:

1. Teacher Perception Score by County
2. Teacher Perception Score by Age
3. Teacher Perception Score by Bungoma and Kirinyaga Counties only.



## Assessing Normality - *Kurtosis & Skewness*

Two measures of normality are *Kurtosis* and *Skewness*.

### **Kurtosis**

- **A** measure of the ‘*peakedness*’ or the ‘*flatness*’ of a distribution.
- A value near zero indicates a shape close to normal
- A positive value indicates a distribution more peaked than normal
- A negative values indicates a shape flatter than normal.
- A kurtosis value between  $\pm 1.0$  normal

# Assessing Normality – Teacher Perception Score

Valid N (listwise) 3

```
SPLIT FILE OFF.
DESCRIPTIVES VARIABLES=Teacher_Perception_Score
  /STATISTICS=MEAN KURTOSIS SKEWNESS.
```

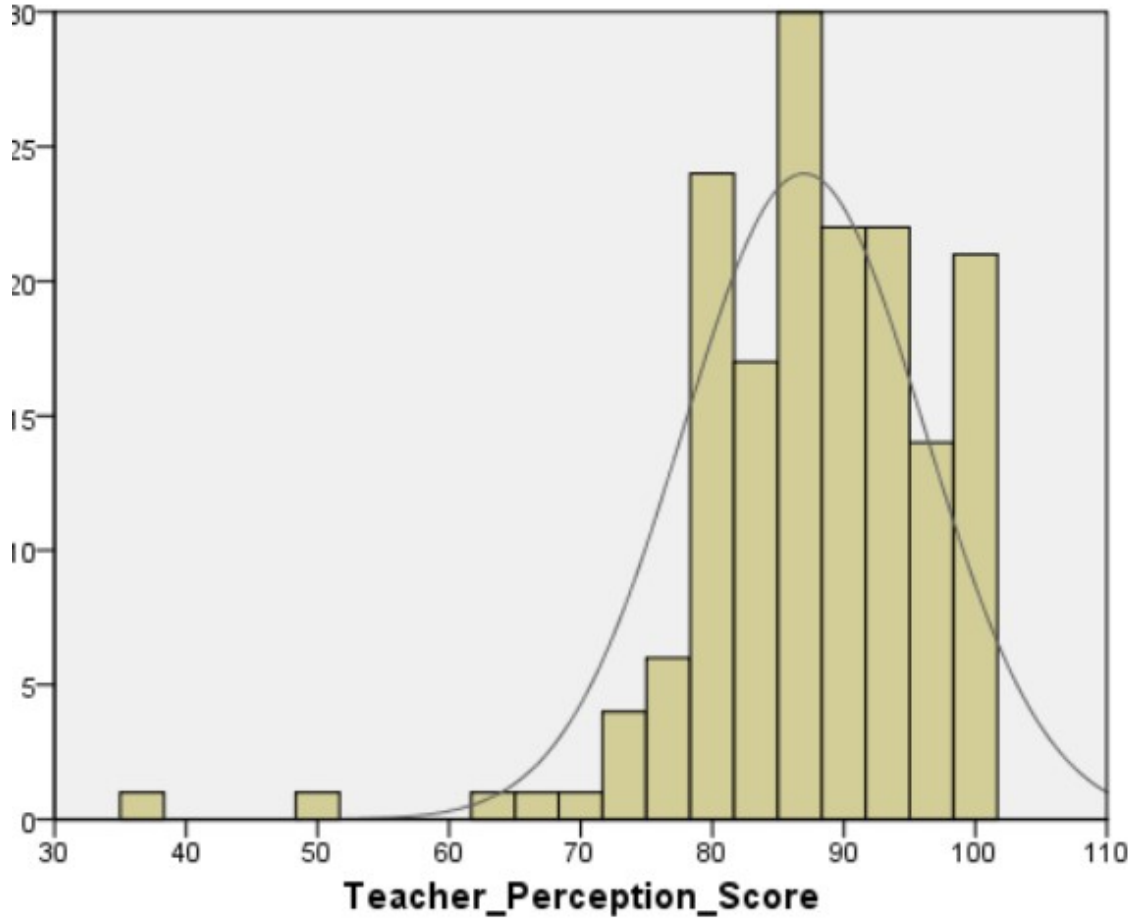
➔ **Descriptives**

**Descriptive Statistics**

|                          | N         | Mean      | Skewness  |            | Kurtosis  |            |
|--------------------------|-----------|-----------|-----------|------------|-----------|------------|
|                          | Statistic | Statistic | Statistic | Std. Error | Statistic | Std. Error |
| Teacher_Perception_Score | 165       | 86.97     | -1.468    | .189       | 5.905     | .376       |
| Valid N (listwise)       | 165       |           |           |            |           |            |

IBM SPSS Statistics Processor is ready | Unicode:ON | H: 239, W: 671 pt.

# Assessing Normality





# Assessing Normality

## Skewness

- Skewness measures the extent a distribution deviates from symmetry around the mean
- A value of zero (0) represents a balanced distribution
- A positive Skewness indicates a greater number of smaller values
- A negative Skewness indicates a greater number of larger values
- A Skewness value between  $\pm 1.0$  is normal



# Assessing Normality – Teacher Perception Score

Valid N (listwise) 3

```
SPLIT FILE OFF.
DESCRIPTIVES VARIABLES=Teacher_Perception_Score
  /STATISTICS=MEAN KURTOSIS SKEWNESS.
```

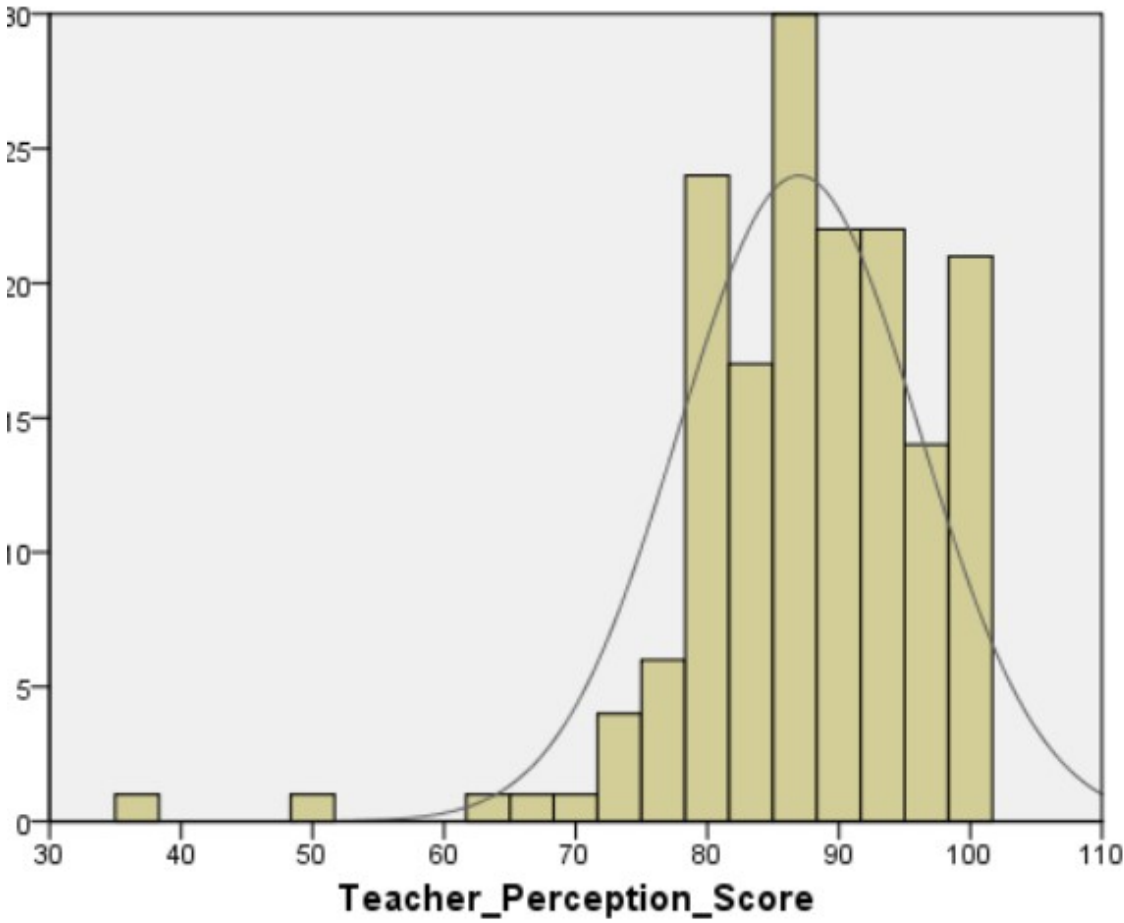
➔ **Descriptives**

**Descriptive Statistics**

|                          | N         | Mean      | Skewness  |            | Kurtosis  |            |
|--------------------------|-----------|-----------|-----------|------------|-----------|------------|
|                          | Statistic | Statistic | Statistic | Std. Error | Statistic | Std. Error |
| Teacher_Perception_Score | 165       | 86.97     | -1.468    | .189       | 5.905     | .376       |
| Valid N (listwise)       | 165       |           |           |            |           |            |

IBM SPSS Statistics Processor is ready | Unicode:ON | H: 239, W: 671 pt

# Assessing Normality



# Assessing Normality – Teacher Perception Score

Valid N (listwise) 3

```

SPLIT FILE OFF.
DESCRIPTIVES VARIABLES=Teacher_Perception_Score
  /STATISTICS=MEAN KURTOSIS SKEWNESS.
    
```

➔ **Descriptives**

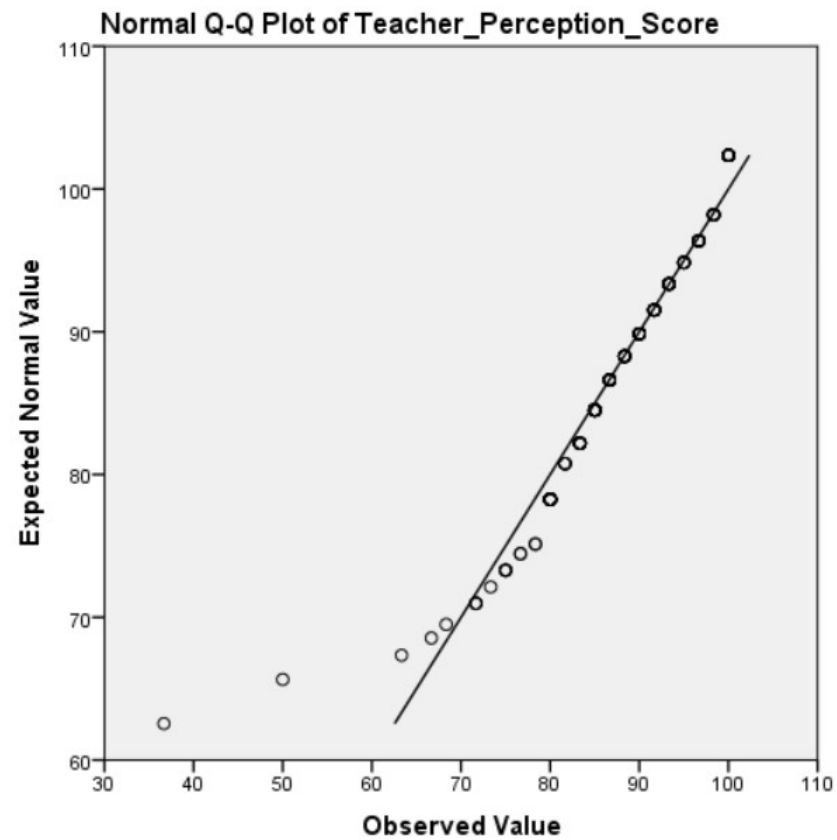
**Descriptive Statistics**

|                          | N         | Mean      | Skewness  |            | Kurtosis  |            |
|--------------------------|-----------|-----------|-----------|------------|-----------|------------|
|                          | Statistic | Statistic | Statistic | Std. Error | Statistic | Std. Error |
| Teacher_Perception_Score | 165       | 86.97     | -1.468    | .189       | 5.905     | .376       |
| Valid N (listwise)       | 165       |           |           |            |           |            |

IBM SPSS Statistics Processor is ready | Unicode:ON | H: 239, W: 671 pt.

# Assessing Normality – Q-Q plots

Teacher\_Perception\_Score





# Using Graphs to Describe & Explore Data

- **SPSS possess two graphing procedures: Legacy Dialogs Graphs & Chart Builder graphs.**
- **Chart Builder was introduced with SPSS version 14**

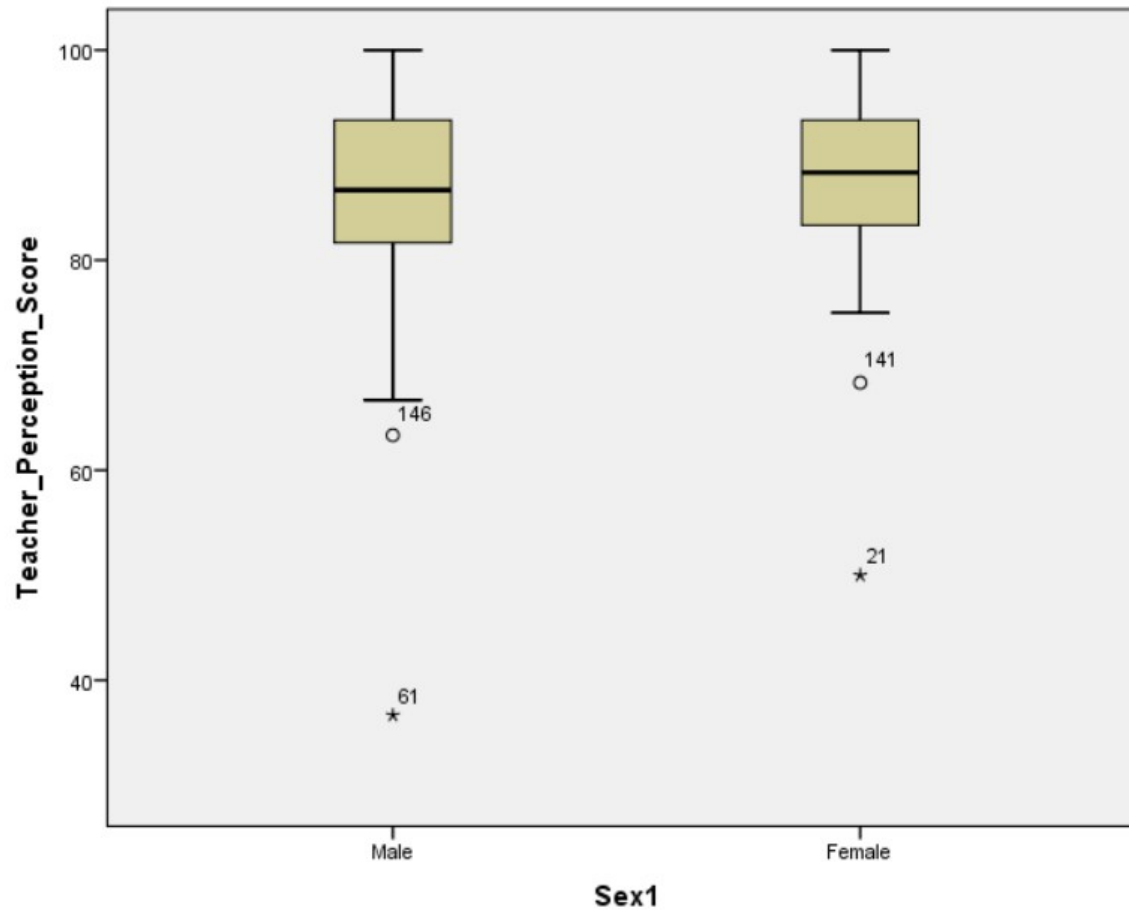


# Using Graphs to Describe & Explore Data

1. Bar graphs
2. Line graphs
3. Pie charts
4. Box plots
5. Histograms
6. Scatter plots

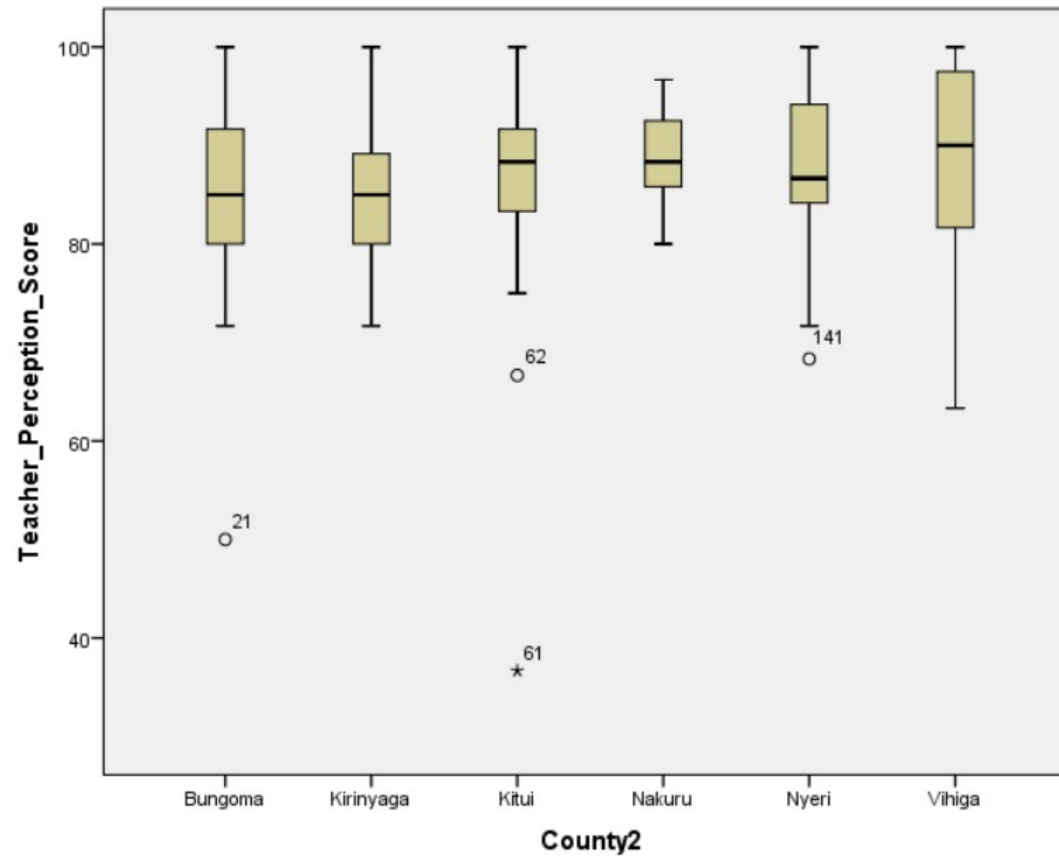
# Using Graphs to Describe & Explore Data – Box Plots

Teacher\_Perception\_Score



# Using Graphs to Describe & Explore Data – Box Plots

Teacher\_Perception\_Score







# Work to Do

Generate the following graph bar graphs;

1. A bar graph of Teacher Perception Score by Gender
2. A box plot of Teacher Perception Score by Gender



## Construction of Composite Scores

- A major challenge in research is that some variables are complex concepts & cannot be measured using one single item
- Researchers therefore design several items that measure such concepts
- Collectively the individual items constitute a scale & their summation generate a measure in form of a composite score
- The quality of the scale is important



# Examples of Composite Scores

## I. **Teacher Perception Score - 12 items**

# Reliability

- For a composite score to be useful, the scale on which it is based must be reliable
- The Cronbach Alpha is one measure of reliability in SPSS
- Measures the internal consistency of the items returns a coefficient ranging from 0-1
- General consensus for a reliable scale is coefficient of 0.7 or above.

# Reliability (Teacher Perception Score)

\*Tracer\_Secondary\_Teacher\_Dataset.sav [DataSet1] - IBM SPSS Statistics Data Editor

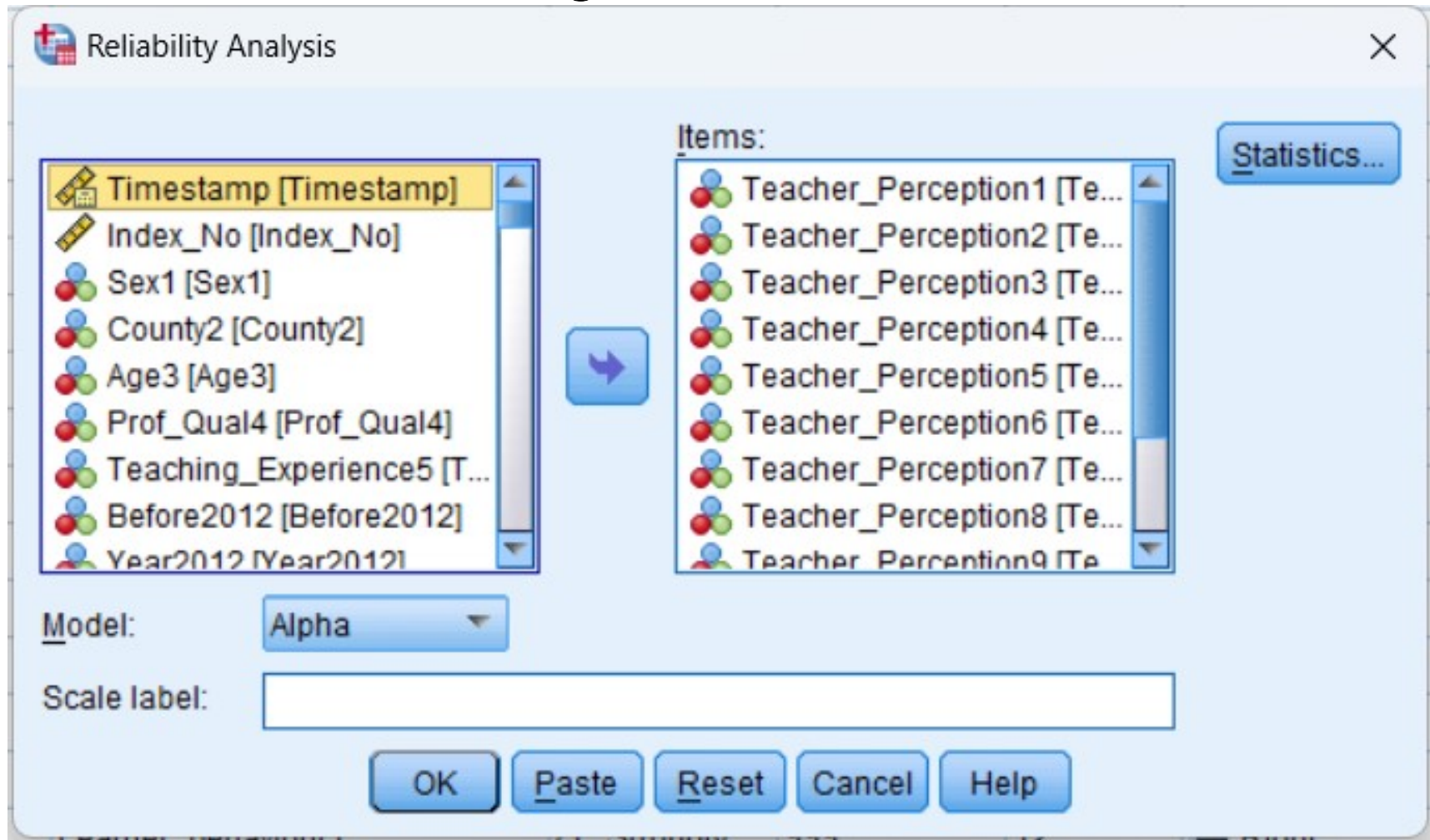
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Reports  
Descriptive Statistics  
Tables  
Compare Means  
General Linear Model  
Generalized Linear Models  
Mixed Models  
Correlate  
Regression  
Loglinear  
Neural Networks  
Classify  
Dimension Reduction  
Scale  
Nonparametric Tests  
Forecasting  
Survival  
Multiple Response  
Missing Value Analysis...  
Multiple Imputation  
Complex Samples  
Simulation...  
Quality Control  
ROC Curve...

|    | Name                 | Label                | Values           | Missing | Columns | Align | Measure | Role  |
|----|----------------------|----------------------|------------------|---------|---------|-------|---------|-------|
| 1  | Timestamp            | Timestamp            | None             | None    | 11      | Right | Scale   | Input |
| 2  | Index_No             | Index_No             | None             | None    | 9       | Right | Scale   | Input |
| 3  | Sex1                 | Sex1                 | {1, Male}...     | 999     | 7       | Right | Nominal | Input |
| 4  | County2              | County2              | {1, Bungom...    | 999     | 8       | Right | Nominal | Input |
| 5  | Age3                 | Age3                 | {1, <25}...      | 999     | 6       | Right | Nominal | Input |
| 6  | Prof_Qual4           | Prof_Qual4           | {1, P1}...       | 999     | 9       | Right | Nominal | Input |
| 7  | Teaching_Experience5 | Teaching_Experience5 | {1, <5}...       | 999     | 12      | Right | Nominal | Input |
| 8  | SMASE_Attendance6    | SMASE_Attendance6    | None             | None    | 11      | Left  | Nominal | Input |
| 9  | Before2012           | Before2012           | None             | None    | 10      | Right | Nominal | Input |
| 10 | Year2012             | Year2012             | None             | None    | 8       | Right | Nominal | Input |
| 11 | Year2013             | Year2013             | None             | None    | 8       | Right | Nominal | Input |
| 12 | Year2014             | Year2014             | None             | None    | 8       | Right | Nominal | Input |
| 13 | Year2015             | Year2015             | None             | None    | 8       | Right | Nominal | Input |
| 14 | Year2016             | Year2016             | None             | None    | 8       | Right | Nominal | Input |
| 15 | Year2017             | Year2017             | None             | None    | 8       | Right | Nominal | Input |
| 16 | Year2018             | Year2018             | None             | None    | 8       | Right | Nominal | Input |
| 17 | Year2019             | Year2019             | None             | None    | 8       | Right | Nominal | Input |
| 18 | Year2020             | Year2020             | None             | None    | 8       | Right | Nominal | Input |
| 19 | Year2021             | Year2021             | None             | None    | 8       | Right | Nominal | Input |
| 20 | Year2022             | Year2022             | None             | None    | 8       | Right | Nominal | Input |
| 21 | Learner_behaviour1   | Learner_behaviour1   | {1, Strongly ... | 999     | 12      | Right | Nominal | Input |
| 22 | Learner_behaviour2   | Learner_behaviour2   | {1, Strongly ... | 999     | 12      | Right | Nominal | Input |
| 23 | Learner_behaviour3   | Learner_behaviour3   | {1, Strongly ... | 999     | 12      | Right | Nominal | Input |
| 24 | Learner_behaviour4   | Learner_behaviour4   | {1, Strongly ... | 999     | 12      | Right | Nominal | Input |
| 25 | Learner_behaviour5   | Learner_behaviour5   | {1, Strongly ... | 999     | 12      | Right | Nominal | Input |
| 26 | Learner_behaviour6   | Learner_behaviour6   | {1, Strongly ... | 999     | 12      | Right | Nominal | Input |
| 27 | Learner_behaviour7   | Learner_behaviour7   | {1, Strongly ... | 999     | 12      | Right | Nominal | Input |
| 28 | Teacher_Perception1  | Teacher_Perception1  | {1, Strongly ... | 999     | 12      | Right | Nominal | Input |

Reliability Analysis...  
Multidimensional Unfolding (PREFSCAL)...  
Multidimensional Scaling (PROXSCAL)...  
Multidimensional Scaling (ALSCAL)...

# Reliability (Teacher Perception Score)



# Reliability (Teacher Perception Score)

## → Reliability

**Scale: ALL VARIABLES**

### Case Processing Summary

|       |                       | N   | %     |
|-------|-----------------------|-----|-------|
| Cases | Valid                 | 165 | 93.2  |
|       | Excluded <sup>a</sup> | 12  | 6.8   |
|       | Total                 | 177 | 100.0 |

a. Listwise deletion based on all variables in the procedure.

### Reliability Statistics

| Cronbach's Alpha | N of Items |
|------------------|------------|
| .892             | 12         |

# 'Work to Do'

1. Using the Teacher dataset, determine the reliability of;
  - a. *Teacher practice score*
  - b. *Learner behaviour score*
  - c. *School support score*
2. In all cases, write a statement on the reliability of the scales as appropriate
3. What other scales will be important to construct in your work? (look through the objectives of the study, questionnaire and data and establish other relevant scales that you will need to construct)





# Inferential Statistics

- **Choosing Bivariate Statistics in Data Analysis**



# Bivariate Correlation

- Most common forms of data analysis
- underlies many other analyses,
- support conclusions after primary analyses
- *Correlations* are a measure of the linear relationship between two continuous variables
- A correlation coefficient has a value ranging from -1 to 1
- Magnitude, direction, significance

# Bivariate Correlation

## Example

Amount of money spent on monthly expenses and amount of money earned per month

## *Hypothesis*

- **H<sub>0</sub>**: There is no relationship between teacher practice score and learner behaviour score
- **H<sub>a</sub>**: There is a relationship between teacher practice score and learner behaviour score

# Bivariate Correlation

Tracer\_Secondary\_Teacher\_Dataset.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Challenge7 filter\_S

Learner\_Behaviour\_Score 999

1 83 999

2 999 999

3 90 999

4 999 999

5 999 999

6 999 999

7 26 26

8 67 49

9 80 51

10 93 51

11 72 51

12 80 54

13 90 54

14 85 57

15 75 57

16 80 60

17 88 60

18 97 60

19 80 60

20 85 60

21 80 63

22 88 66

23 100 66

24 63 66

25 78 66

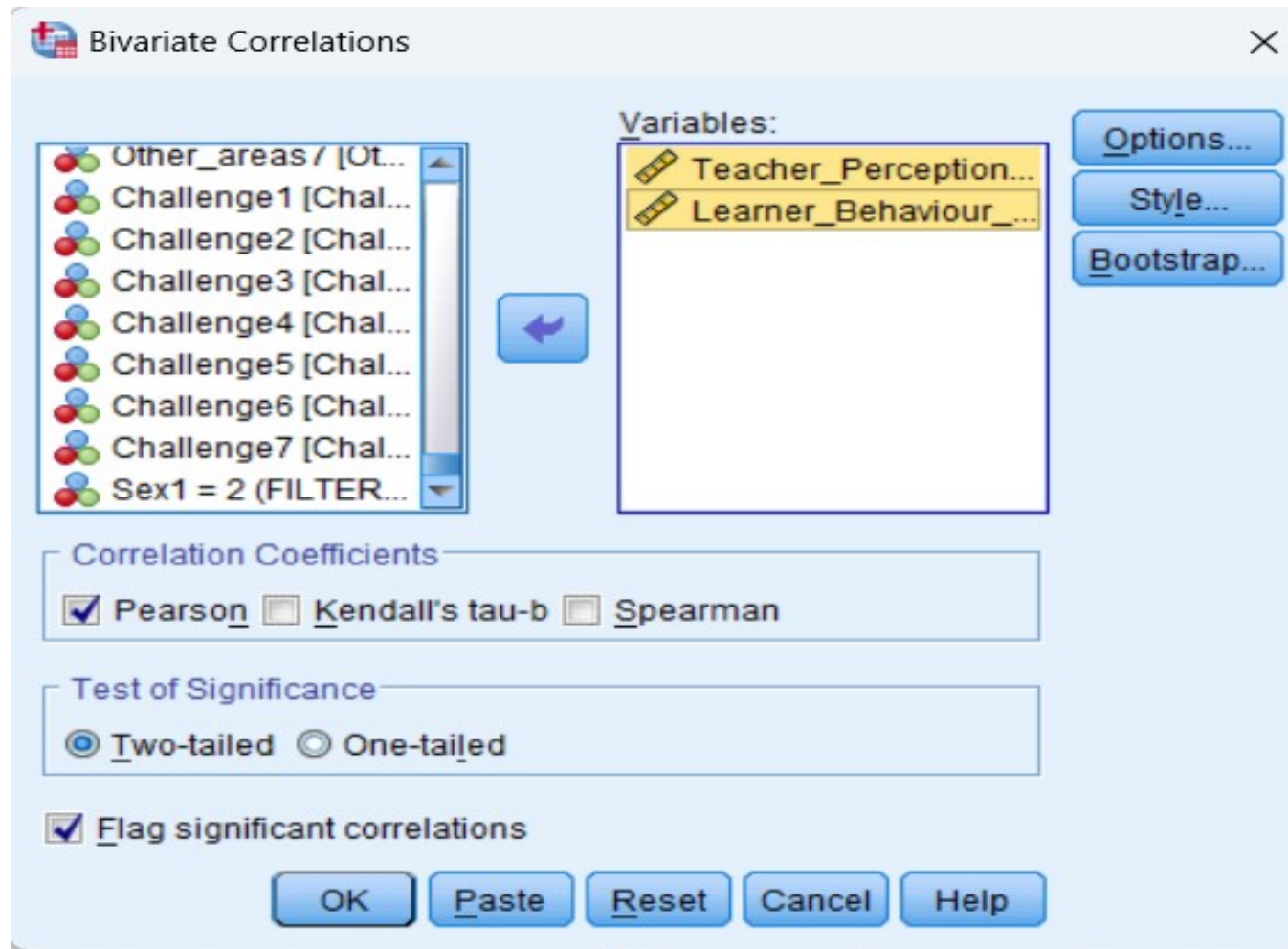
26 80 66

27 72 69

Correlate

- Bivariate...
- Partial...
- Distances...

# Bivariate Correlation



# Bivariate Correlation

IBM SPSS Statistics Processor is ready    Unicode:ON

There is a strong, positive, significant correlation between the teacher perception score and learner behaviour score at  $[r(161)=0.583, p=0.000]$

# 'Work To Do'

1. **Write down the  $H_0$  and  $H_a$  for each of the scores and establish the correlations;**
  - a. *Teacher Perception Score and Teacher Practice*
  - b. *Teacher Perception Score and School Support Score*
2. **Based on the dataset, what other correlations will be relevant in your work?**  
Formulate the hypotheses ( $H_0$  and  $H_a$ )



# Regression

- Regression is used to investigate the effect of one or more predictor variables (independent variable) on an outcome variable (dependent variable)
- How well does one or more independent variables will predict the value of a dependent variable?





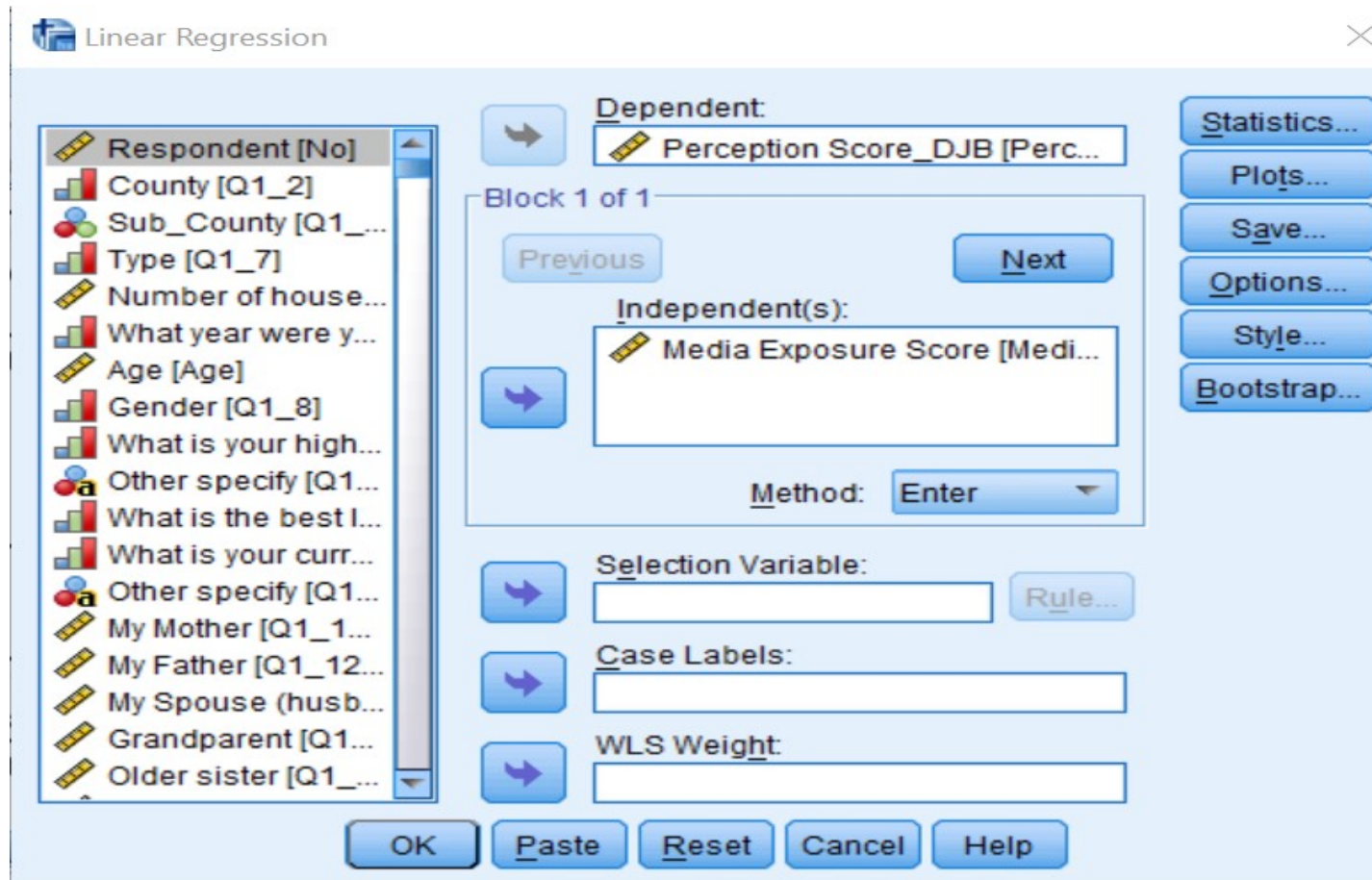
## Regression

Do the School Support Score predict  
Teacher Practice Score?

***H<sub>0</sub>:*** *The School Support Score is not a Significant  
Predictor of the Teacher Practice Score*

***H<sub>a</sub>:*** *The School Support Score is a Significant  
Predictor of the Teacher Practice Score*

# Simple Regression



# Regression

- a. *Perception Score DJB and Shujaaz Info Source Score*
- b. *Shujaaz Info Source Score and Media Exposure Score*
- c. *Media Exposure Score and Mobile Use Score*
- d. *Perception Score DJB and Media Exposure Score*

# Simple Regression

\*WTS Training 2019.spv [Document2] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help

ANOVA

|          |            |     |         |  |  |
|----------|------------|-----|---------|--|--|
| Residual | 182727.411 | 555 | 329.239 |  |  |
| Total    | 185709.476 | 556 |         |  |  |

a. Dependent Variable: Perception Score\_DJB  
b. Predictors: (Constant), Media Exposure Score

**Coefficients<sup>a</sup>**

| Model |                      | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig. |
|-------|----------------------|-----------------------------|------------|---------------------------|--------|------|
|       |                      | B                           | Std. Error | Beta                      |        |      |
| 1     | (Constant)           | 81.402                      | 1.861      |                           | 43.740 | .000 |
|       | Media Exposure Score | .119                        | .040       | .127                      | 3.010  | .003 |

a. Dependent Variable: Perception Score\_DJB

**Residuals Statistics<sup>a</sup>**

IBM SPSS Statistics Processor is ready | Unicode:ON | H: 605, W: 588 pt.



# 'Work To Do'

1. Is the Mobile Phone Use Score a predictor of Media Exposure Score?
2. Is the Media Exposure Score a Predictor of Shujaaz Information Source Score?
  - a. Formulate the hypotheses ( $H_0$  and  $H_a$ )
  - b. Test the hypothesis and make a statement of the results
3. What other regressions are relevant in your work?



# Multiple Regression

*Using a combination of potential predictor variables to determine the most viable model?*



# Multiple Regression

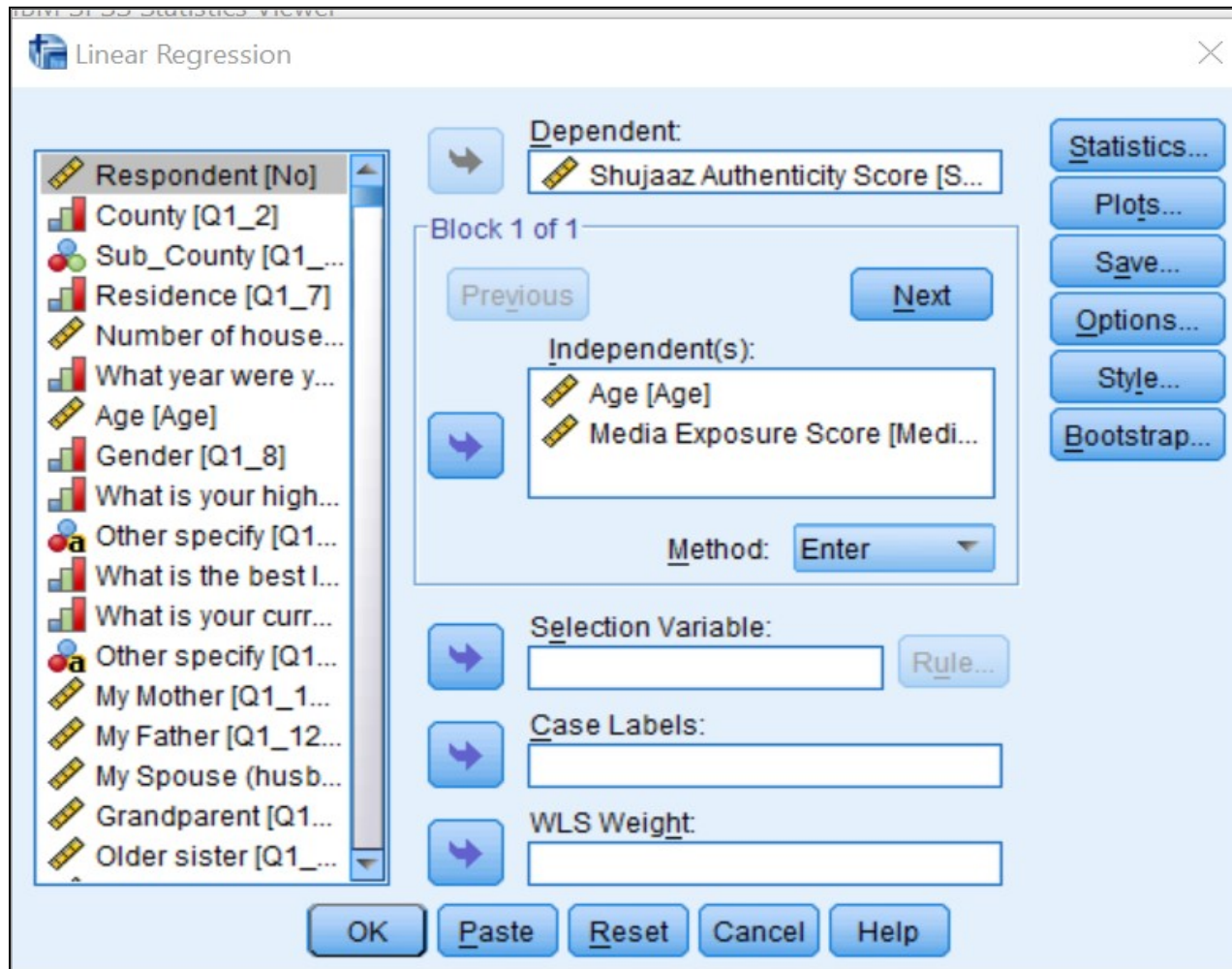
## **WORK To DO**

*Which variables predict Shujaaz Authenticity Score?*

**Ho:** *Age and media exposure score are not significant predictors of Shujaaz Authenticity score*

**Ha:** *Age and media exposure score are significant predictors of Shujaaz Authenticity score*

# Multiple Regression





# Multiple Regression

\*Output1 [Document1] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Output

- Log
- Regression
  - Title
  - Notes
  - Active Dataset
  - Variables Entered
  - Model Summary
  - ANOVA
  - Coefficients

| Model | R                 | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1     | .128 <sup>a</sup> | .016     | .013              | 18.158                     |

a. Predictors: (Constant), Media Exposure Score, Age

**ANOVA<sup>a</sup>**

| Model |            | Sum of Squares | df  | Mean Square | F     | Sig.              |
|-------|------------|----------------|-----|-------------|-------|-------------------|
| 1     | Regression | 3039.218       | 2   | 1519.609    | 4.609 | .010 <sup>b</sup> |
|       | Residual   | 182670.258     | 554 | 329.730     |       |                   |
|       | Total      | 185709.476     | 556 |             |       |                   |

a. Dependent Variable: Shujaaz Authenticity Score

b. Predictors: (Constant), Media Exposure Score, Age

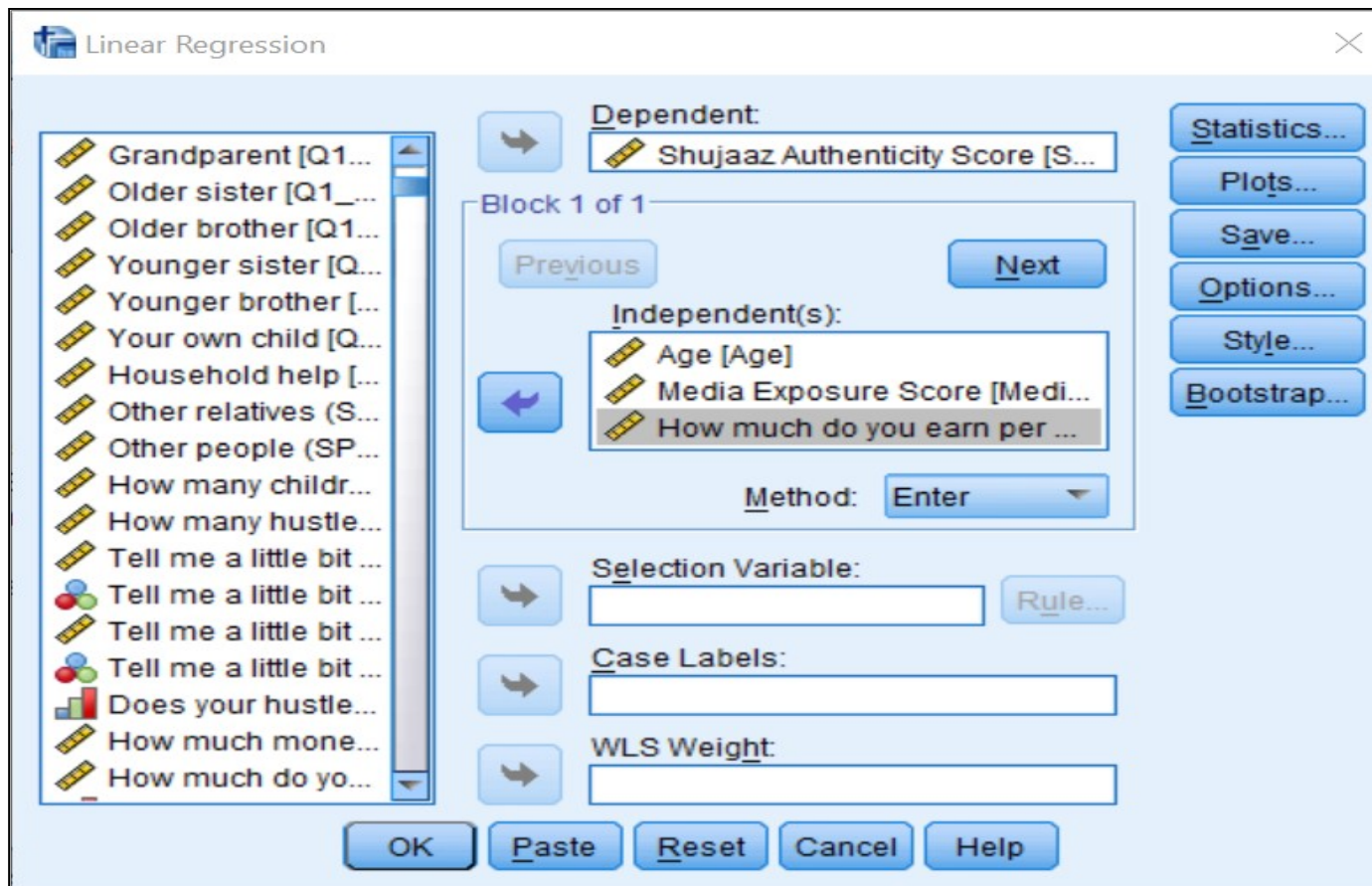
IBM SPSS Statistics Processor is ready Unicode:ON



# What To Do

Conduct multiple regression determine the influence of ***amount of money earned per month, media exposure score, age on Shujaaz authenticity score***

# What To Do





# Factor Analysis

- “Factors” also called “latent variables” or “components” are underlying concepts that are not observable
- Factor analysis make meaning of what is observed as variable
- Also to remove redundant variables
- Principal Component Analysis (PCA) is one of the commonly used method in factor analysis

# What To Do

The screenshot displays the IBM SPSS Statistics Viewer interface. The main window shows the 'Coefficients' table for a regression model. The dependent variable is 'Shujaaz Authenticity Score'. The independent variables are '(Constant)', 'Age', 'Media Exposure Score', and 'How much do you earn per month, on average?'. The table provides unstandardized coefficients (B and Std. Error), standardized coefficients (Beta), t-statistics, and significance levels (Sig.).

| Model |   | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig. |
|-------|---|-----------------------------|------------|---------------------------|--------|------|
|       |   | B                           | Std. Error | Beta                      |        |      |
| 1     | (Constant)                                  | 83.132                      | 6.416      |                           | 12.956 | .000 |
|       | Age   | -.091                       | .335       | -.013                     | -.273  | .785 |
|       | Media Exposure Score                        | .125                        | .041       | .133                      | 3.034  | .003 |
|       | How much do you earn per month, on average? | -3.336E-5                   | .000       | -.016                     | -.346  | .730 |

a. Dependent Variable: Shujaaz Authenticity Score

# Chi-Square Test

- Chi-square is the most frequently used non-parametric test
- Two types of Chi-square ( $\chi^2$ ) tests
- $\chi^2$  test for goodness-of-fit
- $\chi^2$  test of independence



# $\chi^2$ Test of Independence

- Chi-square test for independence is used to evaluate whether two *categorical variables* are independence
- Classic example is gender and risk

## Question

***Is the risk category of the respondent independent of their gender?***



# $\chi^2$ Test of Independence

## Hypothesis

$H_o$ : *Gender and risk category of 15-24 year old Kenyan are independent*

$H_a$ : *Gender and risk category of 15-24 year old Kenyan are not independent*





# $\chi^2$ Test of Independence

In SPSS

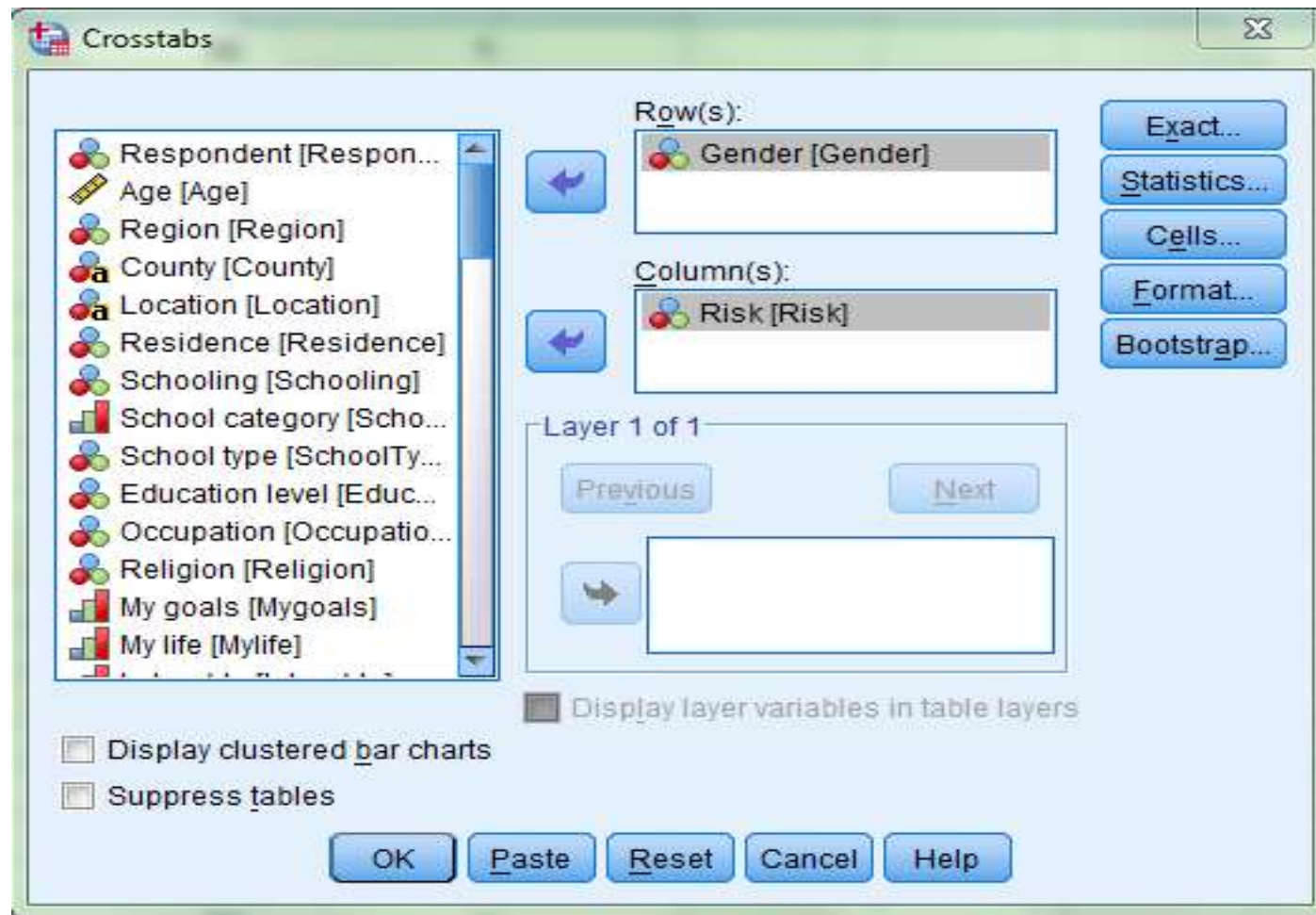
**Analyze**

**Descriptives**

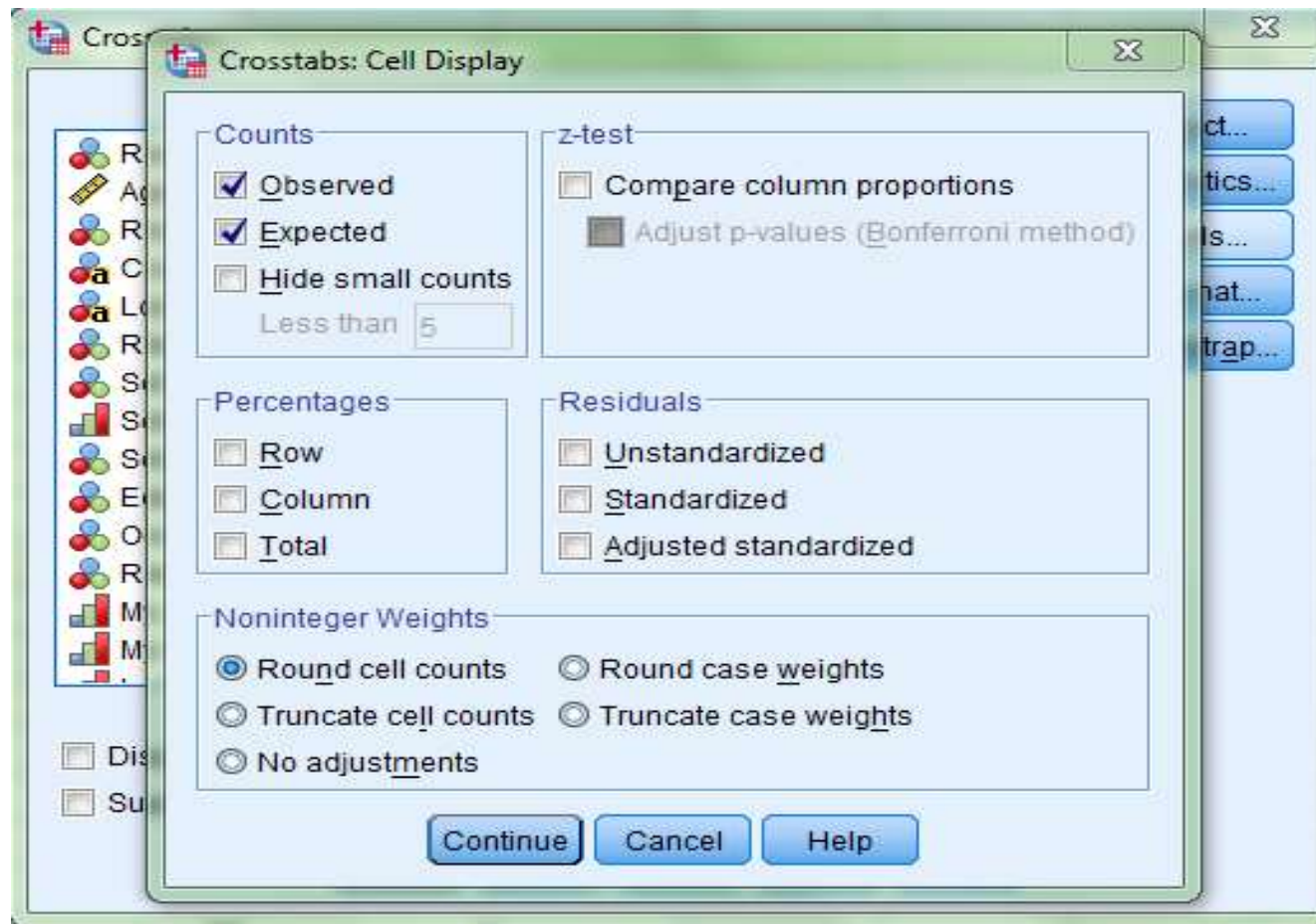
**Crosstabs**

- In the resulting dialog box, select *Gender* and transfer to the **Rows** and select *Risk* and transfer to **Columns**

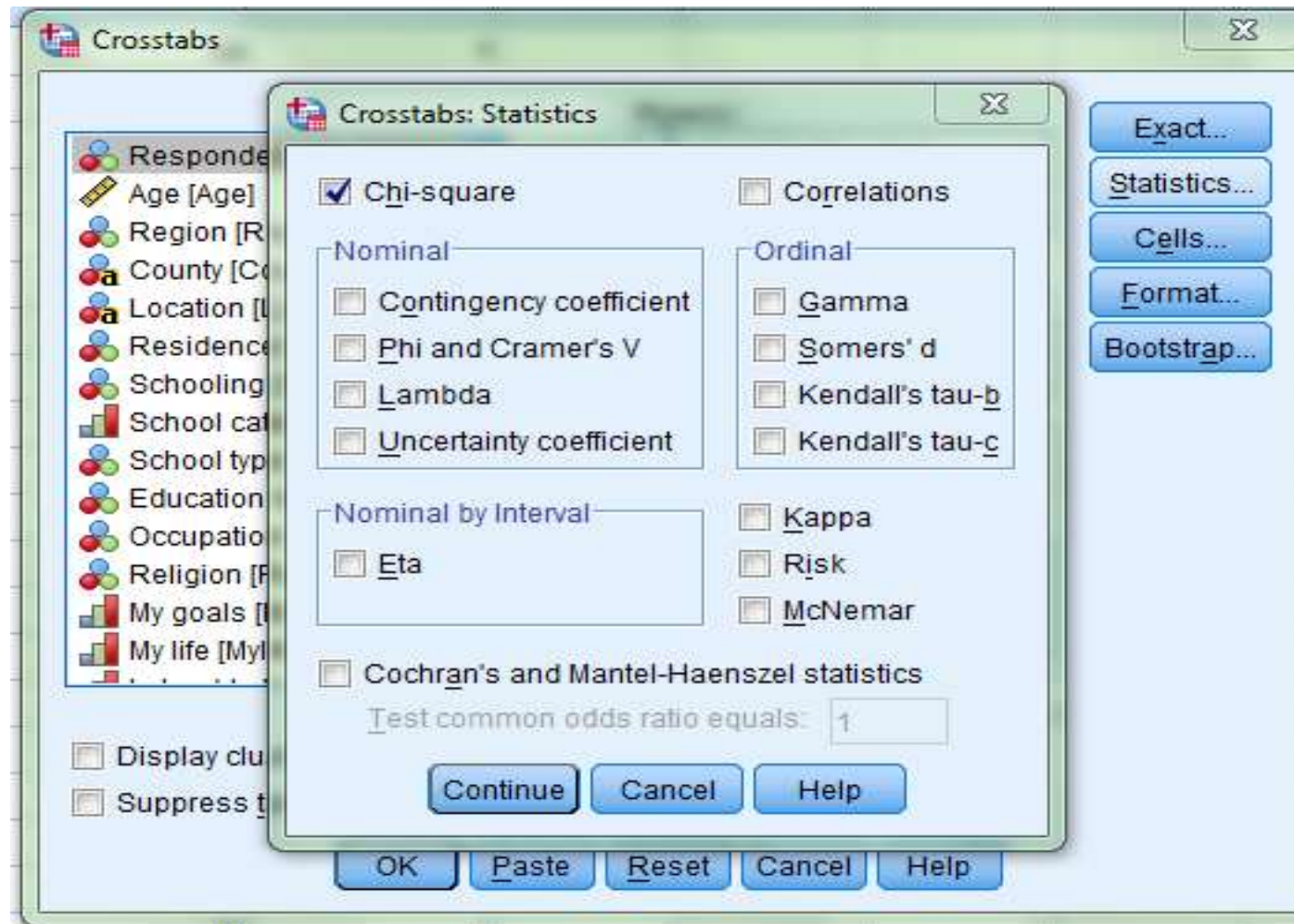
# $\chi^2$ Test of Independence



# $\chi^2$ Test of Independence



# $\chi^2$ Test of Independence



# $\chi^2$ Test of Independence

The screenshot displays the IBM SPSS Statistics Viewer interface. The main window shows the output of a Chi-Square Test of Independence for the relationship between Risk and Gender. The output is organized into two main sections: a Crosstabulation table and a Chi-Square Tests table.

**Risk \* Gender Crosstabulation**

|             |                 | Gender |        |        |
|-------------|-----------------|--------|--------|--------|
|             |                 | Male   | Female | Total  |
| Risk taker  | Count           | 75     | 43     | 118    |
|             | % within Gender | 8.9%   | 6.5%   | 7.9%   |
| Risk aware  | Count           | 447    | 344    | 791    |
|             | % within Gender | 53.3%  | 52.1%  | 52.8%  |
| Risk averse | Count           | 316    | 273    | 589    |
|             | % within Gender | 37.7%  | 41.4%  | 39.3%  |
| Total       | Count           | 838    | 660    | 1498   |
|             | % within Gender | 100.0% | 100.0% | 100.0% |

**Chi-Square Tests**

|                              | Value              | df | Asymp. Sig. (2-sided) |
|------------------------------|--------------------|----|-----------------------|
| Pearson Chi-Square           | 4.137 <sup>a</sup> | 2  | .126                  |
| Likelihood Ratio             | 4.179              | 2  | .124                  |
| Linear-by-Linear Association | 3.667              | 1  | .055                  |
| N of Valid Cases             | 1498               |    |                       |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is .5100.



# $\chi^2$ Test of Independence

## Findings

The results in Table.....show that gender and risk category are independent at  $\chi^2 = 4.137$ ,  $df=2$ ,  $p=0.126$ . Therefore, the  $H_0$  is rejected. In conclusion, gender is not a determinant of the risk category of Kenyans 15-24 years old

***What are the program implications or recommendations?***

# ‘Work To Do’

## Further activities

**Conduct  $\chi^2$  Tests and report on the findings:**

1. Test for the independence of ***Gender*** and ***residence***
2. Test for the independence of ***Gender*** and ***Seen-Shujaaz***
3. Test for the independence of ***Gender*** and ***Read-Shujaaz***
4. Test for the independence of ***Risk*** and ***AgeGroup***
5. Test for the independence of ***Gender*** and ***AgeGroup***



# T-Tests

There are three T-Tests;

- a. one sample T-test
- b. Independent Samples T-Test
- c. Paired Samples T-Test





# Independent Samples T-Tests

The Independent Samples T-Test is for testing mean differences between two groups, for example between male and female



# Independent Samples T-Tests

## Case

The **WTS** want to know whether there are gender differences in Shujaaz Authenticity Score between male and female Kenyan Youth 15-24 years



# Independent Samples T-Tests

## Hypothesis

**H<sub>o</sub>:** *there is no statistically significant mean difference in Shujaaz Authenticity Score between male and female Kenyan Youth 15-24 years*

**H<sub>a</sub>:** *there is a statistically significant mean difference in Shujaaz Authenticity Score between male and female Kenyan Youth 15-24 years*



# Independent Samples T-Tests

In SPSS

**Analyze**

**Compare means**

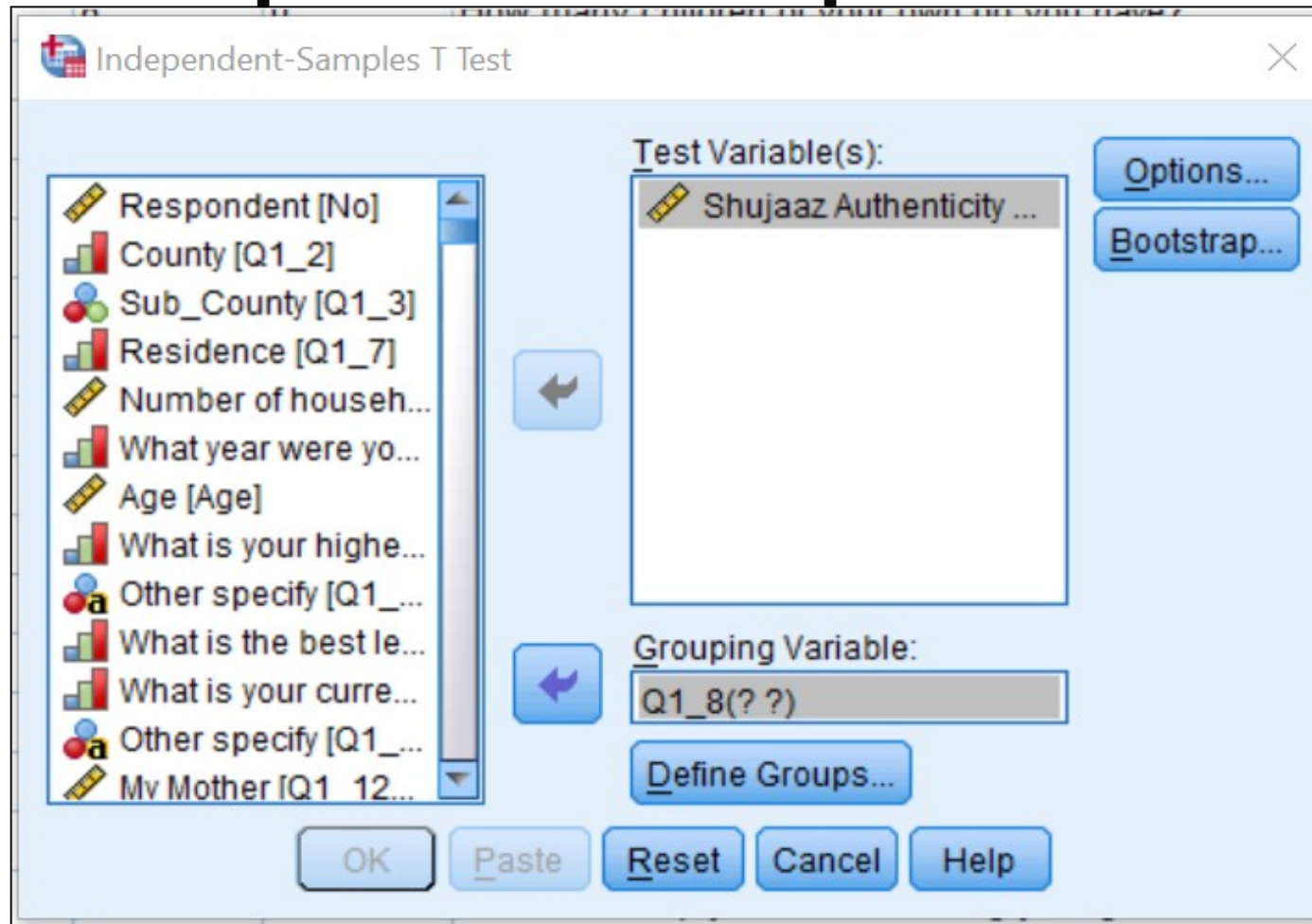
**Independent Samples T-Test**

Select and transfer the dependent variable  
**(Shujaaz Authenticity Score)** into the **'Test Variable'**

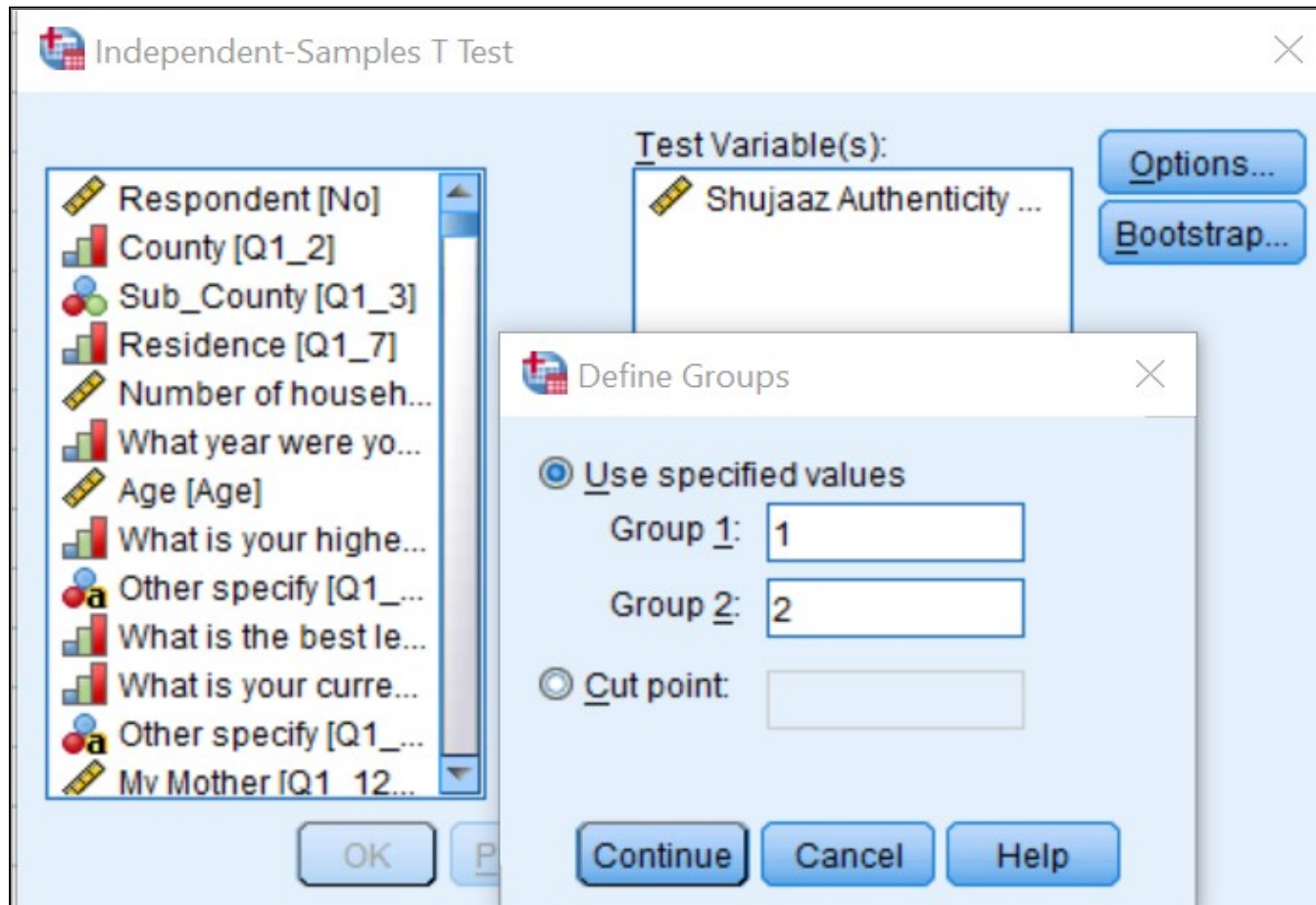
Select and transfer independent variable (gender)  
into the **'Grouping Variable'**

Define Groups (This gives two groups); assign the  
Groups their respective values

# Independent Samples T-Tests



# Independent Samples T-Tests



# Independent Samples T-Tests

The screenshot displays the IBM SPSS Statistics Viewer interface. The main window shows the output of an Independent Samples Test. The left sidebar contains a tree view with categories like Model Summary, ANOVA, Coefficients, Log, Regression, and T-Test. The main area shows a table with the following data:

| Independent Samples Test |   |      |                              |         |                 |                 |                       |
|--------------------------|---|------|------------------------------|---------|-----------------|-----------------|-----------------------|
|                          | Levene's Test for Equality of Variances |      | t-test for Equality of Means |         |                 |                 |                       |
|                          | F                                       | Sig. | t                            | df      | Sig. (2-tailed) | Mean Difference | Std. Error Difference |
| not                      | 9.880                                   | .002 | 2.671                        | 555     | .008            | 4.205           | 1.574                 |
|                          |   |      | 2.552                        | 397.851 | .011            | 4.205           | 1.647                 |

At the top of the window, there is a menu bar (File, Edit, View, Data, Transform, Insert, Format, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, Help) and a toolbar with various icons. The status bar at the bottom indicates 'IBM SPSS Statistics Processor is ready' and 'Unicode:ON'.

# Independent Samples T-Tests

## Findings

The results in Table.....show that male had a higher mean Shujaaz Authenticity Score of 88 compared to female score of 84 and this mean difference between these two groups was statistically significant at  $t=2.671$ ,  $df=555$ ,  $p=0.008$  and the  $H_0$  is rejected. In conclusion, there is a gender difference in the Shujaaz Authenticity Score of Kenyan Youth 15-24 year olds in favour of male youth

***What is the implications of this finding on the program? Any recommendations?***



# 'Work To Do'

## Further T-Test Exercises

Conduct Independent Samples T Tests and report the findings on:

1. Shujaaz Authenticity Score of Male respondents by AgeGroup
2. Shujaaz Authenticity Score of Female respondents by AgeGroup



# Analysis of Variance (ANOVA)

- ANOVA used to compare more than two groups and it is an extension of the t-test developed by R.A. Fisher.
- Hence the **statistics** is 'F'



# Types of ANOVA

Several types of ANOVA that can be conducted depending on the question at hand

1. One Way Between Subjects ANOVA
2. Two Way Between Subjects ANOVA
3. One Way Within Subjects ANOVA
4. Analysis of Covariance (ANCOVA)
5. Repeated Measures ANOVA
6. Multivariate Analysis of Variance (MANOVA)



## One Way Between Subjects ANOVA

- Unlike the Independent Samples T-Test, in One-Way-ANOVA, a comparison can be made between more than two groups
- One-Way ANOVA means that you have exactly one dependent variable (always continuous variable) and exactly one independent variable (always categorical variable)
- One-Way ANOVA is global test that tells a significant difference exist but does not tell where the difference is located



# One Analysis of Variance (ANOVA)

## Example

- Are there significant differences in the Self Efficacy Scores of Respondents in different Risk Categories

# Analysis of Variance (ANOVA)

## Example

**H<sub>o</sub>:** There is no significant different in mean Self Efficacy Scores for Respondents in different Risk Categories

**H<sub>a</sub>:** There is a significant different in mean Self Efficacy Scores for Respondents in different Risk Categories

Or

$$\mathbf{H_o:} \quad \mu_1 = \mu_2 = \mu_3$$

$$\mathbf{H_a:} \quad \mu_1 \neq \mu_2 \neq \mu_3$$



# One Way Between Subjects ANOVA

In SPSS;

**Analyze**

**Compare mean**

**One-Way ANOVA**

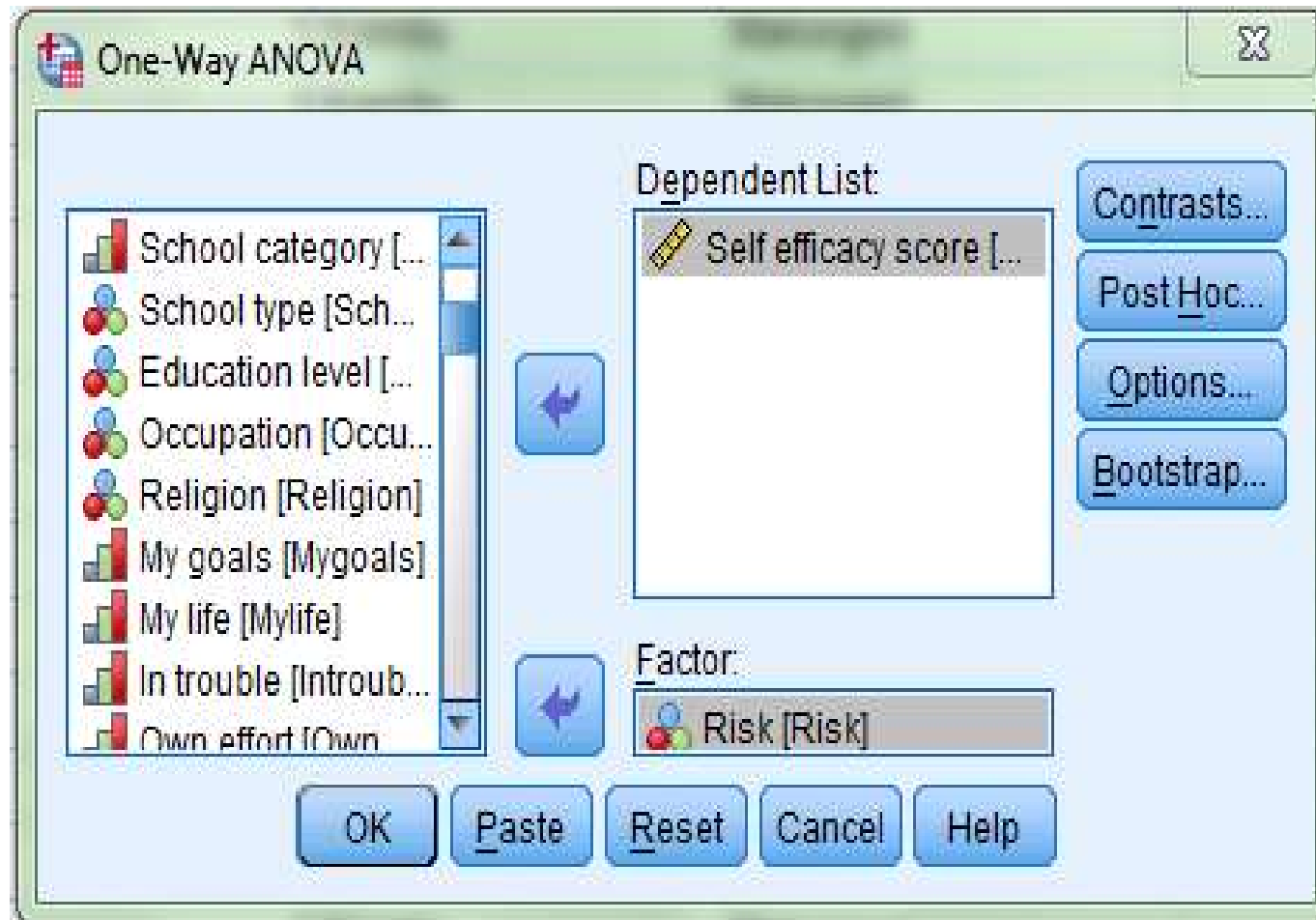
Select **'Self-Efficacy Score'** and transfer to  
**'Dependent List'**

Select **'Risk'** and transfer to **'Factor'**

Click **'Ok'**

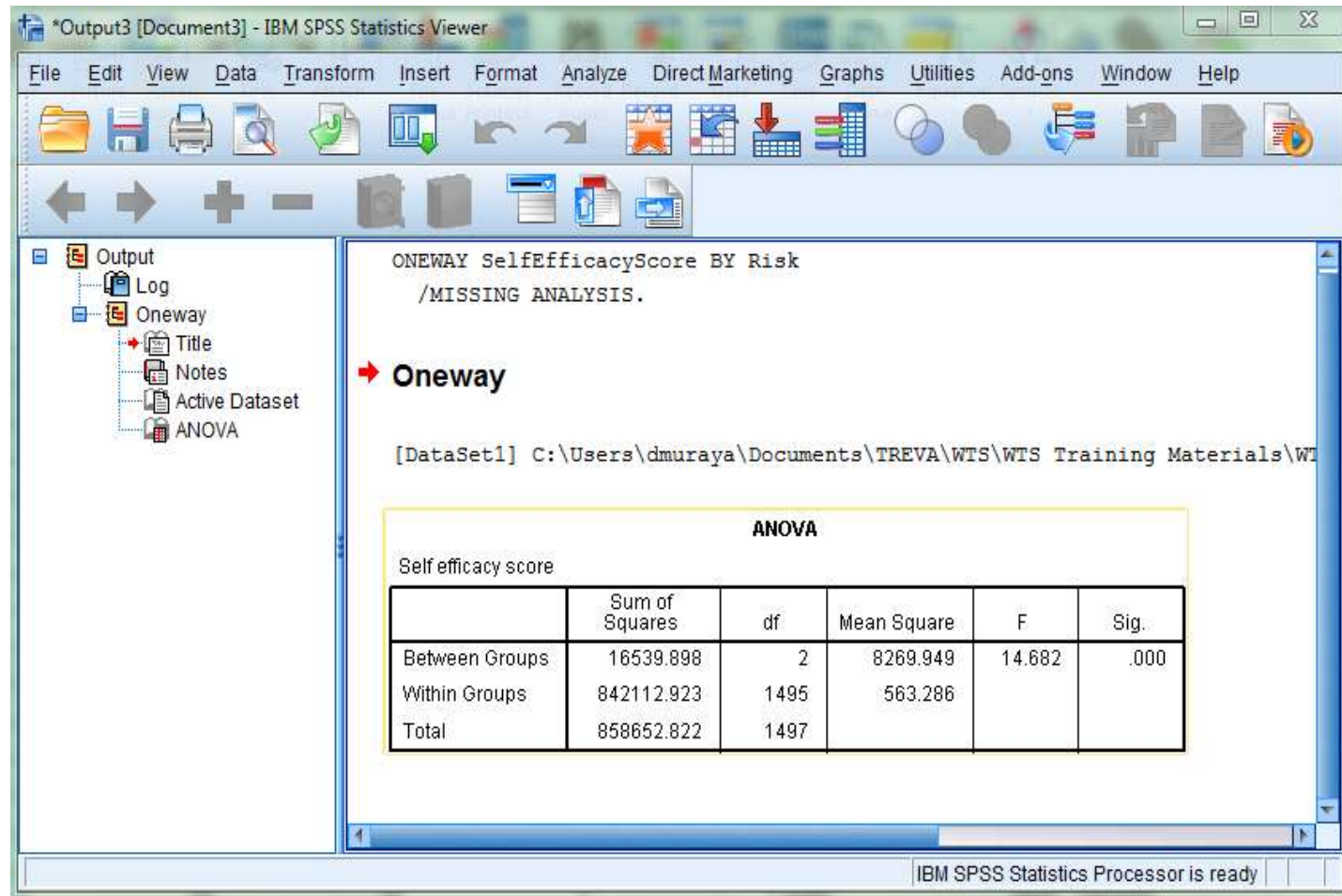
Output

# One Way Between Subjects ANOVA





# One Way Between Subjects ANOVA



The screenshot displays the IBM SPSS Statistics Viewer interface. The main window shows the output of a one-way ANOVA test. The title of the output is "ONEWAY SelfEfficacyScore BY Risk /MISSING ANALYSIS.". Below the title, the word "Oneway" is highlighted with a red arrow. The data source is identified as "[DataSet1] C:\Users\dmuraya\Documents\TREVA\WTS\WTS Training Materials\WT".

The ANOVA table is titled "ANOVA" and shows the results for the "Self efficacy score". The table has six columns: Sum of Squares, df, Mean Square, F, and Sig. The rows represent Between Groups, Within Groups, and Total.

|                | Sum of Squares | df   | Mean Square | F      | Sig. |
|----------------|----------------|------|-------------|--------|------|
| Between Groups | 16539.898      | 2    | 8269.949    | 14.682 | .000 |
| Within Groups  | 842112.923     | 1495 | 563.286     |        |      |
| Total          | 858652.822     | 1497 |             |        |      |

IBM SPSS Statistics Processor is ready

# One Way Between Subjects ANOVA

## Findings

The results in Table....show that there was a statistically significant difference in mean self-efficacy score in the 3 category of respondents at  $F(2, 1495) = 14.682, p < 0.05$ . Therefore the  $H_0$  is rejected and in conclusion the Risk category was a determinant of self-efficacy score.

***Since there were 3 categories of risks, we don't know which category had a higher self-efficacy score***



## Pairwise Post Hoc Multiple Comparison Tests

- *This test is used to locate significant differences in more than one group (global test)*
- *This test compares every pair in the group and whether the difference is significant*
- *Therefore, it calculate the mean difference between every two groups and looks for significance*



# Pairwise Post Hoc Multiple Comparison Tests

In SPSS;

**Analyze**

**Compare mean**

**One-Way ANOVA**

Select **'Self-Efficacy Score'** and transfer to **'Dependent List'**

Select **'Risk'** and transfer to **'Factor'**

**Click 'Post hoc'**

In the **Equal Variances Assumed** select **'LSD'**

Click **'Continue'**, Click **'Ok'**

**Out put**

# Pairwise Post Hoc Multiple Comparison Tests

The screenshot shows the 'One-Way ANOVA: Post Hoc Multiple Comparisons' dialog box in SPSS. The 'Equal Variances Assumed' section is active, with 'LSD' selected. Other options include Bonferroni, Sidak, Scheffe, R-E-G-W F, R-E-G-W Q, S-N-K, Tukey, Tukey's-b, Duncan, Hochberg's GT2, Gabriel, Waller-Duncan, and Dunnett. The 'Type I/Type II Error Ratio' is set to 100, and the 'Control Category' is 'Last'. The 'Test' section has '2-sided' selected. The 'Equal Variances Not Assumed' section is inactive, with options for Tamhane's T2, Dunnett's T3, Games-Howell, and Dunnett's C. The 'Significance level' is set to 0.05. Buttons for 'Continue', 'Cancel', and 'Help' are at the bottom.

One-Way ANOVA: Post Hoc Multiple Comparisons

Equal Variances Assumed

- LSD
- Bonferroni
- Sidak
- Scheffe
- R-E-G-W F
- R-E-G-W Q
- S-N-K
- Tukey
- Tukey's-b
- Duncan
- Hochberg's GT2
- Gabriel
- Waller-Duncan
- Dunnett

Type I/Type II Error Ratio: 100

Control Category: Last

Test:

- 2-sided
- < Control
- > Control

Equal Variances Not Assumed

- Tamhane's T2
- Dunnett's T3
- Games-Howell
- Dunnett's C

Significance level: 0.05

Continue Cancel Help

# Pairwise Post Hoc Multiple Comparison Tests

The screenshot displays the IBM SPSS Statistics Viewer interface. The main window shows the results of an ANOVA test. The ANOVA table is as follows:

|                | Sum of Squares | df   | Mean Square | F      | Sig. |
|----------------|----------------|------|-------------|--------|------|
| Between Groups | 16539.898      | 2    | 8269.949    | 14.682 | .000 |
| Within Groups  | 842112.923     | 1495 | 563.286     |        |      |
| Total          | 858652.822     | 1497 |             |        |      |

Below the ANOVA table, the 'Post Hoc Tests' section is visible, showing 'Multiple Comparisons' for 'Self efficacy score' using the 'LSD' method. The pairwise comparison table is as follows:

| (I) Risk    | (J) Risk    | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval |             |
|-------------|-------------|-----------------------|------------|------|-------------------------|-------------|
|             |             |                       |            |      | Lower Bound             | Upper Bound |
| Risk taker  | Risk aware  | -11.008*              | 2.342      | .000 | -15.60                  | -6.41       |
|             | Risk averse | -12.965*              | 2.394      | .000 | -17.66                  | -8.27       |
| Risk aware  | Risk taker  | 11.008*               | 2.342      | .000 | 6.41                    | 15.60       |
|             | Risk averse | -1.957                | 1.292      | .130 | -4.49                   | .58         |
| Risk averse | Risk taker  | 12.965*               | 2.394      | .000 | 8.27                    | 17.66       |
|             | Risk aware  | 1.957                 | 1.292      | .130 | -.58                    | 4.49        |

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



## Pairwise Post Hoc Multiple Comparison Tests

- *The significant differences were located between risk takers and risk aware and risk takers and risk averse in favour of risk aware and risk averse respectively*
- *No significant differences between risk aware and risk averse groups*



## Options and Contrasts in One-Way ANOVA

- *One-Way ANOVA has 2 additional facilities that are very useful: Options and Contrasts*
- *Options allows 2 important computations:*
  - a. Descriptive statistics** ( Mean, SD, SE, 95% Confidence Limits, Max & Min Scores)
  - b. Test of Homogeneity of Variance**



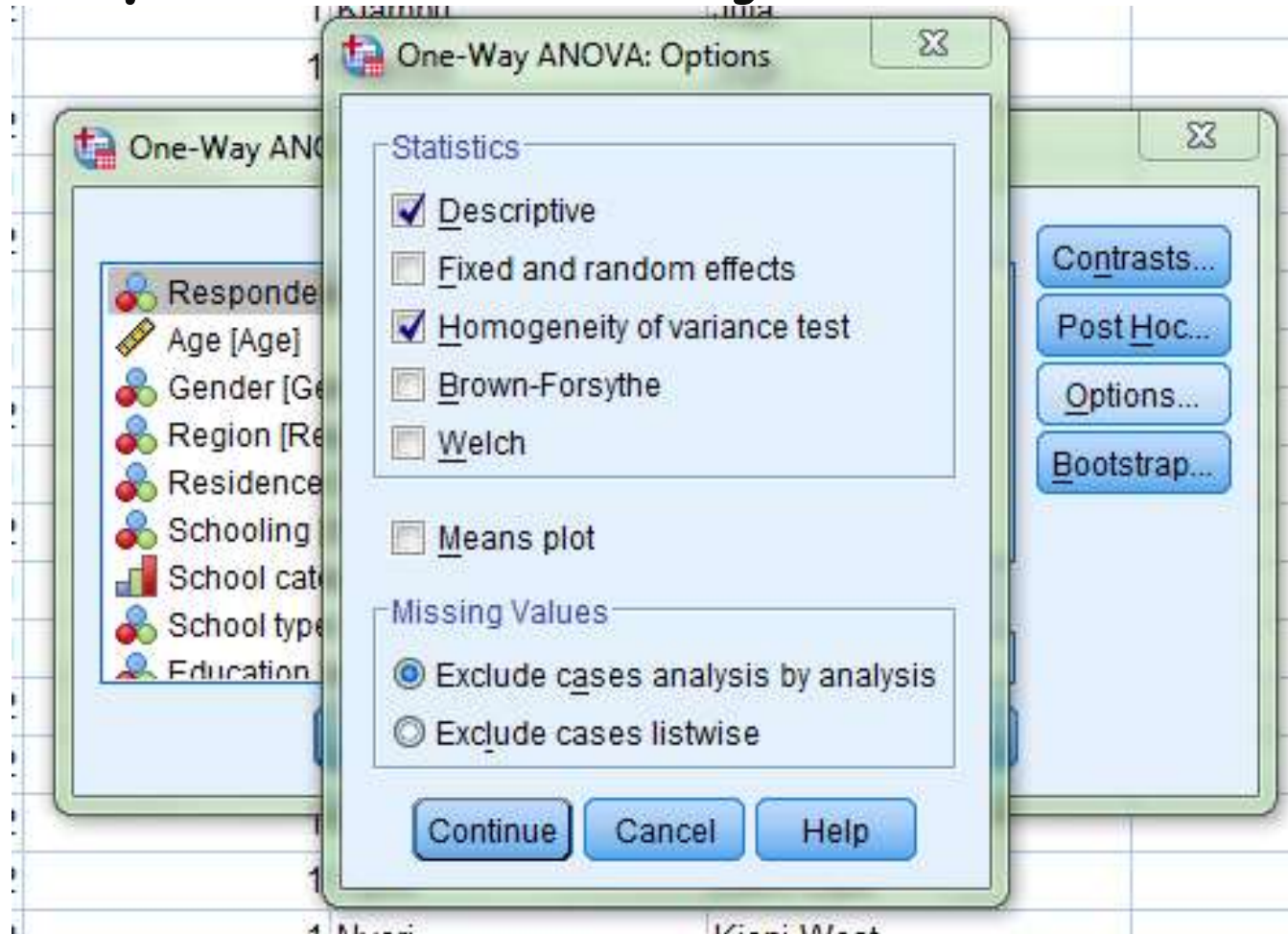


## **‘Options’ in One-Way ANOVA**

*Click: ‘Options’*

*Select: ‘Statistics’ and ‘Homogeneity of  
Variance’*

# 'Options' in One-Way ANOVA





## **‘Options’ in One-Way ANOVA**

*Click:*      **‘Continue’**

*Click:*      **‘OK’**

# 'Options' in One-Way ANOVA

The screenshot displays the IBM SPSS Statistics Viewer interface. The left-hand pane shows a tree view of the output, with 'ANOVA' selected under the 'Oneway' category. The main window displays the following statistical results:

**Descriptives**

Self efficacy score

|             | N    | Mean  | Std. Deviation | Std. Error | 95% Confidence Interval for Mean |             | Minimum | Maximum |
|-------------|------|-------|----------------|------------|----------------------------------|-------------|---------|---------|
|             |      |       |                |            | Lower Bound                      | Upper Bound |         |         |
| Risk taker  | 118  | 40.85 | 20.667         | 1.903      | 37.08                            | 44.62       | 25      | 86      |
| Risk aware  | 791  | 51.86 | 24.969         | .888       | 50.12                            | 53.60       | 25      | 100     |
| Risk averse | 589  | 53.82 | 22.574         | .930       | 51.99                            | 55.64       | 25      | 100     |
| Total       | 1498 | 51.76 | 23.950         | .619       | 50.55                            | 52.98       | 25      | 100     |

**Test of Homogeneity of Variances**

Self efficacy score

| Levene Statistic | df1 | df2  | Sig. |
|------------------|-----|------|------|
| 30.575           | 2   | 1495 | .000 |

**ANOVA**

Self efficacy score

|                | Sum of Squares | df   | Mean Square | F      | Sig. |
|----------------|----------------|------|-------------|--------|------|
| Between Groups | 16539.898      | 2    | 8269.949    | 14.682 | .000 |
| Within Groups  | 842112.923     | 1495 | 563.286     |        |      |



# What To Do 'One-Way ANOVA'

*Determine whether there are statistically significant differences in Self-Efficacy Score based:*

- *Education level*
- *First Saw Shujaaz*
- *Listening Frequency*



## Two-Way Between Subjects ANOVA

- When we want to determine the effect of more than one independent variable on one dependent variable, Two-way ANOVA would be used
- Two-Way ANOVA means that you have exactly one dependent variable (always continuous variable) and exactly two independent variable (always categorical variable)
- For example, it is possible to find out how the Self-Efficacy Score is affected by both First Saw Shujaaz and Read Shujaaz



# Two-Way Between Subjects ANOVA

## Hypothesis

- **H<sub>0</sub>:** There is no significant difference Self-Efficacy Score based on First Saw Shujaaz and Read Shujaaz and their interaction
- **H<sub>a</sub>:** There is significant difference Self-Efficacy Score based on First Saw Shujaaz and Read Shujaaz and their interaction



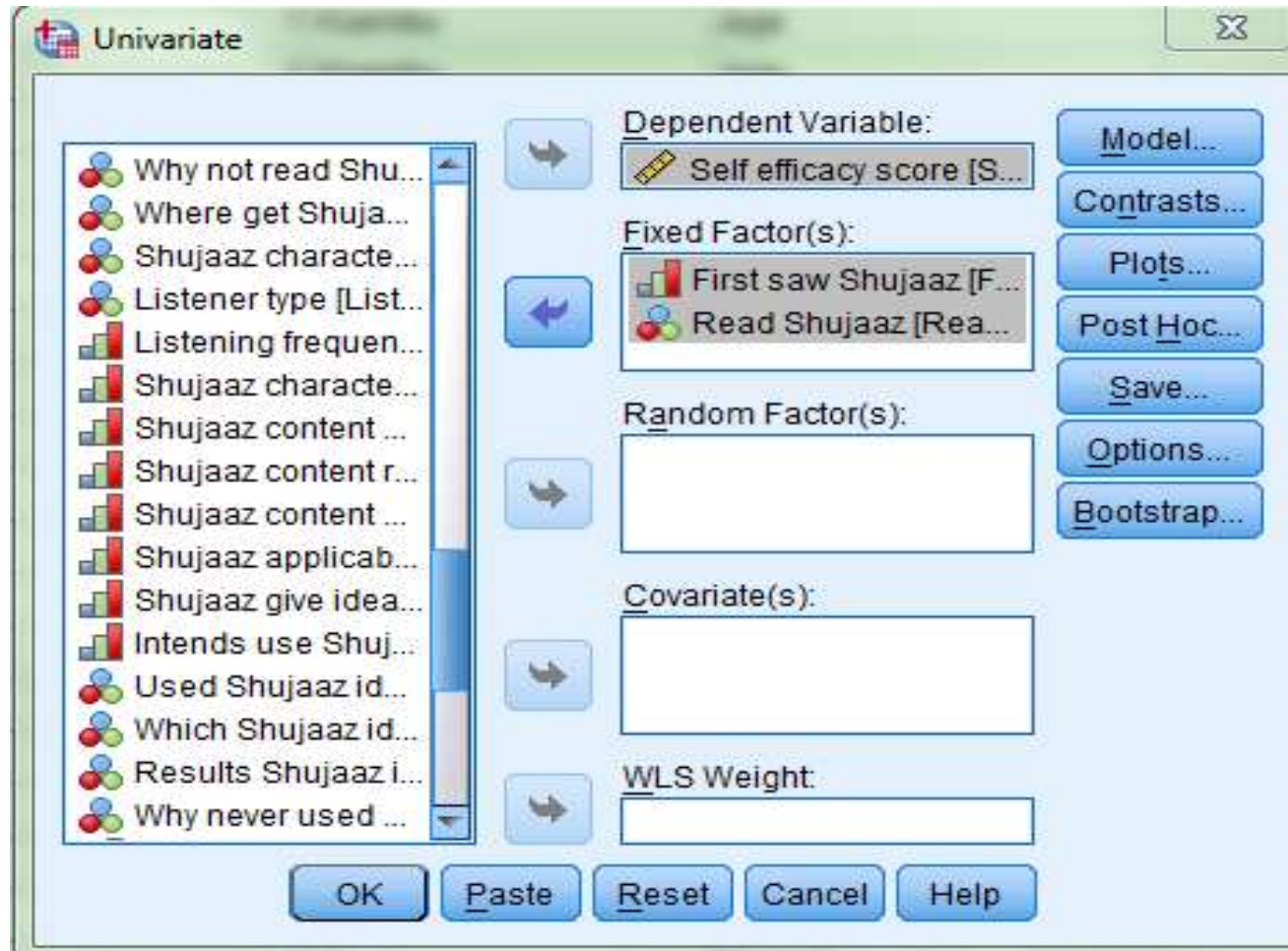
# Two-Way Between Subjects ANOVA

## In SPSS

- Analyze, General Linear Model, Select Univariate
- Select the Self-Efficacy Score (Dependent Variable) and transfer to Dependent Variable List
- Select the First Saw Shujaaz and Read Shujaaz (Independent Variables) and transfer to the Fixed Factors List



# Two-Way Between Subjects ANOVA



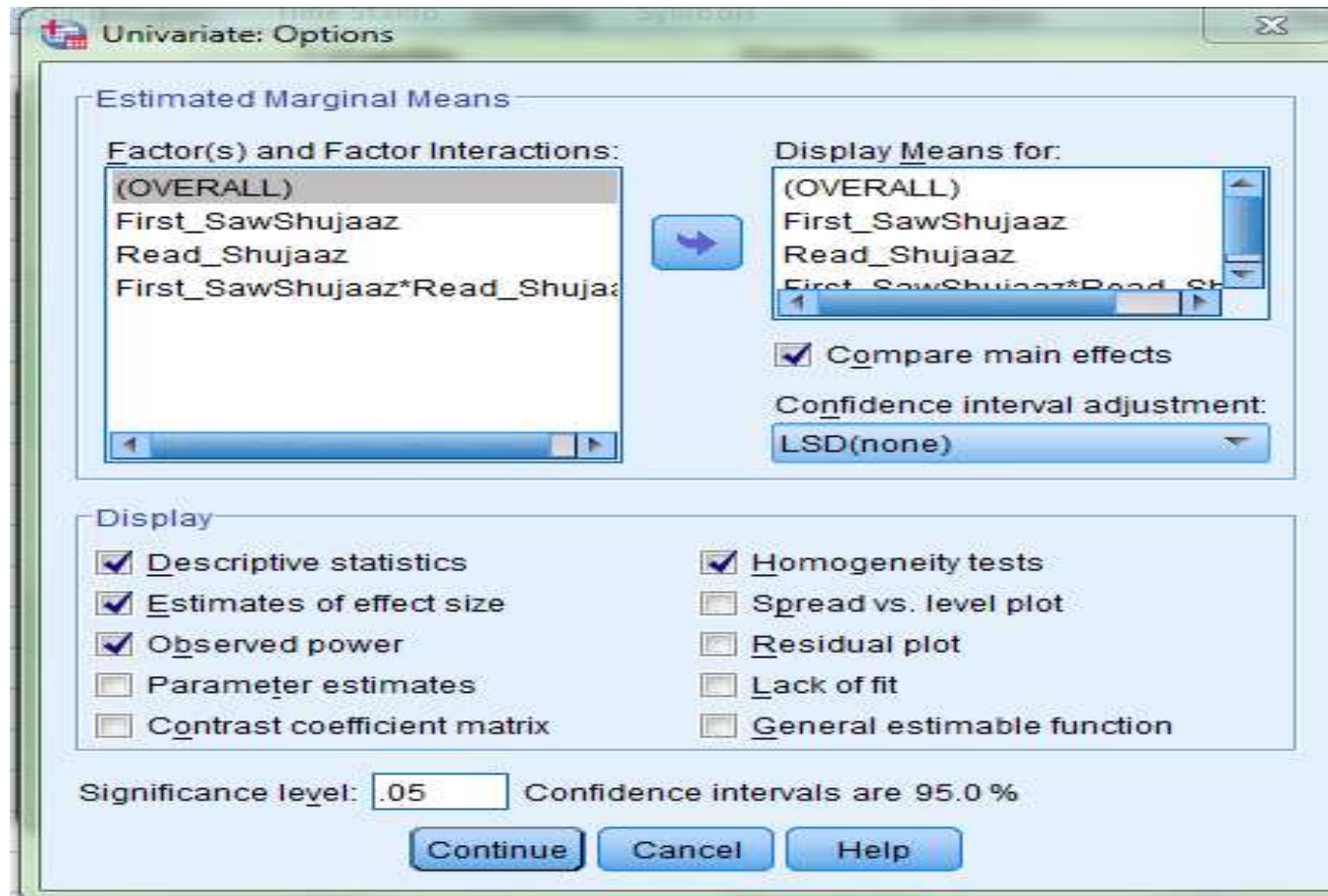


# Two-Way Between Subjects ANOVA

## In SPSS

- Click Options and Select:
  - Descriptives
  - Estimates of effect size
  - Observed power
  - Homogeneity tests
- Click Continue, Click Ok

# Two-Way Between Subjects ANOVA



# Two-Way Between Subjects ANOVA

The screenshot displays the IBM SPSS Statistics Viewer interface. The main window is titled "Univariate Analysis of Variance" and shows the following information:

- Dataset: [DataSet1] C:\Users\dmuraya\Documents\TREVA\WTS\WTS Training Materials\WTS SPSS Training Dataset
- Table: Between-Subjects Factors
- Table: Descriptive Statistics

|                   |   | Value Label           | N   |
|-------------------|---|-----------------------|-----|
| First saw Shujaaz | 1 | About 6 months ago    | 268 |
|                   | 2 | 1-2 years ago         | 264 |
|                   | 3 | More than 2 years ago | 294 |
|                   | 4 | Can't remember        | 672 |
| Read Shujaaz      | 1 | Yes                   | 710 |
|                   | 2 | No                    | 788 |

The interface also shows a tree view on the left with various analysis options like "Estimated Marginal Means", "Pairwise Comparisons", and "Univariate". The status bar at the bottom indicates "IBM SPSS Statistics Processor is ready".

# Two-Way Between Subjects ANOVA

\*Output6 [Document6] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Log

- Univariate Analysis
  - Title
  - Notes
  - Active Dataset
  - Between-Subje
  - Descriptive Sta
  - Levene's Test o
  - Tests of Betwe
  - Estimated Marg
  - 1. Grand M
  - 2. First sav
  - 3. Read Sh
  - 4. First sav

**Descriptive Statistics**

Dependent Variable: Self efficacy score

| First saw Shujaaz     | Read Shujaaz | Mean  | Std. Deviation | N    |
|-----------------------|--------------|-------|----------------|------|
| About 6 months ago    | Yes          | 47.78 | 24.928         | 141  |
|                       | No           | 53.92 | 22.473         | 127  |
|                       | Total        | 50.69 | 23.950         | 268  |
| 1-2 years ago         | Yes          | 52.26 | 23.853         | 149  |
|                       | No           | 56.70 | 24.714         | 115  |
|                       | Total        | 54.19 | 24.286         | 264  |
| More than 2 years ago | Yes          | 55.32 | 25.172         | 213  |
|                       | No           | 51.43 | 24.139         | 81   |
|                       | Total        | 54.25 | 24.912         | 294  |
| Can't remember        | Yes          | 47.65 | 22.763         | 207  |
|                       | No           | 51.26 | 23.420         | 465  |
|                       | Total        | 50.15 | 23.262         | 672  |
| Total                 | Yes          | 50.94 | 24.347         | 710  |
|                       | No           | 52.50 | 23.577         | 788  |
|                       | Total        | 51.76 | 23.950         | 1498 |

IBM SPSS Statistics Processor is ready

# Two-Way Between Subjects ANOVA

\*Output6 [Document6] - IBM SPSS Statistics Viewer

File Edit View Data Transform Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help

3. Read St  
 Title  
 Estim  
 Pairwi  
 Univa  
 4. First sav  
 Log  
 Univariate Analysis  
 Title  
 Notes  
 Active Dataset  
 Between-Subje  
 Descriptive Sta  
 Levene's Test o  
 Tests of Betwe  
 Estimated Marc  
 Title  
 1. Grand M  
 2. First sav  
 Title  
 Estim  
 Pairwi  
 Univa  
 3. Read St  
 Title  
 Estim  
 Pairwi  
 Univa  
 4. First sav

### Variances<sup>a</sup>

Dependent Variable: Self efficacy score

| F     | df1 | df2  | Sig. |
|-------|-----|------|------|
| 2.710 | 7   | 1490 | .009 |

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + First\_SawShujaaz + Read\_Shujaaz + First\_SawShujaaz \* Read\_Shujaaz

### Tests of Between-Subjects Effects

Dependent Variable: Self efficacy score

| Source                          | Type III Sum of Squares | df   | Mean Square | F        | Sig. | Partial Eta Squared | Noncent. Parameter | Observed Power <sup>b</sup> |
|---------------------------------|-------------------------|------|-------------|----------|------|---------------------|--------------------|-----------------------------|
| Corrected Model                 | 12002.185 <sup>a</sup>  | 7    | 1714.598    | 3.017    | .004 | .014                | 21.122             | .941                        |
| Intercept                       | 3186254.383             | 1    | 3186254.383 | 5607.412 | .000 | .790                | 5607.412           | 1.000                       |
| First_SawShujaaz                | 5631.594                | 3    | 1877.198    | 3.304    | .020 | .007                | 9.911              | .756                        |
| Read_Shujaaz                    | 1949.564                | 1    | 1949.564    | 3.431    | .064 | .002                | 3.431              | .457                        |
| First_SawShujaaz * Read_Shujaaz | 3635.772                | 3    | 1211.924    | 2.133    | .094 | .004                | 6.399              | .546                        |
| Error                           | 846650.637              | 1490 | 568.222     |          |      |                     |                    |                             |
| Total                           | 4872164.256             | 1498 |             |          |      |                     |                    |                             |
| Corrected Total                 | 858652.822              | 1497 |             |          |      |                     |                    |                             |

a. R Squared = .014 (Adjusted R Squared = .009)  
 b. Computed using alpha = .05

Preview print this dataset

IBM SPSS Statistics Processor is ready



# ‘What To Do’

Using the WTS Dataset and conduct Two-Way ANOVA and determine the effect of:

1. Gender & Read Shujaaz on Self-Efficacy Score
2. Read Shujaaz & Risk on Self-Efficacy Score
3. Read Shujaaz & Risk on Self-Efficacy Score
4. Occupation & Education level on Self-Efficacy Score



# Binary Logistics Regression

1. The goal is to predict membership in a target group from scores on one or several predictor variables
2. Predictor variable may be quantitative or categorical

## **Assumptions**

1. Outcome variable is binary / dichotomous (Y/N, Rural/Urban)
2. Categories of outcome variable are exhaustive and mutually exclusive
3. Each person is known to be a member of one group or the other





# Binary Logistics Regression

## Reference Group

1. The membership group of interest:
2. Residence: Are females likely to be found in Rural areas compared to males?
3. Question: Does gender predict residence?



# Binary Logistics Regression

## Odds

1. Computed by dividing the number of times on outcome of interest does happen by the number of times when it does not happen
2. Minimum Odds is 0 while maximum is  $\infty$
3. If  $\text{Odd} < 1$ ; target event is less likely and if  $\text{Odd} > 1$ ; target event more likely, if  $\text{Odd}=1$ ; target event has equal chance of happening or not happening

## Odds Ratio

1. Odds for the reference group against the other group and vice versa



**USE STATISTICS  
WISELY FOR DECISION  
MAKING**