# LCDB7011 Quantitative Research Methods

- Week 5 – SPSS IN ACTION
- Dr. Helen LEE

# Content of SPSS

1. SPSS Introduction
2. Data import and Data input
3. Basic data transformation and computation
4. Descriptive analysis
5. Cronbach Alpha (α ) - Reliability Test
6. Association and Correlation analysis
   - Chi-Square
   - Pearson Correlation
7. Normality test

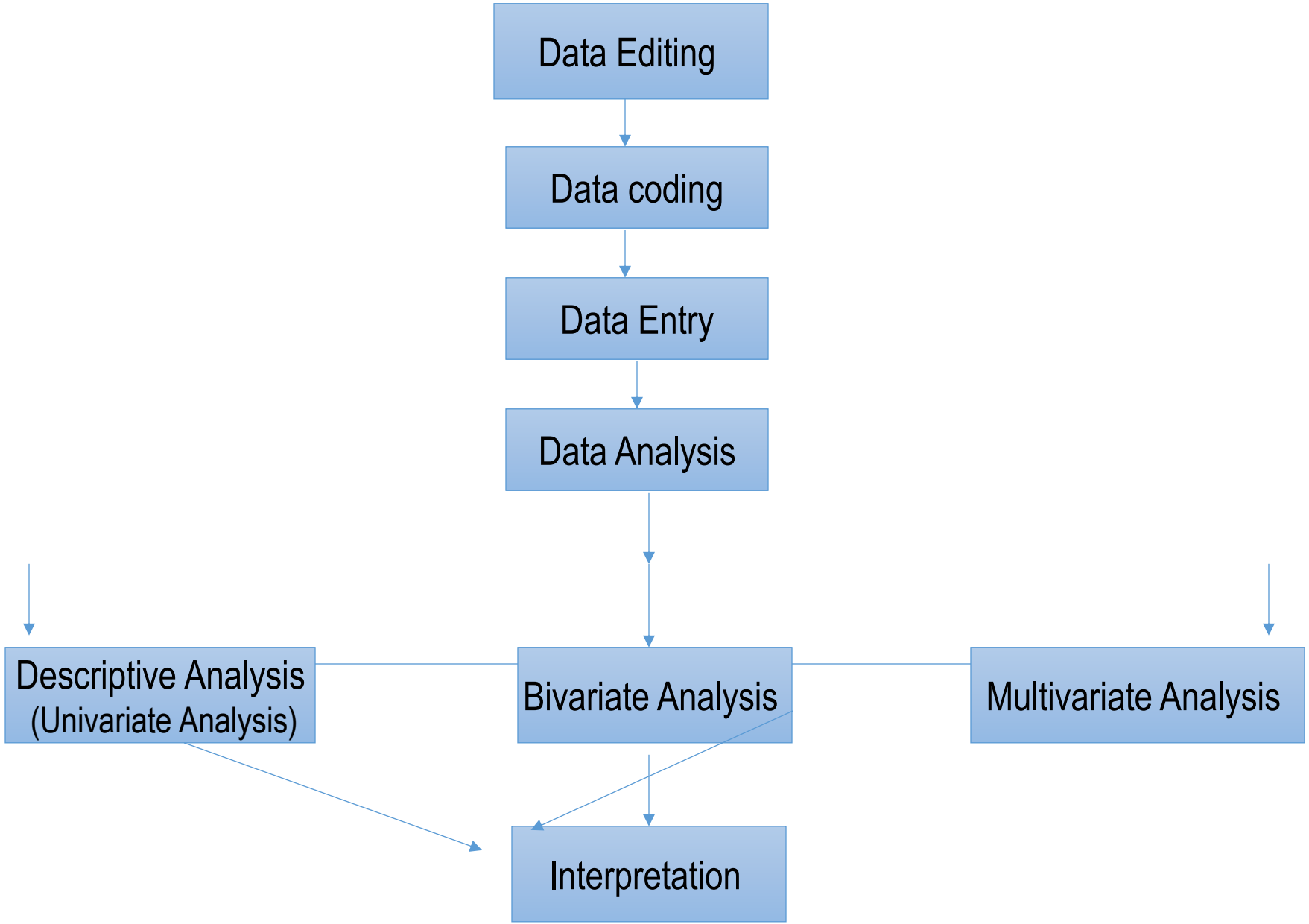
# SPSS IN ACTION

(Version 29.0)

Dr. Helen Lee

# What we will do today

- ✖SPSS interface: data view and variable view
- ✖Entering data into SPSS
- ✖Cleaning and editing data
- ✖Transforming data
- ✖Descriptive statistics
- ✖Charts and basic tests
- ✖Saving Data
- ✖Useful information


Data Editing
Data coding
Data Entry
Data Analysis
Descriptive Analysis
(Univariate Analysis)
Bivariate Analysis
Multivariate Analysis
Interpretation

# SPSS interface

- SPSS Statistics is a software package for statistical analysis. Originally produced by SPSS Inc., it was bought by IBM in 2009. Please download the SPSS from UWTSD The most current version is 29, but SPSS versions are compatible even with much older versions.

- Statistics dataset (.sav) files are backward and forward compatible; even where features have been added, functionality has been retained.

# 2. SPSS Input

# 2. Data input

*Task 1: How to input data in SPSS?*

# Before the data entry

- If you enter data manually, you need a coding frame
- Each case/respondent needs to be allocated a unique ID number.
- You also can import data from Excel or other software such as SurveyMonkey, Google form, and so forth

# Example of a coding Frame

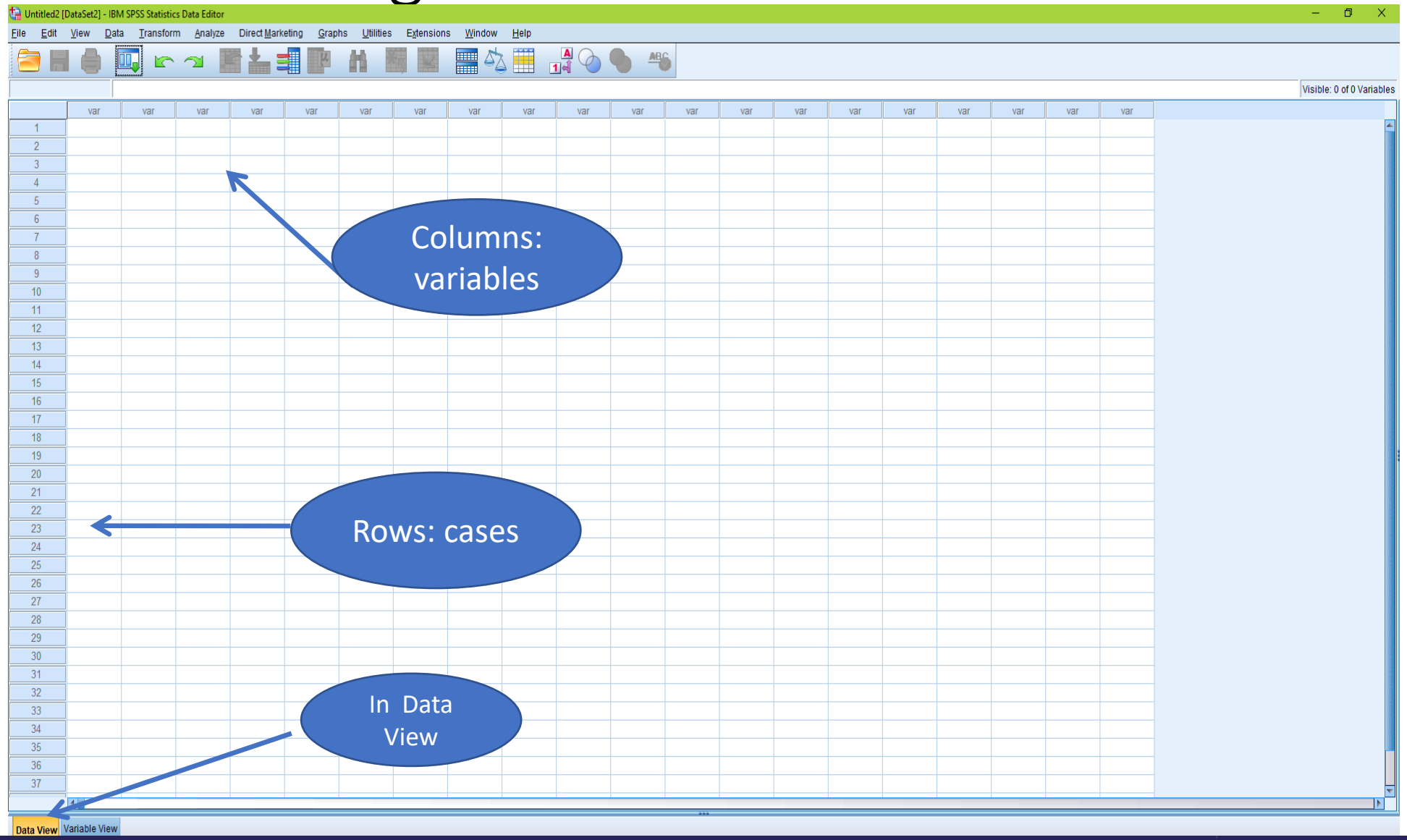
**Coding frame**

- Q1 Male =1, Female= 2
- Q2 Age in years
- Q3 Temporary = 1, Operator = 2, Office = 3, Team leader = 4 Management = 5
- Q4 Salary in £
- Q5 Very dissatisfied = 1, dissatisfied = 2, neutral = 3, satisfied = 4, very satisfied = 5
- Q6 Very low=1, low = 2, medium = 3, high= 4, very high = 5

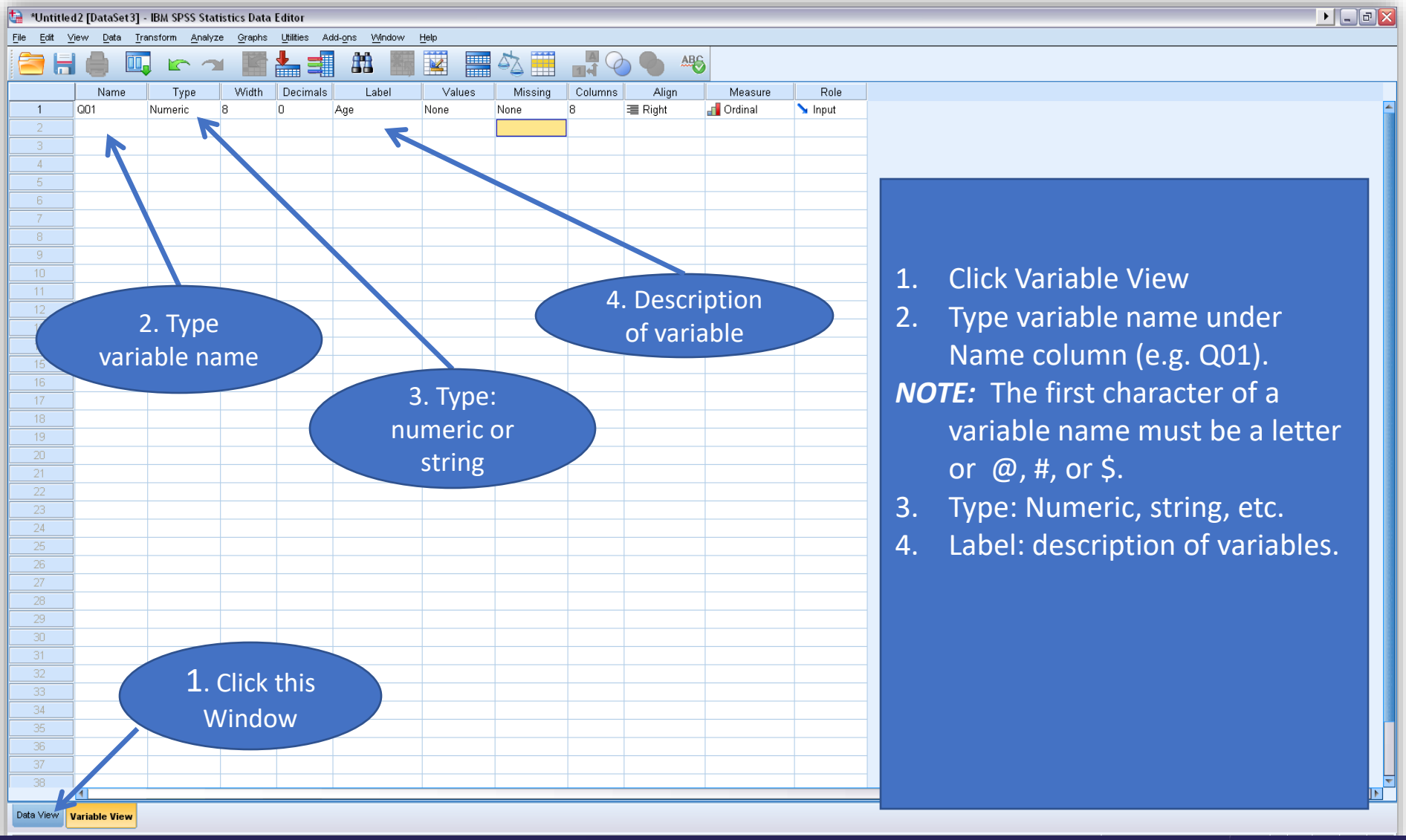
# Views and files in SPSS

- **Data view**
  - The place to enter data
  - Columns: variables
  - Rows: records
- **Variable view**
  - The place to enter variables
  - List of all variables
  - Characteristics of all variables
- **Output files**
  - The files that show your analysis and results
  - They open as separate files

# Entering data in SPSS



# Now Enter the Survey Data

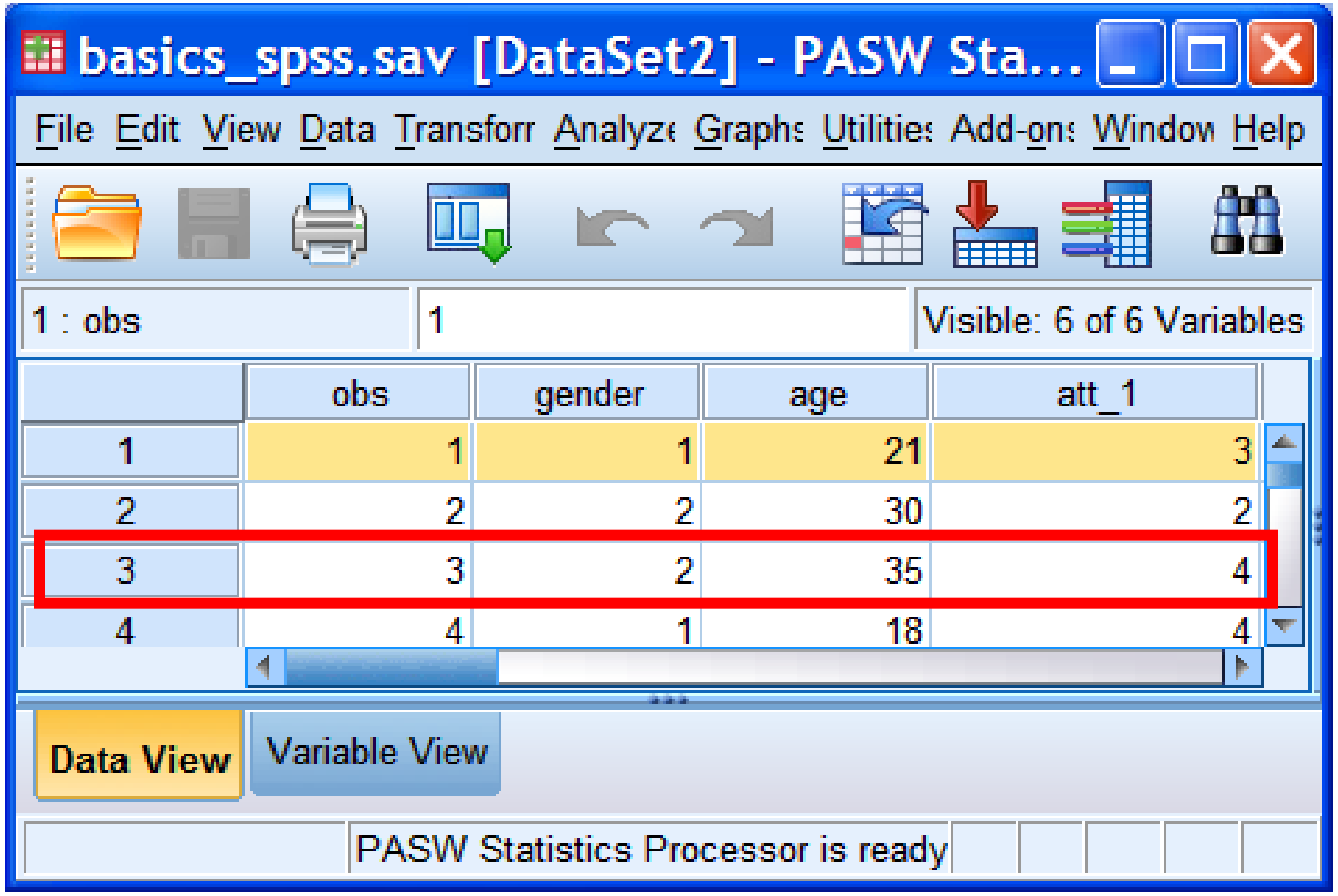



1. Click Variable View
2. Type variable name under Name column (e.g. Q01).

***NOTE:*** The first character of a variable name must be a letter or @, #, or $.

3. Type: Numeric, string, etc.
4. Label: description of variables.

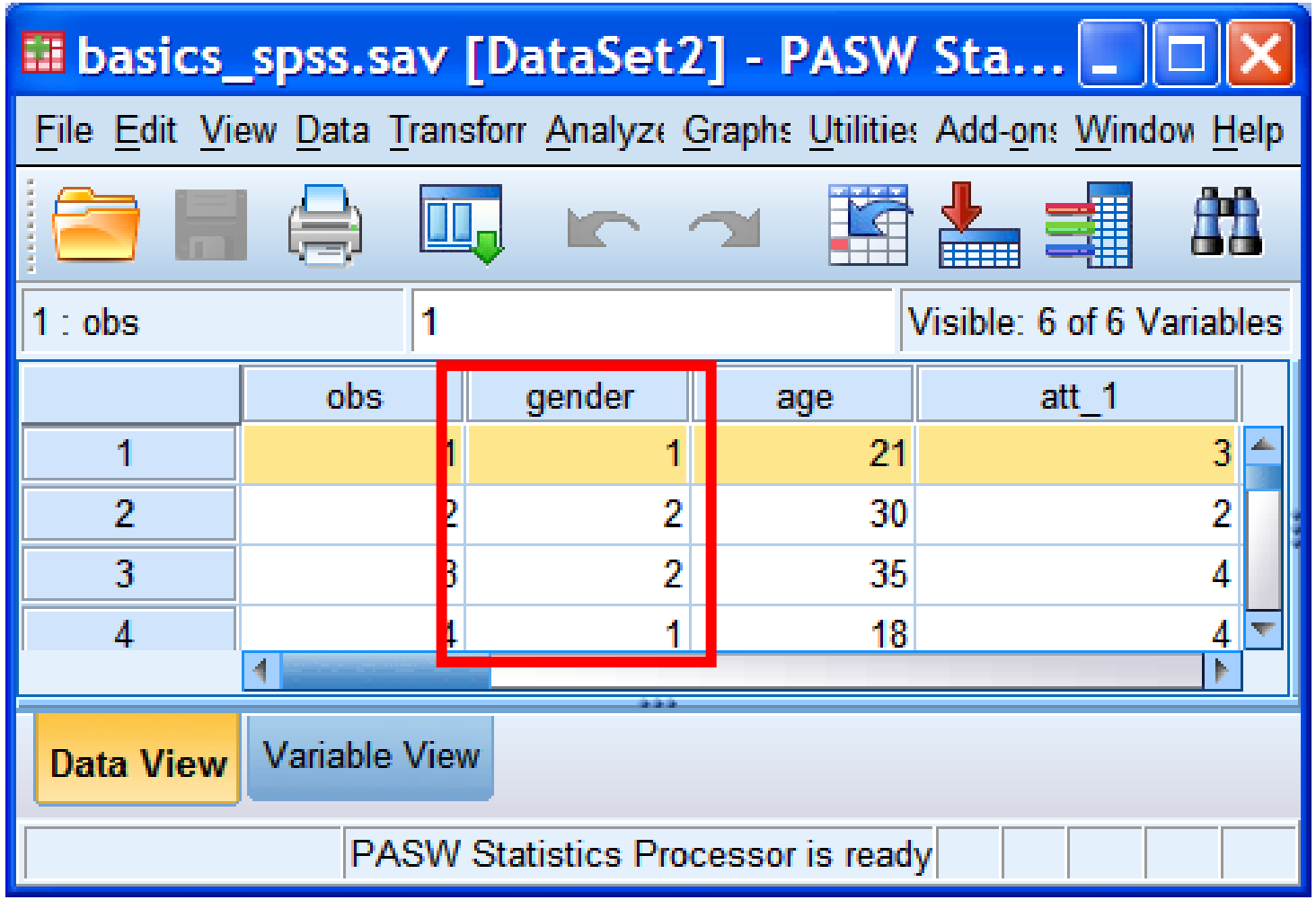
# How to read data?




| | obs | gender | age | att_1 |
|---|---|---|---|---|
| 1 | 1 | 1 | 21 | 3 |
| 2 | 2 | 2 | 30 | 2 |
| 3 | 3 | 2 | 35 | 4 |
| 4 | 4 | 1 | 18 | 4 |

- **In this data file, cases represent individual respondents to a survey**

# How to read data?




| | obs | gender | age | att_1 |
|---|---|---|---|---|
| 1 | 1 | 1 | 21 | 3 |
| 2 | 2 | 2 | 30 | 2 |
| 3 | 3 | 2 | 35 | 4 |
| 4 | 4 | 1 | 18 | 4 |

- Variables represent responses to each question asked in the survey.

SPSS introduction sample MBA large data set.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Extensions Window Help

| | ID | Gender | Age | Role | Salary | Jobsat | Motivation | var | var | var | var | var |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.00 | 1 | 50.00 | 5 | 37000 | 5.00 | 4.00 | | | | | |
| 2 | 2.00 | 1 | 24.00 | 2 | 24000 | 2.00 | 2.00 | | | | | |
| 3 | 3.00 | 2 | 23.00 | 2 | 22500 | 4.00 | 3.00 | | | | | |
| 4 | 4.00 | 1 | 30.00 | 4 | 27500 | 4.00 | 4.00 | | | | | |
| 5 | 5.00 | 2 | 45.00 | 1 | 16000 | 1.00 | 3.00 | | | | | |
| 6 | 6.00 | 2 | 63.00 | 3 | 25000 | 2.00 | 3.00 | | | | | |
| 7 | 7.00 | 1 | 31.00 | 2 | 25000 | 3.00 | 3.00 | | | | | |
| 8 | 8.00 | 1 | 38.00 | 2 | 25000 | 3.00 | 3.00 | | | | | |
| 9 | 9.00 | 2 | 55.00 | 2 | 25000 | 4.00 | 3.00 | | | | | |
| 10 | 10.00 | 1 | 20.00 | 2 | 20000 | 3.00 | 3.00 | | | | | |
| 11 | 11.00 | 1 | 58.00 | 2 | 25000 | 4.00 | 2.00 | | | | | |
| 12 | 12.00 | 1 | 18.00 | 1 | 16000 | 2.00 | 4.00 | | | | | |
| 13 | 13.00 | 1 | 44.00 | 2 | 25000 | 2.00 | 3.00 | | | | | |
| 14 | 14.00 | 1 | 22.00 | 2 | 25000 | 2.00 | 3.00 | | | | | |
| 15 | 15.00 | 2 | 26.00 | 2 | 24000 | 2.00 | 3.00 | | | | | |
| 16 | 16.00 | 1 | 60.00 | 5 | 35000 | 5.00 | 4.00 | | | | | |
| 17 | 17.00 | 1 | 53.00 | 2 | 25000 | 3.00 | 4.00 | | | | | |
| 18 | 18.00 | 1 | 48.00 | 2 | 25000 | 4.00 | 3.00 | | | | | |
| 19 | 19.00 | 2 | 30.00 | 3 | 24000 | 1.00 | 2.00 | | | | | |
| 20 | 20.00 | 1 | 28.00 | 2 | 24000 | 3.00 | 4.00 | | | | | |
| 21 | 21.00 | 1 | 35.00 | 2 | 25000 | 3.00 | 3.00 | | | | | |
| 22 | 22.00 | 1 | 22.00 | 2 | 22000 | 2.00 | 3.00 | | | | | |
| 23 | 23.00 | 1 | 40.00 | 4 | 28000 | 4.00 | 4.00 | | | | | |
| 24 | 24.00 | 1 | 62.00 | 1 | 16000 | 2.00 | 4.00 | | | | | |
| 25 | 25.00 | 2 | 42.00 | 2 | 25000 | 3.00 | 3.00 | | | | | |
| 26 | 26.00 | 1 | 31.00 | 2 | 25000 | 3.00 | 3.00 | | | | | |
| 27 | 27.00 | 1 | 33.00 | 2 | 25000 | 4.00 | 4.00 | | | | | |
| 28 | 28.00 | 1 | 28.00 | 2 | 24000 | 4.00 | 4.00 | | | | | |
| 29 | 29.00 | 2 | 27.00 | 4 | 27500 | 3.00 | 4.00 | | | | | |
| 30 | 30.00 | 2 | 37.00 | 2 | 25000 | 2.00 | 3.00 | | | | | |
| 31 | 31.00 | 1 | 42.00 | 2 | 25000 | 3.00 | 4.00 | | | | | |
| 32 | 32.00 | 1 | 23.00 | 2 | 22500 | 4.00 | 4.00 | | | | | |
| 33 | 33.00 | 1 | 19.00 | 1 | 16000 | 1.00 | 3.00 | | | | | |
| 34 | 35.00 | 1 | 18.00 | 1 | 16000 | 2.00 | 4.00 | | | | | |
| 35 | | | | | | | | | | | | |
| 36 | | | | | | | | | | | | |

Data View | Variable View

In Data View

SPSS introduction sample MBA large data set.sav [DataSet1] - IBM SPSS Statistics Data Editor

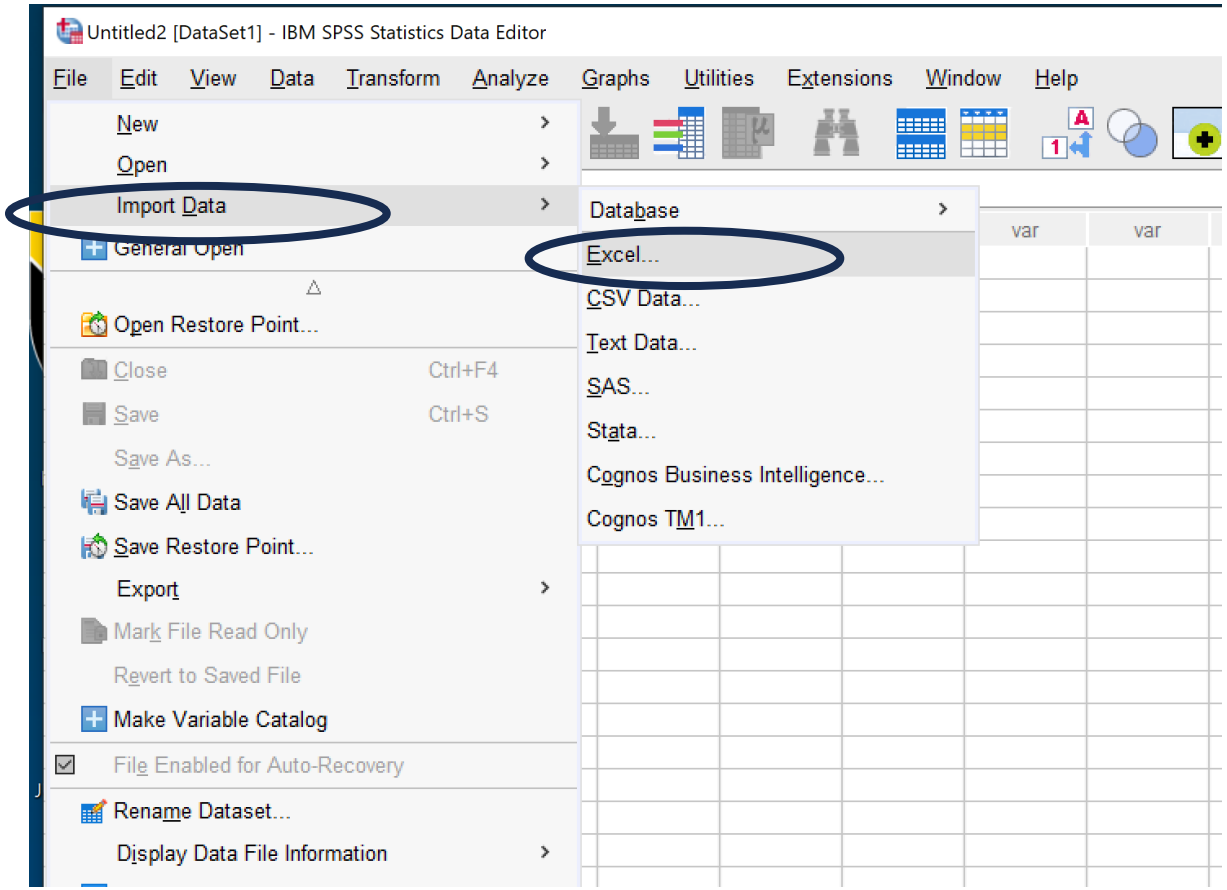
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Extensions Window Help

| | Name | Type | Width | Decimals | Label | Values | Missing | Columns | Align | Measure | Role |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ID | Numeric | 8 | 2 | ID | None | None | 8 | Right | Nominal | Input |
| 2 | Gender | Numeric | 6 | 0 | Gender | {1, Male}... | None | 6 | Right | Nominal | Input |
| 3 | Age | Numeric | 8 | 2 | Age | None | None | 8 | Right | Scale | Input |
| 4 | Role | Numeric | 11 | 0 | Role | {1, Tempora... | None | 11 | Right | Nominal | Input |
| 5 | Salary | Numeric | 6 | 0 | Salary | None | None | 12 | Right | Scale | Input |
| 6 | Jobsat | Numeric | 8 | 2 | Job Satisfaction | {1.00, Very ... | None | 8 | Right | Scale | Input |
| 7 | Motivation | Numeric | 8 | 2 | Motivation | {1.00, Very l... | None | 8 | Right | Scale | Input |

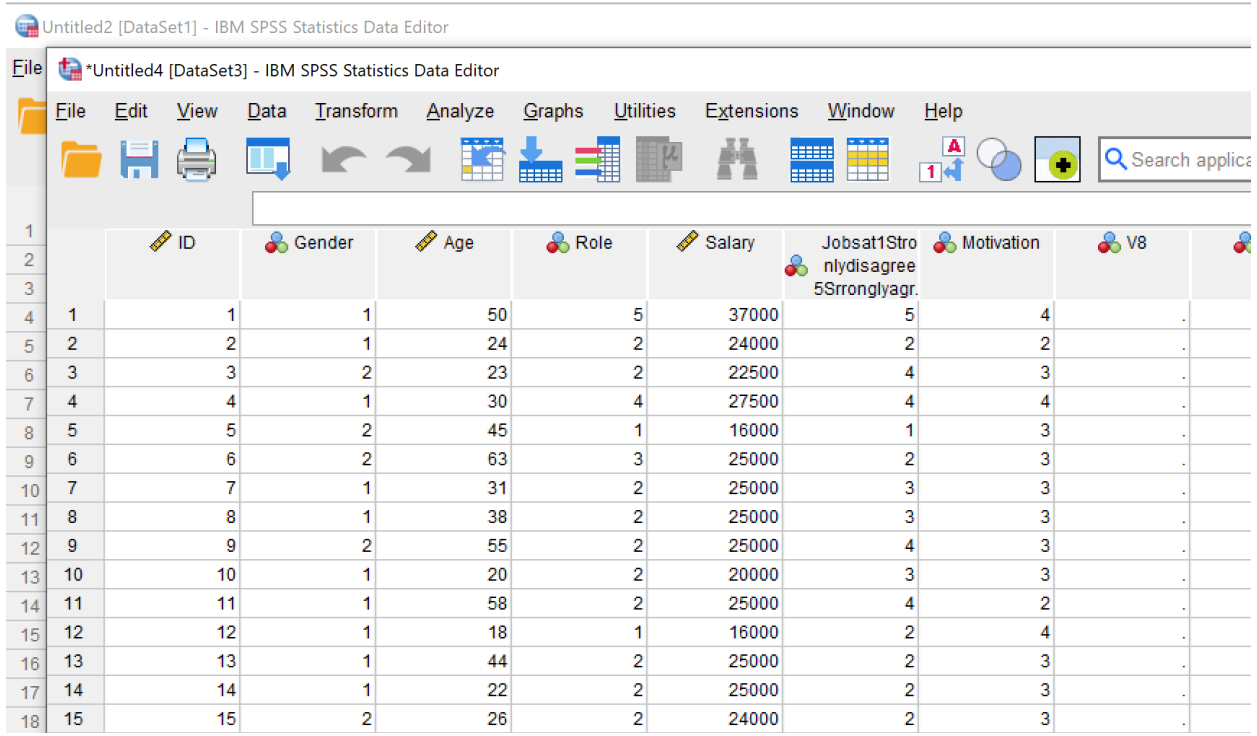
Data View | Variable View

In Variable View

# 2. File Import



# 2. File import




| | ID | Gender | Age | Role | Salary | Jobsat1Stronlydisagree5Srronglyagr. | Motivation | V8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 50 | 5 | 37000 | 5 | 4 | . |
| 2 | 2 | 1 | 24 | 2 | 24000 | 2 | 2 | . |
| 3 | 3 | 2 | 23 | 2 | 22500 | 4 | 3 | . |
| 4 | 4 | 1 | 30 | 4 | 27500 | 4 | 4 | . |
| 5 | 5 | 2 | 45 | 1 | 16000 | 1 | 3 | . |
| 6 | 6 | 2 | 63 | 3 | 25000 | 2 | 3 | . |
| 7 | 7 | 1 | 31 | 2 | 25000 | 3 | 3 | . |
| 8 | 8 | 1 | 38 | 2 | 25000 | 3 | 3 | . |
| 9 | 9 | 2 | 55 | 2 | 25000 | 4 | 3 | . |
| 10 | 10 | 1 | 20 | 2 | 20000 | 3 | 3 | . |
| 11 | 11 | 1 | 58 | 2 | 25000 | 4 | 2 | . |
| 12 | 12 | 1 | 18 | 1 | 16000 | 2 | 4 | . |
| 13 | 13 | 1 | 44 | 2 | 25000 | 2 | 3 | . |
| 14 | 14 | 1 | 22 | 2 | 25000 | 2 | 3 | . |
| 15 | 15 | 2 | 26 | 2 | 24000 | 2 | 3 | . |

# 2. Data Input – Cleaning data

- SPSS
- **How to Use SPSS - An Introduction to SPSS for Beginners (10:15 mins)**

[01 How to Use SPSS - An Introduction to SPSS for Beginners (youtube.com)]

# 2. Data Input – Cleaning data

- Key in values and labels for each variable
- Run frequencies for each variable
- Check outputs to see if you have variables with ‘odd’ values, e.g. 44 instead of 4; or extra cases without data
- Double check missing values.
- Think about recoding variables.

# 2. Data Input – Cleaning data

**Statistics**

| | | Gender | Age | Role | Salary | Job Satisfaction | Motivation |
|---|---|---|---|---|---|---|---|
| N | Valid | 34 | 34 | 34 | 34 | 34 | 34 |
| | Missing | 0 | 0 | 0 | 0 | 0 | 0 |
| Mean | | 1.26 | 36.0294 | 2.26 | 24029.41 | 2.9118 | 3.3235 |
| Median | | 1.00 | 32.0000 | 2.00 | 25000.00 | 3.0000 | 3.0000 |
| Mode | | 1 | 18.00[a] | 2 | 25000 | 2.00[a] | 3.00 |
| Std. Deviation | | .448 | 13.70796 | 1.024 | 4572.548 | 1.08342 | .63821 |
| Sum | | 43 | 1225.00 | 77 | 817000 | 99.00 | 113.00 |

a. Multiple modes exist. The smallest value is shown

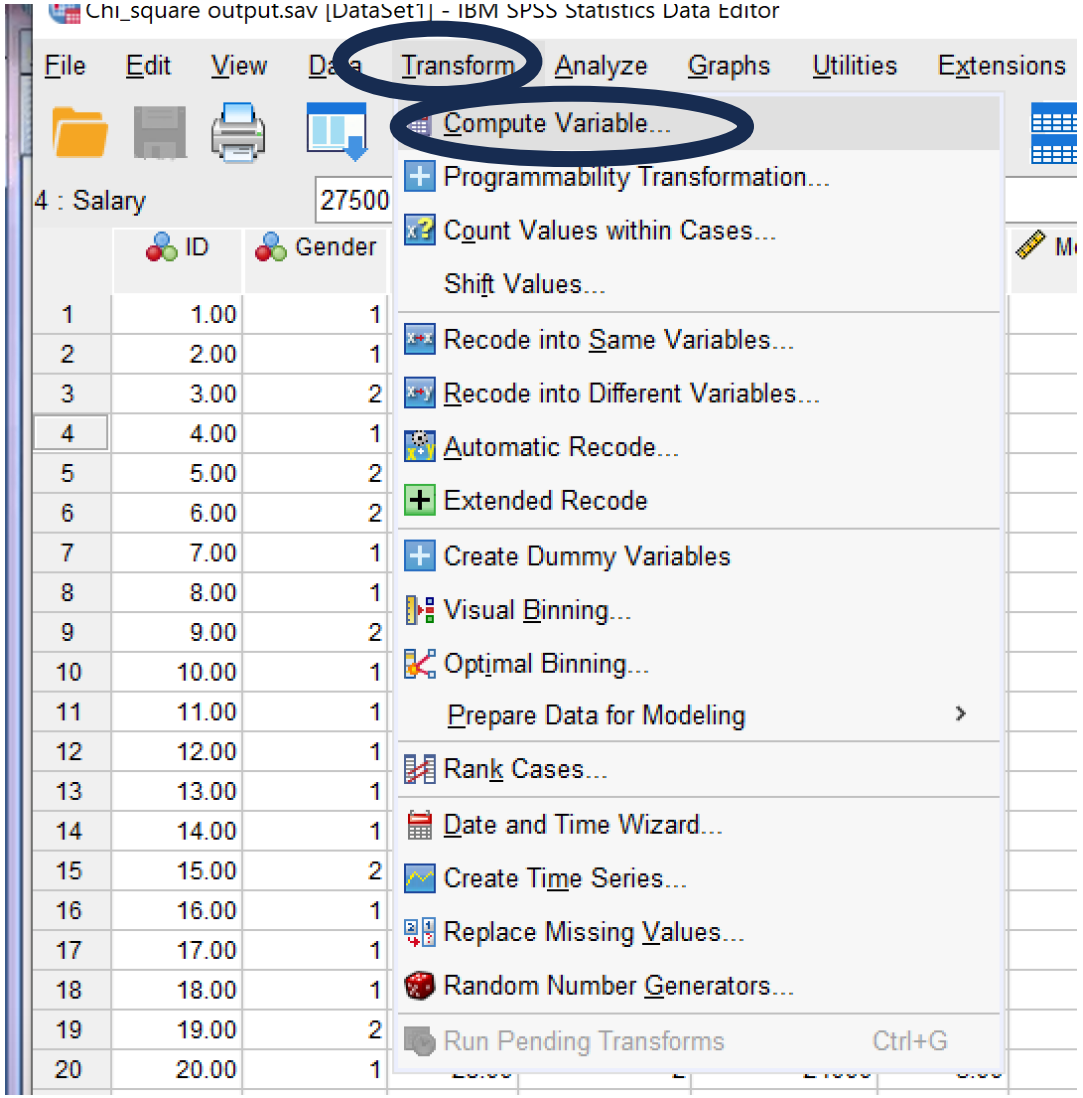
# 2. Data Input – Cleaning data

## Screening for Impermissible Values in the Data

- There are times when respondents simply key in a value wrong or list an invalid response to an inquiry. To test if an answer is outside of an acceptable range, you need to go to your SPSS file and check for Minimum and Maximum Value. To do so
- Go to **Analyze -> Descriptive Statistics -> Descriptives**

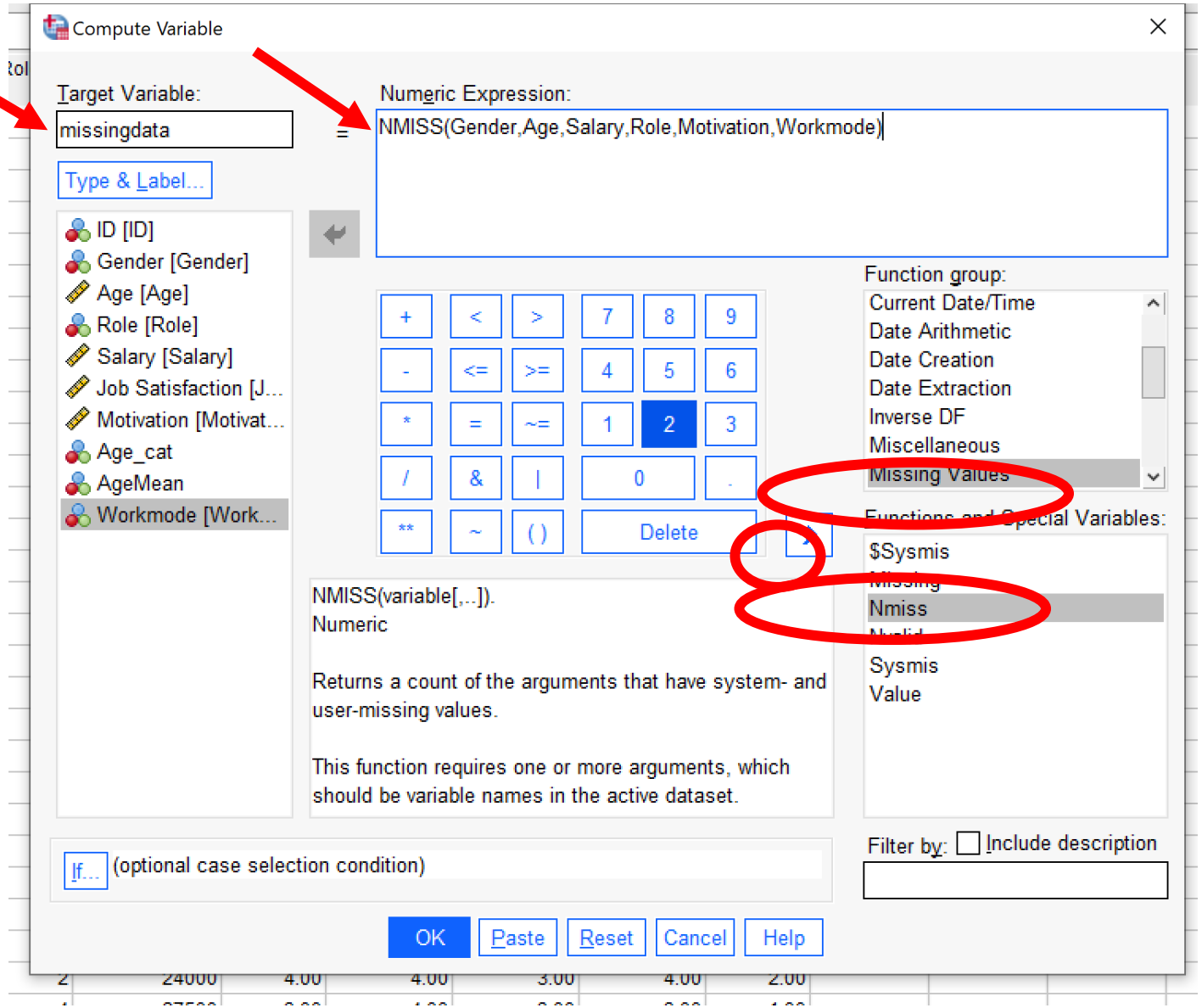
# 2. Data Input – Cleaning data

## Check Missing Data



# 2. Data Input

## Find missing data



*Chi_square output.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

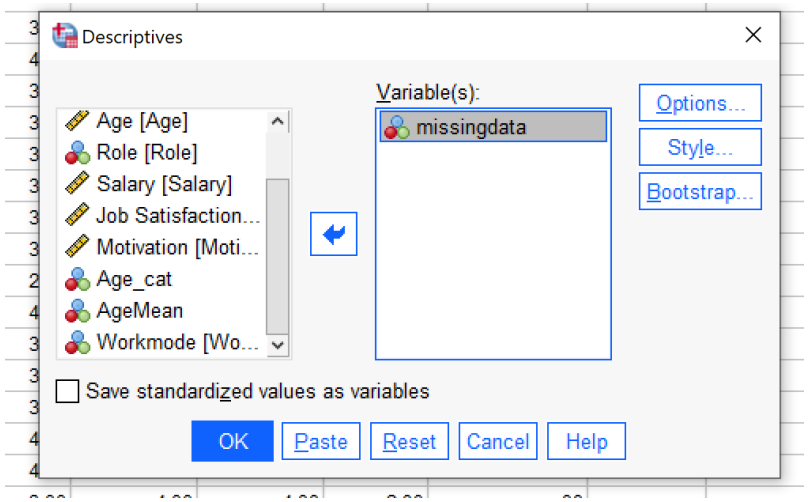
Search application

4 : Salary 27500

| | ID | Gender | Age | Role | Salary | Jobsat | Motivation | Age_cat | AgeMean | Workmode | missingdata | var |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.00 | 1 | 50.00 | 5 | 37000 | 5.00 | 4.00 | 4.00 | 5.00 | 1.00 | .00 | |
| 2 | 2.00 | 1 | 24.00 | 2 | 24000 | 2.00 | 2.00 | 2.00 | 2.00 | 1.00 | .00 | |
| 3 | 3.00 | 2 | 23.00 | 2 | 22500 | 4.00 | 3.00 | 2.00 | 4.00 | 1.00 | .00 | |
| 4 | 4.00 | 1 | 30.00 | 4 | 27500 | 4.00 | 4.00 | 3.00 | 4.00 | 1.00 | .00 | |
| 5 | 5.00 | 2 | 45.00 | 1 | 16000 | 1.00 | 3.00 | 4.00 | 1.00 | 1.00 | .00 | |
| 6 | 6.00 | 2 | 63.00 | 3 | 25000 | 2.00 | 3.00 | 5.00 | 2.00 | 1.00 | .00 | |
| 7 | 7.00 | 1 | 31.00 | 2 | 25000 | 3.00 | 3.00 | 3.00 | 3.00 | 2.00 | .00 | |
| 8 | 8.00 | 1 | 38.00 | 2 | 25000 | 3.00 | 3.00 | 3.00 | 3.00 | 2.00 | .00 | |
| 9 | 9.00 | 2 | 55.00 | 2 | 25000 | 4.00 | 3.00 | 4.00 | 4.00 | 2.00 | .00 | |
| 10 | 10.00 | 1 | 20.00 | 2 | 20000 | 3.00 | 3.00 | 2.00 | 3.00 | 2.00 | .00 | |
| 11 | 11.00 | 1 | 58.00 | 2 | 25000 | 4.00 | 2.00 | 5.00 | 4.00 | 2.00 | .00 | |
| 12 | 12.00 | 1 | 18.00 | 1 | 16000 | 2.00 | 4.00 | 1.00 | 2.00 | 2.00 | .00 | |
| 13 | 13.00 | 1 | 44.00 | 2 | 25000 | 2.00 | 3.00 | 4.00 | 2.00 | 2.00 | .00 | |
| 14 | 14.00 | 1 | 22.00 | 2 | 25000 | 2.00 | 3.00 | 2.00 | 2.00 | 2.00 | .00 | |
| 15 | 15.00 | 2 | 26.00 | 2 | 24000 | 2.00 | 3.00 | 3.00 | 2.00 | 2.00 | .00 | |
| 16 | 16.00 | 1 | 60.00 | 5 | 35000 | 5.00 | 4.00 | 5.00 | 5.00 | 2.00 | .00 | |
| 17 | 17.00 | 1 | 53.00 | 2 | 25000 | 3.00 | 4.00 | 4.00 | 3.00 | 2.00 | .00 | |
| 18 | 18.00 | 1 | 48.00 | 2 | 25000 | 4.00 | 3.00 | 4.00 | 4.00 | 2.00 | .00 | |
| 19 | 19.00 | 2 | 30.00 | 3 | 24000 | 1.00 | 2.00 | 3.00 | 1.00 | 2.00 | .00 | |
| 20 | 20.00 | 1 | 28.00 | 2 | 24000 | 3.00 | 4.00 | 3.00 | 3.00 | 1.00 | .00 | |
| 21 | 21.00 | 1 | 35.00 | 2 | 25000 | 3.00 | 3.00 | 3.00 | 3.00 | 1.00 | .00 | |
| 22 | 22.00 | 1 | 22.00 | 2 | 22000 | 2.00 | 3.00 | 2.00 | 2.00 | 1.00 | .00 | |
| 23 | 23.00 | 1 | 40.00 | 4 | 28000 | 4.00 | 4.00 | 3.00 | 4.00 | 1.00 | .00 | |
| 24 | 24.00 | 1 | 62.00 | 1 | 16000 | 2.00 | 4.00 | 5.00 | 2.00 | 1.00 | .00 | |
| 25 | 25.00 | 2 | 42.00 | 2 | 25000 | 3.00 | 3.00 | 4.00 | 3.00 | 1.00 | .00 | |
| 26 | 26.00 | 1 | 31.00 | 2 | 25000 | 3.00 | 3.00 | 3.00 | 3.00 | 1.00 | .00 | |
| 27 | 27.00 | 1 | 33.00 | 2 | 25000 | 4.00 | 4.00 | 3.00 | 4.00 | 1.00 | .00 | |
| 28 | 28.00 | 1 | 28.00 | 2 | 24000 | 4.00 | 4.00 | 3.00 | 4.00 | 2.00 | .00 | |
| 29 | 29.00 | 2 | 27.00 | 4 | 27500 | 3.00 | 4.00 | 3.00 | 3.00 | 1.00 | .00 | |
| 30 | 30.00 | 2 | 37.00 | 2 | 25000 | 2.00 | 3.00 | 3.00 | 2.00 | 1.00 | .00 | |

# 2. Data Input

## Identify the missing data in SPSS



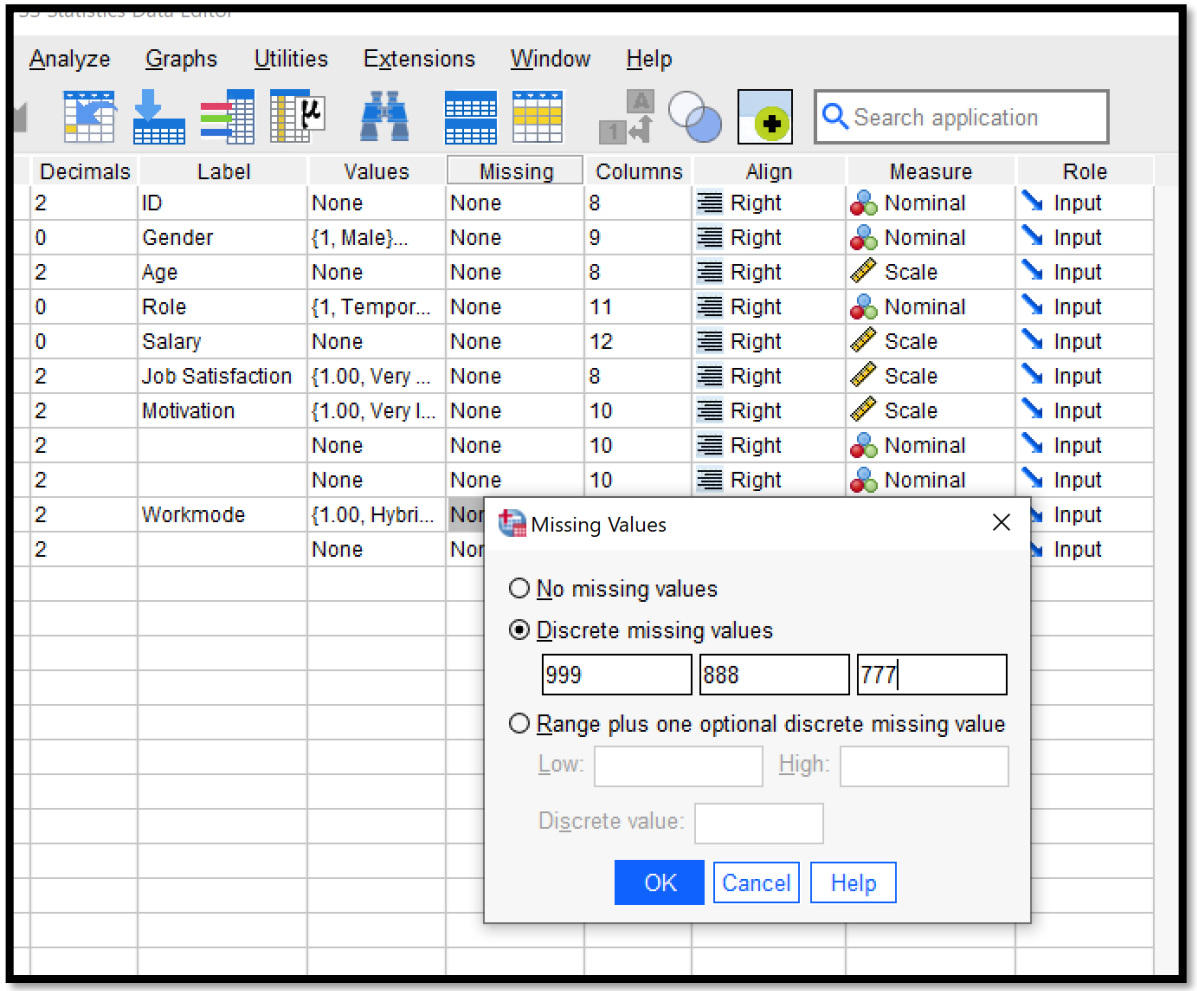

**Descriptives**

**Descriptive Statistics**

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| missingdata | 46 | .00 | .00 | .0000 | .00000 |
| Valid N (listwise) | 46 | | | | |

# 2. Data Input

How to tell SPSS to omit the missing data during the analysis?



# 2. Data input

## Secondary data input

**Table 1: Overseas residents visits to the UK by month and area of residence**

This worksheet contains one table. Freeze panes is turned on.

To turn off freeze panes select the 'View' ribbon then 'Freeze Panes' then 'Unfreeze Panes' or use [Alt W, F]

| Period | North America | Europe | EU | EU15 | Other EU | Other Countries | World Total | Seasonally Adjusted World Total |
|---|---|---|---|---|---|---|---|---|
| 2023 January | 372,000 | 1,972,000 | 1,807,000 | 1,445,000 | 362,000 | 553,000 | 2,898,000 | Not available |
| 2023 February | 217,000 | 1,666,000 | 1,512,000 | 1,215,000 | 298,000 | 309,000 | 2,192,000 | Not available |
| 2023 March | 405,000 | 1,829,000 | 1,649,000 | 1,295,000 | 352,000 | 363,000 | 2,596,000 | Not available |
| 2023 April | 498,000 | 2,203,000 | 1,996,000 | 1,715,000 | 282,000 | 425,000 | 3,126,000 | Not available |
| 2023 May | 684,000 | 2,071,000 | 1,865,000 | 1,484,000 | 379,000 | 545,000 | 3,300,000 | Not available |
| 2023 June | 773,000 | 1,982,000 | 1,811,000 | 1,494,000 | 317,000 | 725,000 | 3,479,000 | Not available |
| 2023 July | 758,000 | 2,354,000 | 2,077,000 | 1,734,000 | 346,000 | 900,000 | 4,012,000 | Not available |
| 2023 August | 659,000 | 2,274,000 | 2,119,000 | 1,773,000 | 345,000 | 877,000 | 3,810,000 | Not available |
| 2023 September | 513,000 | 1,968,000 | 1,799,000 | 1,474,000 | 324,000 | 624,000 | 3,106,000 | Not available |
| 2023 October | 567,000 | 2,453,000 | 2,225,000 | 1,815,000 | 406,000 | 695,000 | 3,715,000 | Not available |
| 2023 November | 372,000 | 1,946,000 | 1,716,000 | 1,360,000 | 347,000 | 476,000 | 2,795,000 | Not available |
| 2023 December | 307,000 | 2,107,000 | 1,924,000 | 1,669,000 | 269,000 | 516,000 | 2,931,000 | Not available |
| Latest three months | 1,250,000 | 6,510,000 | 5,860,000 | 4,840,000 | 1,020,000 | 1,690,000 | 9,440,000 | Not available |

# Exercise on secondary data input

Please run descriptive analysis to tell the difference

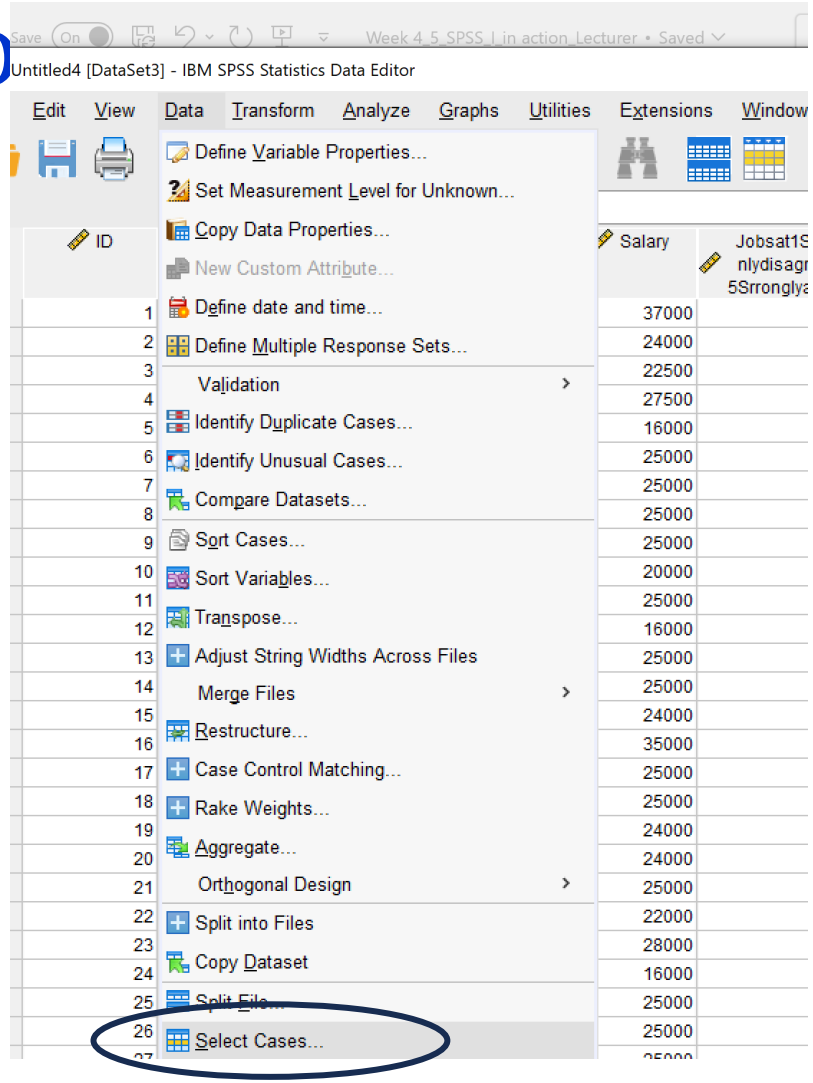
# 2. Data input

1 - Strongly Disagree
2 – Slightly Disagree
3 – Neutral
4 – Slightly Agree
5 – Strongly Agree

| Section A: Customer Orientation of your organization | | | | | | |
|---|---|---|---|---|---|---|
| No. | Statements | Scale | | | | |
| | | 1 | 2 | 3 | 4 | 5 |
| A1 | The customer is at the center of strategic planning in our organization | 6 | 6 | 8 | 45 | 35 |
| A2 | Meeting customer's needs and expectations is the commitment of our organization | 5 | 8 | 8 | 46 | 33 |
| A3 | In our organization there is an established framework for getting customers feedback | 6 | 8 | 6 | 45 | 35 |
| A4 | In order to track customer's expectation in our organization, different process are implemented | 6 | 6 | 8 | 35 | 45 |
| A5 | In our organization customer database are regularly updated | 6 | 9 | 5 | 47 | 33 |
| A6 | Our competitive advantage is based on building and maintaining long-term customer relationships | 6 | 8 | 4 | 42 | 40 |
| A7 | Our organization makes an effort to find out what our key customer needs | 5 | 8 | 8 | 46 | 33 |

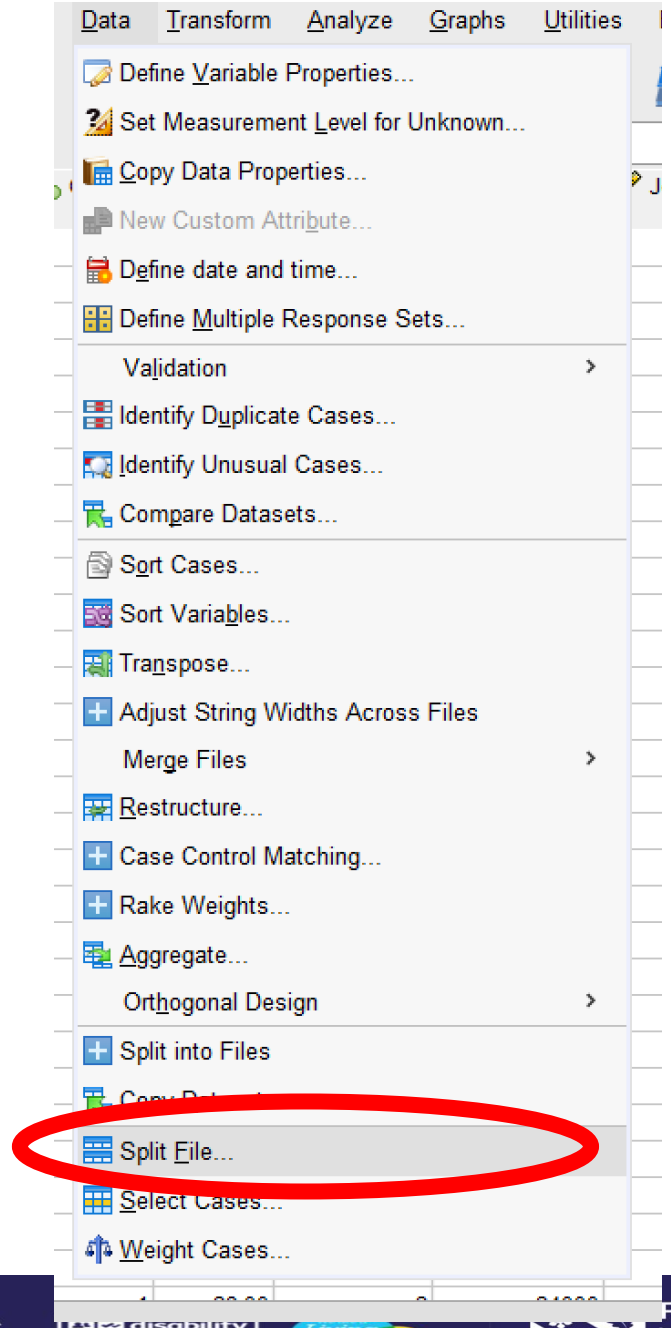
# 3.Basic data transformatio and computation

## Select cases



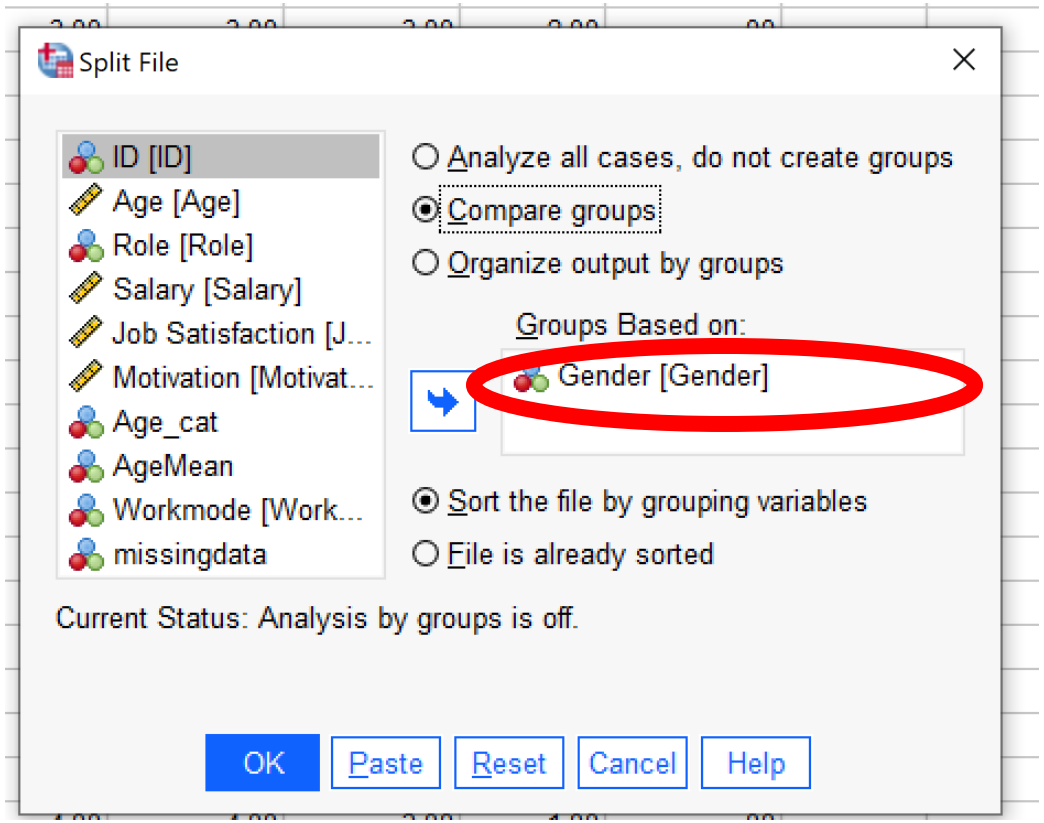
# 3.Basic data transformation and computation

## Spilt file in SPSS



# 3.Basic data transformation and computation

Spilt file in SPSS



# Spilt file-  by Gender

**Jobsat1 Stronlydisagree5Srronglyagree**

| Gender | | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|---|
| 1 | Valid | Strongly disagree | 1 | 4.0 | 4.0 | 4.0 |
| | | Disagree | 7 | 28.0 | 28.0 | 32.0 |
| | | Neutral | 8 | 32.0 | 32.0 | 64.0 |
| | | Agree | 7 | 28.0 | 28.0 | 92.0 |
| | | Strongly Agree | 2 | 8.0 | 8.0 | 100.0 |
| | | Total | 25 | 100.0 | 100.0 | |
| 2 | Valid | Strongly disagree | 2 | 22.2 | 22.2 | 22.2 |
| | | Disagree | 3 | 33.3 | 33.3 | 55.6 |
| | | Neutral | 2 | 22.2 | 22.2 | 77.8 |
| | | Agree | 2 | 22.2 | 22.2 | 100.0 |
| | | Total | 9 | 100.0 | 100.0 | |

# 3.Basic data transformation and computation

## Compare Mean

**Descriptives**

**Descriptive Statistics**

| Gender | | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|---|
| Male | Job Satisfaction | 39 | 1.00 | 5.00 | 3.0256 | 1.06344 |
| | Valid N (listwise) | 39 | | | | |
| Female | Job Satisfaction | 19 | 1.00 | 4.00 | 2.5263 | .84119 |
| | Valid N (listwise) | 19 | | | | |

## Organise Output by groups

**Descriptives**

**Gender = Male**

**Descriptive Statistics[a]**

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| Job Satisfaction | 39 | 1.00 | 5.00 | 3.0256 | 1.06344 |
| Valid N (listwise) | 39 | | | | |

a. Gender = Male

**Gender = Female**

**Descriptive Statistics[a]**

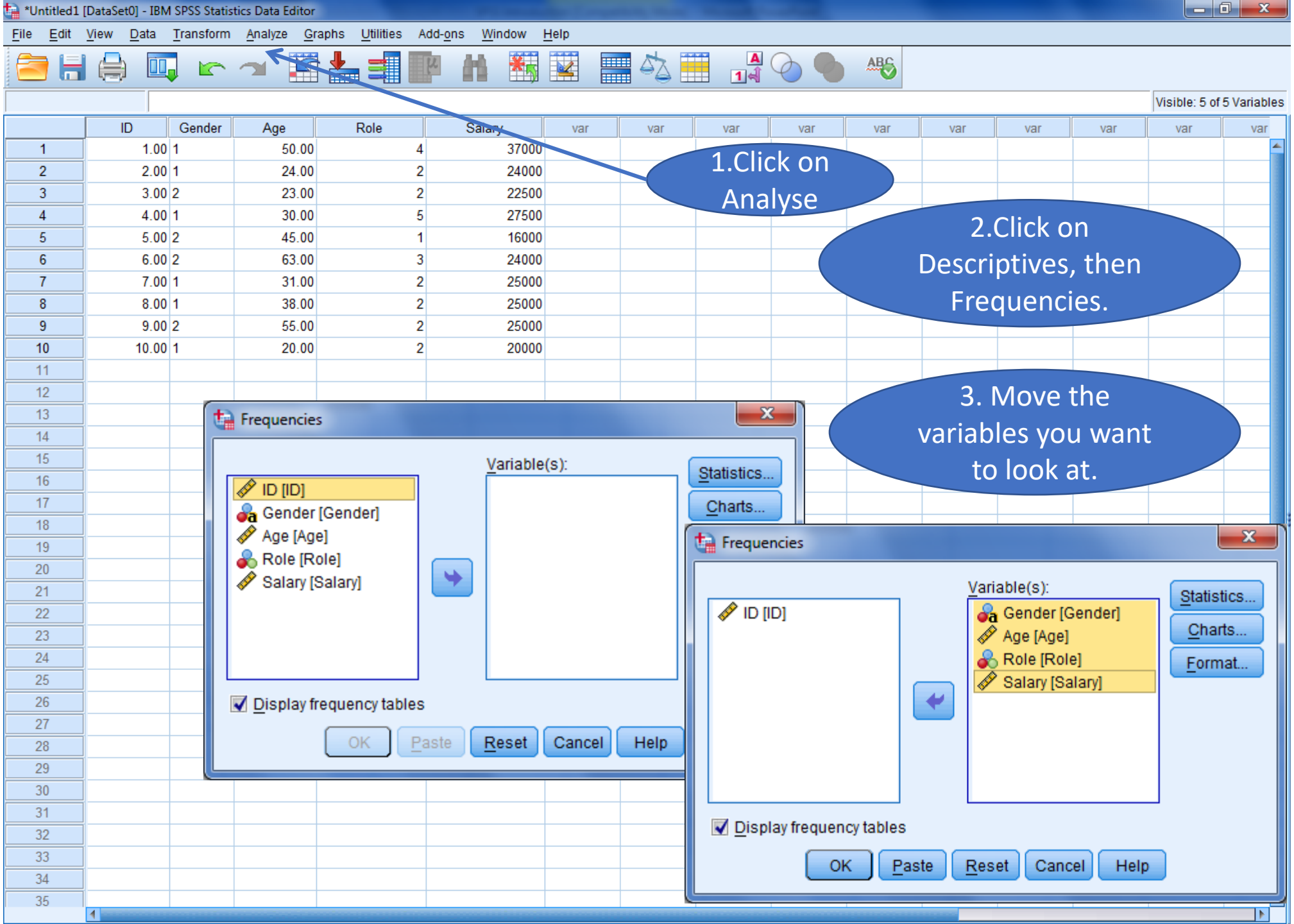
| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| Job Satisfaction | 19 | 1.00 | 4.00 | 2.5263 | .84119 |
| Valid N (listwise) | 19 | | | | |

a. Gender = Female

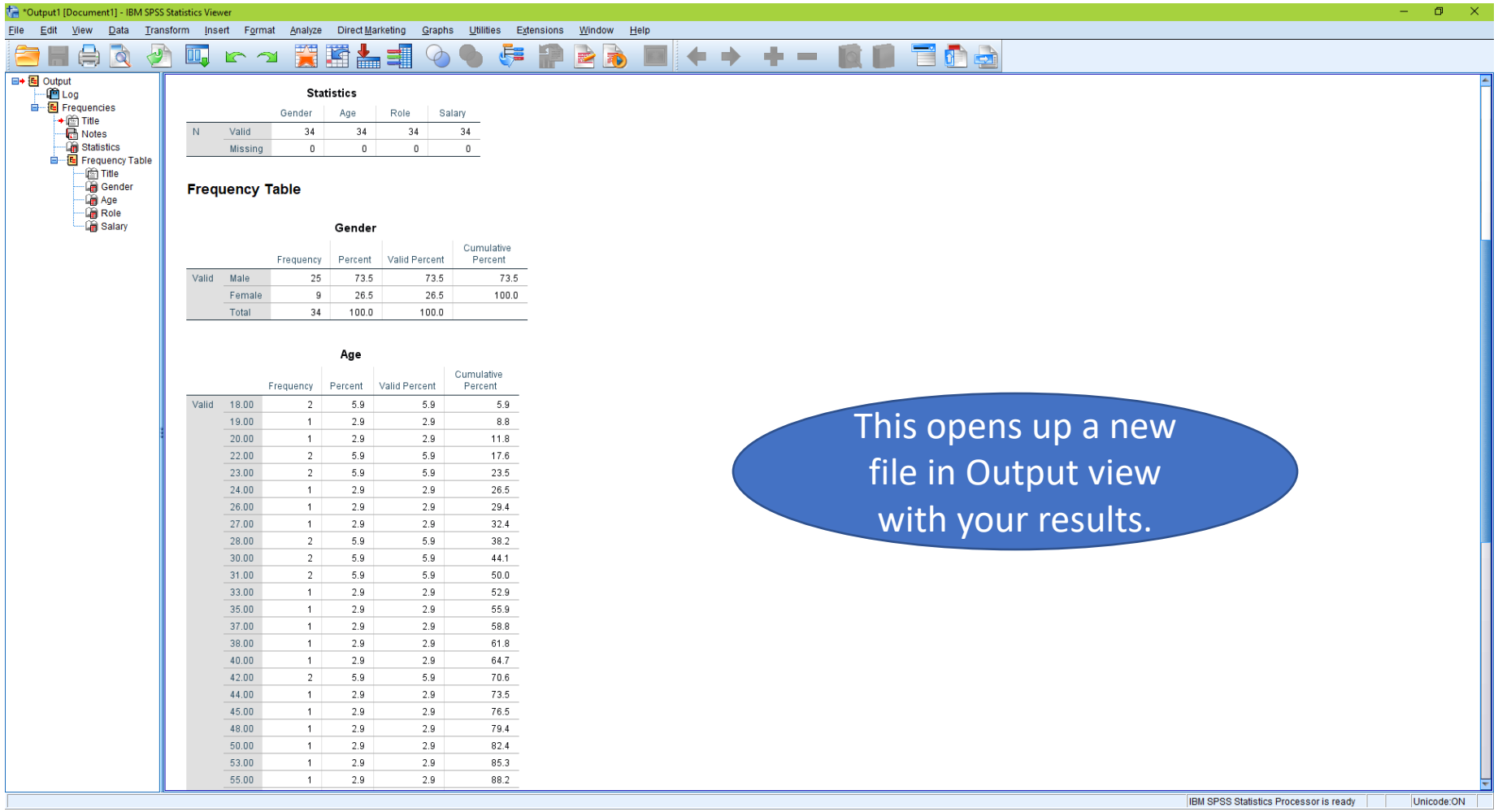
# 3. Basic data transformation and computation

Exercises on
Data handling
Missing Data
Spilt Data

# 4. Descriptive Analysis




| | ID | Gender | Age | Role | Salary |
|---|---|---|---|---|---|
| 1 | 1.00 | 1 | 50.00 | 4 | 37000 |
| 2 | 2.00 | 1 | 24.00 | 2 | 24000 |
| 3 | 3.00 | 2 | 23.00 | 2 | 22500 |
| 4 | 4.00 | 1 | 30.00 | 5 | 27500 |
| 5 | 5.00 | 2 | 45.00 | 1 | 16000 |
| 6 | 6.00 | 2 | 63.00 | 3 | 24000 |
| 7 | 7.00 | 1 | 31.00 | 2 | 25000 |
| 8 | 8.00 | 1 | 38.00 | 2 | 25000 |
| 9 | 9.00 | 2 | 55.00 | 2 | 25000 |
| 10 | 10.00 | 1 | 20.00 | 2 | 20000 |




**Statistics**

| | | Gender | Age | Role | Salary |
|---|---|---|---|---|---|
| N | Valid | 34 | 34 | 34 | 34 |
| | Missing | 0 | 0 | 0 | 0 |

## Frequency Table

**Gender**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | Male | 25 | 73.5 | 73.5 | 73.5 |
| | Female | 9 | 26.5 | 26.5 | 100.0 |
| | Total | 34 | 100.0 | 100.0 | |

**Age**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 18.00 | 2 | 5.9 | 5.9 | 5.9 |
| | 19.00 | 1 | 2.9 | 2.9 | 8.8 |
| | 20.00 | 1 | 2.9 | 2.9 | 11.8 |
| | 22.00 | 2 | 5.9 | 5.9 | 17.6 |
| | 23.00 | 2 | 5.9 | 5.9 | 23.5 |
| | 24.00 | 1 | 2.9 | 2.9 | 26.5 |
| | 26.00 | 1 | 2.9 | 2.9 | 29.4 |
| | 27.00 | 1 | 2.9 | 2.9 | 32.4 |
| | 28.00 | 2 | 5.9 | 5.9 | 38.2 |
| | 30.00 | 2 | 5.9 | 5.9 | 44.1 |
| | 31.00 | 2 | 5.9 | 5.9 | 50.0 |
| | 33.00 | 1 | 2.9 | 2.9 | 52.9 |
| | 35.00 | 1 | 2.9 | 2.9 | 55.9 |
| | 37.00 | 1 | 2.9 | 2.9 | 58.8 |
| | 38.00 | 1 | 2.9 | 2.9 | 61.8 |
| | 40.00 | 1 | 2.9 | 2.9 | 64.7 |
| | 42.00 | 2 | 5.9 | 5.9 | 70.6 |
| | 44.00 | 1 | 2.9 | 2.9 | 73.5 |
| | 45.00 | 1 | 2.9 | 2.9 | 76.5 |
| | 48.00 | 1 | 2.9 | 2.9 | 79.4 |
| | 50.00 | 1 | 2.9 | 2.9 | 82.4 |
| | 53.00 | 1 | 2.9 | 2.9 | 85.3 |
| | 55.00 | 1 | 2.9 | 2.9 | 88.2 |

This opens up a new file in Output view with your results.

# 4. Descriptive Analysis

Ways of running descriptive analysis in SPSS

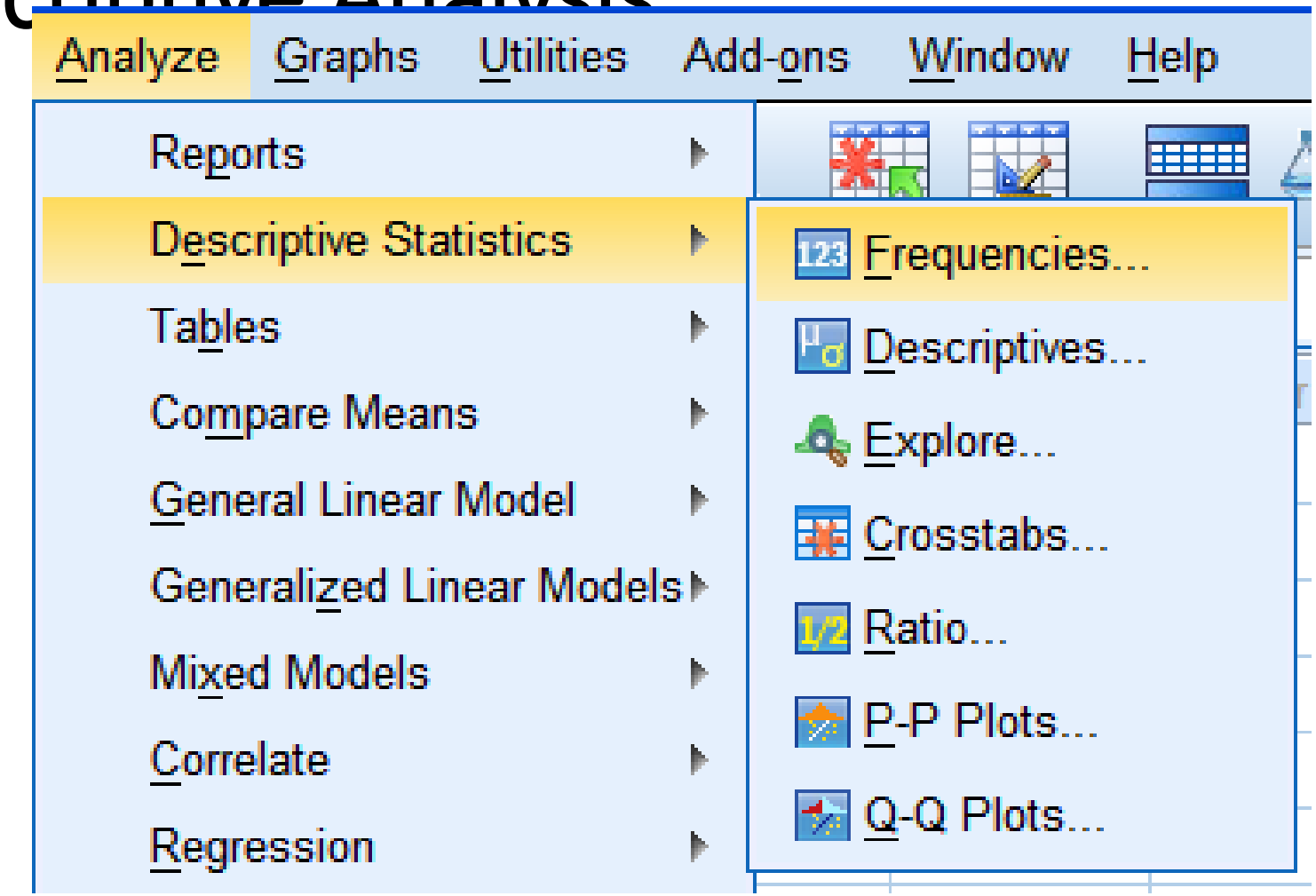
- Frequency
- Descriptive Analysis

# 4. Descriptive Analysis

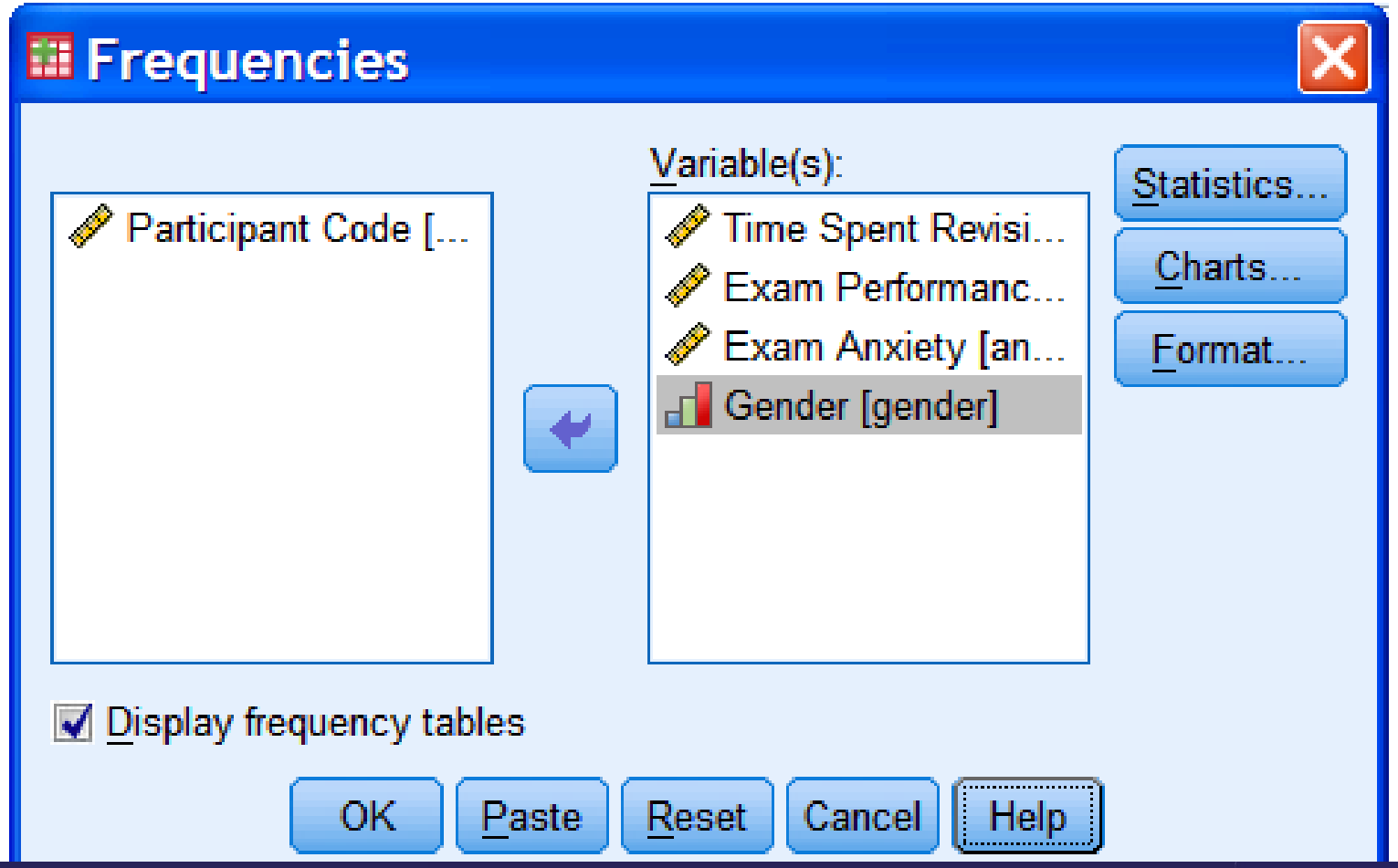
# 4. Descriptive Analysis

- Basic knowledge of the sample and targeted variables in a study
- Summary of data
- Anomalies in data

# 4. Descriptive Analysis



# 4. Descriptive Analysis



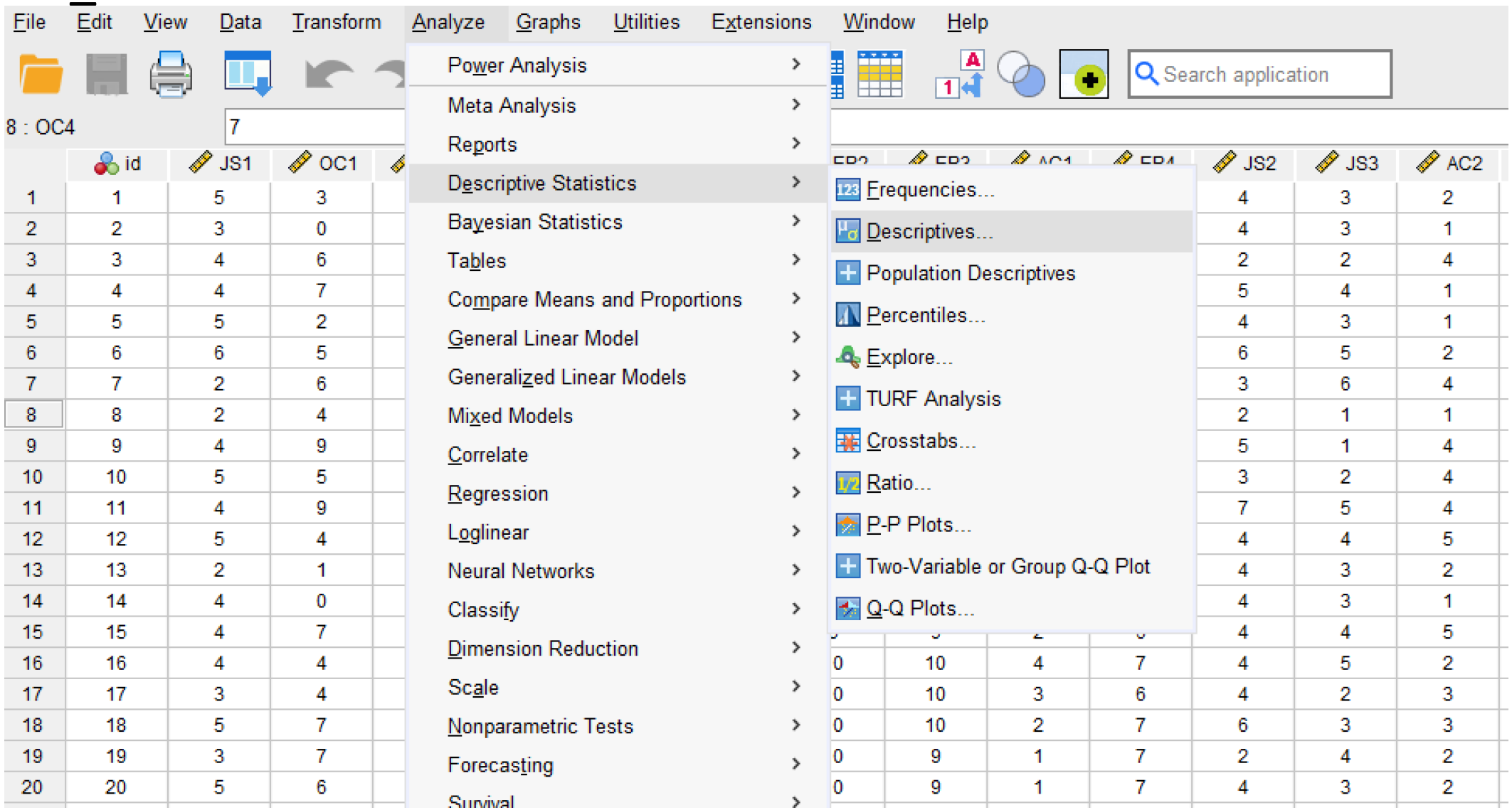
Summaries of individual variables provide an important "first look" at your data. Some of the tasks that these summaries help you to complete are listed below.

- Determining "typical" values of the variables. What values occur most often? What range of values are you likely to see?
- Checking the assumptions for statistical procedures. Do you have enough observations? For each variable, is the observed distribution of values adequate?
- Checking the quality of the data. Are there missing or mis-entered values? Are there values that should be recoded?

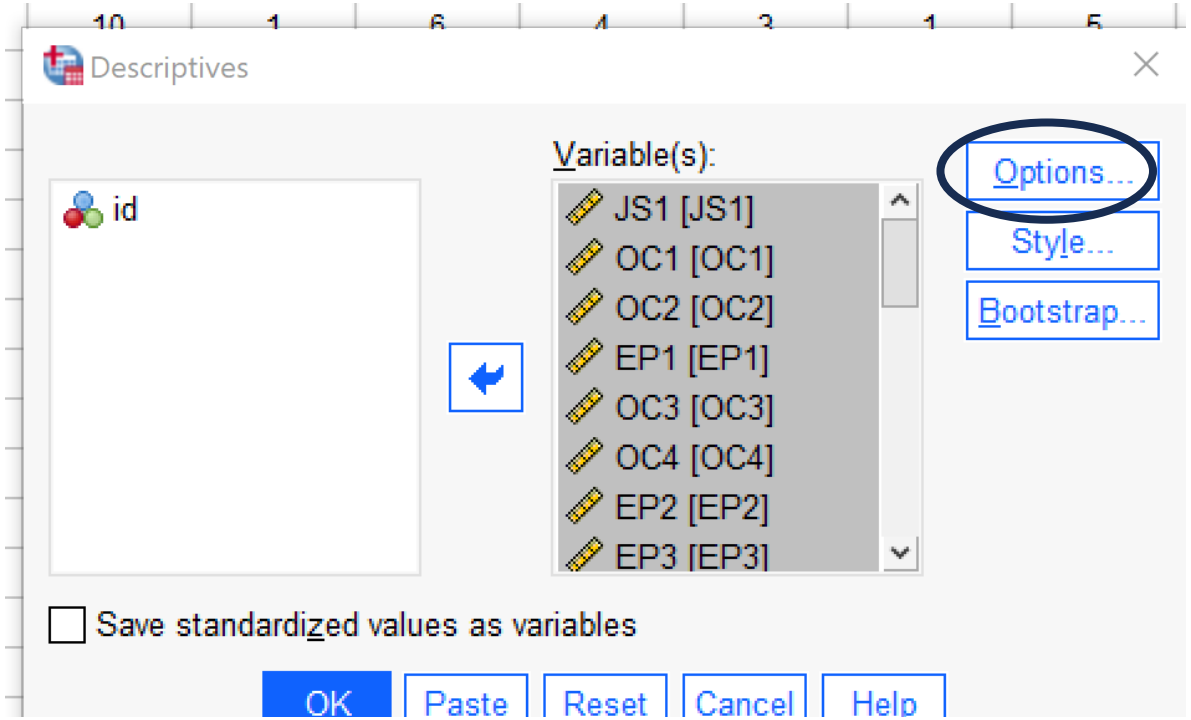
The Frequencies procedure is useful for obtaining summaries of individual variables.

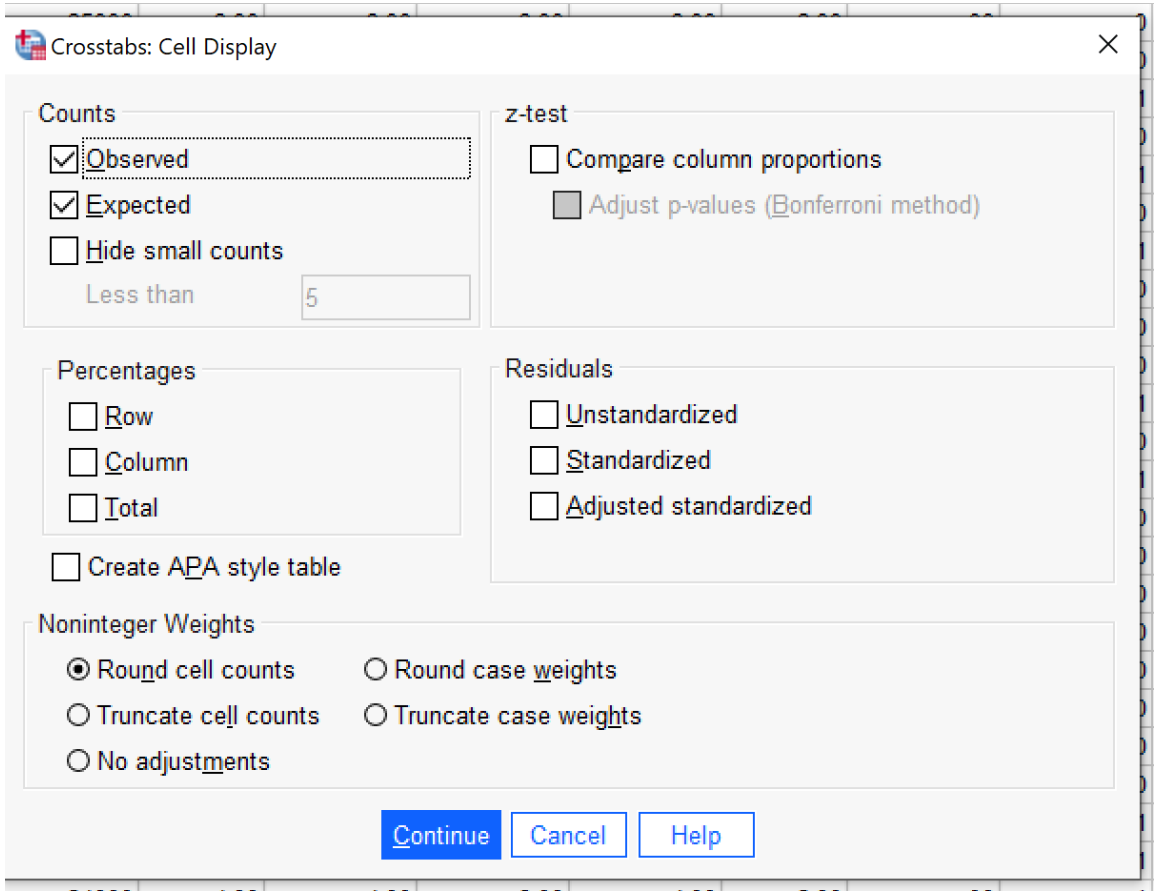
**Note: The variables to be analyzed are measured at nominal, ordinal, and scale levels.**

- HBAT_Full set



- HBAT_Full set



- HBAT_Full set

Descriptive analysis in SPSS

**Descriptive Statistics**

| | N | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|---|
| JS1 | 400 | 1 | 7 | 4.20 | 1.339 |
| OC1 | 400 | 0 | 10 | 4.89 | 2.525 |
| OC2 | 400 | 0 | 10 | 8.42 | 2.186 |
| EP1 | 400 | 0 | 10 | 8.53 | 1.831 |
| OC3 | 400 | 0 | 10 | 8.66 | 1.755 |
| OC4 | 400 | 0 | 10 | 8.41 | 2.053 |
| EP2 | 400 | 0 | 10 | 8.85 | 1.628 |
| EP3 | 400 | 0 | 10 | 8.93 | 1.335 |
| AC1 | 400 | 1 | 5 | 2.76 | 1.394 |
| EP4 | 400 | 1 | 7 | 5.83 | 1.393 |
| JS2 | 400 | 1 | 7 | 4.20 | 1.370 |
| JS3 | 400 | 1 | 6 | 3.21 | 1.316 |
| AC2 | 400 | 1 | 6 | 3.55 | 1.726 |
| SI1 | 400 | 1 | 5 | 4.20 | .871 |
| JS4 | 400 | 1 | 5 | 2.67 | 1.281 |
| SI2 | 400 | 1 | 5 | 4.20 | .877 |
| JS5 | 400 | 0 | 100 | 54.83 | 20.599 |
| AC3 | 400 | 1 | 5 | 2.78 | 1.419 |
| SI3 | 400 | 1 | 5 | 3.47 | 1.016 |
| AC4 | 400 | 1 | 6 | 3.21 | 1.611 |
| SI4 | 400 | 1 | 5 | 3.48 | .968 |
| C1 - Gender Class | 400 | 0 | 1 | .50 | .501 |
| C2 - Employee Type Class | 400 | 0 | 1 | .52 | .500 |
| C3 -Geographic Location Class | 400 | 0 | 1 | .50 | .501 |
| Age | 399 | 24 | 62 | 43.39 | 7.131 |
| EXP - Experience in Years | 399 | .1 | 25.0 | 8.637 | 3.6198 |
| JP - Performance | 399 | 1 | 5 | 4.36 | .974 |
| Valid N (listwise) | 399 | | | | |

- Presentation of demographic information

- Example

# 4 Findings

## 4.1 Demographic information

Table 2 presents the demographic information of the respondents who completed the valid questionnaire. The gender distribution was predominantly female (60.06%). The age range with the highest number of respondents was 18–24 years old, and as age increased, the number of respondents decreased. The proportion of respondents with and without a local place of birth was similar. Lastly, the largest occupation group among the respondents was students (55.17%), followed by enterprise employees (31.03%).

file:///C:/Users/h.lee/Downloads/fpsyg-14-1314223%20(11).pdf

https://onlinelibrary.wiley.com/doi/epdf/10.1002/cb.1974

# 5. Cronbach’s Alpha (α) - Reliability Test

# 5. Cronbach’s Alpha - Reliability Test

- To measure the internal consistency between items in a scale

- To make sure our scales are designed at the same direction

e.g. Not one question within the construct mixed with positively and negatively worded question = Alpha will be negative

=> Reverse code variables if needed

# 5. Cronbach’s Alpha - Reliability Test

- **Reliability Statistics Table**

**Reliability Statistics**

| Cronbach's Alpha | N of Items |
|---|---|
| .948 | 44 |

**Cronbach's Alpha value (0 – 1)**

**>0.7 (Considered good),** >0.6 acceptable

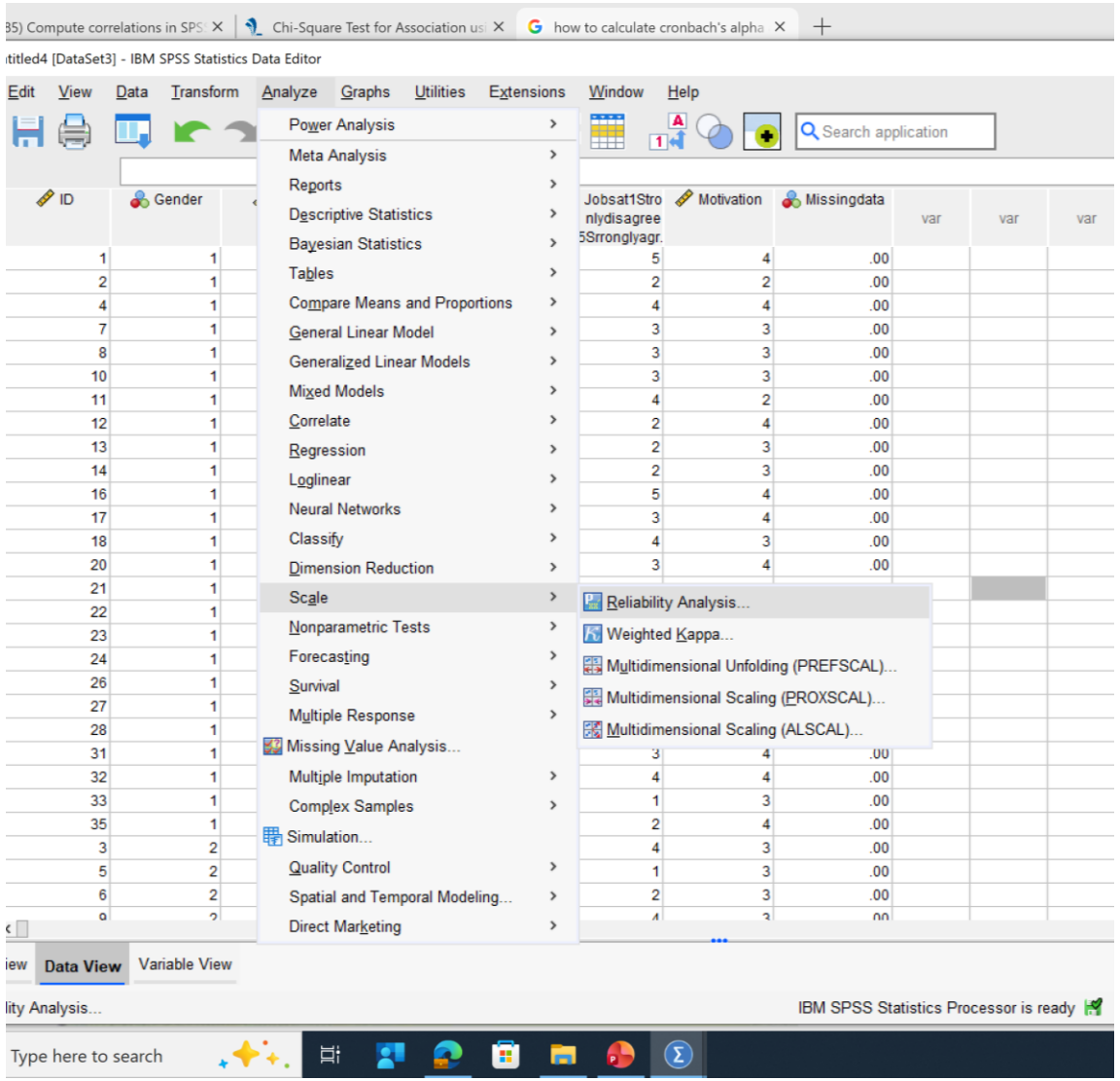
Must be **Cronbach's Alpha value**

α> 0.7 in the assignment

# 5. Cronbach’s Alpha Reliability Test

**Cronbach's Alpha**

**Analyze > Scale > Reliability Analysis**



# 6. Association and Correlation analysis

# Chi Square Test of Independence

- Very versatile basic statistic test to establish differences between groups Compares frequencies of distributions occurring in different categories or groups (categorial data)
- Tests whether group distributions are different
  - *Customer preference of 3 brands of nuts*
- determines the association between two or more variables by counting the frequencies of distribution
  - *Gender relative to study preference (alone or in group)*
- For more on this test see this video: [Introduction to Chi-Squre by Andy Field](/)

# Chi Square

The Chi-square test of independence is used to test hypothesis that two categorical variables are independent of each other.

The Chi-square test of independence is a popular method for testing hypothesis on discrete data. Discrete data can be nominal or ordinal.

$p < .05$ indicates that two variables are not independent of each other. One variable is associated with another.

# Assumptions

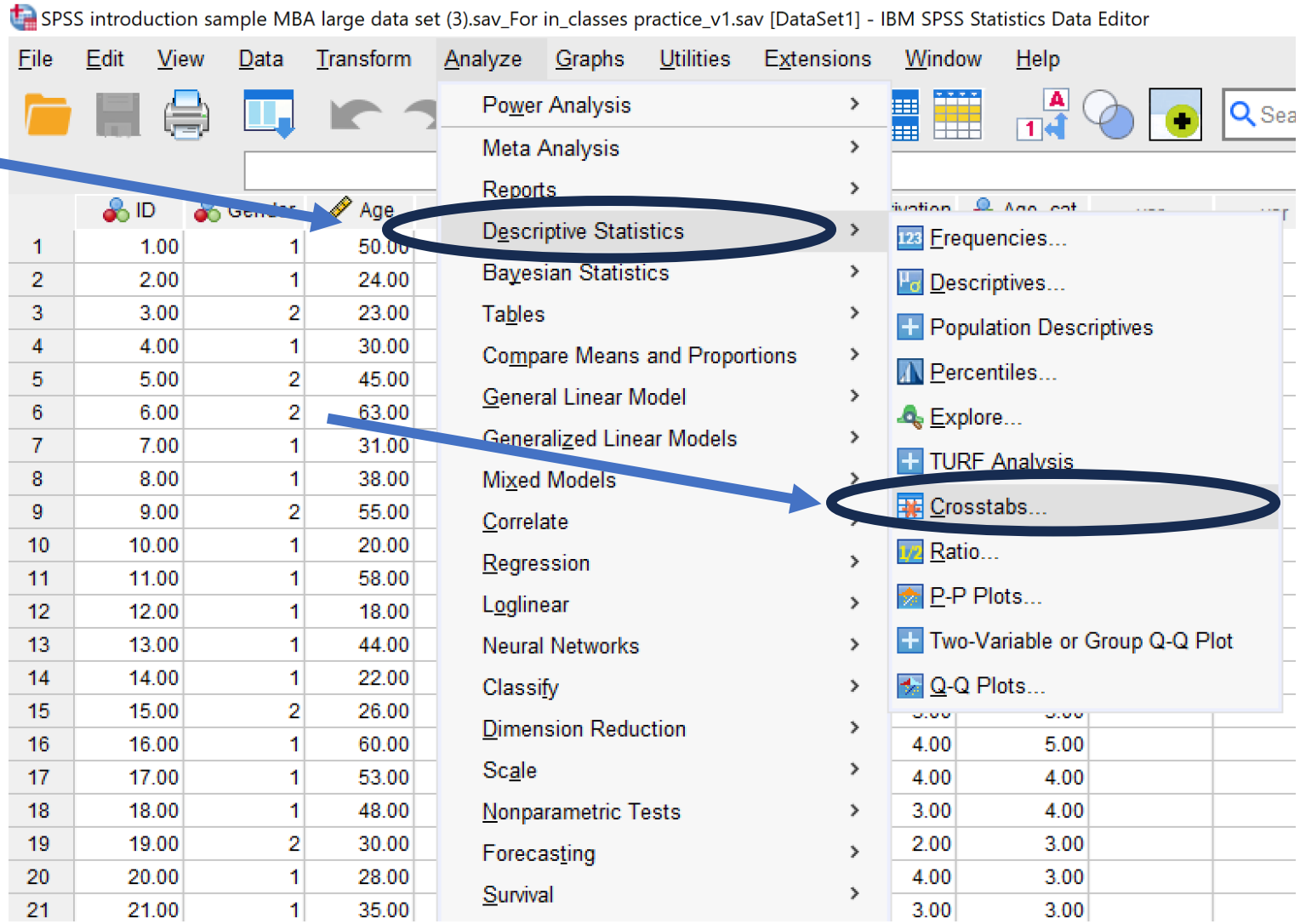
1. The data in the cells should be frequencies, or counts of cases rather than percentages or some other transformation of the data.
2. **Data must be categorical**
3. **Observations must be independent.** This means that a different test must be used if the two groups are related. For example, a different test must be used if the researcher's data consists of paired samples, such as in studies in which a parent is paired with his or her child.
4. The levels (or categories) of the variables are mutually exclusive. That is, a particular subject fits into one and only one level of each of the variables.

- The value of the cell *expecteds* should be 5 or more in at least 80% of the cells.

# Chi Square in SPSS

[Chi-square in SPSS (5:02 mins)](https://www.youtube.com/watch?v=wfIfEWMJY3s)
https://www.youtube.com/watch?v=wfIfEWMJY3s

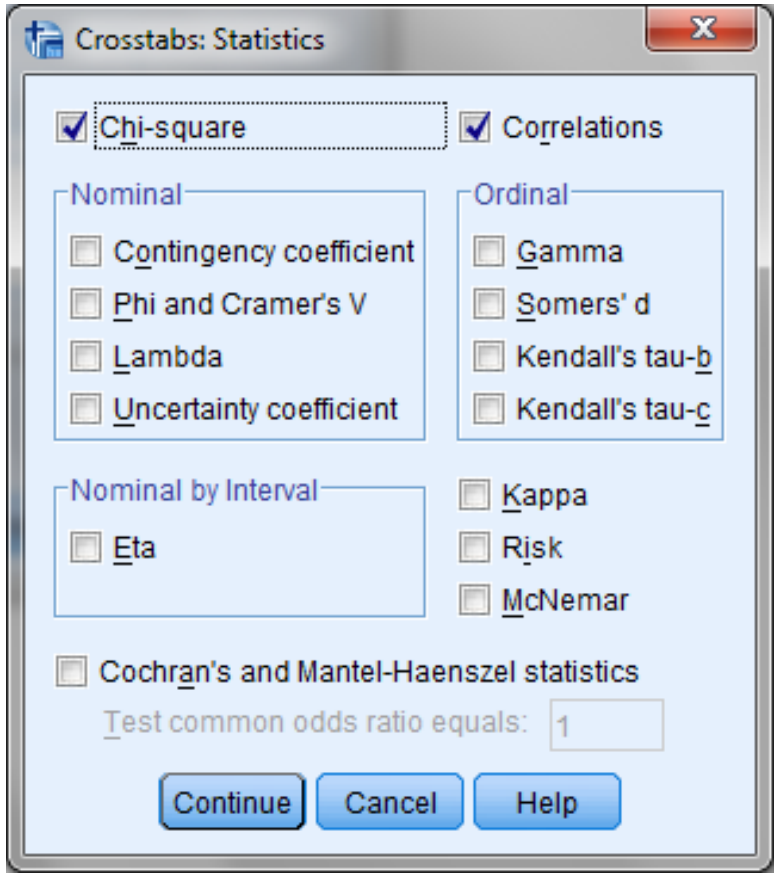
# Chi Square in SPSS

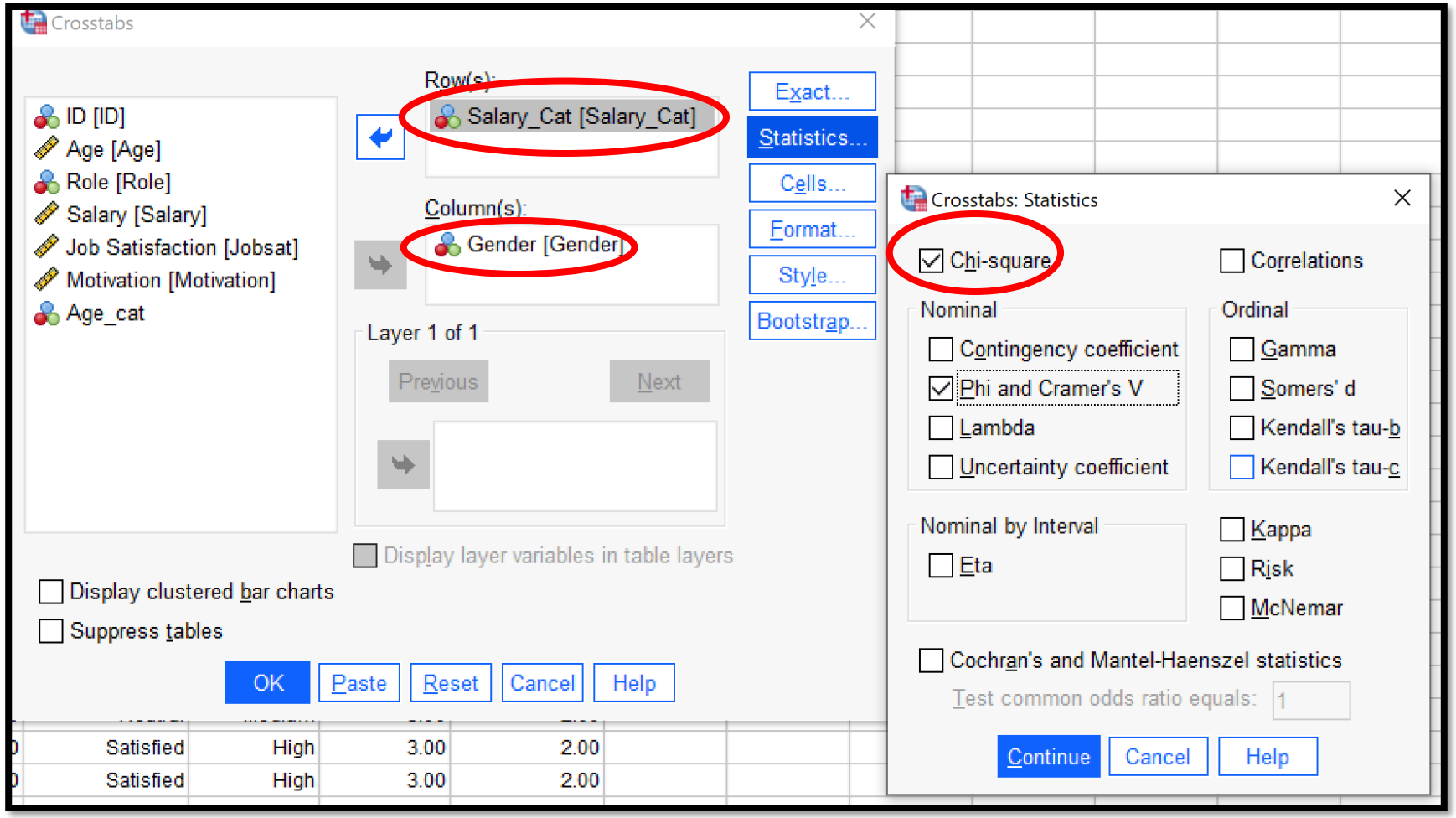


Q: Are there differences in salary between men and women?

1.Click on Analysis, Crosstabs.

4.Click on Statistics and select Chi-Squares




Crosstabs
ID [ID]
Age [Age]
Role [Role]
Salary [Salary]
Job Satisfaction [Jobsat]
Motivation [Motivation]
Age_cat
Row(s):
Salary_Cat [Salary_Cat]
Column(s):
Gender [Gender]
Layer 1 of 1
Previous
Next
Display layer variables in table layers
Display clustered bar charts
Suppress tables
OK
Paste
Reset
Cancel
Help
Exact...
Statistics...
Cells...
Format...
Style...
Bootstrap...
Crosstabs: Statistics
Chi-square
Correlations
Nominal
Contingency coefficient
Phi and Cramer's V
Lambda
Uncertainty coefficient
Ordinal
Gamma
Somers' d
Kendall's tau-b
Kendall's tau-c
Nominal by Interval
Eta
Kappa
Risk
McNemar
Cochran's and Mantel-Haenszel statistics
Test common odds ratio equals: 1
Continue
Cancel
Help
Satisfied High 3.00 2.00
Satisfied High 3.00 2.00

## Gender * Workmode Crosstabulation

| | | | Workmode | | |
|---|---|---|---|---|---|
| | | | Hybrid work mode | Work in premise | Total |
| Gender | Male | Count | 20 | 19 | 39 |
| | | Expected Count | 24.2 | 14.8 | 39.0 |
| | Female | Count | 16 | 3 | 19 |
| | | Expected Count | 11.8 | 7.2 | 19.0 |
| Total | | Count | 36 | 22 | 58 |
| | | Expected Count | 36.0 | 22.0 | 58.0 |

**Case Processing Summary**

| | Cases | | | | | |
|---|---|---|---|---|---|---|
| | Valid | | Missing | | Total | |
| | N | Percent | N | Percent | N | Percent |
| Gender * Workmode | 58 | 100.0% | 0 | 0.0% | 58 | 100.0% |

Answer to question: There are no significant differences in work mode between man and women.

Your output should look like this.

**Chi-Square Tests**

| | Value | df | Asymptotic Significance (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 5.884[a] | 1 | .015 | | |
| Continuity Correction[b] | 4.568 | 1 | .033 | | |
| Likelihood Ratio | 6.378 | 1 | .012 | | |
| Fisher's Exact Test | | | | .021 | .014 |
| Linear-by-Linear Association | 5.782 | 1 | .016 | | |
| N of Valid Cases | 58 | | | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.21.

b. Computed only for a 2x2 table

*Chi- Square* **Report a Chi-Square Test Result (APA)**

$X^2$ (degress of freedom, $N$ = sample size) = chi-square statistic value, $p$ = $p$ value.

# Tips and Hints

- If possible, support your cross-tabulations with chi-square tests which will show you if any differences are statistically significant.
- Further statistical analysis usually depends on establishing if your data is normal. You can test for this in the Descriptives menu by establishing Skewness* (should be between -0.8 and 0.8) and Kurtosis (should be between -3.0 and 3.0) for interval vari ables, or by running the Kolmogorov-Smirnov test in the Non-parametric test section.
- If your data is not normally distributed, do not use standard statically tests but use non-parametric tests.

# Exercise on Work mode and gender

Using Chi-square

- *Set Hypothesis*
- *Run SPSS*
- *Result interpretation*

# Result :

**Chi-Square Tests**

| | Value | df | Asymptotic Significance (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 5.884[a] | 1 | .015 | | |
| Continuity Correction[b] | 4.568 | 1 | .033 | | |
| Likelihood Ratio | 6.378 | 1 | .012 | | |
| Fisher's Exact Test | | | | .021 | .014 |
| Linear-by-Linear Association | 5.782 | 1 | .016 | | |
| N of Valid Cases | 58 | | | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.21.

b. Computed only for a 2x2 table

# p-value interpretation

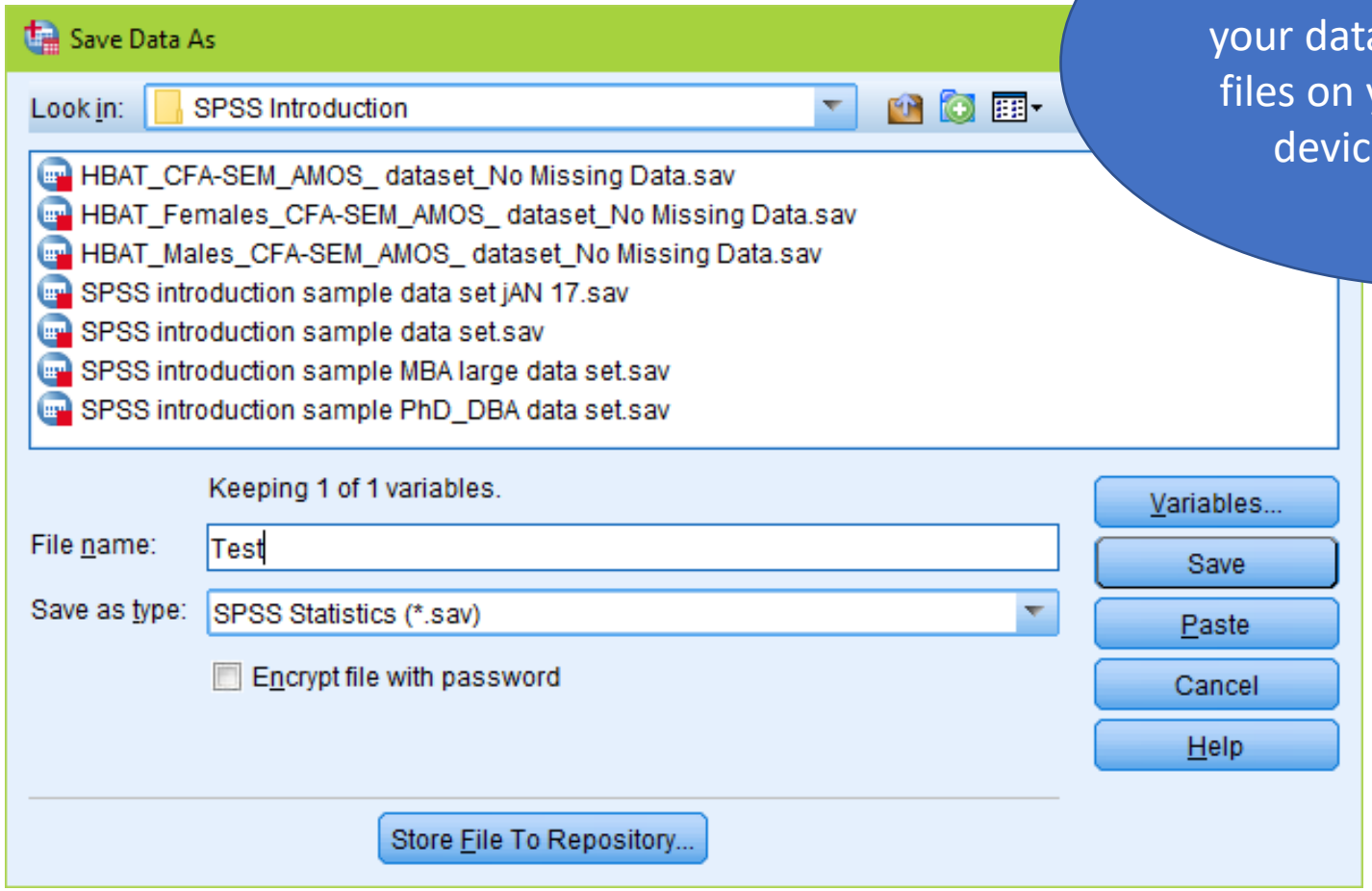
The interpretation of a p-value is a little tricky

- We can usually say that a value is statistically significant if it is below 0.05 and it is highly significant if it is below 0.001
- A .05 *p* value means that we would expect to get the same results 95 times out of a hundred.
- However, sometimes you do not want to find a statistically significant difference! Link back to your hypothesis – sometimes you are looking for a non significant result, e.g. for example on gender pay gap

# Other common tests

- T-tests (2 groups and scaled variables) e.g. is there a significant difference between men and women's salaries?
- ANOVAs (Several groups and scaled variables) e.g. is there a significant difference in salary by job role?
- Correlations e.g. are salary and age correlated?
- Regression (One scaled dependent variable and several scaled independent variables) e.g. is salary predicted by age?

# SAVING DATA




Make sure you save both your data and your output files on your own storage device or computer

# Useful Information

- You can download a 14 day trial of SPSS from the IBM SPSS website. Once the trial expires, you need a license or need to buy the software.
- Contact the IT helpdesk for the university license if you need to, it changes regularly so they need to give you the current one.

- The best book on SPSS is without doubt Andy Field's (2018) Discovering Statistics using IBM SPSS Statistics, 5th edition, Sage (or previous editions). Very widely used and supported by a good website.

| Obs | Gender | Age | Attitude_1 | Attitude_2 | Stay nights |
|---|---|---|---|---|---|
| 1 | Male | 21 | 3 | 2 | 5 |
| 2 | Female | 30 | 2 | 3 | 10 |
| 3 | Female | 35 | 4 | 2 | 3 |
| 4 | Male | 18 | 4 | 1 | n.a. |
| 5 | Female | 22 | 4 | 5 | 5 |
| 6 | Female | 22 | 3 | 5 | n.a. |
| 7 | Male | 26 | 1 | 2 | 2 |

Notes: Attitude_1: The hotels in Hong Kong provide high quality services.
Attitude_2: Taxi drivers are not very friendly to the passengers.
Attitude levels: 5: strongly agree
4: agree
3: neutral
2: disagree
1: strongly disagree

# Data measurement levels

- Nominal: is for mutual exclusive but not ordered categories.
- Ordinal: is a measurement where the order matters but not the difference between values.
- Interval: the difference between two values is meaningful.
- Ratio: has all the properties of an interval variable, and also has a clear definition of 0. The ratio of two measurements is meaningful.

| Scale Type | Permissible Statistics | Admissible Caculations | Example |
|---|---|---|---|
| nominal (or, categorical) | mode, Chi-square | $=, \neq$ | Gender (Female, Male) |
| ordinal | median, percentile | $=, \neq, >, <$ | 1. Grade (A+, A , B+, B,C+,C)<br>2. "bad", "medium", and "good" |
| interval | mean, standard deviation, correlation, regression, analysis of variance | $=, \neq, <, >,$<br>$+, -$ | Temperature (-1, 0, 1) |
| ratio | All statistics permitted for interval scales plus the following: geometric mean, harmonic mean, coefficient of variation, logarithms | $=, \neq, <, >,$<br>$+, -, \times, \div$ | GDP (e.g. 1500 USD) |

Scale

# 6. Correlation analysis

- The dependent variable must be a scale variable, while the grouping variables may be ordinal or nominal.
- With the Explore procedure, you can:
  - Screen data
  - Identify outliers
  - Check assumptions
  - Characterize differences among groups of cases

# 6. Correlation analysis

- Correlation analysis is used to describe the strength and the direction of the linear relationship between two variables.

# 6. Correlation analysis

Correlation is a measure of relationship between two variables. It has wide application in business and statistics. Now what would you do when you have two variables that are needed to be studied together. How would you use SPSS to identify relationship, association, link or bonding of two variables at one time. The analysis of relationship between two variables is termed as correlation analysis. Correlation analysis is used to describe the strength and direction of the linear relationship between two variables.

Following are few scenarios in which we would go for Correlation Analysis

- A university would like to know if social responsibility is associated with university reputation in the community
- A marketing manager would like to know if a hike in prices has anything to do with decrease in product sales
- A HR manager would like to know if an increase in employee pay would result in lower absenteeism
- A social science researcher would like to know if increase in age results in decrease of conflicts at work/home

There are a number of different statistics available from SPSS, depending on the level of measurement and the nature of your data.

- Pearson product-moment correlation coefficient (r)
- Spearman correlation (rho) procedures.

# 6.Correlation coefficient

- Measure how variables are related, for example what is the relationship between *x* and *y*
- Correlation coefficient values range from -1 to +1, indicating both strength and direction:
  - Positive values indicate a positive relationship, both values increase (or decrease) together
  - Negative values indicate a negative relationship, as one variable increases the other decreases
  - The closer to 0 the weaker the relationship, the closer to – 1 or + 1 the stronger the relationship

# *Correlation coefficient*

- r ranges from +1 to -1
- r = +1 a perfect positive linear relationship
- r = -1 a perfect negative linear relationship
- r = 0 indicates no correlation at all

# Interpretation of the Correlation Coefficient

- Merely computation of correlation does not have any significance, it is important that we know what is meant by the coefficient, what it tells about the data. To have an answer, generally, the coefficient of correlation is interpreted in verbal description. Correlation Coefficient can easily be interpreted using the following table.

| Size of Correlation | Interpretation |
|---|---|
| 1 | Perfect Positive/Negative Correlation |
| ± .90 to ± .99 | Very High Positive/Negative Correlation |
| ± .70 to ± .90 | High Positive/Negative Correlation |
| ± .50 to ± .70 | Moderate Positive/Negative Correlation |
| ± .30 to ± .50 | Low Positive/Negative Correlation |
| ± .10 to ± .30 | Very low Positive/Negative Correlation |
| ± .0 to ± .10 | Markedly Low and Negligible Positive/Negative Correlation |

# 6. Correlation analysis

## Pearson Correlation

Pearson r is designed for interval/ratio level (continuous) variables. It can also be used if you have one continuous variable (Employee Commitment) and one dichotomous variable (e.g. Type of Organization: Public/Private). The correlation is expressed in the form of a coefficient. Symbolized as *r*, a correlation coefficient is normally reported as a decimal number somewhere between +1.00 and -1.00.

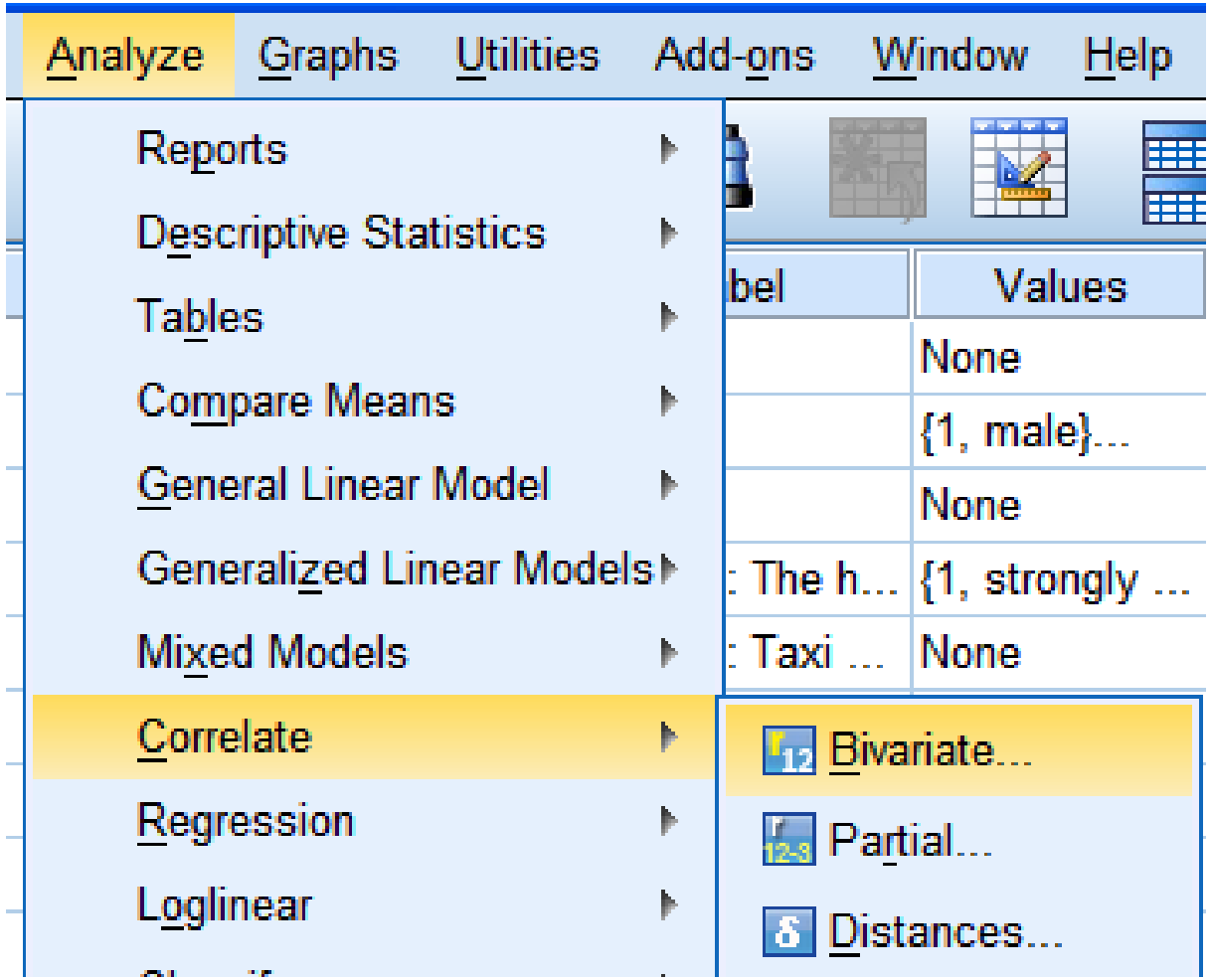
# Exercise - Relationship between Job Satisfaction and Salary

# Exercise - Relationship between Job Satisfaction and Salary

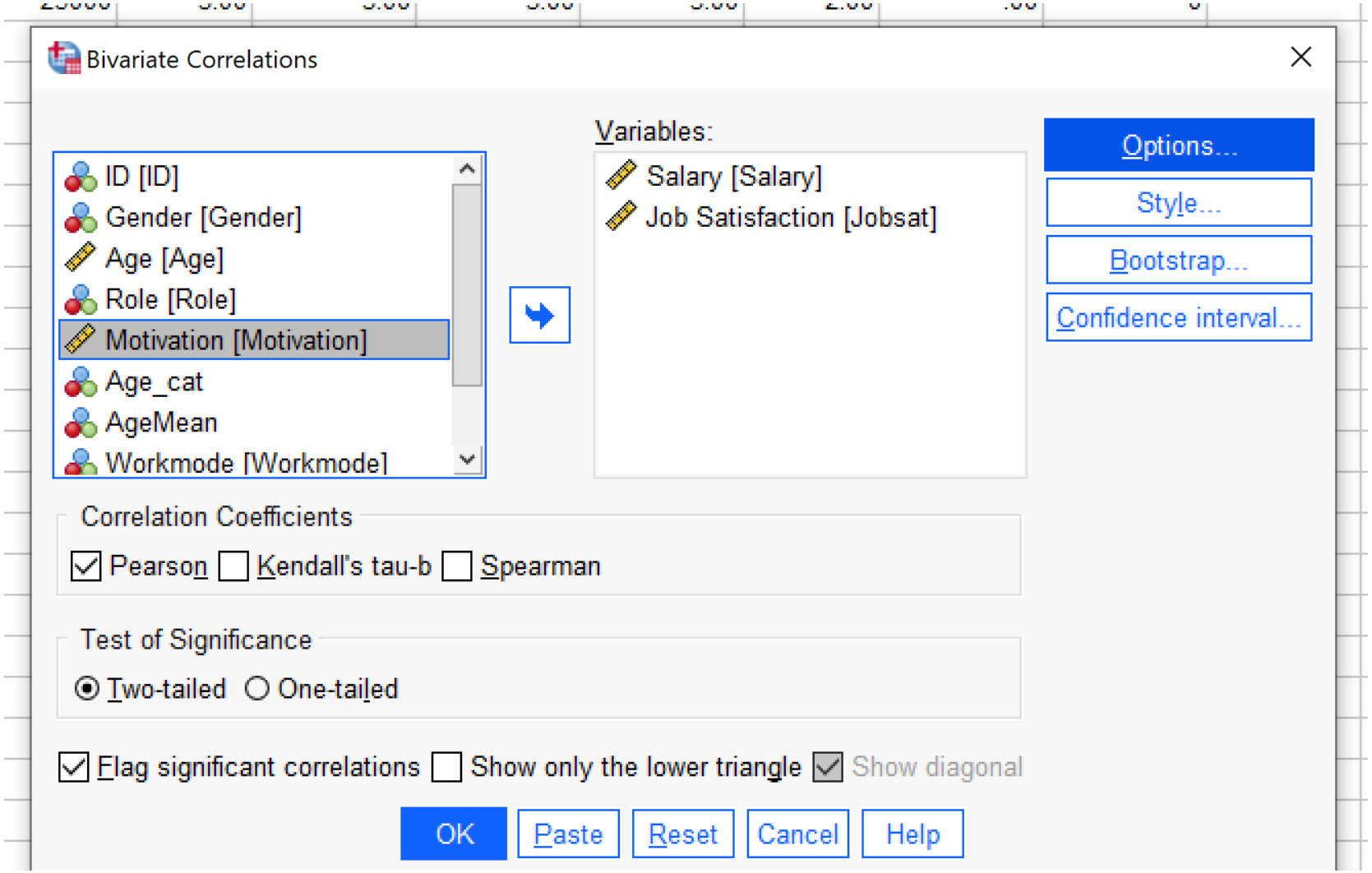
H0 : There is no significant relationship between salary job satisfaction

H1: There is significant relationship between salary and job satisfactionand

# 6. Correlation analysis



# 6. Pearson’s correlation



# Job Satisfaction and Salary

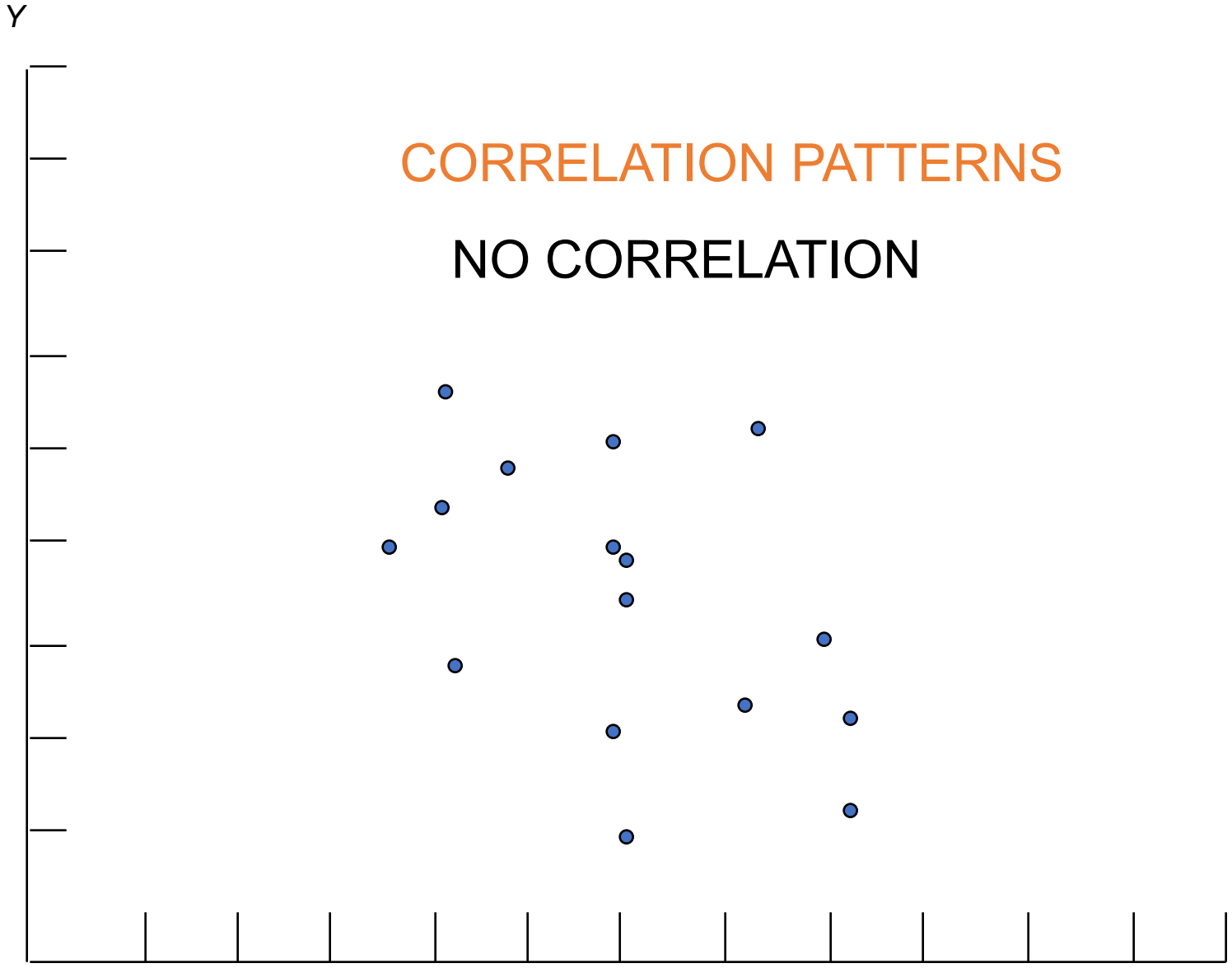
**Correlations**

| | | Salary | Job Satisfaction |
|---|---|---|---|
| Salary | Pearson Correlation | 1 | .643** |
| | Sig. (2-tailed) | | <.001 |
| | N | 58 | 58 |
| Job Satisfaction | Pearson Correlation | .643** | 1 |
| | Sig. (2-tailed) | <.001 | |
| | N | 58 | 58 |

**. Correlation is significant at the 0.01 level (2-tailed).

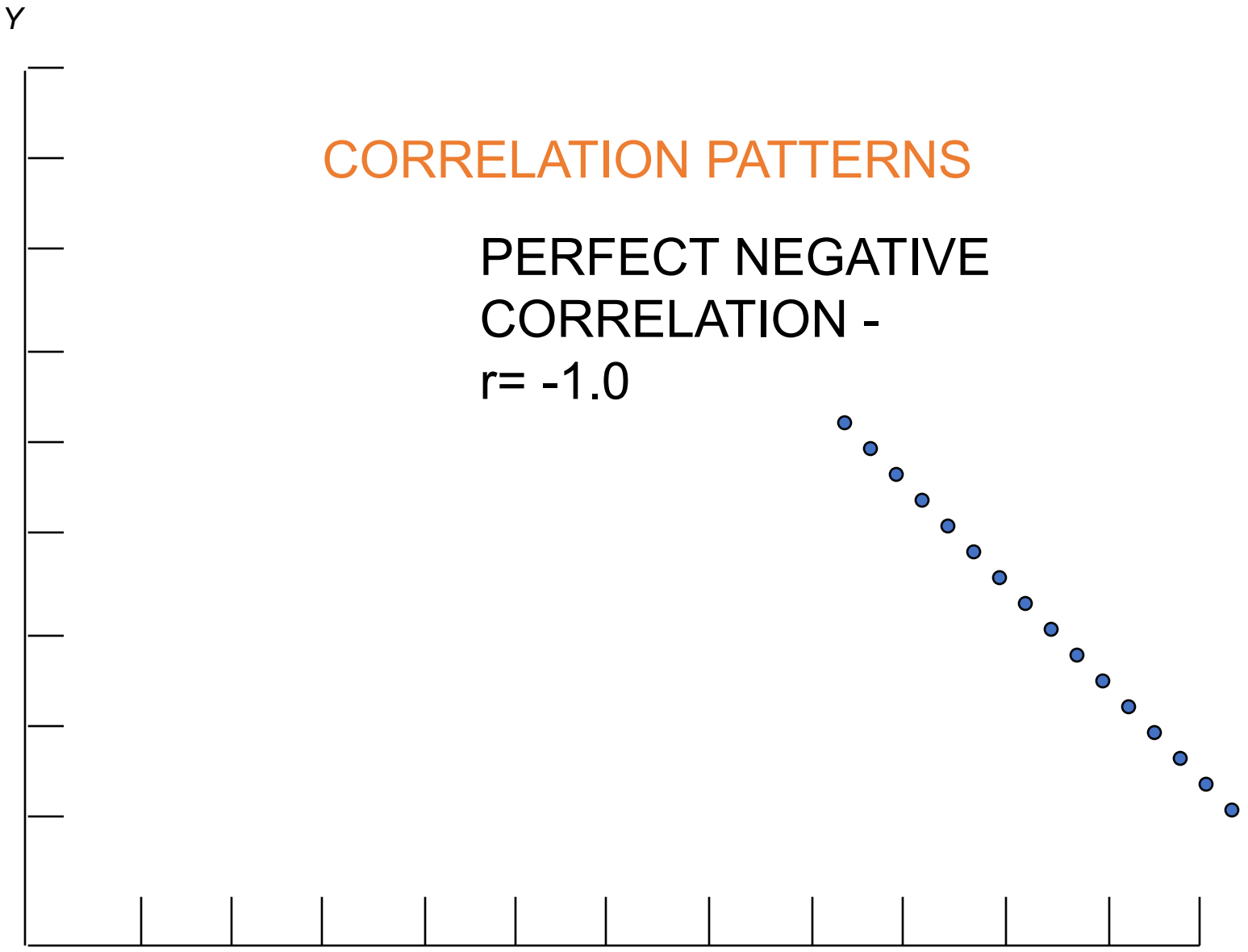
# CORRELATION PATTERNS

## NO CORRELATION



# CORRELATION PATTERNS

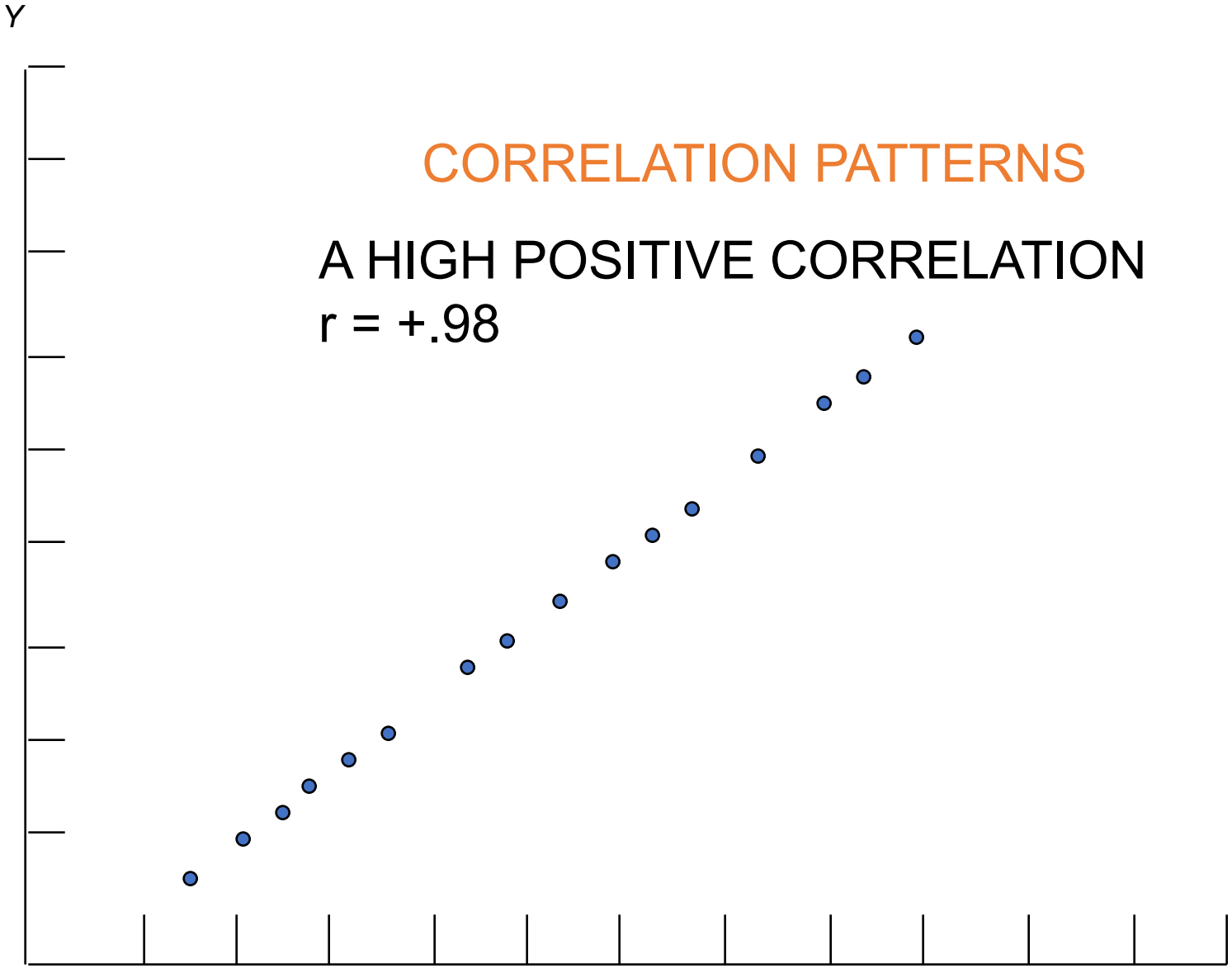
PERFECT NEGATIVE CORRELATION -
r= -1.0



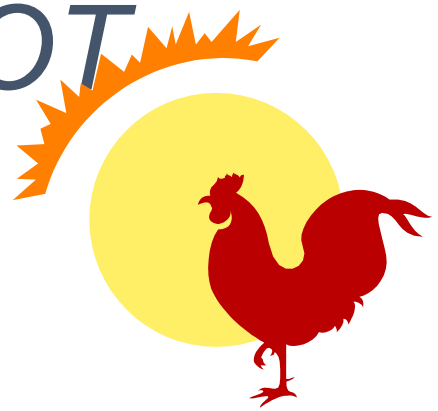
# CORRELATION PATTERNS

A HIGH POSITIVE CORRELATION

r = +.98



# *CORRELATION DOES NOT MEAN CAUSATION*



- High correlation
- Rooster’s crow and the rising of the sun
- Rooster does not cause the sun to rise
- Teachers’ salaries and the consumption of liquor
- Co-vary because they are both influenced by a third variable

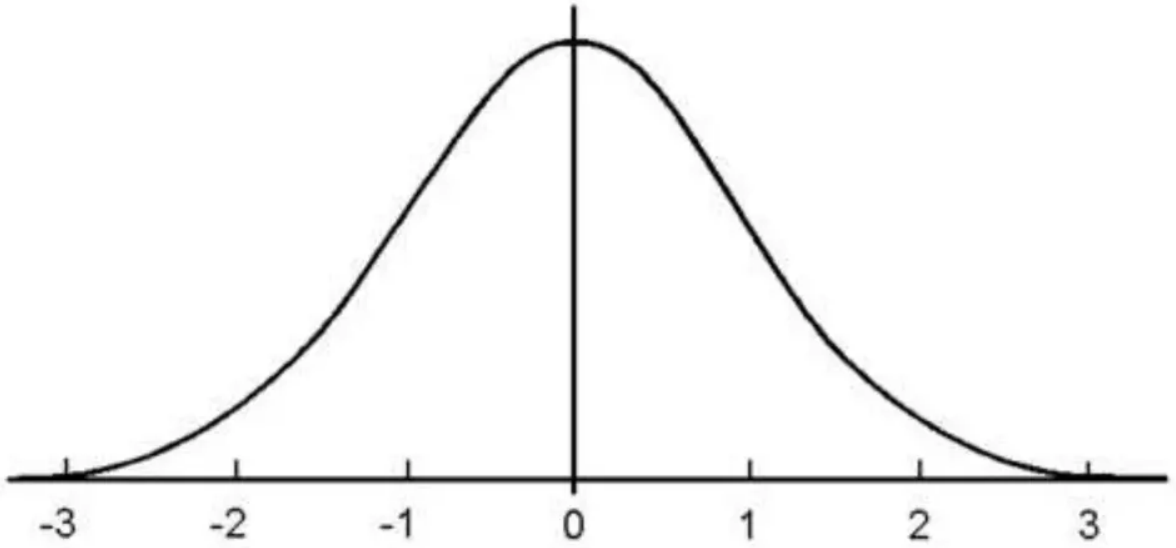
Pearson correlation of job satisfaction and salary was found to be moderately positive and statistically significant ($r=.643$, $p<.001$). Hence, H1 was supported. This indicates that an increase in salary would facilitate higher level of job satisfaction

# 7. Normality Test

# Normal Distribution Curve

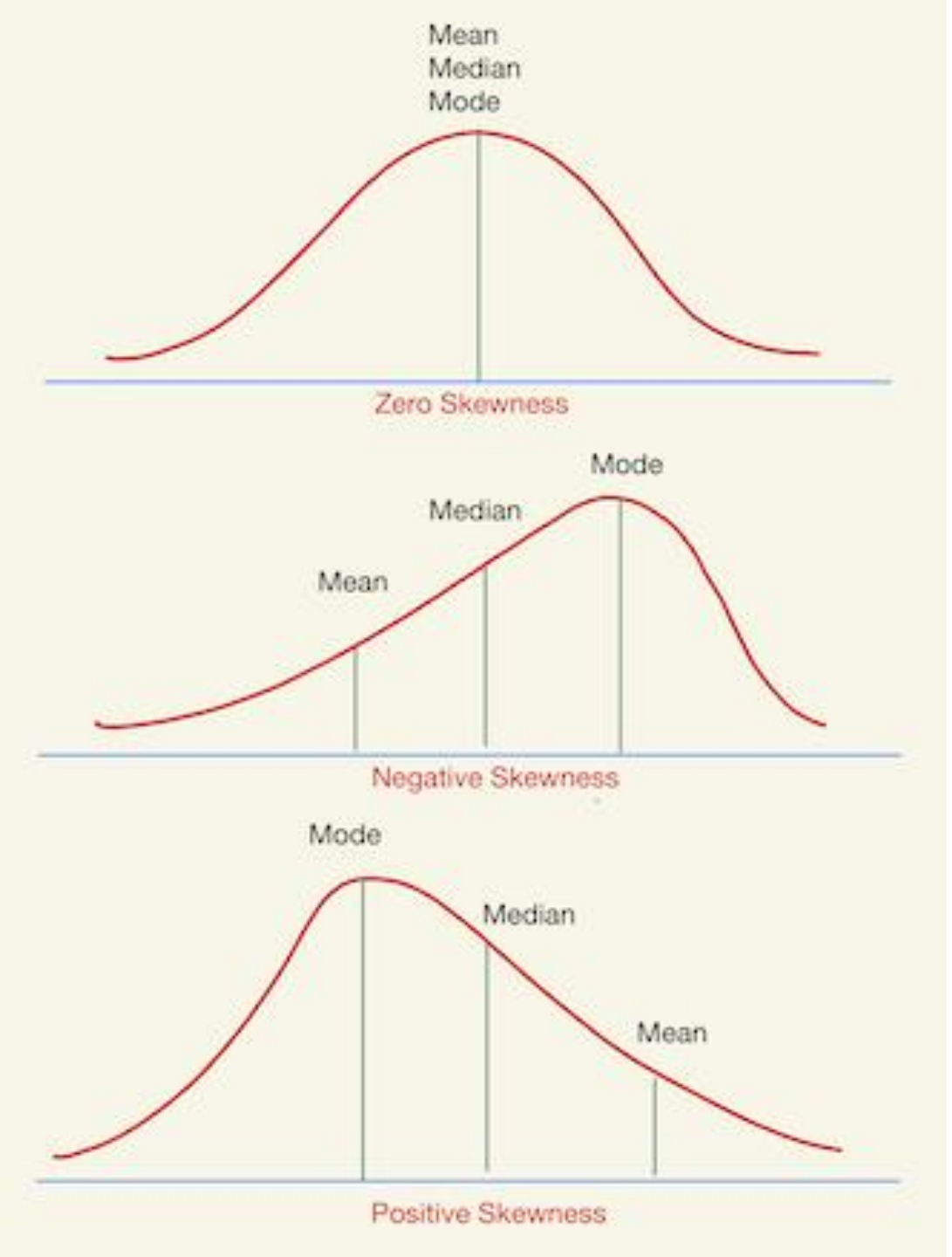
The term “Normal Distribution Curve” is used to describe the mathematical concept called normal distribution.

It refers to the shape that is created when a line is plotted using the data points for an item that meets the criteria of ‘Normal Distribution’



# Normality Test

- **Skewness** is a measure of the asymmetry of a distribution. A distribution is asymmetrical when its left and right side are not mirror images.



# Normality Test

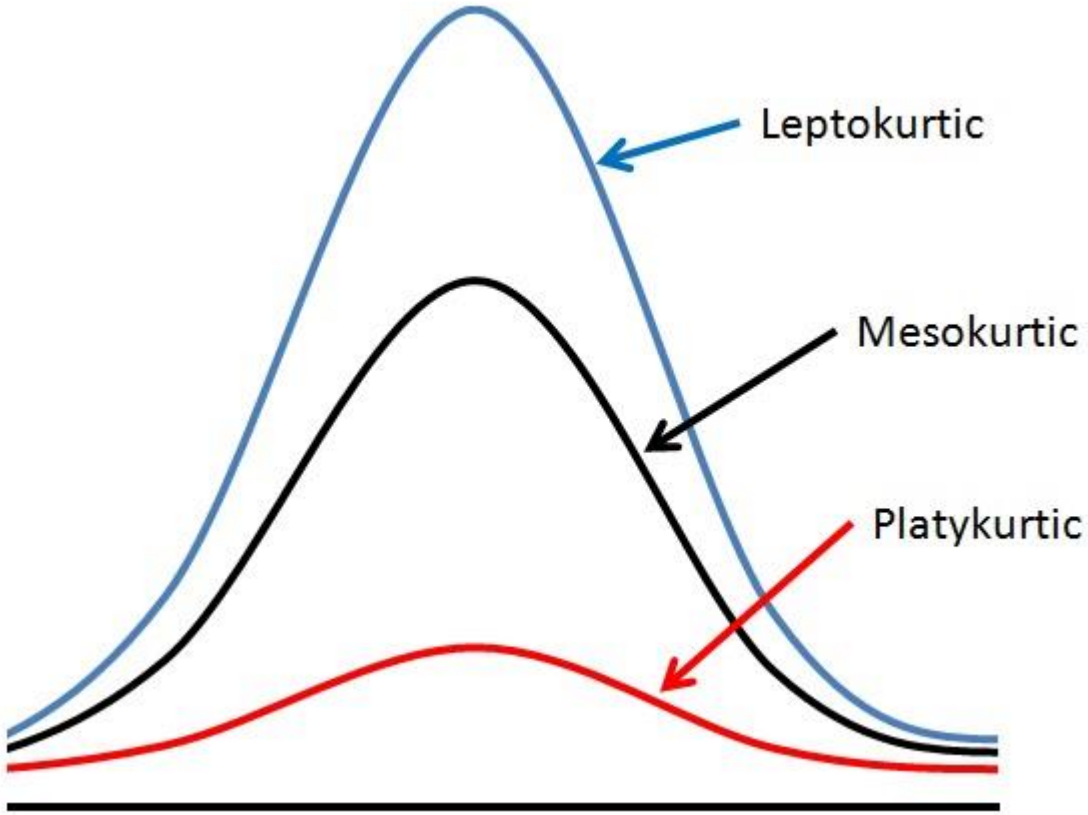
**Skewness**: Skewness measures the symmetry of a variable's distribution. If the distribution stretches toward the right or left tail, it's skewed. Negative skewness indicates more larger values, while positive skewness indicates more smaller values.
A skewness value between -1 and +1 is excellent, while -2 to +2 is generally acceptable. Values beyond -2 and +2 suggest substantial nonnormality (Hair et al., 2022, p. 66).

•Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). **A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)** (3 ed.). Thousand Oaks, CA: Sage.

# Normality Test

**Kurtosis measures**The spread (tails) of the distribution

In the rare scenario where both skewness and kurtosis are zero, the pattern of responses is considered a normal distribution.

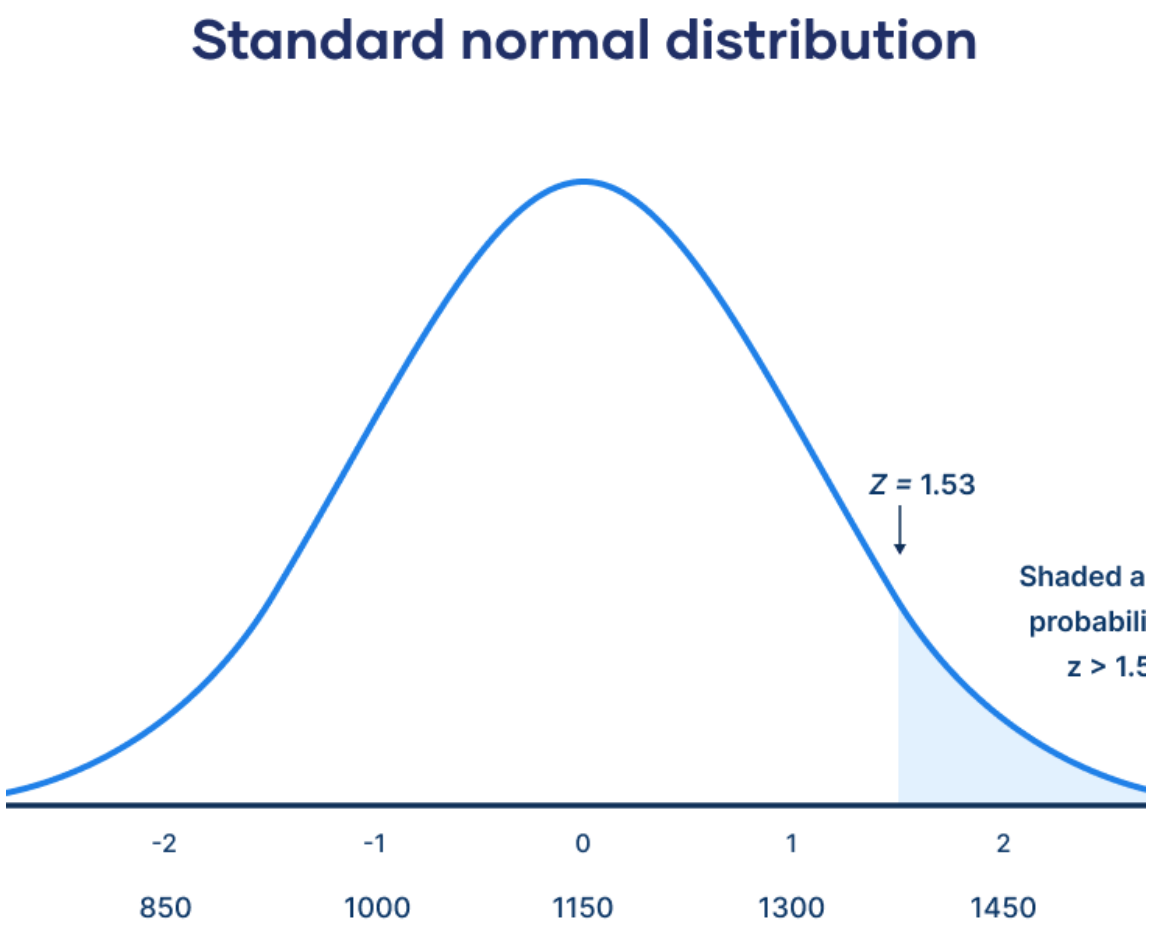



A kurtosis greater than +2 suggests a too peaked distribution, while less than -2 indicates a too flat one. When skewness and kurtosis are close to zero, it's considered a normal distribution (Hair et al., 2022, p. 66).

# 7. Normality test

Normality test in SPSS

Analyze - Descriptive Statistics -Explore...



# 7. Normality Test

If normality requirements are influenced by

i) Sample size and

ii) Outliers

# 7. Normality Test – Implication

- If the normality assumption is violated, non-parametric test should be used.

| arametric test | Non-Parametric equivalent |
|---|---|
| Paired t-test | Wilson Signed Rank test |
| Unpaired t-test | Mann-Whitney U test |
| Pearson correlation | Spearman correlation |
| way Analysis of variance | Kruskal Wallis Test |

Q & A

# Wrap Up

# LCDB 7011 Quantitative Research Methods : One-to-One Consultation (18/03) at IQ102

Please attend the One-to-one consultation session according to your selected timeslot

| Time | Name of student (S/I) |
|---|---|
| 09:00 – 09:25 | Natalie Brailsford 2536525 |
| 09:30 - 09:55 | Hina Tokeer (2419580) |
| 10:00 – 10:25 | Marvin Malthus Gomes (2507129) |
| 10:30 - 10:55 | Guangzheng Wu (2501013) |
| 11:00 - 11:25 | Pricilla Moses (1818519) |
| 11:30 - 11:55 | Nusrat jahan jeba(2508904) |
| 12:00 - 12:25 | Panktibahen chaudhari(2504333) |
| 12:30 - 12:55 | Ummay Huney (2513079) |
| 13:00 - 13:25 | Sumaiya Minnat (2506343) |
| 13:30 – 13:55 | Sadia islam (2511736) |

# References


- Byrne, D., (2002) Interpreting quantitative data, London: Sage.
- Argyrous., G., (2011) Statistics for research: with a guide to SPSS 3rd ed., Los Angeles: Sage.
- Field, A., (2018) Discovering statistics using IBM SPSS statistics 5th ed., London: Sage
- Field, A., (2013) Discovering Statistics Using SPSS for Windows: Advanced Techniques for Beginners (Introducing Statistical Methods series), Sage Publications Ltd
- Berenson, M., Levine, D., Szabat, K., Stephan, D., (2019) Basic business statistics: concepts and applications. 14th Ed. Global Edition, Upper Saddle River: Pearson
- Burns., R., Burns., R., (2008) Business research methods and statistics using SPSS, London, Sage.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). **A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)** (3 ed.). Thousand Oaks, CA: Sage