

College of Basic Education
Department of English
PhD Applied Linguistics



زانكۆی سه لاهه دین - هه ولییر
Salahaddin University-Erbil

Advanced Statistics Using SPSS

Dr. Rizgar Maghded Ahmed

Asst. Prof. in Applied Statistics

E-mail: rizgar.ahmed@su.edu.krd

2023-2024

Definitions of Statistics:

- It is a science which helps us to **collect, analyze** and **present data** systematically.
- It is the process of **collecting, processing, summarizing, presenting, analysing** and **interpreting of data in order to study** and **describe a given problem.**
- Statistics is the **art of learning from data.**
- Statistics may be regard as (i)the study of populations, (ii) the study of variation, and (iii) the study of methods of the reduction of data.

Basic Definitions → *Statistics steps in Research:*

- 1. Problem or Phenomenon**
- 2. Collect data**
- 3. Organize data**
- 4. Analyze data**
- 5. Result**
- 6. Interpret → Information - Data**
- 7. Draw conclusions(Make Decision)**

Importance of Statistics:

- It simplifies of data.
- Helps to get **concrete** information about any problem;
- Helps for **reliable** and **objective decision** making;
- It presents **facts** in a **precise & definite** form;
- It facilitates comparison(**Measures of central tendency and measures of dispersion**);
- It facilitates Predictions (**Regression analysis is the most commonly used methods towards prediction.**)
- It helps in formulation of suitable policies;

Application areas of statistics

Some of the diverse fields in which Statistical methodology has extensive applications are:

➤ **Engineering:**

Improving product design, testing product performance, determining reliability and maintainability, working out safer systems of flight control for airports, etc.

➤ **Business:**

Estimating the volume of retail sales, designing optimum inventory control system, producing auditing and accounting procedures, improving working conditions in industrial plants, assessing the market for new products.

➤ Quality Control:

Determining techniques for evaluation of quality through adequate sampling, in process control, consumer survey and experimental design in product development etc.

* Realizing its importance, large organizations are maintaining their own Statistical Quality Control Department *.

➤ Economics:

Measuring indicators such as volume of trade, size of labor force, and standard of living, analyzing consumer behavior, computation of national income accounts, formulation of economic laws, etc.

* Particularly, Regression analysis extensively uses in the field of Economics*.

➤ **Health and Medicine:**

Developing and testing new drugs, delivering improved medical care, preventing diagnosing, and treating disease, etc. Specifically, inferential Statistics has a tremendous application in the fields of health and medicine.

➤ **Biology:**

Exploring the interactions of species with their environment, creating theoretical models of the nervous system, studying genetically evolution, etc.

➤ **Psychology:**

Measuring learning ability, intelligence, and personality characteristics, creating psychological scales and abnormal behavior, etc.

➤ **Sociology:**

Testing theories about social systems, designing and conducting sample surveys to study social attitudes, exploring cross-cultural differences, studying the growth of human population, etc.

Type of Statistics

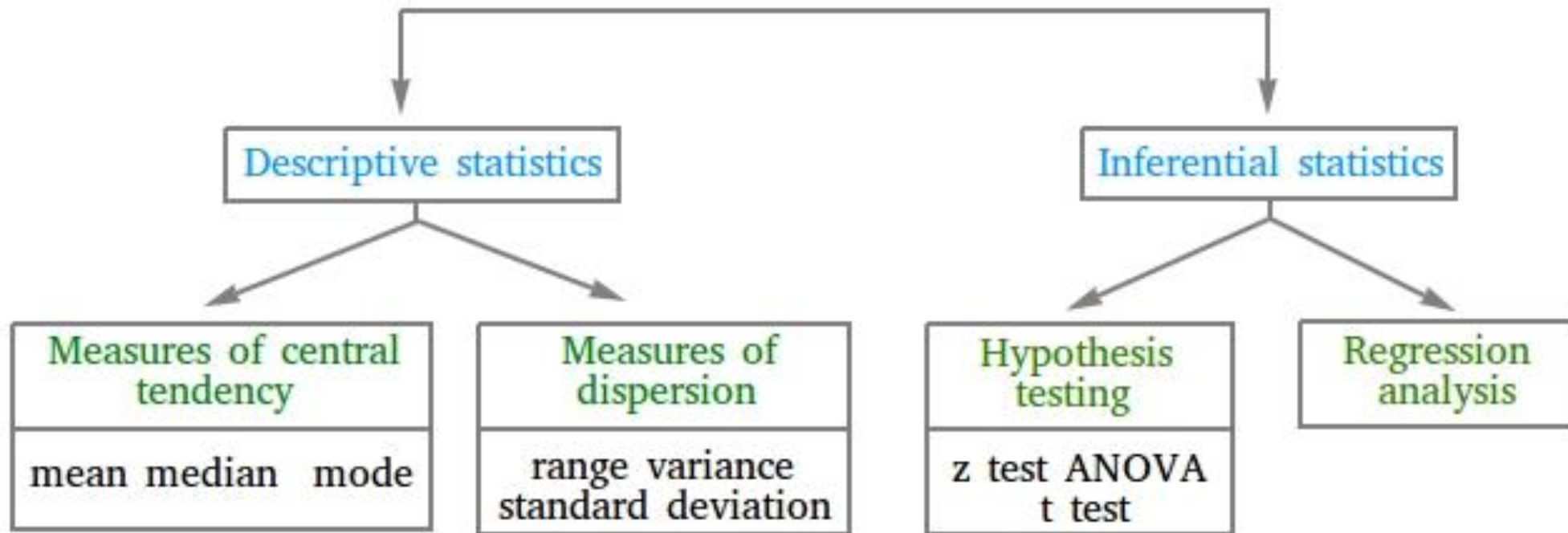
There are two main branches of statistics:

1. Descriptive statistics
2. Inferential statistics

1.Descriptive Statistics - Through graphs or tables, or numerical calculations, descriptive statistics uses the data to provide descriptions of the population.

2.Inferential Statistics - Based on the data sample Generalized from the population, inferential statistics makes the predictions and inferences.

Types of statistics



Main terms in Statistics

- **Data:** Is a set of numbers, Those values represent by the thing called a variable.
- **Variable:** A variable is a factor or characteristic that can take on different possible values or outcomes. Example: Income, height, weight, sex, age, etc of a certain group of individuals are examples of variables. A variable can be qualitative or quantitative (numeric).

Type of Variables:

i. **Quantitative data:** data of a certain group of individuals which is expressed numerically.

Example: Heights, Weights, Ages and, etc of a certain group of individuals.

ii. **Qualitative data:** data of a certain group of individuals that is not expressed numerically as it is.

Example: Colors, Languages, Nationalities, Religions, health, poverty etc of a certain group of individuals

*Employee data.sav [DataSet1] - IBM SPSS Statistics Data Editor

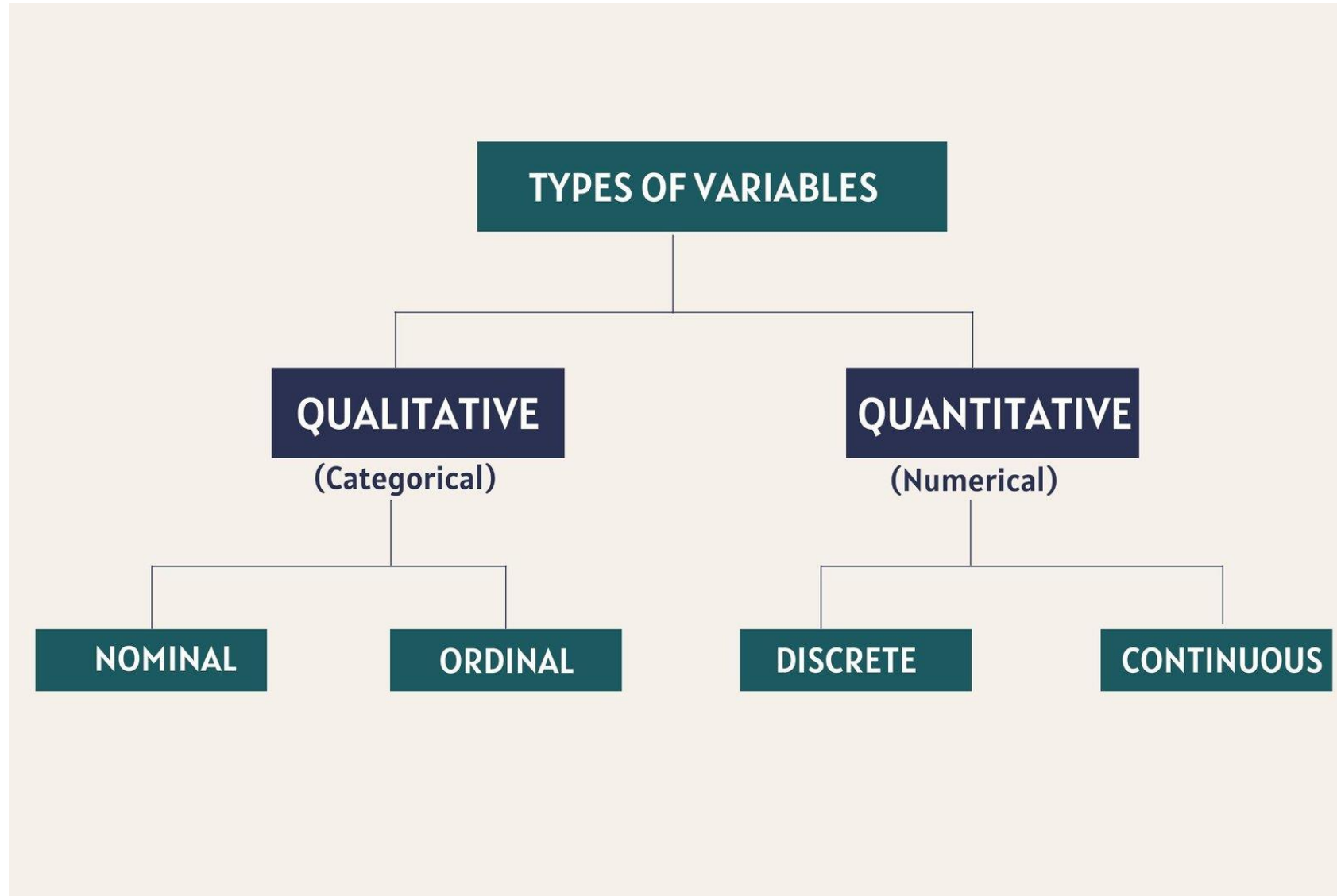
File Edit View Data Transform Analyze Graphs Custom Utilities Add-ons Window Help

Visible: 10 of 10 Variables

	id	gender	bdate	educ	jobcat	salary	salbegin	jobtime	p
1	1	Male	02/03/1952	15	Manager	\$57,000	\$27,000	98	
2	2	Male	05/23/1958	16	Clerical	\$40,200	\$18,750	98	
3	3	Female	07/26/1929	12	Clerical	\$21,450	\$12,000	98	
4	4	Female	04/15/1947	8	Clerical	\$21,900	\$13,200	98	
5	5	Male	02/09/1955	15	Clerical	\$45,000	\$21,000	98	
6	6	Male	08/22/1958	15	Clerical	\$32,100	\$13,500	98	
7	7	Male	04/26/1956	15	Clerical	\$36,000	\$18,750	98	
8	8	Female	05/06/1966	12	Clerical	\$21,900	\$9,750	98	
9	9	Female	01/23/1946	15	Clerical	\$27,900	\$12,750	98	
10	10	Female	02/13/1946	12	Clerical	\$24,000	\$13,500	98	
11	11	Female	02/07/1950	16	Clerical	\$30,300	\$16,500	98	
12	12	Male	01/11/1966	8	Clerical	\$28,350	\$12,000	98	
13	13	Male	07/17/1960	15	Clerical	\$27,750	\$14,250	98	
14	14	Female	02/26/1949	15	Clerical	\$35,100	\$16,800	98	

Data View Variable View

IBM SPSS Statistics Processor is ready Unicode:ON



Types of Qualitative Variables

1- Nominal Variable:

There is no order among characteristics of variable.

Example:

- 1- Gender (male , female).
- 2- Blood group (A , B , AB , O)

2- Ordinal Variable:

There is order among characteristics of variable.

Example:

- 1- Response to treatment (poor, fair, good)
- 2- Likert Scale(Strongly disagree- Disagree-Neutral-Agree-Strongly agree)

Types of Quantitative Variables

1 - Discrete Variables:

Usually obtained by counting.

Example:

- 1- The number of decayed teeth per child in an elementary school.
- 2- The number of bacteria on a plate.
- 3- The number of heart beats within a specified time interval.

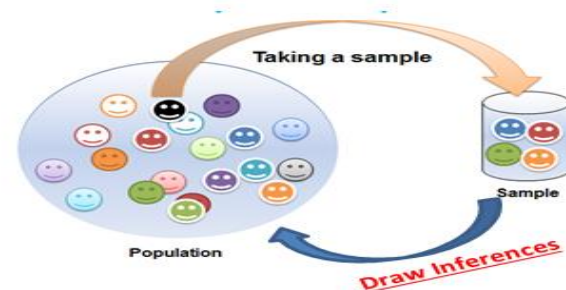
2 - Continuous Variables:

Usually obtained by measurement.

Example:

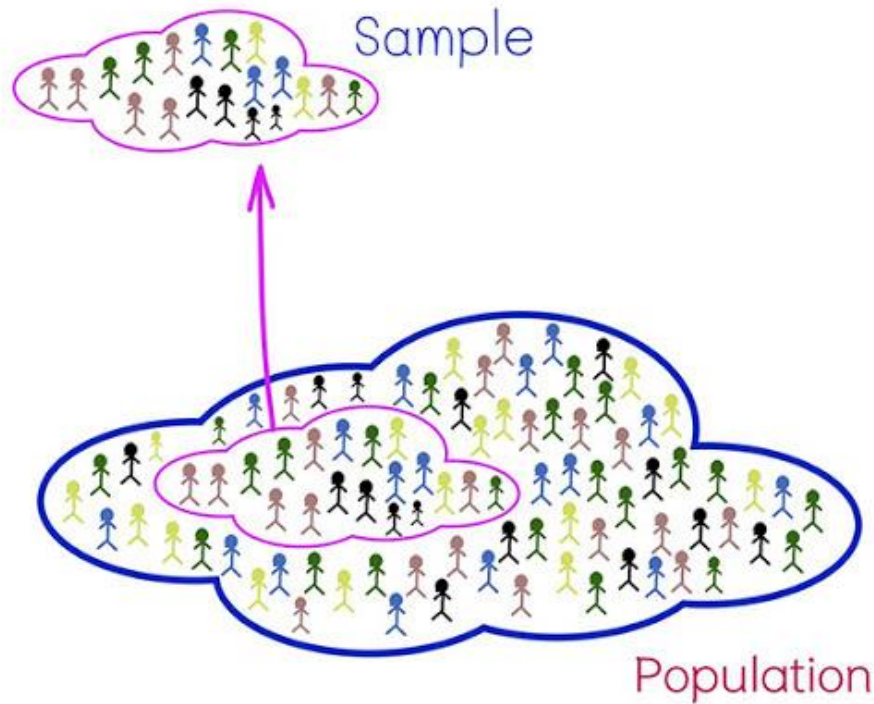
- 1- blood pressure
- 2- Age (54 , 65.6 , 43 ,).
- 3- Height (154 , 165.6 , 143 ,).

- **Population:** A complete set of observation (data) of the entire group of individuals under consideration e.g. The number of students in this class, the population in Addis Ababa etc. A population can be finite or infinite.
- **Sample:** A set of data drawn from population containing a part which can reasonably serve as a basis for valid generalization about the population. Or a sample is a portion of a population selected for further analysis.
- **Sample size:** The number of items under investigation.



- **Survey (experiment):** it is a device of obtaining the desired data. Two types of survey:
 1. **Census Survey:** A way of obtaining data referring the entire population which is said to provide a total coverage of the population.
 2. **Sample Survey:** A way of obtaining data referring a portion of the entire population which is said to provide only a partial coverage of the population.

Population and Sample

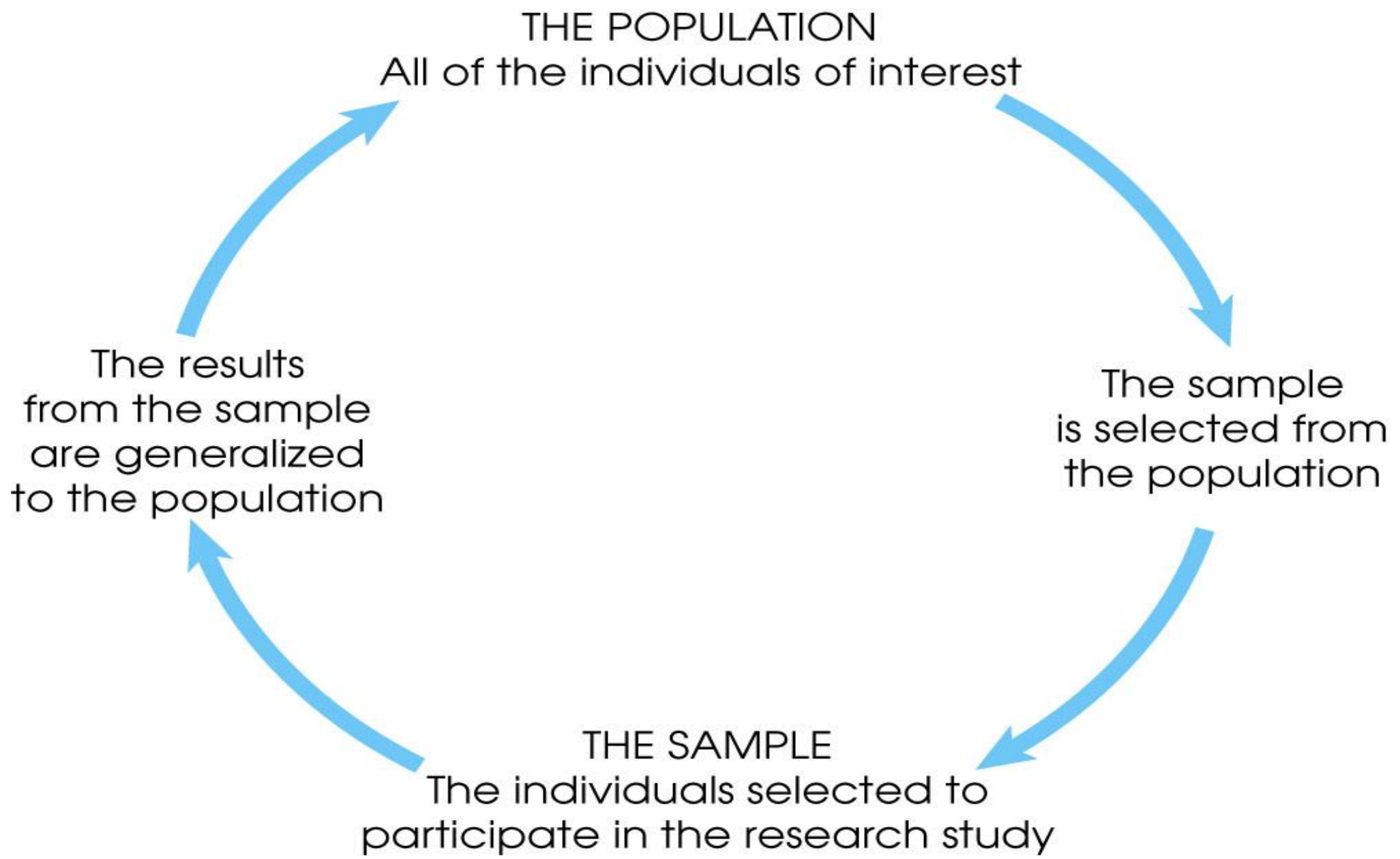


POPULATION

- The measurable quality is called a parameter.
- The population is a complete set.
- Reports are a true representation of opinion.
- It contains all members of a specified group.

SAMPLE

- The measurable quality is called a statistic.
- The sample is a subset of the population.
- Reports have a margin of error and confidence interval.
- It is a subset that represents the entire population.



Dr.Rizgar Maghded Ahmed

Sampling methods

Sampling methods can be:

1- Random Sampling :Is a sampling technique where each sample has an equal probability of getting selected (each member of the population has an equal chance of being selected)

2- Non-Random Sampling

is a sampling technique where the sample selected will be based on factors such as convenience, judgement and experience of the researcher and not on probability.

A. Random sampling methods

1. **Simple Random Sample** (each sample of the same size has an equal chance of being selected)
2. **Stratified Sample** (divide the population into groups called strata and then take a sample from each stratum)
3. **Systematic Sample** (randomly select a starting point and take every n-th piece of data from a listing of the population)
4. **Cluster Sample** (divide the population into strata and then randomly select some of the strata. All the members from these strata are in the cluster sample.)

B. Nonprobability Sampling

Every element in the population **does not have an equal probability of being chosen**. The process of inclusion in the sample is based on the judgment of the person selecting the sample.

Types of Nonprobability Sampling

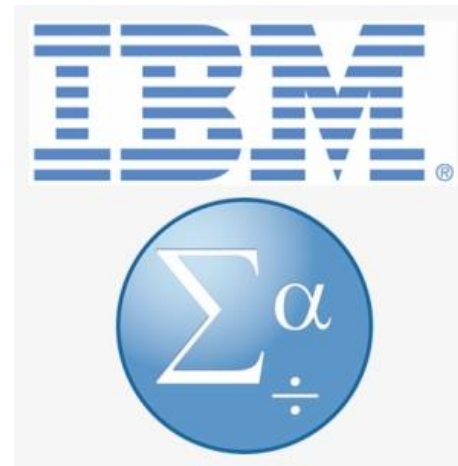
1. Judgment Sampling
2. Convenience Sampling
3. Quota Sampling
4. Snowball Sampling

SPSS means “**Statistical **Package for the **Social **Sciences**” and was first launched in 1968. Since SPSS was acquired by IBM in 2009, it's officially known as IBM SPSS Statistics but most users still just refer to it as “SPSS”.******

SPSS -Quick Overview Main Features

SPSS is software for editing and analyzing all sorts of data. These data may come from basically any source: scientific research, a customer database, Google Analytics or even the server log files of a website. SPSS can open all file formats that are commonly used for structured data such as:

- spreadsheets from MS Excel;
- plain text files (.txt or .csv);
- relational (SQL) databases;
- Stata and SAS.



SPSS Windows

SPSS has three windows for working with data:

The Data Editor Window (.sav)

shows data in two forms:

Data view

Variable view

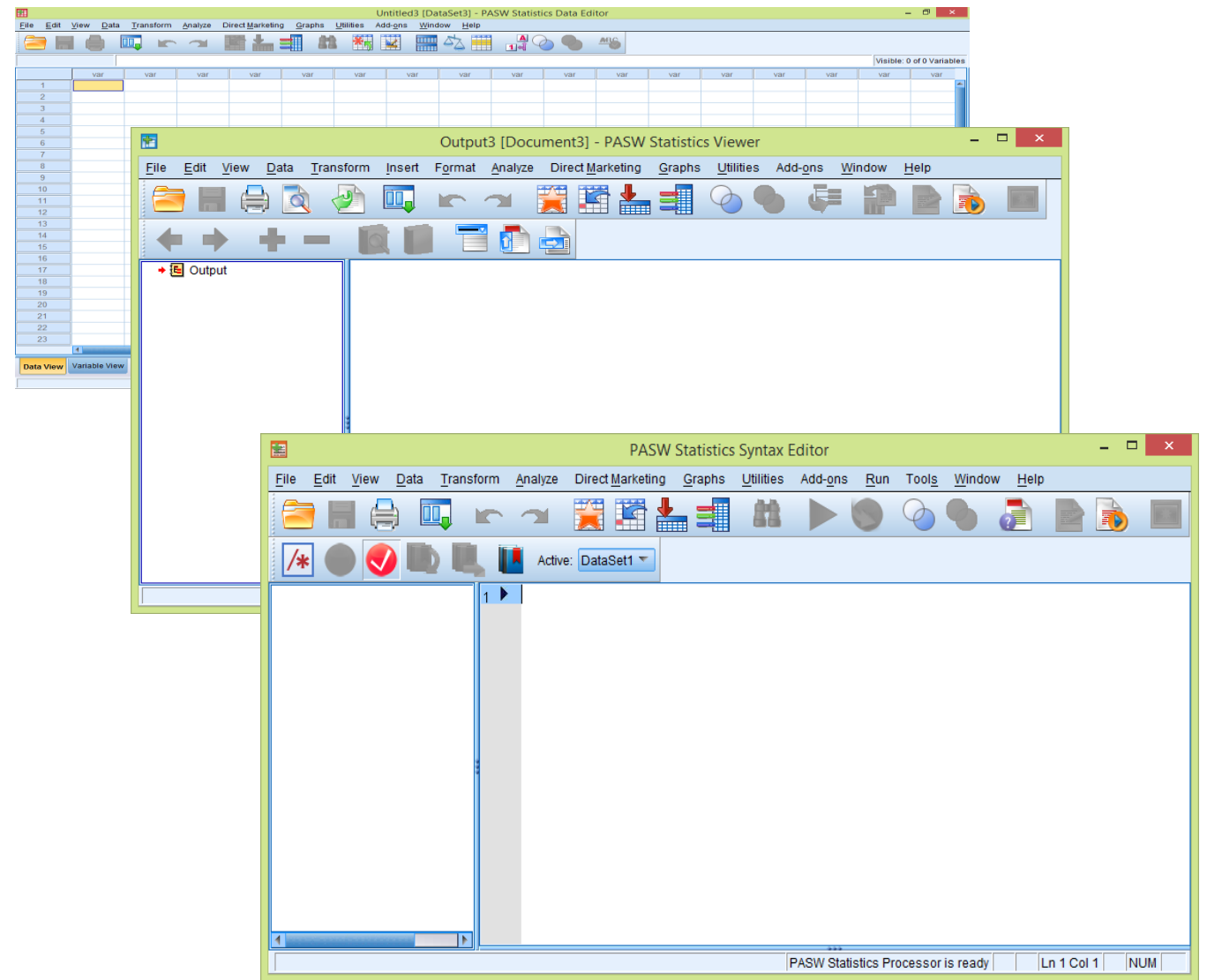
The Output Viewer Window (.spv)

shows results of data analysis

The Syntax Editor Window (.sps)

shows the syntax command script

This also where you can type and run your own syntax commands.



The Data Editor Window (*sav)

Variable view

The place to enter variables

Rows define the variable characteristics:

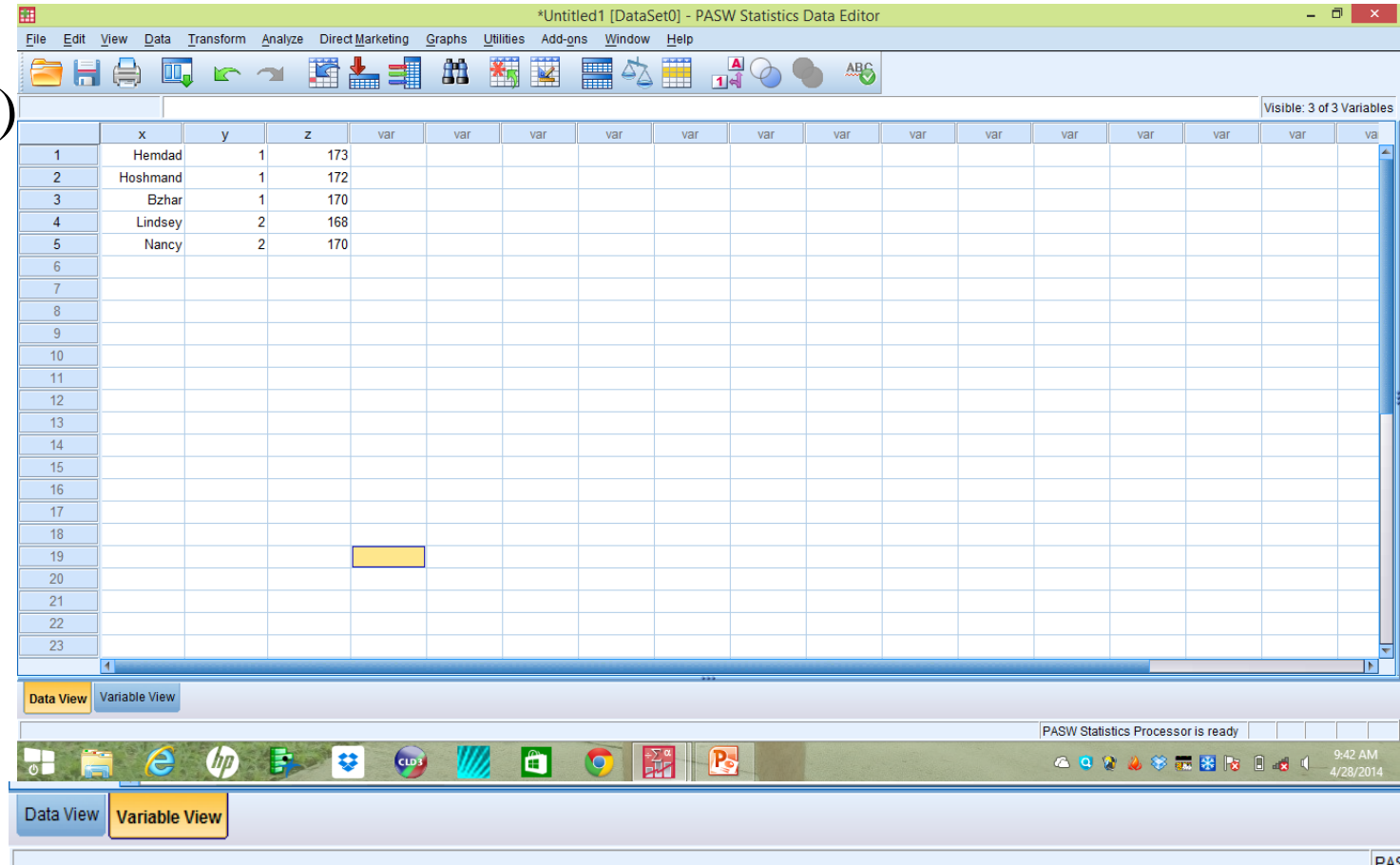
Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, Role

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	x	String	8	0	Name	None	None	8	Right	Nominal	Input
2	y	Numeric	8	0	Gender	{1, Male}...	None	8	Right	Scale	Input
3	z	Numeric	8	0	Height	None	None	8	Right	Scale	Input
4											
5											
6											

The Data Editor Window (*sav)

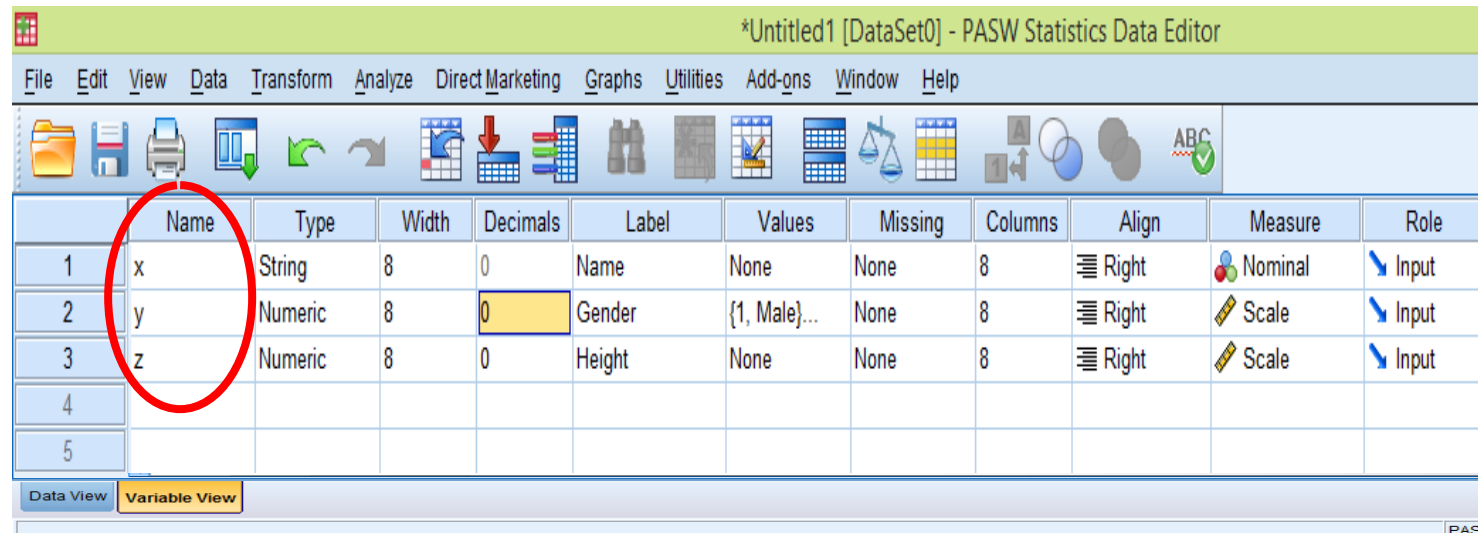
Data view

The place to enter data
Rows are cases (records)
Columns are variables



Variable View window: Name

- The first character of the variable name must be alphabetic
- Variable names must be unique, and have to be less than 64 characters.
- Spaces and special characters (eg !, ?, ' , and *) cannot be used.
- variable names cannot end with a period
- Reserved keywords cannot be used as variable names – these are ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO and WITH; (lower and upper case)



The screenshot shows the SPSS Variable View window for a dataset named '*Untitled1 [DataSet0]'. The window title bar includes the menu options: File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. Below the menu is a toolbar with various icons for file operations, data manipulation, and analysis. The main area is a table with the following columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. The table contains three rows of variable definitions:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	x	String	8	0	Name	None	None	8	Right	Nominal	Input
2	y	Numeric	8	0	Gender	{1, Male}...	None	8	Right	Scale	Input
3	z	Numeric	8	0	Height	None	None	8	Right	Scale	Input
4											
5											

At the bottom of the window, there are two tabs: 'Data View' and 'Variable View', with 'Variable View' currently selected. The SPSS logo is visible in the bottom right corner.

Variable View window: Type

- The **Type** column is showing **Numeric** for all rows. This means that numeric (number) values will be expected in the dataset relating to these variables.
 - Click on the 'type' box. The two basic types of variables that you will use are numeric and string. This column enables you to specify the type of variable.

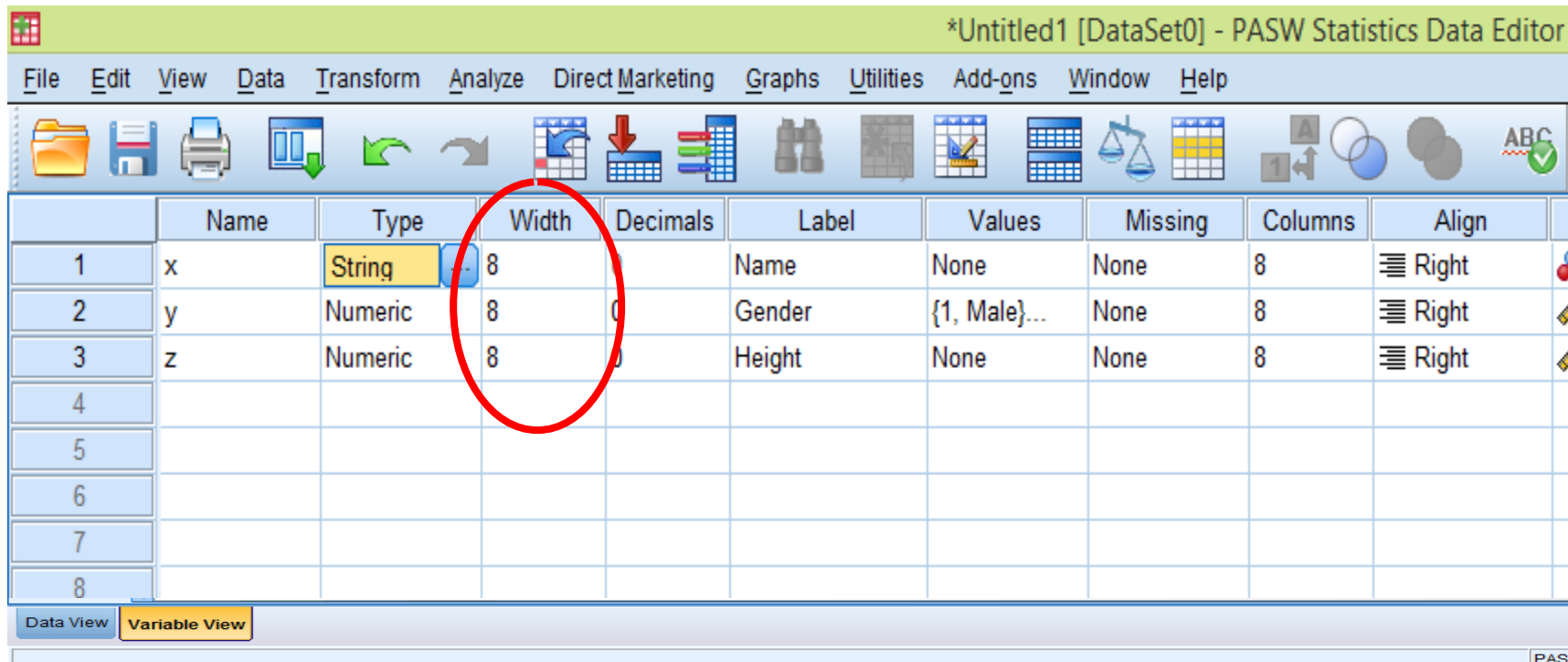
The screenshot shows the SPSS Variable View window for a dataset named '*Untitled1 [DataSet0]'. The window displays a table with columns: Name, Type, Width, Decimals, Label, Values, and Missing. The rows are:

Name	Type	Width	Decimals	Label	Values	Missing
x	String	8	0	Name	None	None
y	Numeric	8	0	Gender	{1, Male}...	None
z	Numeric	8	0	Height	None	None

The 'Type' column for rows 'y' and 'z' is circled in red. A dialog box titled 'Variable Type' is open in the foreground, showing the 'Numeric' radio button selected. The dialog also includes options for 'Comma', 'Dot', 'Scientific notation', 'Date', 'Dollar', 'Custom currency', and 'String'. The 'Width' is set to 8 and 'Decimal Places' is set to 0. The dialog has 'OK', 'Cancel', and 'Help' buttons.

Variable View window: Width

Width allows you to determine the number of characters SPSS will allow to be entered for the variable



*Untitled1 [DataSet0] - PASW Statistics Data Editor

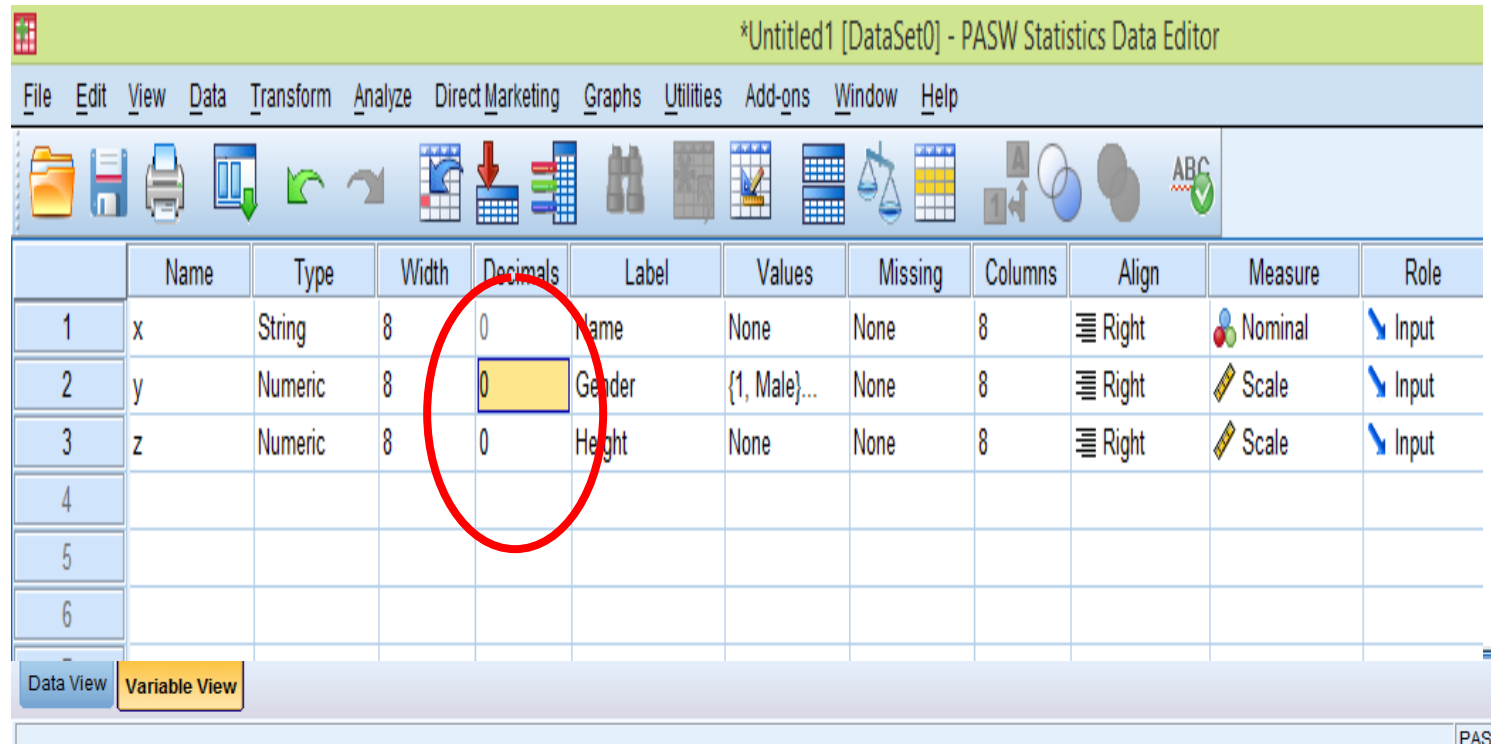
File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
1	x	String	8		Name	None	None	8	Right
2	y	Numeric	8	0	Gender	{1, Male}...	None	8	Right
3	z	Numeric	8	0	Height	None	None	8	Right
4									
5									
6									
7									
8									

Data View Variable View

Variable View window: Decimals

- Number of decimals
- It has to be less than or equal to 16



The screenshot shows the Variable View window in PASW Statistics Data Editor. The window title is '*Untitled1 [DataSet0] - PASW Statistics Data Editor'. The menu bar includes File, Edit, View, Data, Transform, Analyze, Direct Marketing, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations, data manipulation, and analysis. The main area is a table with the following columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. The table contains three rows of variables: 'x' (String, Width 8, Decimals 0, Label 'Name', Measure 'Nominal'), 'y' (Numeric, Width 8, Decimals 0, Label 'Gender', Measure 'Scale'), and 'z' (Numeric, Width 8, Decimals 0, Label 'Height', Measure 'Scale'). The 'Decimals' column for variable 'y' is highlighted with a red circle. At the bottom left, there are buttons for 'Data View' and 'Variable View'. The status bar at the bottom right shows 'PASW'.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	x	String	8	0	Name	None	None	8	Right	Nominal	Input
2	y	Numeric	8	0	Gender	{1, Male}...	None	8	Right	Scale	Input
3	z	Numeric	8	0	Height	None	None	8	Right	Scale	Input
4											
5											
6											

Variable View window: Label

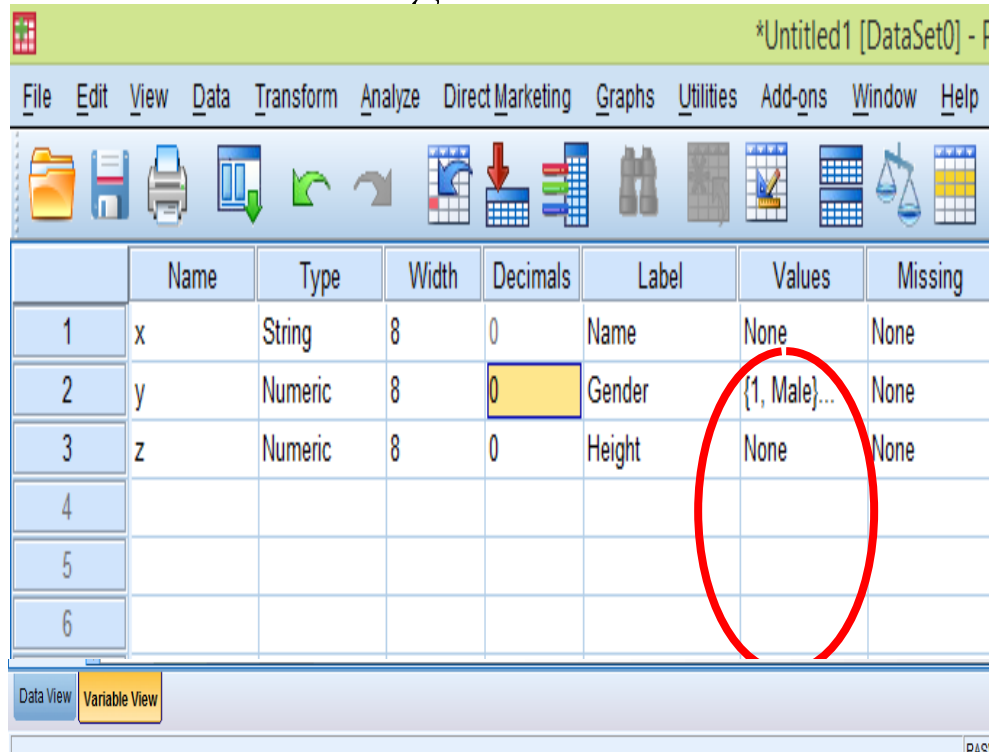
- You can specify the details of the variable
- You can write characters with spaces up to 256 characters
- If you want to specify where a new line appears in a label, type `\n` within the text and SPSS will wrap the label at this point.

The screenshot shows the SPSS Variable View window for a dataset named '*Untitled1 [DataSet0]'. The window displays a table with columns for Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. The 'Label' column is circled in red. The variables are defined as follows:

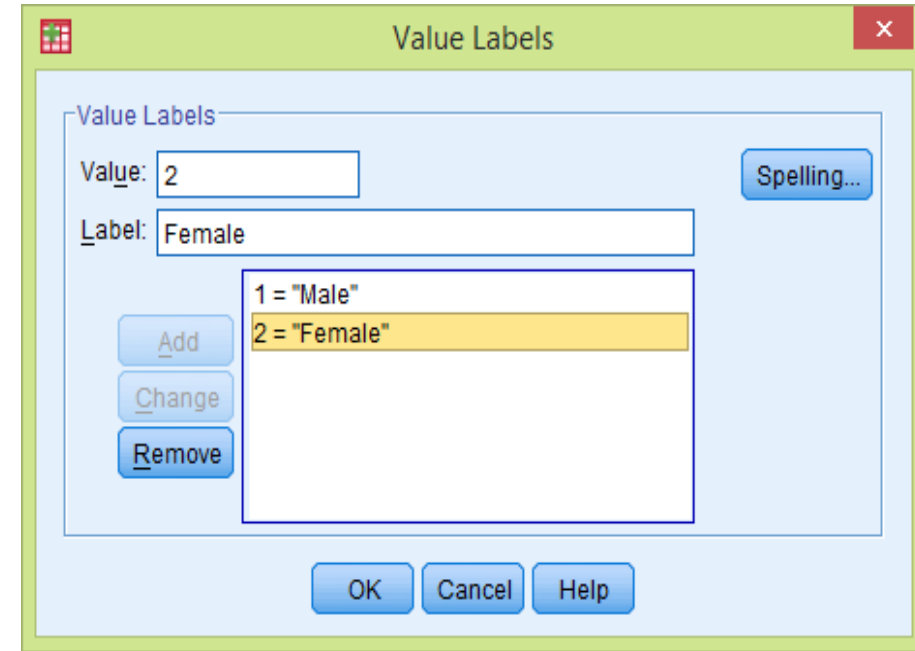
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	x	String	8	0	Name	None	None	8	Right	Nominal	Input
2	y	Numeric	8	0	Gender	{1} Male}...	None	8	Right	Scale	Input
3	z	Numeric	8	0	Height	None	None	8	Right	Scale	Input
4											
5											
6											

Variable View window: Value labels

- This is used and to suggest which numbers represent which categories when the variable represents a category
- For the value, and the label, you can put up to 60 characters.
- In the value's blank put the number (code) and label's blank put the name of code
- After defining the values click add and then click OK.



	Name	Type	Width	Decimals	Label	Values	Missing
1	x	String	8	0	Name	None	None
2	y	Numeric	8	0	Gender	{1, Male}...	None
3	z	Numeric	8	0	Height	None	None
4							
5							
6							



Value Labels

Value: 2

Label: Female

1 = "Male"
2 = "Female"

Add Change Remove Spelling...

OK Cancel Help

Variable View window : Missing Value

- Missing values are used to define user-specified missing information.
 - No response
 - Refused to answer
 - Data entry mistakes

There are two types of missing values in SPSS:

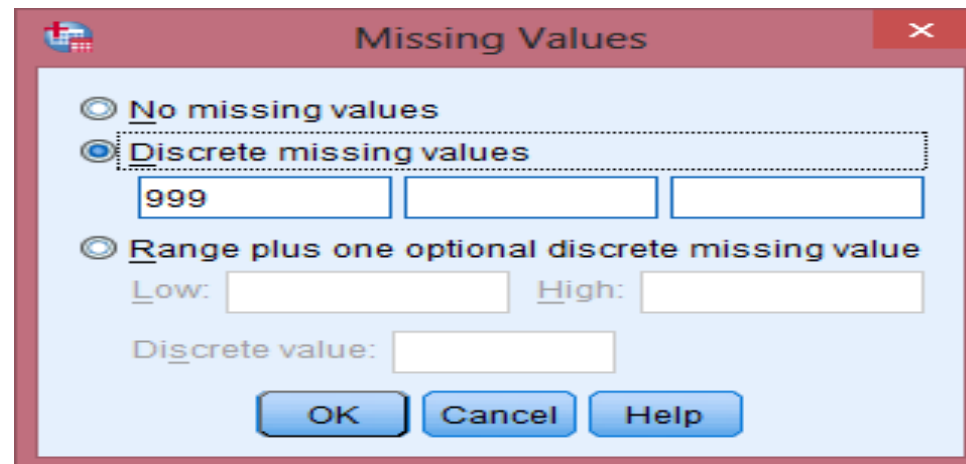
System-missing values: By default, for non-string variables, an empty cell is defined as a system missing value and does not need to be further declared.

User-defined missing data are values that the researcher can tell SPSS to recognize as missing. For example, 99, 999, and 9999 are common user-defined missing value.

How use a 999 code as missing values in your variable which has missing values

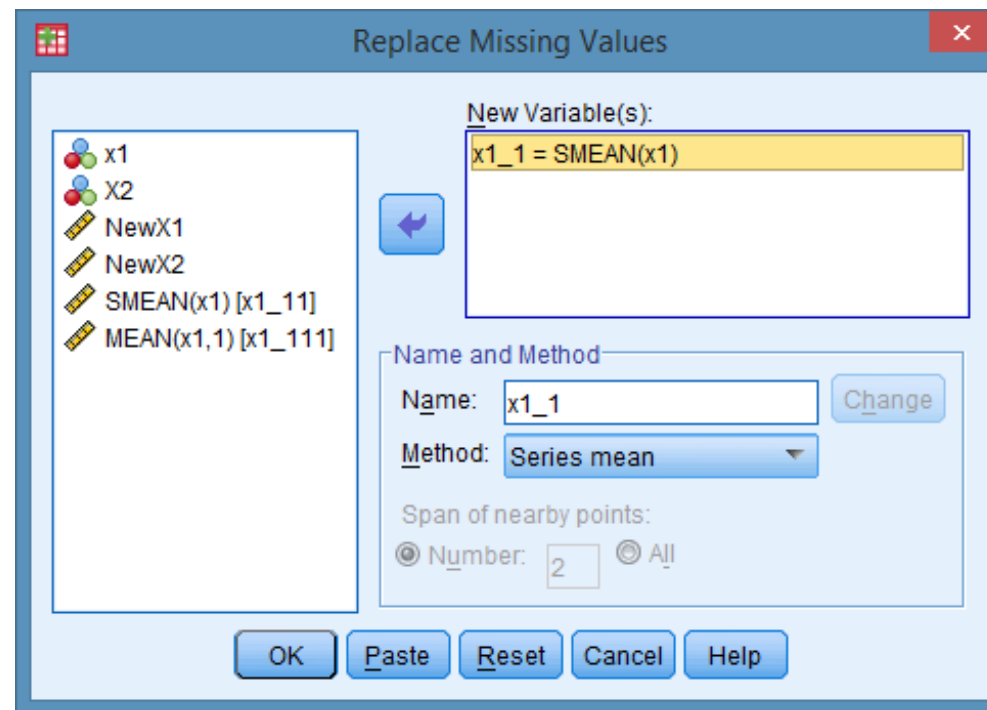
Look at your variables in VARIABLE VIEW that has missing values

- Go to Transform → Recode into same variable → Click on System or user-missing → plug the number 999 in the Value box → Continue & Ok
- Select that variable's missing cell by clicking on the blue box in the right corner.
- Click DISCRETE MISSING VALUES
- Enter 999 to define this variable's missing value



Replacing missing values

- click TRANSFORM
- click REPLACE MISSING VALUES
- select the variable with missing values and move it to the right using the arrow
- SPSS will rename and create a new variable with your filled in data.
- click METHOD to select what type of method you would like SPSS to use when replacing missing values.
- click OK and view your new data in data view



Variable View window: Measurement

- The almost final column is concerned with the measurement scale properties of your variable. In statistics certain procedures are only appropriate for variables measured on specific scales of measurement. The measurement characteristics recognized by SPSS are as follows:

Scale :	to represent a numeric variable that can take discrete or continuous values along a range (e.g., age, weight, income)
Ordinal :	“ordered” – categories that can be ranked (e.g., level of education)
Nominal:	“non ordered” – categories that cannot be ranked (e.g., Color of eye)

Variable View window: Role of variables

- The final column is concerned with role your variable is going to take in the analysis. This is a new column for version 18. In statistics certain procedures are only appropriate for certain types of variable.

The roles recognized by SPSS are as follows:

- Input: this is variable can be used as an independent predictor.
- Target: this is the outcome of the analysis
- Both: this can be either target or input
- None: no role assigned
- Partition: this variable can be used to partition the data, such as a variable which defines a test or training data set.
- Split: This is included for compatibility with other PASW programmers