SPSS TRAINING MANUAL FOR BEGINNERS

By Grace Hands Research Institute

GRACE HANDS
RESEARCH INSTITUTE

Grace Hands Research Institute, Highflyers In Data Analytics And Research

OUR SERVICES

DATA ANALYTICS
- Data Cleaning And Mining
- Data Visualization
- Critical Thinking (Data Reasoning)
- Data Analysis Reporting (Data Presentation)
- Analyzing Of Big Data etc

TRAINING
- Thesis Writing & Proposal Writing
- Research Analysis
  SPSS, EVIEW, STATA etc
- Data Analytics
  Excel, Power Bi, python etc

RESEARCH
- Writing And Correction Of Full Thesis
- Proposal Development
- Designing Of Questionnaire
- Data Collection (Field Work)
- Analyzing Of Research Data
- Plagiarism Test And Level Reduction
- Business Plan Writing Etc

SPSS

ADEWUYI DAVID
LEAD TRAINER

OFFER PARTNERSHIP
Partner With Our Brand And Earn On Our Monthly Training @ Zero Cost

☎ +2348181969, +2348055295969 Gracehandsresearchinstitute.com ✉ gracehands123 @gmail.com

In our data-rich age, understanding how to analyze and extract true meaning from our business's digital insights is one of the primary drivers of success.

Despite the colossal volume of data we create every day, a mere 0.5% is actually analyzed and used for data discovery, improvement, and intelligence.

While that may not seem like much, considering the amount of digital information we have at our fingertips, half a percent still accounts for a vast amount of data.

With so much data and so little time, knowing how to collect, curate, organize, and make sense of all of this potentially business-boosting information can be a minefield – but online data analysis is the solution.

WHAT IS DATA ANALYSIS?

Data analysis is the process of collecting, cleaning, and analyzing data to extract insights that support decision-making. There are several methods and techniques to perform analysis depending on the industry and the aim of the analysis.

All these various methods for data analysis are largely based on two core areas: quantitative methods and qualitative methods in research. But in this training we will be looking at quantitative research analysis.

WHY IS DATA ANALYSIS IMPORTANT?

In organization

Before we go into detail about the categories of data analysis along with its methods and techniques, you must understand the potential that analyzing data can bring to organizations.

Let’s start with customers, arguably the most crucial element in any business. By using data analysis to get a 360° vision of all aspects related to customers, you can understand which channels they use to communicate with you, their demographics, interests, habits, purchasing behaviors, and more.

In the long run, it will drive success to your marketing strategies, allow you to identify new potential customers, and avoid wasting resources on targeting the wrong people or sending the wrong message. You can also track customer satisfaction by analyzing your client’s reviews or your customer service department’s performance.

From a management perspective, you can also benefit from analyzing your data as it helps you make business decisions based on facts and not simple intuition. For example, you can understand where to invest your capital, detect growth opportunities, predict your incomes, or tackle uncommon situations before they become problems.

Like this, you can extract relevant information from all areas in your organization, and with the help of a dashboard software, present the data in a professional and interactive way to different stakeholders.

The software most used here in organization are Excel, power BI, SPSS, Stata and python but there are otwr software

In Reseach

Data analysis helps you to make meaning of data collected from the field or research carried out in the laboratory. It helps in testing and solving any hypothesis formulated during research. In research the softwares we use are SPSS, STATA, EVIEW, SAS etc
But here in this training we will be looking at SPSS both in the research world and organizational world

TYPES OF DATA

We Have Primary Data And Secondary Data

Primary data is information collected through original or first-hand research. For example, surveys and focus group discussions. On the other hand, secondary data is information which has been collected in the past by someone else. For example, researching the internet, newspaper articles and company reports.

The objective of any study defines whether primary or secondary data is to be collected. For example, if a company intends to enter into women's apparel and wants the India market size, then it can resort to secondary data like industry reports & newspaper articles while if it wants to study the preference of consumers for a new type of fabric/style, then it must conduct primary research like surveys.

Usually, collection of primary data is costly & more time-consuming than secondary data but it serves a specific need and control biases.

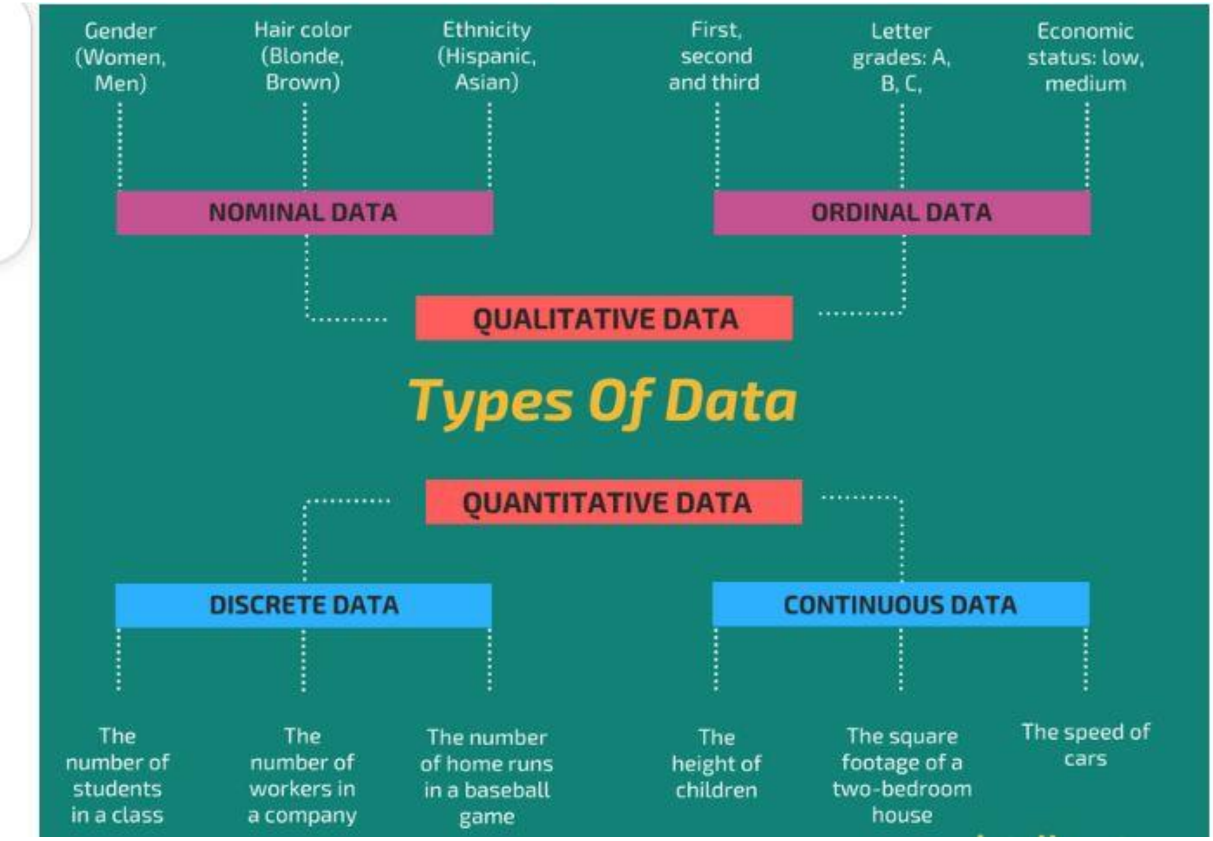
In reseach

Data analysis helps you to make meaning of data collected from the field or research carried out in the laboratory. It helps in testing and solving any hypothesis formulated during research

Primary data is information collected through original or first-hand research. For example, surveys and focus group discussions. On the other hand, secondary data is information which has been collected in the past by someone else. For example, researching the internet, newspaper articles and company reports.

The objective of any study defines whether primary or secondary data is to be collected. For example, if a company intends to enter into women's apparel and wants the India market size, then it can resort to secondary data like industry reports & newspaper articles while if it wants to study the preference of consumers for a new type of fabric/style, then it must conduct primary research like surveys.

Usually, collection of primary data is costly & more time-consuming than secondary data but it serves a specific need and control biases.


Gender (Women, Men)
Hair color (Blonde, Brown)
Ethnicity (Hispanic, Asian)
First, second and third
Letter grades: A, B, C,
Economic status: low, medium
NOMINAL DATA
ORDINAL DATA
QUALITATIVE DATA
Types Of Data
QUANTITATIVE DATA
DISCRETE DATA
CONTINUOUS DATA
The number of students in a class
The number of workers in a company
The number of home runs in a baseball game
The height of children
The square footage of a two-bedroom house
The speed of cars

QUANTITATIVE DATA

Quantitative data seems to be the easiest to explain. It answers key questions such as “how many, “how much” and “how often”.
Quantitative data can be expressed as a number or can be quantified. Simply put, it can be measured by numerical variables.

Quantitative data are easily amenable to statistical manipulation and can be represented by a wide variety of statistical types of graphs and charts such as line, bar graph, scatter plot, and etc.

Examples of quantitative data:

👉Scores on tests and exams e.g. 85, 67, 90 and etc.

👉The weight of a person or a subject.

👉Your shoe size.

👉The temperature in a room.

There are 2 general types of quantitative data: discrete data and continuous data. We will be explainig them now

Discrete vs Continuous Data

As we mentioned above discrete and continuous data are the two key types of quantitative data.
In statistics, marketing research, and data science, many decisions depend on whether the basic data is discrete or continuous.

Discrete data

Discrete data is a count that involves only integers. The discrete values cannot be subdivided into parts.
For example, the number of children in a class is discrete data. You can count whole individuals. You can’t count 1.5 kids.

To put in other words, discrete data can take only certain values. The data variables cannot be divided into smaller parts.
It has a limited number of possible values e.g. days of the month.

Examples of discrete data:

- The number of students in a class.
- The number of workers in a company.The number of home runs in a baseball game.
- The number of test questions you answered correctly

Continuous data

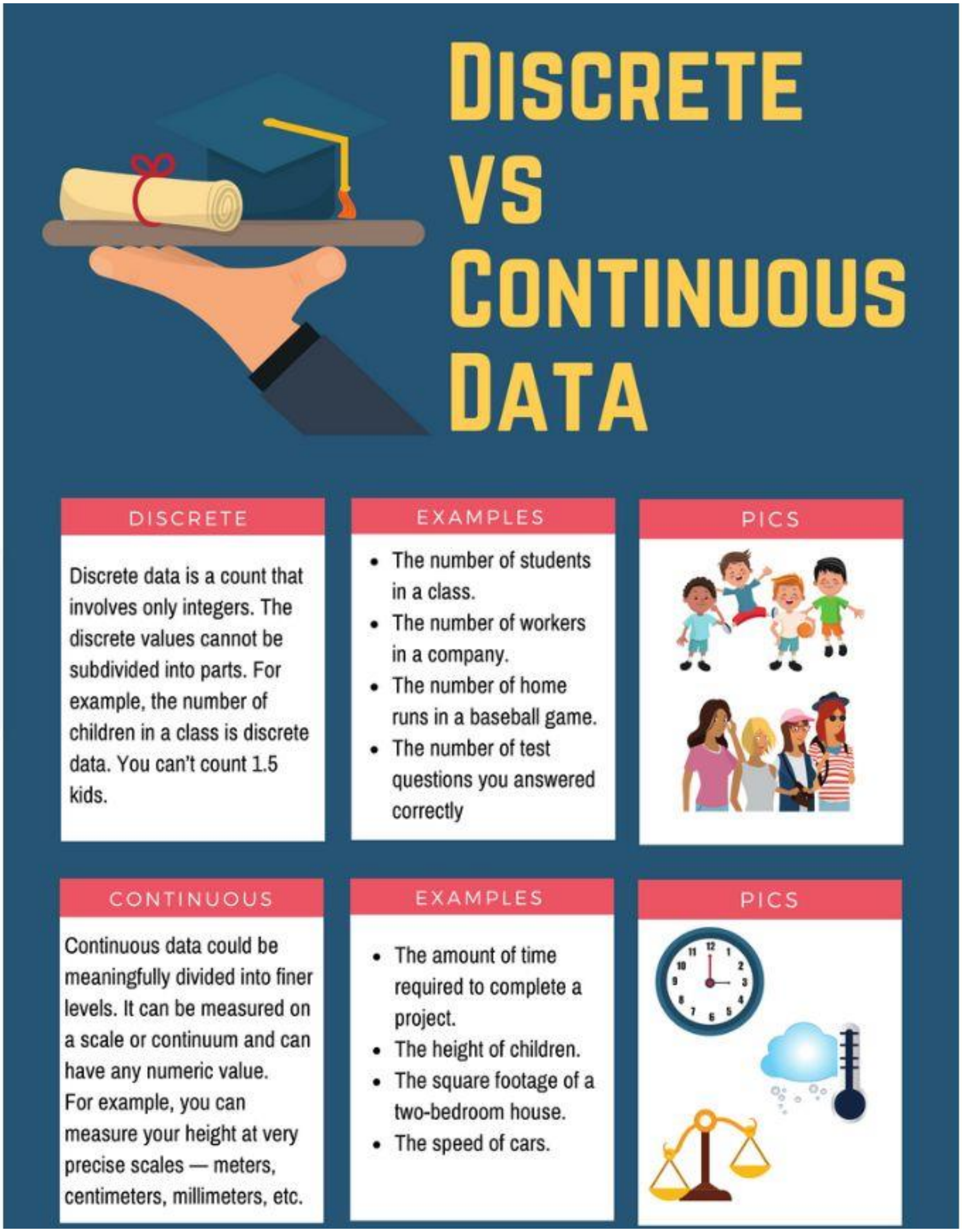
Continuous data is information that could be meaningfully divided into finer levels. It can be measured on a scale or continuum and can have almost any numeric value.

For example, you can measure your height at very precise scales — meters, centimeters, millimeters and etc.
You can record continuous data at so many different measurements – width, temperature, time, and etc. This is where the key difference from discrete types of data lies.

The continuous variables can take any value between two numbers. For example, between 50 and 72 inches, there are literally millions of possible heights: 52.04762 inches, 69.948376 inches and etc.
A good great rule for defining if a data is continuous or discrete is that if the point of measurement can be reduced in half and still make sense, the data is continuous.

Examples of continuous data:

- The amount of time required to complete a project.
- The height of children.
- The square footage of a two-bedroom house.
- The speed of cars.



DISCRETE VS CONTINUOUS DATA

DISCRETE

Discrete data is a count that involves only integers. The discrete values cannot be subdivided into parts. For example, the number of children in a class is discrete data. You can't count 1.5 kids.

EXAMPLES

- The number of students in a class.
- The number of workers in a company.
- The number of home runs in a baseball game.
- The number of test questions you answered correctly

PICS

CONTINUOUS

Continuous data could be meaningfully divided into finer levels. It can be measured on a scale or continuum and can have any numeric value. For example, you can measure your height at very precise scales — meters, centimeters, millimeters, etc.

EXAMPLES

- The amount of time required to complete a project.
- The height of children.
- The square footage of a two-bedroom house.
- The speed of cars.

PICS

Qualitative data can’t be expressed as a number and can’t be measured. Qualitative data consist of words, pictures, and symbols, not numbers.

Qualitative data can answer questions such as “how this has happened” or and “why this has happened”.

Examples of qualitative data:

☞Colors e.g. the color of the sea

☞Your favorite holiday destination such as Hawaii, ☞New Zealand and etc.

☞Names as John, Patricia,…..

☞Ethnicity such as American Indian, Asian, etc.

There are 2 general types of qualitative data: nominal data and ordinal data. We will explain them now

Nominal data

Nominal data is used just for labeling variables, without any type of quantitative value. The name ‘nominal’ comes from the Latin word “nomen” which means ‘name’.

The nominal data just name a thing without applying it to order. Actually, the nominal data could just be called “labels.”

Examples of Nominal Data:

☞Gender (Women, Men)

☞Hair color (Blonde, Brown, Brunette, Red, etc.)

☞Marital status (Married, Single, Widowed)

☞Ethnicity (Hispanic, Asian)

As you see from the examples there is no intrinsic ordering to the variables.
Eye color is a nominal variable having a few categories (Blue, Green, Brown) and there is no way to order these categories from highest to lowest.

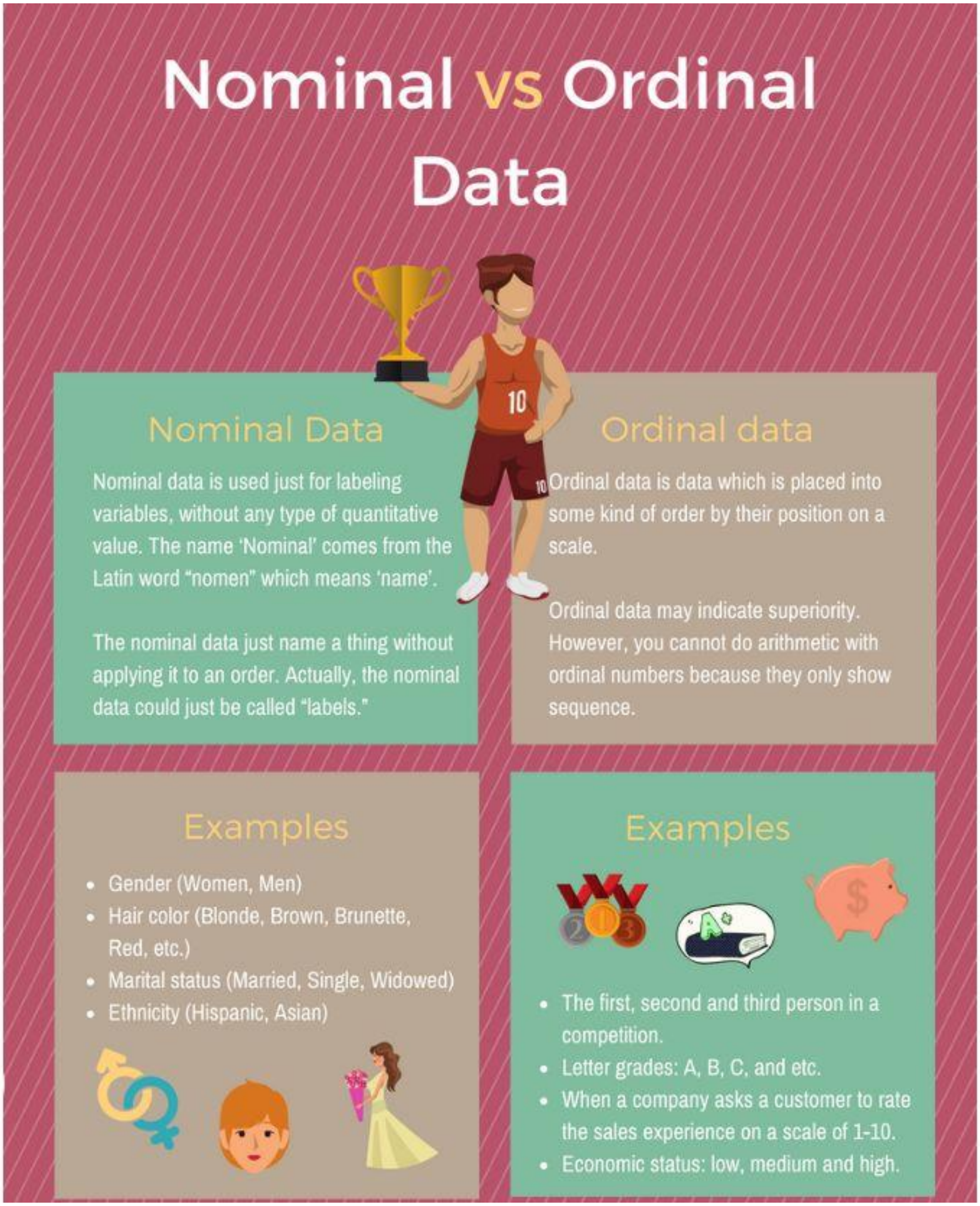
Ordinal data

Ordinal data shows where a number is in order. This is the crucial difference from nominal types of data.
Ordinal data is data which is placed into some kind of order by their position on a scale.
Ordinal data may indicate superiority.

However, you cannot do arithmetic with ordinal numbers because they only show sequence.
Ordinal variables are considered as “in between” qualitative and quantitative variables.

In other words, the ordinal data is qualitative data for which the values are ordered.
In comparison with nominal data, the second one is qualitative data for which the values cannot be placed in an ordered.

Now a diagrammatic representation of the differences between quantitative and qualitative data is below



Nominal vs Ordinal Data

Nominal Data

Nominal data is used just for labeling variables, without any type of quantitative value. The name 'Nominal' comes from the Latin word "nomen" which means 'name'.

The nominal data just name a thing without applying it to an order. Actually, the nominal data could just be called "labels."

Ordinal data

Ordinal data is data which is placed into some kind of order by their position on a scale.

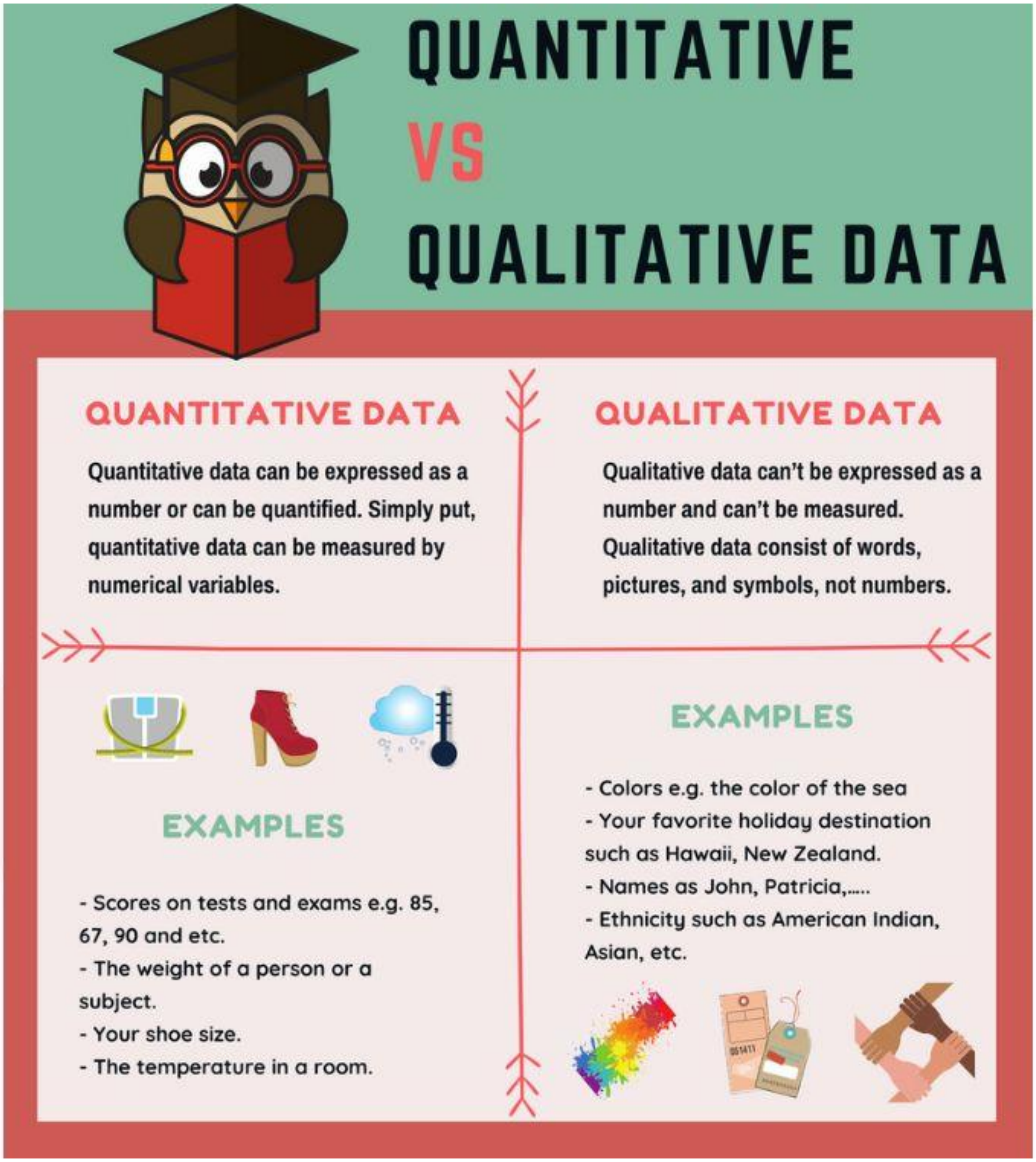
Ordinal data may indicate superiority. However, you cannot do arithmetic with ordinal numbers because they only show sequence.

Examples

- Gender (Women, Men)
- Hair color (Blonde, Brown, Brunette, Red, etc.)
- Marital status (Married, Single, Widowed)
- Ethnicity (Hispanic, Asian)

Examples

- The first, second and third person in a competition.
- Letter grades: A, B, C, and etc.
- When a company asks a customer to rate the sales experience on a scale of 1-10.
- Economic status: low, medium and high.



QUANTITATIVE VS QUALITATIVE DATA

QUANTITATIVE DATA

Quantitative data can be expressed as a number or can be quantified. Simply put, quantitative data can be measured by numerical variables.

EXAMPLES

- Scores on tests and exams e.g. 85, 67, 90 and etc.
- The weight of a person or a subject.
- Your shoe size.
- The temperature in a room.

QUALITATIVE DATA

Qualitative data can't be expressed as a number and can't be measured. Qualitative data consist of words, pictures, and symbols, not numbers.

EXAMPLES

- Colors e.g. the color of the sea
- Your favorite holiday destination such as Hawaii, New Zealand.
- Names as John, Patricia,.....
- Ethnicity such as American Indian, Asian, etc.

Below, we've outlined the six steps you'll need to follow to analyze your data:

HOW TO ANALYZE YOUR DATA

You'll need to implement a data analysis process to get the most out of your data. While it can be complex to perform data analysis, depending on the type of data you're analyzing, there are some hard and fast rules that you can follow. They include setting goals, collecting, cleaning, and analyzing data, then visualizing it in striking dashboards to make it easy to spot patterns and trends.

1. Define Your Goals: Setting clear objectives is key and will help determine the type of data that you'll need to collect and analyze. This is applicable to both organization and research world

In organization you set your objectives at the beginning of the year it is called target and this will affect the target given to your staffs
In research, objectives are set in chapter one, introduction, after selecting the topic of the research

2. Collect Your Data: Data is everywhere, and you'll want to bring it all into one place ready for analysis. Whether you're collecting quantitative or qualitative data, Excel is a great platform for storing your data both in organizational settings and research world.

You can gather data in research world by distributing of questionnaires or interview methods but in organization you collect it by just inputting customer information in excel

3. Clean Your Data: Now whatever data is collected may not be useful or irrelevant to your aim of Analysis, hence it should be cleaned. The data which is collected may contain duplicate records, white spaces or errors. The data should be cleaned and error free. This phase must be done before analysis because based on data cleaning, your output of Analysis will be closer to your expected outcome.

4. Analyze Your Data: Once the data is collected, cleaned, and processed, it is ready for Analysis. As you manipulate data, you may find you have the exact information you need, or you might need to collect more data. During this phase, you can use data analysis tools and software which will help you to understand, interpret, and derive conclusions based on the requirements.

5. Data interpretation: After analyzing your data, it's finally time to interpret your results. You can choose the way to express or communicate your data analysis either you can use simply in words or maybe a table or chart. Then use the results of your data analysis process to decide your best course of action.

6. Draw Conclusions: Gain actionable insights and make data-based decisions by digging into your data from every angle. This is what your boss or your supervisor is waiting for.

WHAT IS SPSS

SPSS is short for Statistical Package for the Social Sciences, and it’s used by various kinds of researchers for complex statistical data analysis.

The SPSS software package was created for the management and statistical analysis of social science data. It was originally launched in 1968 by SPSS Inc., and was later acquired by IBM in 2009.

Officially dubbed IBM SPSS Statistics, most users still refer to it as SPSS. As the world standard for social-science data analysis, SPSS is widely coveted due to its straightforward and English-like command language and impressively thorough user manual.

SPSS is used by market researchers, health researchers, survey companies, government entities, education researchers, marketing organizations, data miners, and many more for processing and analyzing survey data, such as you collect with an online survey platform like Alchemer.

Most top research agencies use SPSS to analyze survey data and mine text data so that they can get the most out of their research and survey projects.

There are a handful of statistical methods that can be leveraged in SPSS, including:

☞Descriptive statistics, including methodologies such as frequencies, cross-tabulation, and descriptive ratio statistics.

☞Bivariate statistics, including methodologies such as analysis of variance (ANOVA), means, correlation, and nonparametric tests.

Numeral outcome prediction such as linear regression.

☞Prediction for identifying groups, including methodologies such as cluster analysis and factor analysis.

SPSS TERMINOLOGY

Cases: Cases represent independent observations, experimental units, or subjects. For example, if the data are based on a survey of college students, then each row in the data would represent a specific college student who participated in the study.

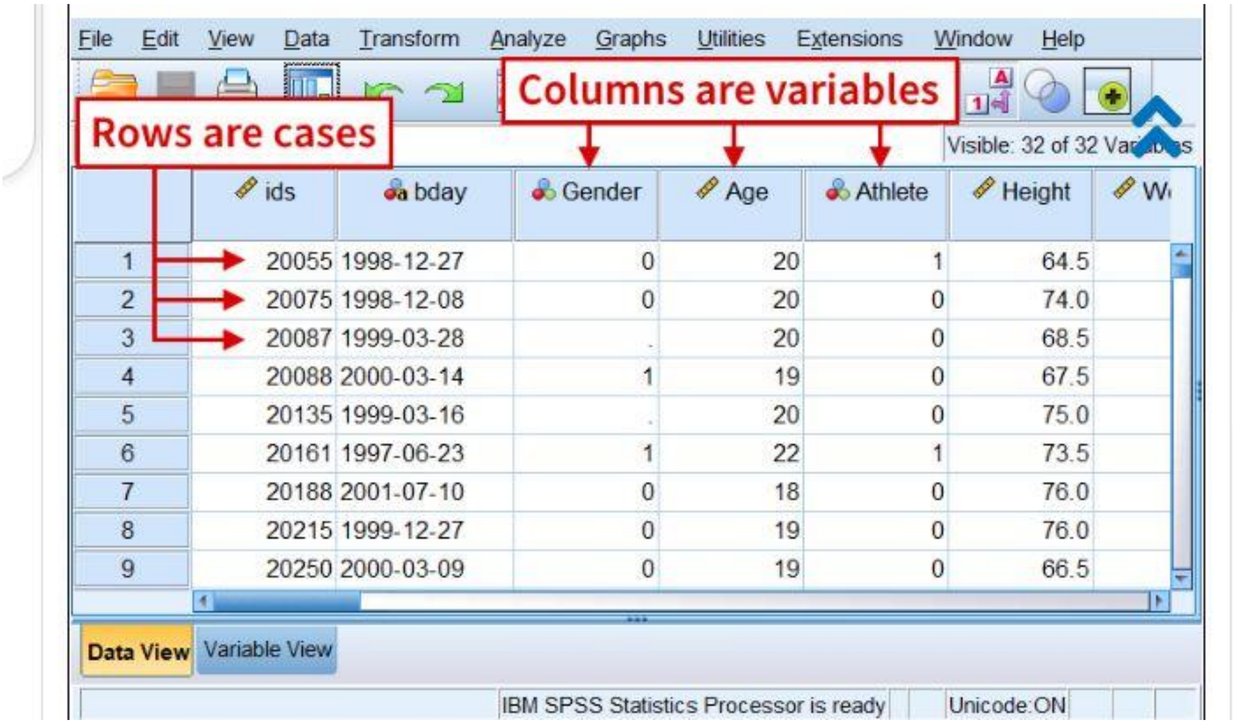
When you view data in SPSS, each row in the Data View represents a case.

Variables: Variables are attributes, characteristics, or measurements that describe cases.

For example, your data might include information such as each college student’s date of birth, gender, or class rank.

Each of these pieces of information is a variable that describes each case (college student).

When you view data in SPSS, each column represents a variable.



Rows are cases

Columns are variables

	ids	bday	Gender	Age	Athlete	Height	W
1	20055	1998-12-27	0	20	1	64.5	
2	20075	1998-12-08	0	20	0	74.0	
3	20087	1999-03-28	.	20	0	68.5	
4	20088	2000-03-14	1	19	0	67.5	
5	20135	1999-03-16	.	20	0	75.0	
6	20161	1997-06-23	1	22	1	73.5	
7	20188	2001-07-10	0	18	0	76.0	
8	20215	1999-12-27	0	19	0	76.0	
9	20250	2000-03-09	0	19	0	66.5	

.

Coding: This is an act of representing variables with short names, letters or combination of letters and number

Inputting: This is called data entering. This is where you input the answers given on the questionnaire

Respondents: These are people that filled the questionnaire or used for the research survey

UNDERSTANDING SPSS INTERFACE

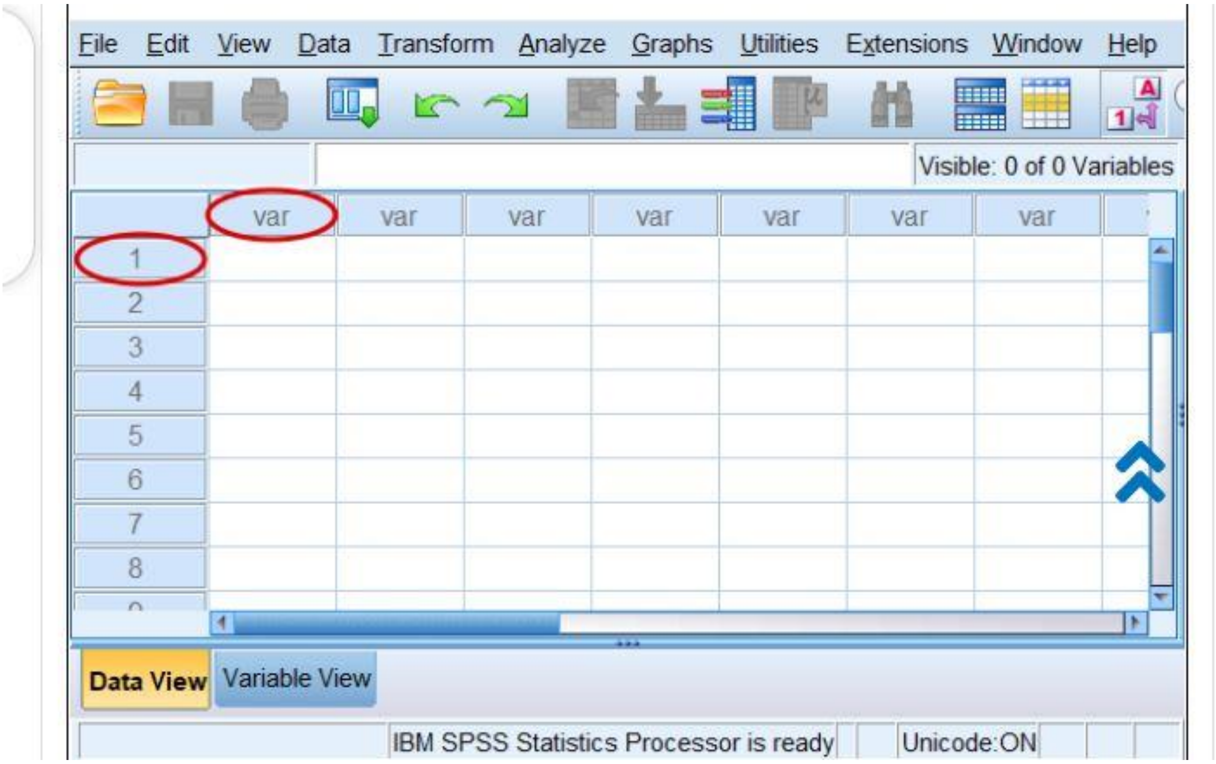
Now the first thing we will talk about here is the variable view, then we move to the data view after the output view

Now these are most important views we have in excel or let me say the most used views

Variable view : Variable View tab displays information about the variables in your data. This is where we coding

Data view: this is where you do your inputting of data

Output view: This shows the result of your analysis and savings



If u open your system you will see this interface

What you are looking at here is the data view

First you will notice Var at the top of each column that means variables

So it means your your data view can also display your variables

Its very key you know this bcous it will go along way while you are inputting your data

CREATING A NEW DATA SET

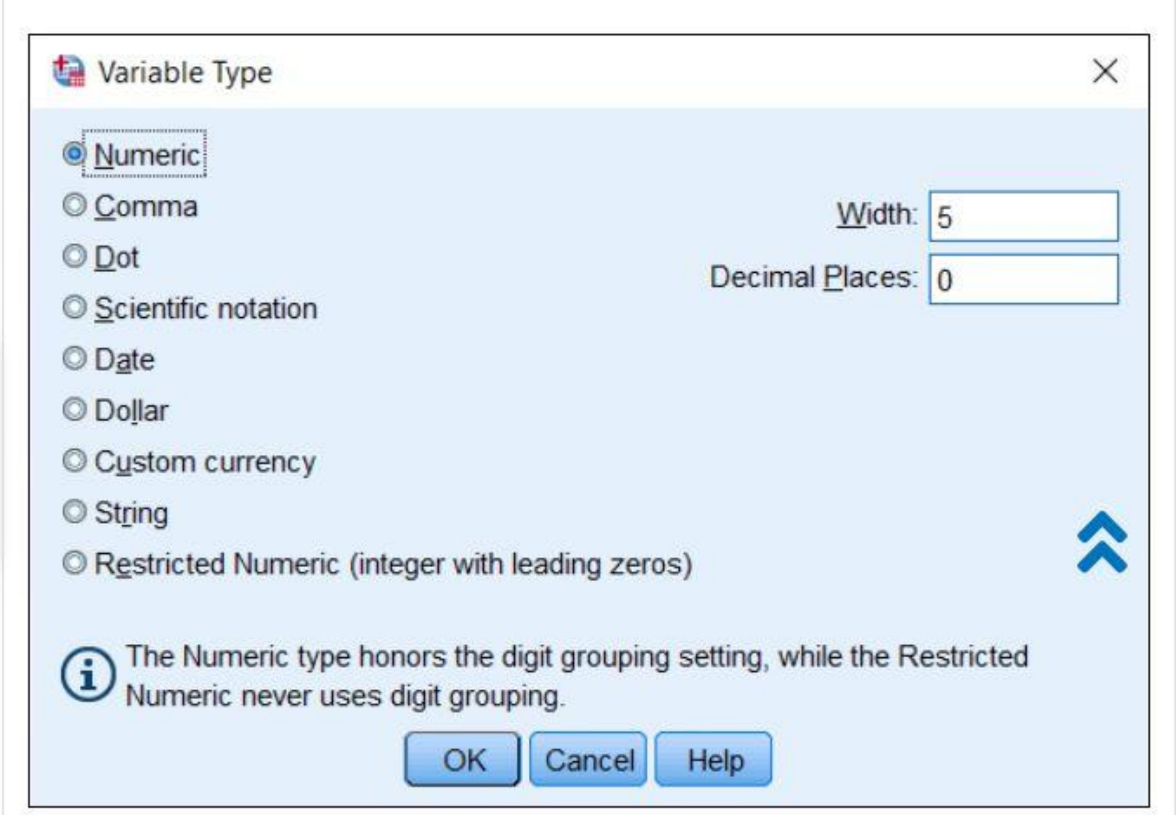
You create a new data set first by coding which can only be done in variable view. There are three steps that must be followed to create a new data set in SPSS. The steps are listed below

STEP 1: **DEFINING VARIABLES IN A NEW DATA SET**

NAME

This field describes the name of the variable being defined. To change the name, place the cursor in this field and type the name. The variable name must begin with a letter of the alphabet and cannot exceed 8 characters. Spaces are not allowed within the variable name. Each variable name must be unique.

VARIABLE TYPES



Variable Type

● Numeric
○ Comma
○ Dot
○ Scientific notation
○ Date
○ Dollar
○ Custom currency
○ String
○ Restricted Numeric (integer with leading zeros)

Width: 5

Decimal Places: 0

The Numeric type honors the digit grouping setting, while the Restricted Numeric never uses digit grouping.

Numeric variables: Variables that deals with numbers majorly e.g house old size, age, income etc

Example: Continuous variables that can take on any number in a range (e.g., height in centimeters and weight in kilograms) should be treated as numeric variables.

Example: Counts (e.g., number of people living in a household) should be treated as numeric variables with zero decimal places. Certain mathematical calculations are valid when applied to count variables (e.g., mean and standard deviation), but some statistical procedures requiring continuous numeric variables may not be (e.g., the dependent variable in a linear regression), depending on the distribution of the variable.

Example: Nominal categorical variables that have been coded numerically (e.g., recording a subject's gender as 1 if male or 2 if female) should be treated as numeric variables with zero decimal places. In this situation, the Measure setting must be defined as Nominal. This type of numeric variable should never be used in mathematical calculations, nor used in any statistical procedure requiring continuous numeric variables (e.g. the dependent variable of a linear regression).

Example: Ordinal categorical variables that have been coded numerically (e.g., a questionnaire item with responses 1=Small, 2=Medium, 3=Large) should be treated as numeric variables with zero decimal places.

String variable: Use when you want to type letters (words). For example, peoples' names, breeds of dog, occupations. You can also include numbers or symbols, but they will be treated by SPSS as text. For example, zip codes are numeric but you may want to treat them as text (i.e. you don't actually want to perform calculations on them like 90210 * 10 !).

Comma: Numeric variables that are separated every three places by a comma. For example, 100,000.00 or 999,988,565.21.

Dot: Similar to comma, but the dot is used to separate the three places and a comma is used to indicate a decimal. For example. 100.000,00 and 999.988.565,21. Not used in the UK or USA, but common in some other countries.

You are most likely to use either string variables, or numeric variables.

WIDTH

The number of digits displayed for numerical values or the length of a string variable.

To set a variable's width, click inside the cell corresponding to the “Width” column for that variable. Then click the "up" or "down" arrow icons to increase or decrease the number width.

DECIMALS:

The number of digits to display after a decimal point for values of that variable. Does not apply to string variables. Note that this changes how the numbers are displayed, but does not change the values in the dataset.

To specify the number of decimal places for a numeric variable, click inside the cell corresponding to the “Decimals” column for that variable. Then click the “up” or “down” arrow icons to increase or decrease the number of decimal places.

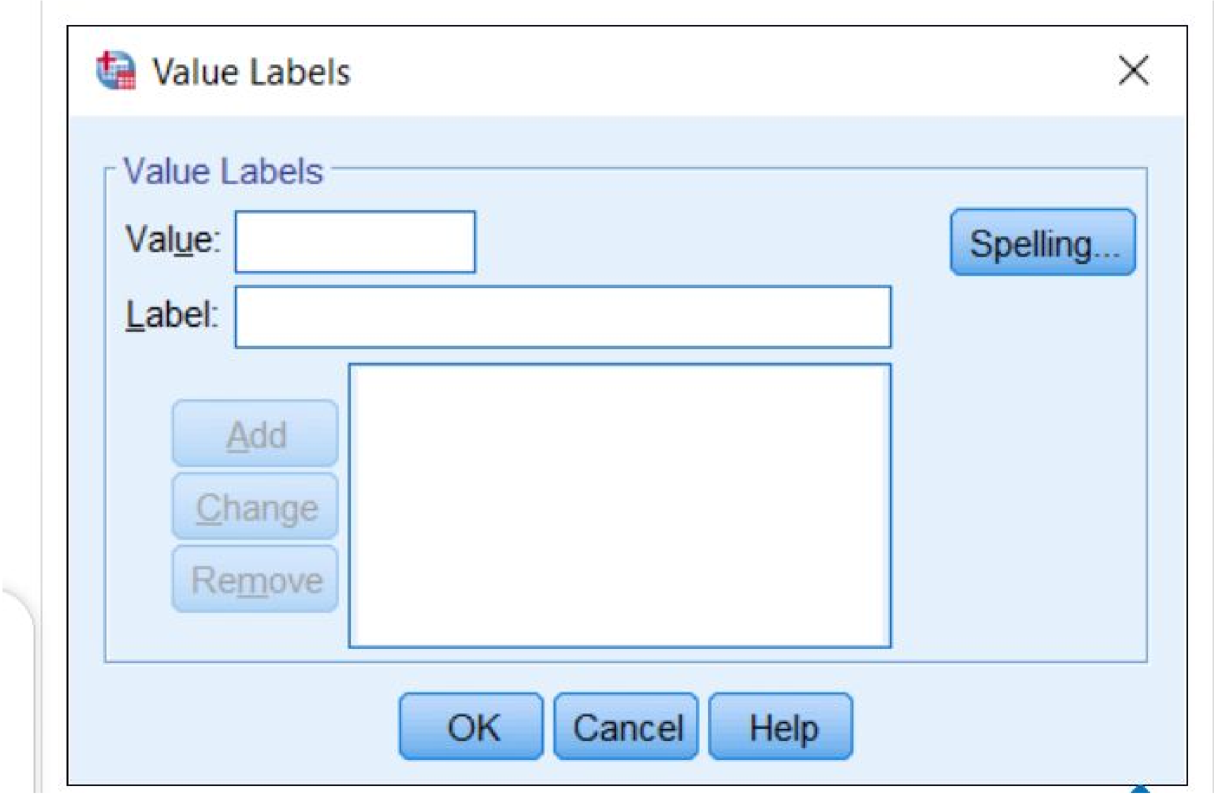
Example: If you specify that values should have two decimal points, they will display as 1.00, 2.00, 3.00, and so on.

LABEL

The full name for the variable that can be up to 120 characters long and can include spaces (which variable names cannot). If a variable label is entered, the label will be printed on charts and reports instead of the name, making them easier to understand.

VALUES

For coded categorical variables, the value label(s) that should be associated with each category abbreviation. Value labels are useful primarily for categorical (i.e., nominal or ordinal) variables, especially if they have been recorded as codes (e.g., 1, 2, 3). It is strongly suggested that you give each value a label so that you (and anyone looking at your data or results) understands what each value represents.

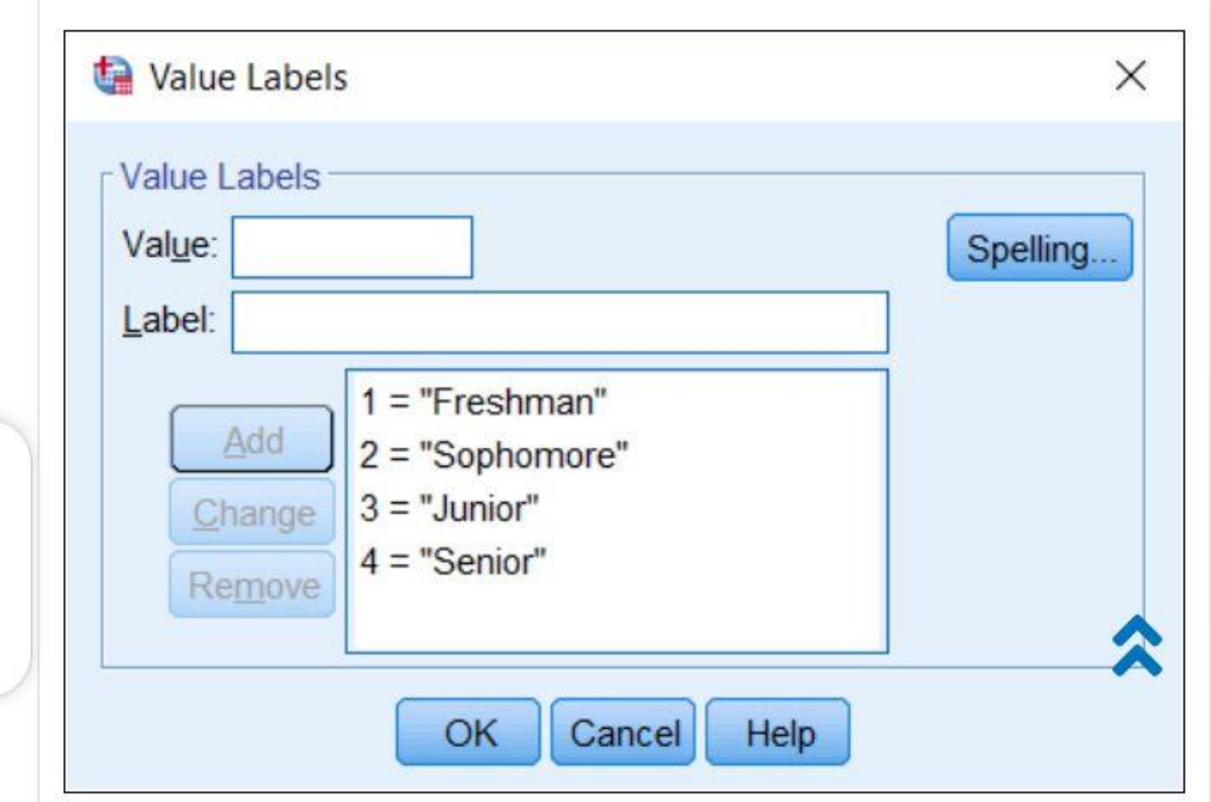


When value labels are defined, the labels will display in the output instead of the original codes.Note that defining value labels only affects the labels associated with each value, and does not change the recorded values themselves.

Example: In the sample dataset, the variable Rank represents the student's class rank. The values 1, 2, 3, 4 represent the categories Freshman, Sophomore, Junior, and Senior, respectively. Let's define the category labels for the Rank variable in the sample data.

Under the column “Values,” click the cell that corresponds to the variable whose values you wish to label. If the values are currently undefined, the cell will say “None.” Click the square “…” button. The Value Labels window appears.

Type the first possible value (1) for your variable in the Value field. In the Label field type the label exactly as you want it to display (e.g., "Freshman"). Click Add when you are finished defining the value and label. Your variable value and label will appear in the center box. Repeat these steps for each possible value for your variable. When all of the labels have been defined, the Value Labels window should look like this



1 = "Freshman"
2 = "Sophomore"
3 = "Junior"
4 = "Senior"

Click OK at the bottom of the window.

If you wish to change or remove a value and label that you have added to the center dialog box, do the following:

To change a specific value or label, highlight the value/label in the center text box in the Value Labels window. Now the selected value/label will be highlighted yellow. Make changes to the selected value or label as needed. Click Change. The changes will be applied to the value/label you highlighted.

To remove a specific value/label, highlight the value/label in the center text box. Click Remove. The selected value/label will be removed from the center text box.

MISSING VALUES

This field indicates which subset of the data will not be included in the data set. To change this field, click on the Missing Values… button. This will open the Define Missing Values: dialog box. Enter the appropriate information into the fields. When done, click on the Continue button.

COLUMNS

The width of each column in the Data View spreadsheet. Note that this is not the same as the number of digits displayed for each value. This simply refers to the width of the actual column in the spreadsheet.

To set a variable's column width, click inside the cell corresponding to the “Columns” column for that variable. Then click the “up” or “down” arrow icons to increase or decrease the column width.

ALIGN

The alignment of content in the cells of the SPSS Data View spreadsheet. Options include left-justified, right-justified, or center-justified.

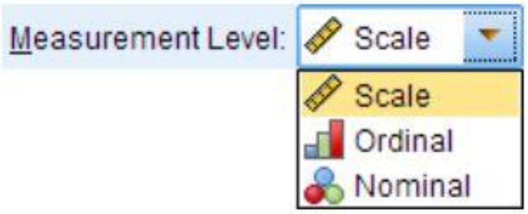
To set the alignment for a variable, click inside the cell corresponding to the "Align" column for that variable. Then use the drop-down menu to select your preferred alignment: Left, Right, or Center.

MEASURE

The level of measurement for the variable (e.g., nominal, ordinal, or scale).

Some procedures in SPSS treat categorical and scale variables differently. By default, variables with numeric responses are automatically detected as “Scale” variables. If the numeric responses actually represent categories, you must change the specified measurement level to the appropriate setting.

To define a variable's measurement level, click inside the cell corresponding to the “Measure” column for that variable. Then click the drop-down arrow to select the level of measurement for that variable: Scale, Ordinal, or Nominal.



ROLE

The role that a variable will play in your analyses (i.e., independent variable, dependent variable, both independent and dependent). Some options in SPSS allow you to pre-select variables for particular analyses based on their defined roles. Any variable that meets the role requirements will be available for use in such analyses. You can choose from the following roles for each variable:

Input: The variable will be used as a predictor (independent variable). This is the default assignment for variables.

Target: The variable will be used as an outcome (dependent variable).

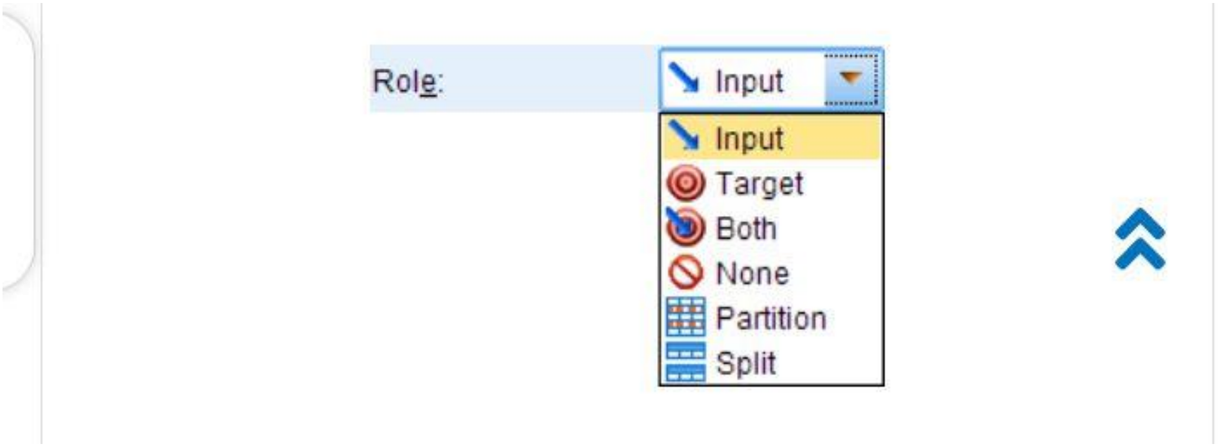
Both: The variable will be used as both a predictor and an outcome (independent and dependent variable).

None: The variable has no role assignment.

Partition: The variable will partition the data into separate samples.

Split: Used with the IBM® SPSS® Modeler (not IBM® SPSS® Statistics).

To define a variable's role in your analysis, click inside the cell corresponding to the “Role” column for that variable. Then use the drop-down menu to select the role that variable will take: Input, Target, Both, None, Partition, or Split.


Role: Input

STEP TWO : DATA CODING AND INPUTTING

Click the Variable View tab. Type the name for your first variable under the Name column. You can also enter other information about the variable, such as the type (the default is “numeric”), width, decimals, label, etc.

Type the name for each variable that you plan to include in your dataset.

In this example, I will type “School_Class” since I plan to include a variable for the class level of each student (i.e., 1 = first year, 2 = second year, 3 = third year, and 4 = fourth year). I will also specify 0 decimals since my variable values will only include whole numbers. (The default is two decimals.). Basically I will fill the entire row for this first variable.

STEP THREE : DATA SET SAVING

You jus learn how to save your data set very fast. Ones you save your saving will display on the out put view. In fact every time you save it will display on the this view. This means if it does not display you have not saved the data set

Two easy ways to save

1) Pressing control S

2) click the file on SPSS interface, click **save as** if you are saving for the first time but if not just click **save**

DATA MANIPULATION

Data manipulation is defined as the process of changing or altering data in order to make it more readable and organized. For example, data can be arranged alphabetically to help the owner quickly find useful information.

For organization

Manipulation of data allows businesses to efficiently use the data for predicting trends, understanding customer behavior, increasing productivity, reducing costs

In the research it allows for easy understanding of the data and performing data analysis with it

In performing data manipulation in SPSS, I will be discussing two methods with us

These are the methods I usually use a lot

They are sorting and transformation

Sorting

These two needs practical demonstration which we will have access to on the practical videos but let me say something about it

Sorting is act of arranging or rearranging of data for the purpose of identifying a particular data or variable

You can sort in ascending or descending manner

Or

In lowest to highest and highest to lowest

Now to Data Transformation

Under data transformation what we will be looking at is computing

This is the process of making multiple data to become one singlar data

So if you have 10 variables in you data set and you don't want to start analyzing with them one by one

You can compute them, the variables will become one variable

The one variable can then be use any analysis you need to carry out

DATA CLEANING

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled.

The steps to clean a dataset may vary slightly depending on the research methodology, and if the resulting data is largely quantitative or qualitative. However, the below represent some of the most commonly used steps in the data cleaning process.

Steps To Data Cleaning

1. Remove Duplicate and Incomplete Cases:

Datasets may sometimes include duplicate cases if a respondent accidentally took a survey twice, data were combined from multiple sources, or there was an error when retrieving the dataset. Depending on the data collection tool you use, it is also possible that an initial dataset includes incomplete cases, for example, if a survey respondent took only half of the questions.

The first step in data cleaning is to remove any duplicate or incomplete cases so that you are examining a set of unique and complete cases.

2. Remove Oversample:

In many cases, particularly when conducting survey research, a researcher may collect more responses than they need. For example, you may be aiming to gather 500 completed survey responses, of which 250 identify as female and 250 identify as male, but end up gathering 700 completed responses, 300 who are female and 400 who are male. As you have 50 extra female completes and 150 additional male completes, you will need to cut back the data so that the sample is equally representative of male and female respondents.

Researchers should use a randomized method to remove any oversample in order to meet the sample requirements so that each respondent has an equal chance of being included or excluded in the final dataset.

3. Ensure Answers are Formatted Correctly:

dData may come in several different formats when it is first accessed – for example, a multiple-choice survey question about ice cream flavors may list answers as numbers 1, 2, and 3 when in the question they represented text choices Vanilla, Chocolate, and Strawberry. Depending on how the data will be analyzed, researchers may want to replace the numerical data with the textual data.

If data is being combined from multiple surveys or data sources, there could be two different words used to represent the same thing – for example ‘Not sure’ vs ‘Unsure’. To avoid these cases being represented as unique answers, they should be combined.

4. Remove Nonsense Answers and Unreadable Data:

Datasets may include nonsense answers, such as those including symbols or other words or numbers which do not make sense in the context of the question or field. It is also possible when importing data from multiple files that some data may be unreadable. These cases should be removed from the final dataset.

5. Code Open Ended Data

The coding of open-ended response data is an entire process unto itself, however, it is also an important part of data cleaning. Datasets that include open-ended data can be particularly time-consuming to clean as data can be lengthy, unrelated to the question at hand, or hard to decipher. In order to glean statistical insights from qualitative responses, open-ended responses may be coded into categories, a process that involves first reviewing all responses manually to create categories, and then going through open-ended data and actually placing it into the categories. If the dataset is in multiple languages, this step may also include translating responses into the language the analysis will be conducted in.

6. Check Data Consistency:

Check on logic relations and ensure there are correlated data sets, and there are no inconsistencies including contradictions and gaps in data. In case of any inconsistencies between the data and the questionnaire, the issues should be flagged in order for you to decide a way forward or if data should be excluded from the final dataset.

7. Perform Final Quality Assurance Checks:

After you have gone through the above steps, researchers will still want to perform manual quality assurance checks of the data before starting data analysis. This final step should examine the dataset in its entirety, looking to see if there are anomalies in data for any individual question or data point and double-checking that data is formatted correct

STATISTICS

Statistics is a term used to summarize a process that an analyst uses to characterize a data set. If the data set depends on a sample of a larger population, then the analyst can develop interpretations about the population primarily based on the statistical outcomes from the sample. Statistical analysis involves the process of gathering and evaluating data and then summarizing the data into a mathematical form.

Statistics is a form of mathematical analysis that uses quantified models, representations and synopses for a given set of experimental data or real-life studies.

Statistics studies methodologies to gather, review, analyze and draw conclusions from data

DESCRIPTIVE STATISTICS

Descriptive statistics does the following with data

1. Getting mean

2. Getting median

3. Getting mode

4. Getting variance

5. Getting range

Etc

SPSS can give you answers to all these within seconds, Its far better than ur calculator

Statistics in data analytics is divided into two

Descriptive and inferential

Everything I explained earlier falls under descriptive statistics

Inferential statistics: This is one of the major analytics techniques that has a researcher you must know and it requires a lot of practice. Inferential statistics, are chi square, ANOVA, Regression, correlation etc

They are used for analysis of hypothesis and complex objectives

I hope we have heard those words before

They are under inferential statistics

TESTING FOR NORMALITY

An assessment of the normality of data is a prerequisite for many statistical tests because normal data is an underlying assumption in parametric testing. There are two main methods of assessing normality: graphically and numerically.

This "quick start" guide will help you to determine whether your data is normal, and therefore, that this assumption is met in your data for statistical tests.

The approaches can be divided into two main themes: relying on statistical tests or visual inspection. Statistical tests have the advantage of making an objective judgement of normality, but are disadvantaged by sometimes not being sensitive enough at low sample sizes or overly sensitive to large sample sizes. As such, some statisticians prefer to use their experience to make a subjective judgement about the data from plots/graphs. Graphical interpretation has the advantage of allowing good judgement to assess normality in situations when numerical tests might be over or under sensitive, but graphical methods do lack objectivity. If you do not have a great deal of experience interpreting normality graphically, it is probably best to rely on the numerical methods

DESIGNING OF A QUESTIONNAIRE

Now when designing your questionnaire note the following:

👉First your questionnaire must follow your objectives

See the flow

👉Your research questions must come from your statement of problem

👉Your objectives must come from your research questions

👉Your questionnaire must also follow your objectives too

If you have three objectives you will be having 4 sections in your questionnaire

👉Section A are questions on the personal details of your respondents like age, marital status, etc. All questionnaires must have this.

👉Section B are questions on your first objective

👉Section C should be on your second objective

👉Section D should be based on your third objective

Do we get it?

Please let's note this

Make sure the questions under each section actually answer those particular objectives. Don't be like a lady I once worked with that after I have distributed her questionnaire and was about to start her analysis that was I realized that her questionnaire cannot answer her objectives which are the basics for any project.

Notes: These is the beginners class lecture note

The step by step practical videos explained everything discussed here in depth and its 3 hours +. Make sure you download it and watch it

At the intermediate class we will be looking at the following, don't miss it.

iNTERMEDIATE TRAINING

☞What is pilot research

☞Reliability testing

☞INFERENTIAL STATISTICS

☞Understanding hypothesis in research

☞Understanding of P value and F value

☞Chi square test + interpretation

☞ANOVA test + interpretation

☞ANCOVA test + interpretation

☞Correlation analysis + interpretation

☞Trend analysis / time series analysis + interpretation