RESEARCH METHODOLOGY
AGG 3381
DATA ANALYSIS

Dr. Malunga Syacumpi

CONTENT

- What is data analysis
- Descriptive statistics
- Inferential statistics
- Associations between two variables: bivariate analysis
- Analysing three or more variables: multivariate analysis
- Recap on course expectations

WHAT IS DATA ANALYSIS

- Data analysis is the process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making
- It is the systematic, cleaning, transforming, describing, modeling, and interpreting data, generally employing statistical techniques.
- It is an important part of both scientific research and business, where demand has grown in recent years for data-driven decision making.
- Data analysis techniques are used to gain useful insights from datasets, which can then be used to make operational decisions or guide future research.

KEY TERMS IN QUANTITATIVE RESEARCH

TERM	DEFINITION
Concept	The building blocks of theory and represent the points around which research is conducted. These are categories for the organization and of ideas and observations (eg. Organisational performance; fiscal discipline)
Unit of Analysis	The most elemental part of what is being studied or observed (eg. Individuals, households, cases)
Indicator	An indicator is something that is devised or already exists that is used as a measure for a concept.
Variable	Concepts, characteristics or properties that can vary from one unit of analysis to another. ***(Dependent Variable :*** *Variable whose change the researcher wishes to explore* *(I**ndependent Variable:*** *Variable that helps to explain the change in dependent variable)*
Hypothesis	An empirical statement which seeks to test the relationship between at least two variables
Levels of measuring Variables	Variables can either be***:*** ***Nominal***: Has qualitative categories that can be ranked in terms of magnitude/ degree (Race) **Ordinal :** Qualitative categories that can be ranked in terms of magnitude/ degree (Education) **Interval/ratio:** Has quantitative values or numbers (age, money,) **Dichotomous:** A variable with only two categories
Reliability	Consistency of a measure of a concept (consistent, stable,)
Validity	Is an issue of whether an indicator devised to measure a concept really measures the concept.
Quantitative Data Analysis	It a process of converting raw numbers into meaning using mathematical, rational and critical thinking. It is It can be ***Univariate, Bivariate*** *or* ***Multivariate***

QUANTITATIVE DATA ANALYSIS

- There are generally two types of statistics that you will encounter in quantitative data analysis

1. **Descriptive statistics**

2. **Inferential statistics**

DESCRIPTIVE STATISTICS

- Also referred to as **Univariate Statistics**
- They are often used to describe variables (Charts, graphs, tables)
- Their analysis involves one variable at a time
  1. Frequency tables
  2. Measures of central tendency
  3. Measures of variability (dispersion/deviation)
- Researchers often perform these descriptive statistics before beginning any type of data analysis.

1. FREQUENCY TABLES

- Detailed description of categories/values for one variable. A frequency table often includes:
  - Absolute frequency
  - Relative frequency
  - Cumulative frequency

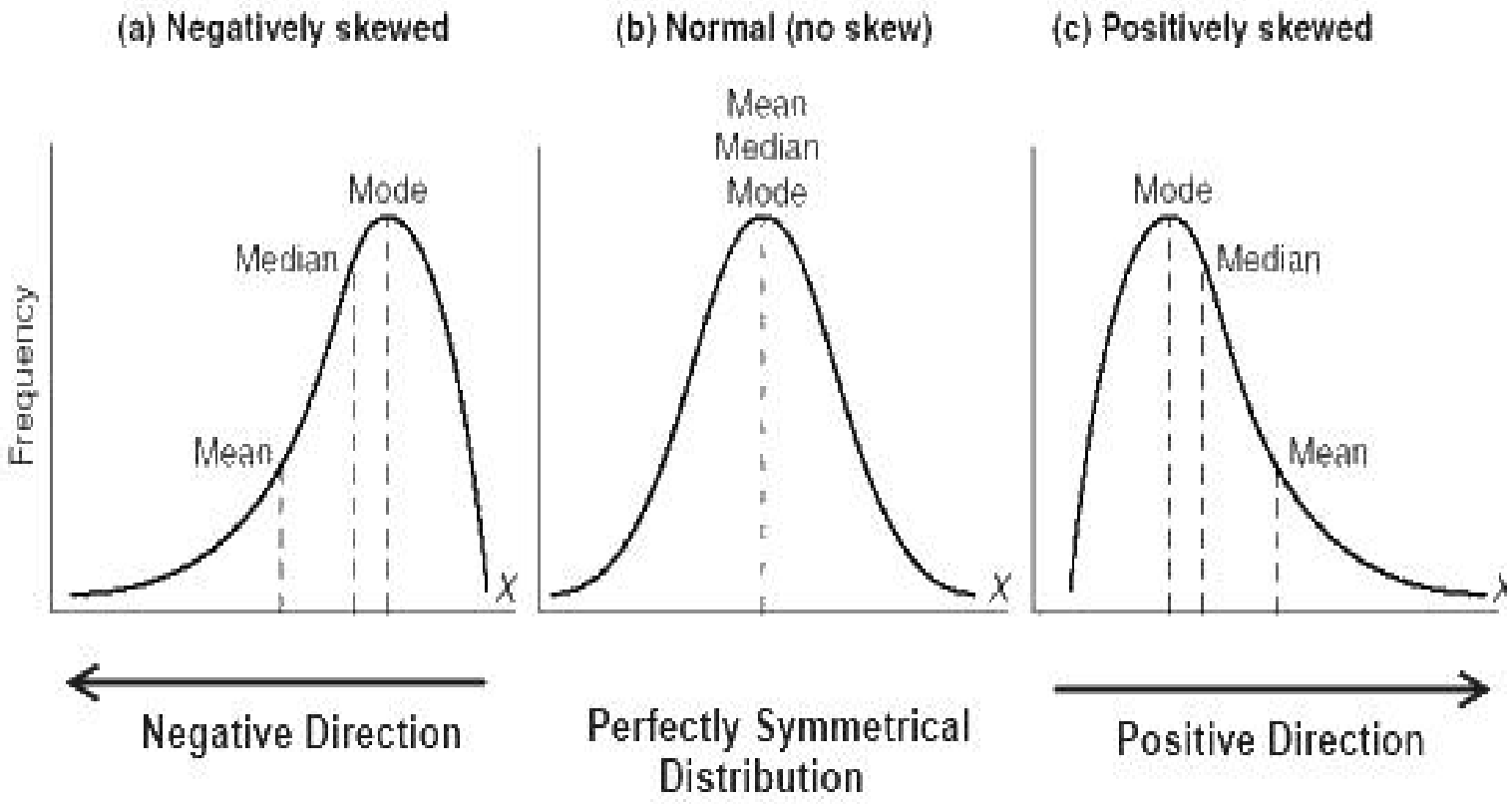
favourite means of transportation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Car	38	38,0	38,0	38,0
	Plane	40	40,0	40,0	78,0
	Coach	12	12,0	12,0	90,0
	Train	10	10,0	10,0	100,0
	Total	100	100,0	100,0	

2. MEASURES OF CENTRAL TENDANCY

- Measures of central tendency provide the most occurring or middle value for each variable. There are three measures of central tendency;
  - ***Mode:*** value occurring most frequently. It is rarely of any practical use for numerical data.
  - ***Median:*** (also referred to as the 50th percentile) is the middle value in a sample of ordered values. Half the values are above the median and half are below the median.
  - ***Mean:*** is obtained by summing the values of all the observations and dividing by the number of observations.
- A comparison of the mean, median and mode can reveal information about skewness. The mean, median and mode are similar when the distribution is symmetrical. When the distribution is skewed the median is more appropriate as a measure of central tendency.

MEASURES OF CENTRAL TENDENCY



3. MEASURES OF VARIABILITY (DISPERSION/DEVIATION

- Measure of variability is defined as the dispersion (or deviation) away from the mean of each variable.
- They only exist for interval level variable.
- The three measures of variability include :
    1. *Range :* *T*he difference between the highest and the lowest numbers
    2. *Standard Deviation:* A measure of how the average score deviates or spreads away from the me(defined as the square root of the variance)
    3. *Variance: t*he variance basically means standard deviation squared

INFERENTIAL STATISTICS

- Statistics used to draw a conclusion (inference) on a larger population.
- The two main areas of inferential statistics include:
  1. ***Estimating Parameters:*** This means taking statistics from your sample data and using it to say something about a population parameter (eg. Population mean)
  2. ***Hypothesis Testing:*** This is where you use sample data to answer research questions (eg. You might be interested in knowing if class attendance improve exam performance)
- The analyses involves two or more variable (Bivariate and Multivariate)

BIVARIATE ANALYSIS

- It is one of the simplest form of quantitative analysis
- Analysis of two variable at a time in order to determine whether or not the two variables are related. (X, Y)
- It is helpful for testing simple hypothesis of association
- It is a means of searching for evidence that variation in one variable coincides with variation in another variable.
- It can be used to help determine to what extent it becomes easier to predict a value for one variable (dependent) if we know the value of the other variable (independent)
- It is however dependent on the nature of the two variables being analysed
- There are three types of bivariate analysis: ***Correlation, Covariance and Linear Regression***

MULTIVARIATE ANALYSIS

- Simultaneous analysis of three or more variables
- There are three main contexts within which it can be applied
  1. ***Spurious (False) Relationship***: There appears to be a relationship between variables but the relationship is not real. (each variable is related to a third variable)
  2. ***Intervening Variable:*** There appears to be a relationship because of a third variable (hence the relationship is not direct)
  3. ***Moderation Variable:*** There appears to be a relationship but this relationship is moderated by another variable. (eg Fitness levels and time spent in the gym)

MORE DEFINITIONS

TERM	DEFINITION
Normal Curve	
Confidence Interval	These are two numbers which represent the higher and lower limits of a statistic, coefficient and parameter. It measures the certainty or uncertainty of the sampling method. They are used to estimate the range within which the value of the population parameter lies
Standard Error	It is the estimated standard deviation. It is a test for significance and can determine incorrect use of a statistical method .
Sampling Error	The probability
Confidence level	The percentage or probability of correctly rejecting the null hypothesis
Statistical Test	Mathematical tools for analysing quantitative data generated in a research study
Margin of Error	A measure for the precision of results. It tells you to what degree your results differ from real-world results. A smaller margin of error suggests that a survey's results are more precise
Variance	A statistical measurement of the spread between numbers in a data set
Statistical significance	A measure of how confident we are that our findings can be generalized to the lager population

TESTS FOR STATISTICAL SIGNIFICANCE

- Tests for statistical significance are used if the researcher is interested in comparing two or more groups, and determining whether the difference is statistically significant or is as a result of a sampling error.
- Can only be used in relation to samples that have been drawn using probability sampling
- The tests include:
    1. ***Two sample T-tests***
    2. ***Paired T- test***
    3. ***ANOVA (Analysis of Variance)***
    4. ***Correlation***

- Demonstrate an understanding of quantitative research methods;
- Demonstrate an understanding of the relationships between theory, conceptualisation, method, hypothesis or research questions, evidence and measurement;
- Differentiate between different types of surveys and demonstrate how they are used in research (business, social, or economic)
- Demonstrate knowledge and understanding of the derivation and attributes of survey data, including levels of measurement;
- Understand the role of sampling in survey research and the underlying theory that enables generalisation from random samples;
- Exhibit awareness of key issues related to objectivity, generalisability and utility of socio-economic research;
- Explain the key issues of data collection and analysis and communicate ideas in a succinct and clear manner;
- Describe different sample designs and how these can be applied in a practical context;
- Perform basic techniques for exploratory data analysis using SPSS;
- To interpret the output of secondary analysis accurately and critically;
- Use and interpret descriptive and inferential statistics for quantitative data; and
- Draw, verify and evaluate the quality of conclusions and produce an appropriate report.