RESEARCH METHODOLOGY

IMAYAVENDAN.V, B.E., MBA(IB).,

Mail ID: businessstudies@imayavendan.com

www.imayavendan.com

www.imayavendan.com

RESEARCH

- Organized set of activities

- Study and develop a model/procedure

- Find the results of a realistic problem

- Supported by literature and data

- Optimize the objectives

- Make recommendations/interferences

RESEARCH METHODOLOGY

- System of models, procedures, techniques
- To find results of a research problem

TYPES OF RESEARCH

- Exploratory
  - Literature Survey
  - Experience Survey
- Conclusive
  - Descriptive
  - Experimental
- Modelling
  - Symbolic
  - Mathematical
  - Simulation
- Algorithmic

EXPLORATORY RESEARCH

- Initial research

- Analyses data and explores possibility of obtaining relationships between variables

- Unaware of the end applications

Literature Survey

- Study involving a collection of literatures in a selected area
- Researcher has limited experience
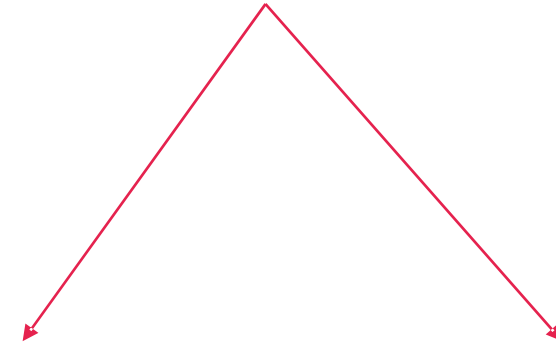- Critical examination and comparison is made.

Experience Survey

- Survey of experiences of experts/specialists in a specific field.
- Acts as database for future research

CONCLUSIVE RESEARCH

- Tests the hypotheses of a research problem formulated by exploratory research.

- Draws definite conclusions for implementation.

Descriptive Research

- Carried out with specific objectives
- Results in definite conclusions
- Conclusions are descriptive

Experimental Research

- Used to study the effect of a set of factors on the response variable of a system of study.
- Conducted in a controlled environment
- Analyzed using ANOVA

RESEARCH PROCESS

- Problem definition
- Objectives of the research
- Research design
- Data collection
- Data analysis
- Interpretation of the results
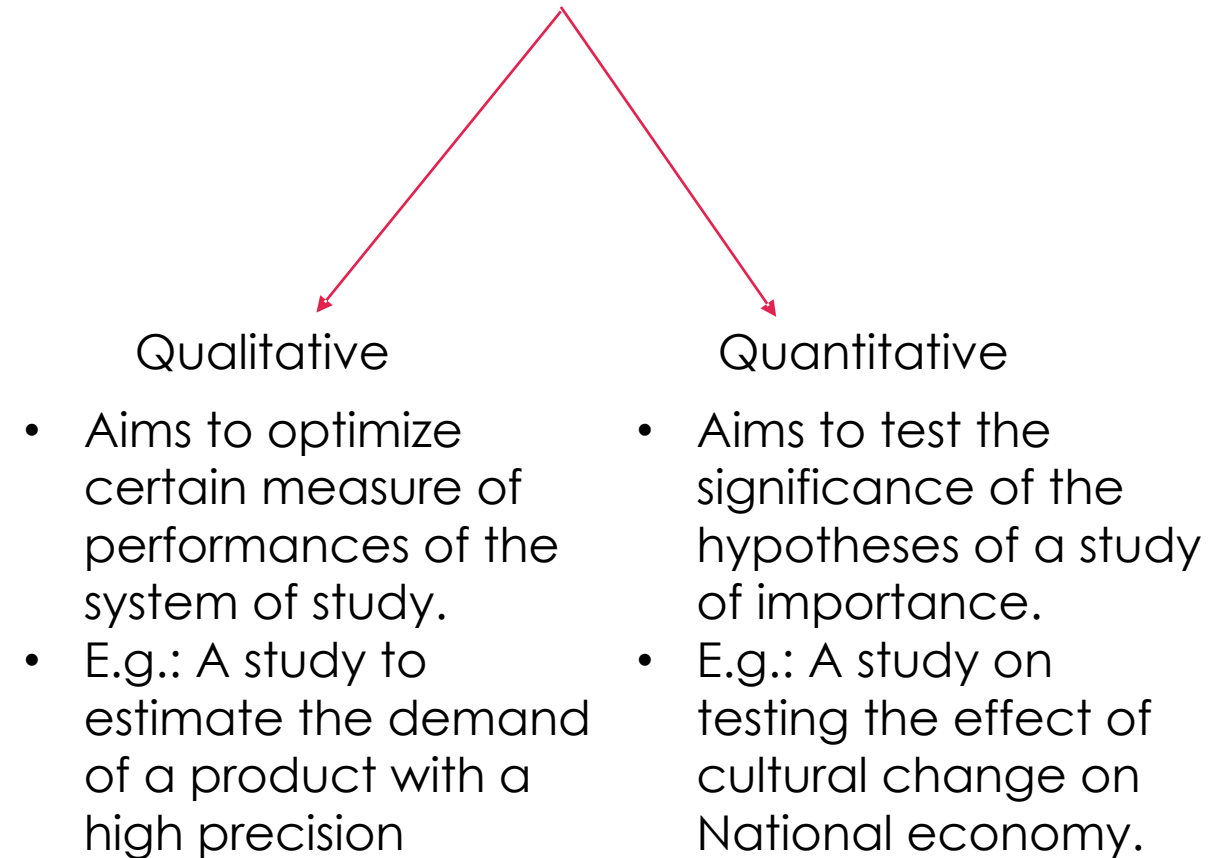- Validation of the results

PROBLEM DEFINITION

A research problem should be

- Identified and defined without ambiguity

- Precise and clear

- Additional value

OBJECTIVES OF THE RESEARCH

- In the process of identifying the objectives, the following are to be finalized
  - Research Questions – problems not resolved till date.
    - E.g.: Purpose, Place, Present state, means etc.,
  - Hypotheses – Assumption about a population of the study
    - E.g.: Age has no influence over achieving sales target of a salesperson
  - Boundary of the study – Properly defined

Qualitative
- Aims to optimize certain measure of performances of the system of study.
- E.g.: A study to estimate the demand of a product with a high precision

Quantitative
- Aims to test the significance of the hypotheses of a study of importance.
- E.g.: A study on testing the effect of cultural change on National economy.

RESEARCH DESIGN

- Once the project is identified and defined, the next stage is to design the research.

- It provides a complete guidelines for data collection.

- The essence of research design include:

  - Selection of research approach
  - Design of Sampling Plan
  - Design of Experiment
  - Design of Questionnaire

SELECTION OF RESEARCH APPROACH

- Two classifications: Exploratory, Conclusive for survey based

- Modelling research – Used to find the best result which should possess:
    - Objective function and
    - Set of constraints

- Algorithmic: Optimal or Near optimal solution

Based on the type of reality, the fittest approach is to be selected.

DESIGN OF SAMPLING PLAN

- Mechanism by which the sample units of a study are selected from sampling frame of the population.

- Affects the cost and time to conduct the study

SAMPLING – BASIC CONCEPTS

- Population – Entire spectrum of a system of interest

- Sampling: Process of selecting a subset of randomized number of members of the population of a study and collecting data about their attributes.

- Limited members of population selected – Sampling Units

- Advantages of sampling:
  - Less time taken
  - Less cost incurred
  - Physical impossibility of complete enumeration
  - More accuracy

SAMPLING METHODS

- Probabilistic:
  - Each unit of the population has a probability of being selected.
  - More rigorous and free from biases
- Non- Probabilistic:
  - Instances of certain units having zero probability of selection.
  - Presence of factors of judgement, biases and convenience of interviewers
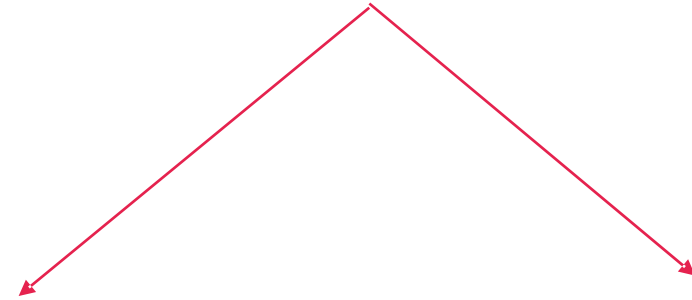
Probabilistic

1. Simple random sampling
2. Systematic sampling
3. Stratified random sampling
4. Cluster sampling
5. Multi-stage sampling

Non - probabilistic

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

PROBABILITY SAMPLING – SIMPLE RANDOM SAMPLING

- There are two ways of performing simple random sampling:
  - With Replacement: Each unit of the population has the equal probability of being selected as an unit of the sample,

    Probability of a unit being selected = 1/N

    Where N is the total number of units

  - Without Replacement: In this method of sampling, each unit of the population has a varying probability of being selected as an unit of the sample,

    Probability of a unit, n, being selected   = 1/ N-(n-1)

    Where N is the total number of units

PROBABILITY SAMPLING – SYSTEMATIC SAMPLING

- This is a special kind of random sampling in which the selection of the first unit of the sample from the population is based on randomization. The remaining units of the sample are selected from the population at a fixed interval of n, where n is the sample size.

Sampling Interval Width = N/n

- The advantage of this sampling method is that, once the first unit of the sample is randomly determined all the sampling units can be easily obtained. But, the sampling units starting from 2 to n are dependent on the first sampling unit which lacks complete randomization of the selection of the sampling units from the population.

PROBABILITY SAMPLING – STRATIFIED SAMPLING

- Stratified sampling is an improvised sampling over simple random sampling and systematic sampling. This sampling will have more statistical efficiency. In this sampling method, the population is divided into a specified set of strata such that the members within each stratum have similar attributes but the members between strata have dissimilar attributes. This means that each stratum is homogeneous when compared to the population.

- If we use the same proportion of sampling units from each stratum, then it is known as proportional stratified sampling; otherwise, the sampling is known as disproportional stratified sampling.

PROBABILITY SAMPLING – CLUSTER SAMPLING

- Cluster sampling is a sampling technique in which the population is divided into different clusters such that the members within each cluster are dissimilar (heterogeneous) in terms of their attributes, but different clusters are similar to each other. This leads to the inference that each cluster can be treated as a small population which possess all the attributes of the population. Hence, in cluster sampling, any one of the clusters is randomly selected and all the units of that cluster are selected (sampled) to arrive at inference about the population.

PROBABILITY SAMPLING – MULTI STAGE SAMPLING

- In a large scale survey covering the entire nation/subcontinent, the size of the sampling frame will be too large which leads to more time and cost of the study. In such study, multi-stage sampling technique helps designing a smaller sampling frame which will make the study practicable in terms of cost and time.

- The multi-stage sampling employs more than one stage to sample the population depending upon the reality. The combination of the types of sampling techniques to be used in the specified number of stages is unique to the reality.

NON-PROBABILITY SAMPLING – CONVENIENCE SAMPLING

- Convenience sampling This is a non-probability sampling method in which the interviewers will decide the choice of sampling units based on their convenience.

- In most of the situations, the following may be true:
  - The sampling units may be distributed sparsely.
  - Many respondents will refuse to fill the questionnaires.
  - Some respondents will not cooperate in filling the questionnaires.
  - Some of the interviewers may not be serious in selecting the sampling units as per the assumed sampling plan.

- The sampling units for this type of sampling are selected from a telephone directory, newspaper subscribers list, departmental stores, etc.

NON-PROBABILITY SAMPLING – JUDGEMENT SAMPLING

- The judgement sampling is a non-probability sampling in which the sampling units are selected on the advice of some expert or by the intuition/opinion of the researcher himself. In the first case, an expert who is familiar with the sampling frame guides the researcher in selecting the sampling units from the sampling frame. In the second case, the researcher applies his/her intuitive judgement and previous experience in selecting the sampling units from the sampling frame.

- In this sampling method, there is more chance of personal biases. If the exercise is done more seriously, it would lead to better results. This sampling is also called as purposive sampling because the sampling units of the sample are identified from the population selectively which prevents the inclusion of other sampling units in the sample. This sampling technique is applied for sampling the population related to rare events in which the members of the population are not equally qualified to become the members of the sampling frame and hence the members of the sample.

NON-PROBABILITY SAMPLING – QUOTA SAMPLING

- The quota sampling is a non-probability sampling in which the population is classified into a number of groups based on some criterion, say age of the members of population, viz., old age, middle age and young age.

- Later, while selecting the required number of sampling units from each category, one can use any one of the other non-probability sampling methods, viz., convenience sampling or judgement sampling.

- Though this method comes under non-probability sampling methods, in the first phase, certain amount of proportionate selections of sampling units from different strata of the population are made which makes the sample as a representative sub-population of the main population.

NON-PROBABILITY SAMPLING – SNOWBALL SAMPLING

- The snowball sampling is a restrictive multi-stage sampling in which initially certain number of sampling units (respondents) are randomly selected. Later, additional sampling units are selected based on referral process.

- This means that the initially selected respondents provide addresses of additional respondents for the interviewers. Initial respondents may be randomly selected, for example, from the information contained in telephone directories. Later, additional respondents can be included in the sample based on the references made by those initial respondents.

- This is an inexpensive and convenient non-probability sampling method which suits the situations where the development of sampling frame is a difficult and time-consuming task.

DATA COLLECTION

- Data can be classified into primary data and secondary data.

- The data which are collected from the field under the control and supervision of an investigator is known as primary data. This type of data are generally afresh and collected for the first time.

  - **Eg: Data of a study to determine the morale of the employees in a company.**

- If data are collected from journals, magazines, government publications, annual reports of companies, etc., then such data are called as secondary data. In each of these sources of data, the process of data collection has already been done by the respective organization/agency.

METHODS OF COLLECTING PRIMARY DATA

- The different methods which are used for primary data collection are:
    - Observation method,
    - Personal interview,
    - Telephone interview and
    - Mail survey.

OBSERVATION METHOD

- In the observation method, the investigator will collect data through personal observations.

- Consider an example in which an investigator collects the data about the organizational climate in an organizations through direct observation. In this study, the investigator will speak to the employees at different levels of the organization, observe the behaviour of the employees to assess the organizational climate.

- Generally, observation method of data collection deals with the recording of the behaviour of respondents/sampling units.

- In this method, the investigator will observe the behaviour of the respondents in disguise.

- This method helps capture the behaviour of customers directly. But, it is a time consuming and costly exercise. Also, it suffers from personnel biases of investigators which will distort the findings.

PERSONAL INTERVIEW

- Personal interview is a survey method of data collection which employs a questionnaire.
- The components of the personal interview are
  - The researcher,
  - The interviewer,
  - Interviewee and the
  - Interview environment.
- Under the guidance of the researcher several interviewers will be sent with questionnaires to meet the respondents/interviewees of the survey for seeking responses to the questions in the questionnaire under the defined interview environment.

TELEPHONIC INTERVIEW

- Telephone interview is considered to be a cost effective and dominant data collection method because of the following reasons:

  - The travel time of interviewers is totally eliminated
  - The cost of travel of interviewers is also eliminated
  - Greater possibility of reaching the customers all over the geography
  - Total time of conducting the interview of the sample is least when compared to other methods
  - Greater possibility of random selection of respondents among the population having telephone connection
  - There is a greater probability of reaching the respondent unlike the nonresponse problems of personal interview.

TELEPHONIC INTERVIEW - DRAWBACKS

Though the telephone interview has many advantages, it has the following drawbacks:

- Impossible to employ visual aids
- Exclusion of population who are not having telephone connection
- Interviewee may discontinue conversation in the mid-way unlike the personnel interview in which there is greater probability of completing the interview fully because of the rapport developed between the interviewer and the interviewee through personal contact.

MAIL SURVEY

- Mail survey is a data collection method in which questionnaires are mailed to potential respondents who in turn fill and return them at their convenience.
- This method has the following advantages:
  - Less cost of data collection
  - Less time of data collection
  - Wider coverage of population
  - Better accuracy of data
  - Absence of interviewer's bias

MAIL SURVEY -DRAWBACKS

Drawbacks of Mail Survey include,

- The identity of the respondents is not known to the interviewers.
- The questionnaire may be filled in by the assistant/family members of the respondent.
- Some respondents may not return filled-in questionnaires.
- There may be delay from the part of the respondents in returning the filled-in questionnaires.

QUESTIONNAIRE DESIGN

- A questionnaire consists of a set of well-formulated questions to probe and obtain responses from respondents.

- The following steps are considered for effective designing of questionnaire:
  - Identifying Research Issues
  - Formulating of questions and format
  - Deciding question wording
  - Appropriate sequence
  - Pre-testing of questionnaire
  - Review for improvements

IDENTIFICATION OF RESEARCH ISSUES

- If the researcher is not clear about the kind of measurements to be made, then the questionnaire will suffer from ambiguous questions.

- So the answers to these questions will have no value for the study. This amounts to garbage-in and garbage-out.

- So, the researcher should define different hypotheses for various issues of the research which in turn help him/her to decide about the measurements to be made.

FORMULATION OF QUESTIONS AND FORMAT

- If the types of measurements to be made are clearly known, then the researcher can formulate necessary questions relating to each measurement.

- The format of a question can be of three types:
  - Open-ended questions,
  - Close-ended questions and
  - Structured questions.

OPEN ENDED QUESTIONS

- In this type of question, the interviewer writes the answer of the respondent verbatim. In this type, there is no pre-coded classification of answers to the question.

- The open-ended questions may bring unconventional answers which are not known to the researcher.

- This is alternatively known as unstructured question. Though it has certain benefits, the answers to unstructured questions may not be realistic as expected if the respondents are not articulate and educated.

- Eg: What are your reasons for working in this company?

CLOSE – ENDED QUESTIONS

- Close-ended questions are structured questions. It has two formats:
  - Questions with multiple responses out of which the respondents have to select one or more choices.
    - Eg. What is your education qualification?
      A) 10th/12th **B)** Graduate **C)** Post-Graduate **D)** Doctorate
  - Questions with rating scale with discrete responses or continuous range.
    - Eg. You like the working environment of the company
      A) Strongly Agree **B)** Agree **C)** Neutral **D)** Disagree **E)** Strongly Disagree
- An extreme form of close-ended question has only two responses. The respondent has to select one of these two responses which is demonstrated with reference to a study on consumer survey. This question has dichotomous responses.
  - Eg. Are you satisfied with the working environment of the company?
    A) Yes   **B)** No

DECIDING QUESTION WORDING

- The wording of each question should fully convey its meaning to the respondents.
- The degree of the level of understanding varies from person to person.
- If questions are not easily understood by the respondents, they will respond with random answers to questions, which may distort the research findings.

APPROPRIATE SEQUENCE OF QUESTIONS

- After formulating necessary questions, they should be arranged in an appropriate sequence such that respondents are able to have better grasp of them and provide accurate responses.

- A preferred sequence of questions is listed here:
  - First few questions relating to personal details of the respondents.
  - Introductory questions to establish a rapport with the respondents.
  - Simple and direct questions to build an opinion that the questionnaire is simple to answer.
  - Focussed questions relating to objectives and issues of the research.
  - Main body of the questionnaire should contain questions to obtain needed information.

PRE-TESTING OF QUESTIONNAIRE

- Once the questionnaire is ready, it should be pre-tested through a pilot survey involving the respondents in the proposed sampling frame.

- This exercise is mainly intended to test the degree of understanding the meaning of the questions, difficulty in understanding the questions by the respondents if the meaning of the questions are conveyed correctly, to check the relevance of the questions, to ascertain the interest of the respondents, etc.

REVIEW OF QUESTIONNAIRE FOR IMPROVEMENTS

- The purpose of pre-testing of questionnaire is to obtain information to improve its content, format and sequence. Based on the information, the questionnaire should be revised in its format, content and sequence for final use in the survey.

SCALING – TYPES OF SCALES

- The types of scales which are used in Behavioural Sciences are:

  - Nominal scale,
  - Ordinal scale,
  - Interval scale and
  - Ratio scale.

NOMINAL SCALE

- Nominal scale This scale has no order, distance or origin. The numbers used in this scale are used as tags or symbols to recognize or identify objects.

- This scale can also be used to represent yes/no, like/dislike, male/female type of responses.

- Only rudimentary mathematical operations can be performed on such code numbers. At the maximum, one can count the numbers and make any inference accordingly.

- Further, contingency table summarizing the membership of persons in two categories can be formed.

ORDINAL SCALE

- Ordinal scale It is known as ranking scale which possesses the property of order.

- Using this scale, one can rank objects based on certain characteristic or attribute of the objects.

- For example, ranking the products of a company according to the satisfaction of customers

- An ordinal scale has all the information of a nominal scale for the same situation which means that equivalent entities will have the same rank.

- The data using ordinal scale can be used to compute only the positional measures like, mode and median.

INTERVAL SCALE

- This scale has the properties of order and distance.

- Using this scale, one can judge the difference between two objects.

- It has arbitrary zero point and constant unit of measurement.

- The numbers on the scale are placed at equidistance.

- The mathematical operations, addition, subtraction and calculation of mean are possible with any interval scaled data.

RATIO SCALE

- Ratio scale has a unique zero point. Such scales are used in physical sciences to measure length, breadth, weight, volume, etc.

- Equal ratio among the scale values will correspond to equal ratio among the entities, which are being measured.

- This means that a positive multiplicative constant will help converting feet into into inches, kilogram into grams, kilogram into pounds, etc.

- This scale has the properties of order and distance and origin.

- All statistical operations can be performed on any ratio scaled data.

SCALING RESPONDENTS

There are two methods of scaling respondents:

- Summated scale (Likert scale) and
- Q-sort scale.

SUMMATED SCALE

- The summated scale was proposed by Rensis Likert.
- In this scale, a set of items is proposed with respect to a particular attitudinal object, like the role of educational institutions in employment market.
- In this case, the researcher may be interested to scale the characteristic, like public's attitude on the above-mentioned attitudinal object.
- The attitudinal object will have a set of items which are described in the form of statements/phrases. The items can be classified into favourable items and unfavourable items.
- Each favourable item is given a numerical weight like, +2, +1, 0, – 1, – 2.
- Each unfavourable item will have a numerical weight in the reverse order as – 2, – 1, 0, +1, +2.
- Each respondent is asked to assign/tick a value for each item. In the summated model, the total score of a respondent is obtained by adding the scores on individual items.

Q-SORT SCALE

- The Q-sort scale was developed by William Stephenson.

- This scale has some commonality with the summated scale.

- In this scale, respondents are asked to sort a given set of statements (items) and classify them into a predetermined number of categories (piles), usually 11.

- The number of statements (items) to be included in each category is also defined in advance.

TESTS OF HYPOTHESES

- Hypothesis is an assumption about a population.

- There are two types of hypothesis:
    - Null hypothesis - It is an assumption (preferred assumption) about a population. It is denoted by H0.
    - Alternate hypothesis - If the null hypothesis (H0) is rejected, then the opposite of the null hypothesis must be true. The hypothesis representing the opposite of the null hypothesis is called alternate hypothesis. It is denoted by H1.

SOME CONCEPTS

- Significance level (a) The significance level is the probability with which the null hypothesis (H0) will be rejected due to sampling error, though H0 is true.

- Based on the explanation given above, the error can be classified into Type-I error (producer's risk) and Type-II error (consumer's risk).
  - Type-I error It is the probability of rejecting the null hypothesis when it is true. It is denoted by a. In quality control, it is termed as producer's risk, because it is the probability of rejecting a good lot.
  - Type-II error It is the probability of accepting the null hypothesis when it is false. It is denoted by b. In quality control, it is termed as consumer's risk, because it is the probability of accepting a bad lot.

WHEN TO USE A PARTICULAR STATISTICAL TEST - UNIVARIATE DESCRIPTIVE

- ***Central Tendency:*** A summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution.

- The measures of central tendencies are
  - Mean
  - Median
  - Mode

MEAN, MEDIAN, MODE

- **Mean**
  - The mathematical average
  - The formula is $\sum X/N$
  - Eg: mean age = age of person one + age of person two + age of person three, etc./number of people

- **Median**
  - The center value
  - The formula is $(N+1)/2$
  - Eg: 6 people with ages 21, 22, 24, 23, 19, 21
  - Line them up in order form lowest to highest 19, 21, 21, 22, 23, 24 and take the centre value -> median =21.5

- **Mode**
  - The most commonly occurring value
  - Eg: 6 people with ages 21, 22, 21, 23, 19, 21 - mode = 21

STANDARD DEVIATION AND VARIANCE

- **Variance**
  - A measure of how spread out a distribution is
  - It is computed as the average squared deviation of each number from its mean

- **Standard Deviation**
  - How much scores deviate from the mean
  - It is the square root of the variance
  - It is the most commonly used measure of spread

BI- AND MULTIVARIATE INFERENTIAL STATISTICAL TESTS - DIFFERENCES OF GROUPS

- **Chi Square**
  - Compares observed frequencies to expected frequencies
  - Eg: Is the distribution of sex and voting behaviour due to chance or is there a difference between the sexes on voting behaviour?

- *t*-Test
  - Looks at differences between two groups on some variable of interest
  - The IV must have only two groups (male/female, undergrad/grad)
  - Eg: Do males and females differ in the amount of hours they spend shopping in a given month?

- **ANOVA**
  - Tests the significance of group differences between two or more groups
  - The IV has two or more categories
  - Only determines that there is a difference between groups, but doesn't tell which is different
  - Eg: Do SAT scores differ for low-, middle-, and high-income students?

BI- AND MULTIVARIATE INFERENTIAL STATISTICAL TESTS – DIFFERENCES OF GROUPS

- **ANCOVA**
  - Same as ANOVA, but adds control of one or more covariates that may influence the DV
  - Eg: Do SAT scores differ for low-, middle-, and high-income students after controlling for single/dual parenting?

- **MANOVA**
  - Same as ANOVA, but you can study two or more related DVs while controlling for the correlation between the DV
  - If the DVs are not correlated, then separate ANOVAs are appropriate
  - Eg: Does ethnicity affect reading achievement, math achievement, and overall scholastic achievement among 6th graders?

- **MANCOVA**
  - Same as MANOVA, but adds control of one or more covariates that may influence the DV
  - Eg: Does ethnicity affect reading achievement, math achievement, and overall scholastic achievement among 6 graders th after controlling for social class?

BI- AND MULTIVARIATE INFERENTIAL STATISTICAL TESTS - RELATIONSHIPS

- **Correlation**
  - Used with two variables to determine a relationship/association
  - Do two variables covary?
  - Does not distinguish between independent and dependent variables
  - Eg: Amount of damage to a house on fire and number of firefighters at the fire.

- **Regression**
  - Used with an independent variable and a dependent variable
  - Used for prediction
  - Eg: Independent Variable: Drug use, Dependent Variable: Suicidal tendencies

WHEN TO USE WHAT TEST

Statistical Analyses	Independent Variables		Dependent Variables		Control Variables	Question Answered by the Statistic
	Number of Ivs	Data Type	Number of Dvs	Data Type		
Chi- Square	1	Categorical	1	Categorical	0	Do differences exist between groups?
t-Test	1	Dichotomous	1	Continuous	0	Do differences exist between 2 groups on one DV?
ANOVA	1 +	Categorical	1	Continuous	0	Do differences exist between 2 or more groups on one DV?
ANCOVA	1 +	Categorical	1	Continuous	1+	Do differences exist between 2 or more groups after controlling for CVs on one DV?
MANOVA	2 +	Categorical	2+	Continuous	0	Do differences exist between 2 or more groups on multiple DVs?
MANCOVA	1+	Categorical	2+	Continuous	1+	Do differences exist between 2 or more groups after controlling for CVs on multiple Dvs?
Correlation	1	Dichotomous or Continuous	1	Continuous	0	How strongly and in what direction (i.e., +, -) are the IV and DV related?
Multiple Regression	2+	Dichotomous or Continuous	1	Continuous	0	How much variance in the DV is accounted for by linear combination of the IVs? Also, how strongly related to the DV is the beta coefficient for each IV?
Path Analysis	2+	Continuous	1+	Continuous	0	What are the direct and indirect effects of predictor variables on the DV?
Logistic Regression	1+	Categorical or Continuous	1	Dichotomous	0	What is the odds probability of the DV occurring as the values of the IVs change?

THANK YOU!