Research Methodology

Mrs.C.Sathya lakshmi M.Sc., M.Phil,

Vice Principal, Holy Cross Home Science College

INTRODUCTION

- The research is defined as a logical, scientific and systematic search for new and useful information on a specific topic.
- It refers to an investigation of finding solution to specific and social problems, through systematic analysis.

- Research methodology simply refers to the practical "how" of any given piece of research.

- More specifically, it's about **how** a researcher **systematically designs a study** to ensure valid and reliable results that address the research aims and objectives.

OBJECTIVES

- The main aim of research is find out the truth which is hidden and which has not been discovered yet.
- To serve the society by solving social problems.
- To overcome or solve the problems occuring in our everyday life.

- To get research degree for better employment or promotion.
- To develop new scientific method, concept and theories to solve the scientific and non-scientific problems.
- To discover new things and tests the existing facts.

IMPORTANCE RESEARCH

- Research provides basis for many government policies.
- E.g.: preparation of budget, solution for social problem etc.
- It is an important tool in all pharmaceutical and other business industries for higher growth.

- Isolation, identification and characterization of new living organisms, materials.
- It gives guidelines for solving problems.
- The process such as human genome project, gene therapy, superconductivity have been discovered through research.
- It helps to study the applications of existing theories and concepts.

- Research gives the solution for problems of large and small scale industries.
- Helps to learn how to use libraries and other information resources.
- They help to find answers to social problems.
- Research is one of the important parameter for international university ranking.

CHARACTERISTICS OF RESEARCH

- A systematic approach must be followed for accurate data. Rules and procedures are an integral part of the process that set the objective.

- Researchers need to practice ethics and a code of conduct while making observations or drawing conclusions.

- Research is based on logical reasoning and involves both inductive and deductive methods.

- The data or knowledge that is derived is in real time from actual observations in natural settings.
- There is an in-depth analysis of all data collected so that there are no anomalies associated with it.
- Research creates a path for generating new questions. Existing data helps create more opportunities for research.

- Research is analytical in nature. It makes use of all the available data so that there is no ambiguity in inference.

- Accuracy is one of the most important aspects of research. The information that is obtained should be accurate and true to its nature. For example, laboratories provide a controlled environment to collect data.

CHARACTERISTICS OF RESEARCH

- The following are the characteristics of research :-

1) Research is a creative process and includes investigation or inquiry, discovery or invention or experimentation to find out the truth with evidence contained in the product.
2) Research is systematic.
3) Research is logical.
4) Research is purposeful.
5) Research is accurate.
6) A research into any problem is not the end of it but it is the beginning of new search which always raises a new question.
7) It gathers new knowledge and sometimes it replaces the old knowledge.
8) Knowledge enables man to understand, explain, control, and hence cope with any situation. This means the human mind will always search for knowledge. It is universal and true yesterday, today and tomorrow.
9) Research collects data.
10) Research involves hypothesis or testing of hypothesis.

Critical Action Research (CAR)

Developed by Paulo Freire, who wanted to use education to free the disadvantaged from oppression	Similar to PAR; "critical" signifies addition of ideological element to research -- i.e., orientational research	Attempts to take an emancipatory stance, striving for immediate social change and increasing social justice	Focuses on reduction of inequality of income and wealth and/or reduction of discrimination	Often includes attempt of "consciousness raising" of individuals and groups that have minimal power of society

Characteristics of a Good Research

- **Clearly defined purpose**
- **Detailed process**
  - Providing procedures used to conduct the study
- **Well planned research design**
  - Results should be as objective as possible
- **Maintain high ethical standards**
  - Being responsible and concerned for the welfare of the participants

STEPS FOR MAKING A GOOD RESEARCH

- Raising a Question.
- Suggest Hypothesis.
- Literature Review.
- Literature Evaluation.
- Acquire Data.
- Data Analysis.
- Data Interpretation.
- Hypothesis Support.

CHARACTERISTICS OF GOOD DESIGN

- Understanding the processes by which nurses make diagnoses.
- There are many research questions of interest to nurses for which highly structured designs are unsuitable.
- TIP: Although techniques of research control are mechanisms for controlling bias, there are situations in which too much control can introduce bias.
- For example, if researchers tightly control the ways in which key study variables can manifest themselves, it is possible that the true nature of those variables will be obscured.
- When the key concepts are phenomena that are poorly understood or dimensions of which have not been clarified, then a design that allows some flexibility is better suited to the study aims.

ADVANTAGES AND DISADVANTAGES

Advantages	**Disadvantages**
Researcher can have control over variables.	Can produce artificial results.
Humans perform experiments anyway.	Results may only apply to one situation and may be difficult to replicate.
Gain insight into methods of instruction.	Subject to human error.
Use to determine what is best for population.	Human response can be difficult to measure.

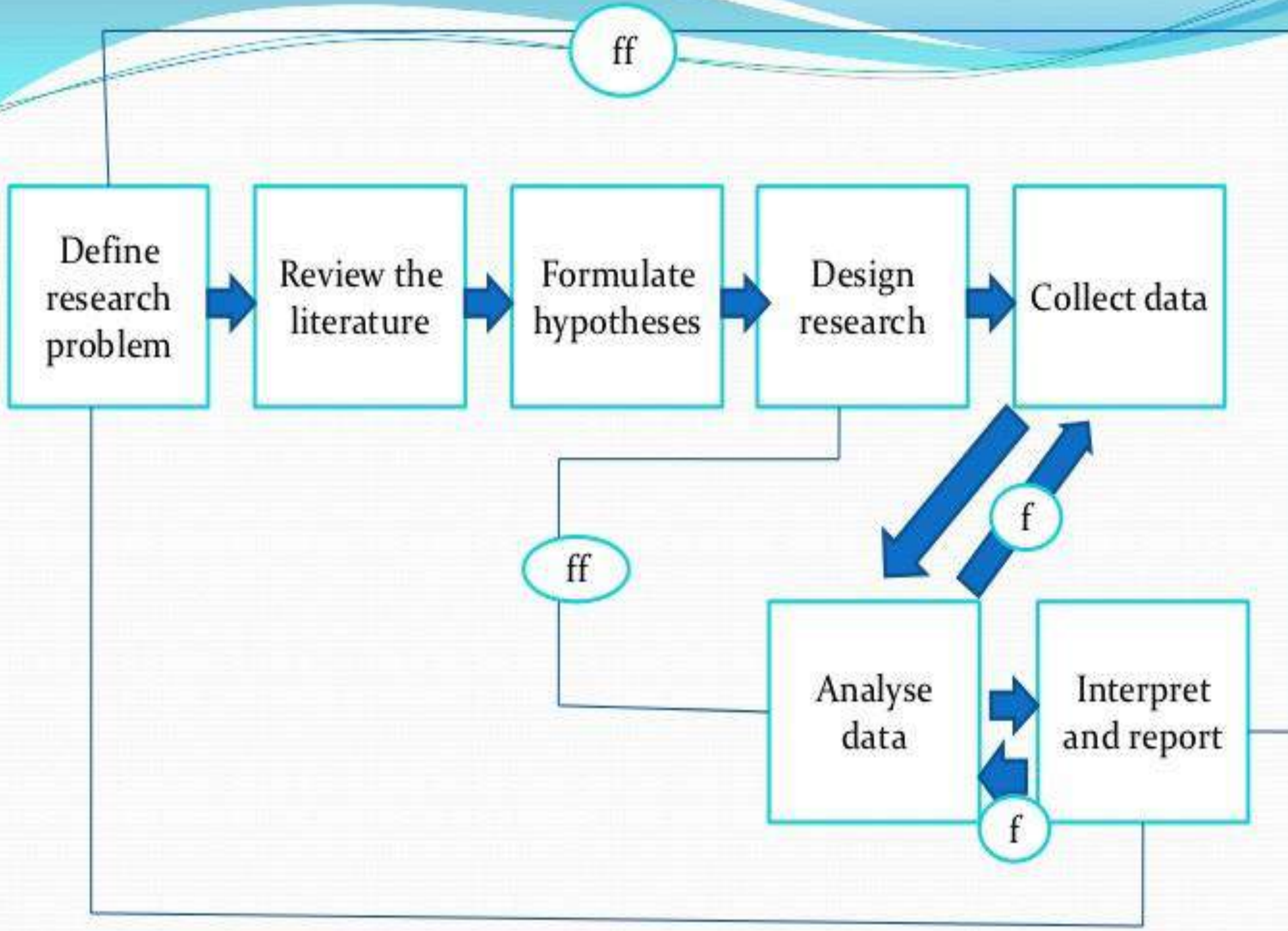
STEPS IN RESEARCH PROCESS

DEFINE-RESEARCH

The process of gathering information for the purpose of initiating, modifying or terminating a particular investment or group of investments.

7 STEPS OF RESEARCH PROCESS

- Step One: Define research problem
- Step Two: Review of literature
- Step Three: Formulate hypotheses
- Step Four: Preparing the research design
- Step Five: Data collection
- Step Six: Data analysis
- Step Seven: Interpretation and report writing



Where f = feed back(helps in controlling the sub system
ff= feed forward(serves the vital function of providing criteria for evaluation

Step One: Define Research Problem

There are two types of research problem, viz., those

- relate to states of nature
- relationship between variables.

Essentially two steps are involved in define research problem, viz.,

- understanding the problem thoroughly and
- rephrasing the same into meaningful terms from an point of view.

Step Two: Review of Literature

- Once the problem is define, a brief summary of it should be written down. It is compulsory for a research worker writing a thesis for a Ph.D. degree to write a synopsis of topic and submit it to necessary committee or the research board for approval.

Step Three: Formulate Hypothesis

- Formulate hypothesis is tentative assumption made in order to draw out and test its logical or empirical consequences. Hypothesis should be very specific and limited to the piece of research in hand because it has to be tested.

The role of the hypothesis is to guide the researcher by delimiting the area of research and to keep him on the right track.

Step Four: Preparing the Research Design

- The function of research design is to provide for the collection of relevant evidence with minimal expenditure of effort, time and money.

- Research purpose may be grouped into four categories, viz., (1) Exploration, (2) Description,(3) Diagnosis, and (4) Experimentation.

Step Five: Data Collection

- Primary data can be collected through:

  By Observation

  Through personal interview

  Through telephone interview

  By mailing of questionnaries

  Through Schedules

Step Six: Data Analysis

The analysis of data requires a number of closely related operations such as establishment of categories.

This stage mainly include :

1. Coding
2. Editing
3. Tabulation

Step Seven: Interpretation and Report Writing

Researcher has to prepare the report of what has been done by him.

Writing of report includes:

1. the preliminary pages;
2. the main text, and
3. the end matter.

HYPOTHESIS

DEFNITIONS:

- Hypothesis is considered as an intelligent guess or prediction, that gives directional to the researcher to answer the research question.
- Hypothesis or Hypotheses are defined as the formal statement of the tentative or expected prediction or explanation of the relationship between two or more variables in a specified population

- A hypothesis is a formal tentative statement of the expected relationship between two or more variables under study.

- • A hypothesis helps to translate the research problem and objective into a clear explanation or prediction of the expected results or outcomes of the study.

CONTRIBUTIONS OF HYPOTHESIS

It provides clarity to the research problem and research objectives.

- It describes, explains or predicts the expected results or outcome of the research.
- It indicates the type of research design.
- It directs the research study process.

It identifies the population of the research study that is to be investigated or examined.

It facilitates data collection, data analysis and data interpretation

FUNCTIONS

- It enables an investigator to start his research work.
- It may lead to formulations of another hypothesis.
- It leads to interpret results drawing conclusions related to original purpose.

TYPES OF HYPOTHESIS

- Simple
- Complex
- Empirical
- Null
- Alternative
- Logical
- statistical

SIMPLE HYPOTHESIS

- Simple hypothesis is that one in which there exits relationship between two variables one is called independent variable or cause and the other is dependent variable or effect.
- Ex. Smoking leads to cancer
- The higher ratio of unemployment leads to crimes.

COMPLEX HYPOTHESIS

- Complex hypothesis is that one in which as relationship among variables exists.
- In this type dependent and independent variables are more than two
- Ex. Smoking and other drugs leads to cancer, tension, chest infections etc.
- The higher ration of unemployment poverty illiteracy leads to crimes like dacoit etc.

EMPIRICAL HYPOTHESIS

- **Empirical** which means it is based on evidence.
- In scientific method the word "**empirical**" refers to the use of working **hypothesis** that can be tested using observation and experiment.
- **Empirical** data is produced by experiment and observation.

QUESTION FORM OF HYPOTHESIS

- It Is the simplest form of empirical hypothesis.
- In simple case of investigation and research are adequately implemented by resuming a question.
- Ex. how is the ability of 9^{th} class students in learning moral values?

NULL HYPOTHESIS

- Null the hypothesis that there is no significant difference between specified populations, any observed difference being due to sampling or experimental error.
- It is denoted by H0

ALTERNATE HYPOTHESIS

- The **alternative hypothesis**, denoted by H_1 or H_a,
- Is the **hypothesis** that sample observations are influenced by some non-random cause.

DIRECTIONAL HYPOTHESIS

- Directional Hypothesis predicts the direction of the relationship between the independent and dependent variable.
- Example- High quality of nursing education will lead to high quality of nursing practice skills.
- Girls ability of learning moral science is better than boys.

STATISTICAL HYPOTHESIS

- A hypothesis which can be verified statistically called statistical hypothesis.

- The statement would be logical or illogical but if statistic verifies it, it will be statistical hypothesis..

NON DIRECTIONAL HYPOTHESIS

- Non -directional Hypothesis predicts the relationship between the independent variable and the dependent variable but does not specific the directional of the relationship.
- Example- teacher student relationship influence student's learning.
- There is no significant difference between 9^{th} class boys and girls abilities of learning moral values.

CASUAL HYPOTHESIS

- Causal Hypothesis predicts a cause and effects relationship or interaction between the independent variable and dependent variable.
- This hypothesis predicts the effect of the independent variable on the dependent variable

ASSOCIATIVE HYPOTHESIS

- Associative Hypothesis predicts an associative relationship between the independent variable and the dependent variable.
- When there is a change in any one of the variables, changes also occurs in the other variable

CHARACTERISTICS

- ❖A Hypothesis must be Capable of Verification.
- ❖A Hypothesis must be Related to the Existing Body of Knowledge.
- ❖A Hypothesis Needs to be Precise, Simple and Specific

Abstract of a Research

- An abstract is a summary of the research paper. It is often found at the beginning of a research work so that readers will immediately have an overview of the entire research. It quickly gives the reader the essence of the report.

Contents of an Abstract

- Research statement or statement of the problem
- Research setting, population or the participants(number, age, course and gender) and data gathering method
- Summary of the findings
- Conclusions
- Recommendations

Example 1:

This research aimed to determine the effects of electronic cigarette to the health of the users. This is important to a lot of smokers who are already addicted to smoking. This research will provide the information if the electronic cigarette is safer to use than the traditional way of smoking.

To gather the needed data, different references were used such as books, magazines, newspapers and the internet. An interview with ten respondents which includes students, vendors, salesmen and bystanders, was also done to ask their experiences in using the EC (electronic cigarette).

All information gathered show that the electronic cigarette poses more health risks on the part of the users because of its chemical components. The users are more prone to diseases such as lung cancer and emphysema.

It is concluded that electronic cigarette can never be a good replacement for the tobacco cigarette.

Thus, it is recommended that smokers should stop using the EC but instead use the tobacco with moderation.

Example 2:

This research sought to find information that provides the causes and effects of compulsive hoarding to the students, to identify the common things or objects they hoard and what satisfaction do they get from hoarding these things.

The data were obtained from different source materials such as books, magazines, pamphlets and from different websites. Other data were gathered from the interviewees consisting of students, faculty members, dormitory mates and relatives who are known to be compulsive hoarders.

Information gathered show that said respondents are merely collectors and not hoarders. They simply love collecting items as a hobby. They find enjoyment when they see them properly arranged and on display to the eyes delight.

Therefore, collectors should not be mistaken to be hoarders. Also, it is not an abnormality when there are people who collect a miscellaneous of things.

Writing Summary, Conclusions and Recommendations

- The summary is the condensed version of the main points highlighted in the research. Hence, the verbs in the past tense are required. It includes a synopsis of the following:

  1. Objectives of the study
  2. The subjects who served as respondents
  3. The research instrument such as the interview

Summary

- The research aimed to determine the effects of electronic cigarette to the health of the users.

  Specifically, the study sought to answer the following problems:

  1. What are the general features of the electronic cigarette that create the impression that it is safer to use it than the tobacco cigarette?

  2. What are its components that pose health risks to smokers?

  3. What is the general observation of those who used the electronic cigarette?

  4. What is the personal assessment of the users in terms of the satisfaction they get from using the EC than using the TC?

The respondents of the research consisted of three groups: six students, a vendor and a salesman and two bystanders/shoppers.

The interview served as the main instrument in validating the claim that electronic cigarette poses health risks to smokers.

Findings/Conclusions

- From the analysis of data, the findings were as follows:

1. The general features of the electronic cigarette as described by the salesmen and vendors were not a guarantee to keep the smokers free from getting lung cancer or emphysema.

2. The components of the electronic cigarette

3. The general observation of the users when they started using the electronic cigarette includes the difficulty in breathing, lost appetite…

4. Their general impression is that tobacco cigarette is better than the electronic cigarette.

Recommendations

From the findings and conclusions of the research, the following recommendations are hereby given:

- ***Conduct*** a thorough research on the potential risks of electronic cigarette to the new users or smokers.

- ***Consolidate*** all findings and if found really dangerous to people's health, ask the government to ban the selling of electronic cigarette.

Sampling Design

Types of Sampling

* LEARNING OBJECTIVES

* Learn the reasons for sampling

* Develop an understanding about different sampling methods

* Distinguish between probability & non probability sampling

* Discuss the relative advantages & disadvantages of each sampling methods

- “Scientific research is systematic, controlled, empirical, and critical investigation of natural phenomena guided by theory and hypotheses about the presumed relations among such phenomena.”
  - Kerlinger, 1986

What is research?

Important Components of Empirical Research

- Problem statement, research questions, purposes, benefits
- Theory, assumptions, background literature
- Variables and hypotheses
- Operational definitions and measurement
- Research design and methodology
- Instrumentation, sampling
- Data analysis
- Conclusions, interpretations, recommendations

SAMPLING

* A sample is "a smaller (but hopefully representative) collection of units from a population used to determine truths about that population" (Field, 2005)
* Why sample?
  * Resources (time, money) and workload
  * Gives results with known accuracy that can be calculated mathematically
* The sampling frame is the list from which the potential respondents are drawn
  * Registrar's office
  * Class rosters
  * Must assess sampling frame errors

* SAMPLING……

* What is your population of interest?
  * To whom do you want to generalize your results?
    * All doctors
    * School children
    * Indians
    * Women aged 15-45 years
    * Other
* Can you sample the entire population?

* SAMPLING…….

* 3 factors that influence sample representative-ness
  * Sampling procedure
  * Sample size
  * Participation (response)

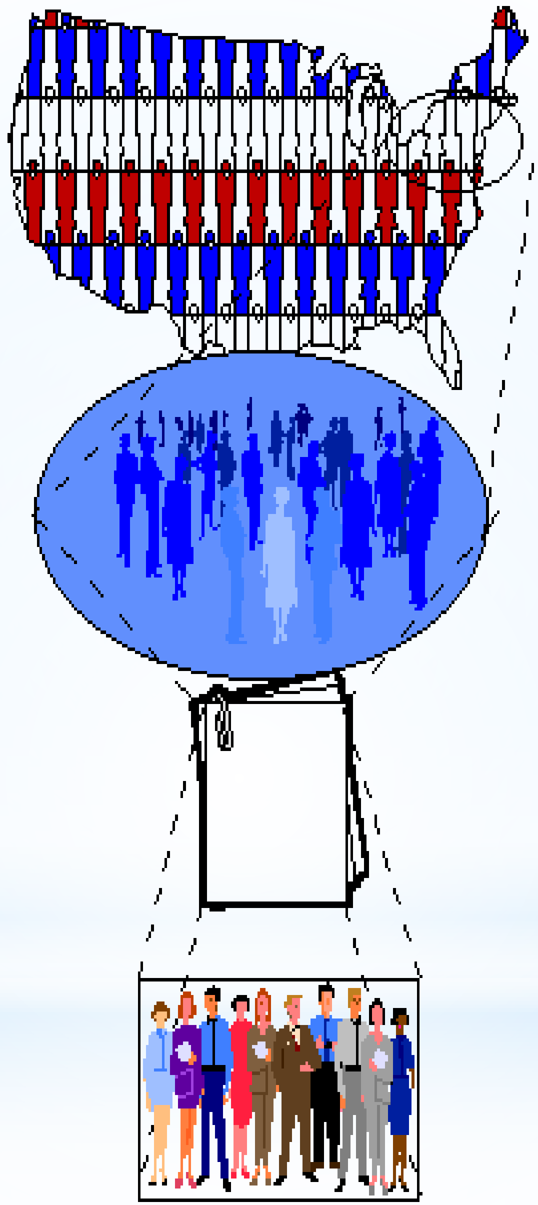
* When might you sample the entire population?
  * When your population is very small
  * When you have extensive resources
  * When you don't expect a very high response

Who do you want to generalize to? | The Theoretical Population

What population can you get access to? | The Study Population

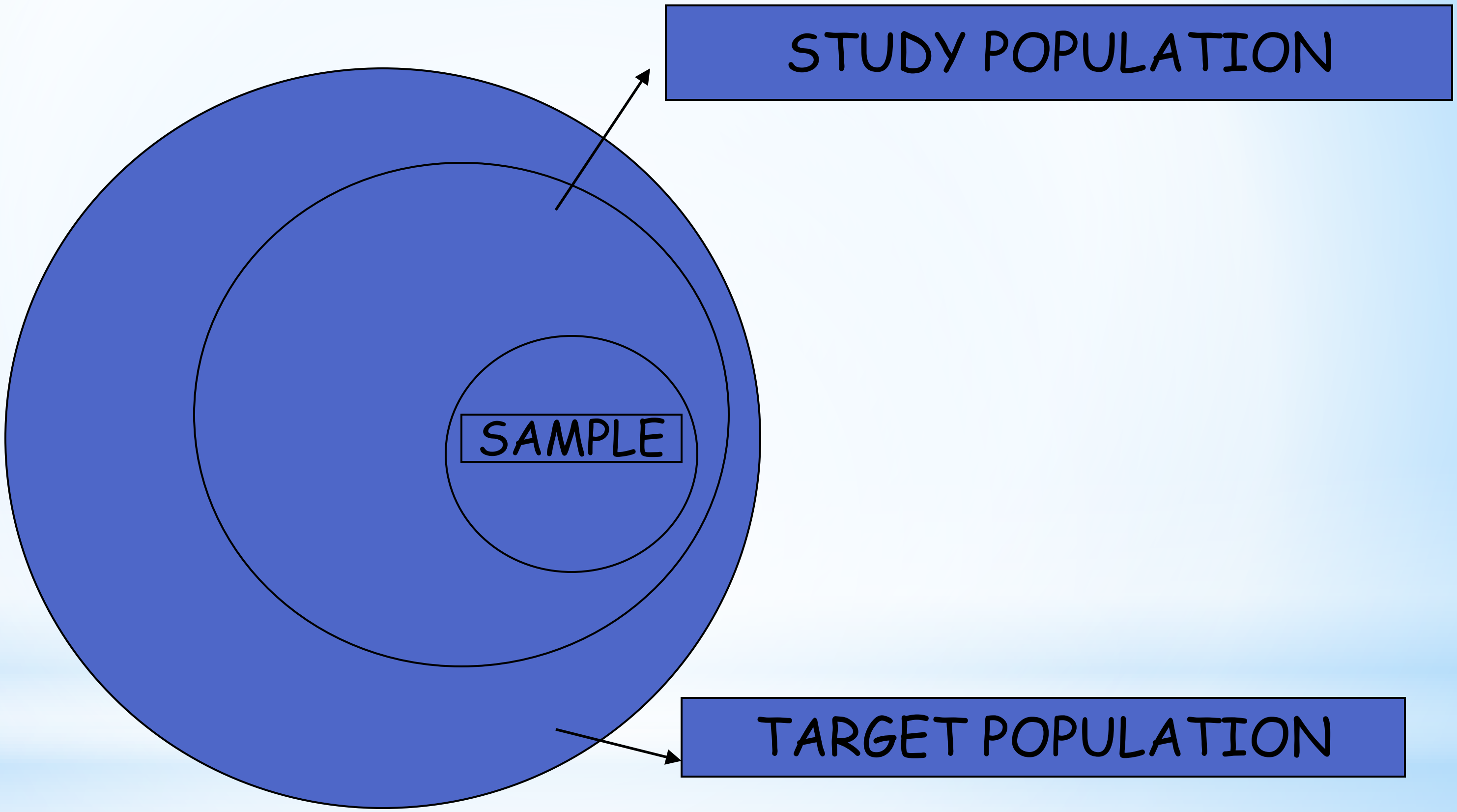
How can you get access to them? | The Sampling Frame

Who is in your study? | The Sample



SAMPLING BREAKDOWN

* SAMPLING…….



Types of Samples

- **Probability (Random) Samples**
- Simple random sample
  - Systematic random sample
  - Stratified random sample
  - Multistage sample
  - Multiphase sample
  - Cluster sample
- **Non-Probability Samples**
  - Convenience sample
  - Purposive sample
  - Quota

Process

- The sampling process comprises several stages:
  - Defining the population of concern
  - Specifying a sampling frame, a set of items or events possible to measure
  - Specifying a sampling method for selecting items or events from the frame
  - Determining the sample size
  - Implementing the sampling plan
  - Sampling and data collecting
  - Reviewing the sampling process

*Population definition

* A population can be defined as including all people or items with the characteristic one wishes to understand.
* Because there is very rarely enough time or money to gather information from everyone or everything in a population, the goal becomes finding a representative sample (or subset) of that population.

*Population definition…….

* Note also that the population from which the sample is drawn may not be the same as the population about which we actually want information. Often there is large but not complete overlap between these two groups due to frame issues etc .
* Sometimes they may be entirely separate - for instance, we might study rats in order to get a better understanding of human health, or we might study records from people born in 2008 in order to make predictions about people born in 2009.

* SAMPLING FRAME

* In the most straightforward case, such as the sentencing of a batch of material from production (acceptance sampling by lots), it is possible to identify and measure every single item in the population and to include any one of them in our sample. However, in the more general case this is not possible. There is no way to identify all rats in the set of all rats. Where voting is not compulsory, there is no way to identify which people will actually vote at a forthcoming election (in advance of the election)
* As a remedy, we seek a *sampling frame* which has the property that we can identify every single element and include any in our sample .
* The sampling frame must be representative of the population

*PROBABILITY SAMPLING

* A **probability sampling** scheme is one in which every unit in the population has a chance (greater than zero) of being selected in the sample, and this probability can be accurately determined.

* . When every element in the population *does* have the same probability of selection, this is known as an 'equal probability of selection' (EPS) design. Such designs are also referred to as 'self-weighting' because all sampled units are given the same weight.

* PROBABILITY SAMPLING…….

* Probability sampling includes:
* Simple Random Sampling,
* Systematic Sampling,
* Stratified Random Sampling,
* Cluster Sampling
* Multistage Sampling.
* Multiphase sampling

* NON PROBABILITY SAMPLING

* Any sampling method where some elements of population have *no* chance of selection (these are sometimes referred to as 'out of coverage'/'undercovered'), or where the probability of selection can't be accurately determined. It involves the selection of elements based on assumptions regarding the population of interest, which forms the criteria for selection. Hence, because the selection of elements is nonrandom, nonprobability sampling not allows the estimation of sampling errors..

* *Example: We visit every household in a given street, and interview the first person to answer the door. In any household with more than one occupant, this is a nonprobability sample, because some people are more likely to answer the door (e.g. an unemployed person who spends most of their time at home is more likely to answer than an employed housemate who might be at work when the interviewer calls) and it's not practical to calculate these probabilities.*

* Nonprobability Sampling includes: Accidental Sampling, Quota Sampling and Purposive Sampling. In addition, nonresponse effects may turn *any* probability design into a nonprobability design if the characteristics of nonresponse are not well understood, since nonresponse effectively modifies each element's probability of being sampled.

*NONPROBABILITY SAMPLING…….

SIMPLE RANDOM SAMPLING

- Applicable when population is small, homogeneous & readily available
- All subsets of the frame are given an equal probability. Each element of the frame thus has an equal probability of selection.
- It provides for greatest number of possible samples. This is done by assigning a number to each unit in the sampling frame.
- A table of random number or lottery system is used to determine which units are to be selected.

*SIMPLE RANDOM SAMPLING……. .

* Estimates are easy to calculate.
* Simple random sampling is always an EPS design, but not all EPS designs are simple random sampling.

* **Disadvantages**
* If sampling frame large, this method impracticable.
* Minority subgroups of interest in population may not be present in sample in sufficient numbers for study.

REPLACEMENT OF SELECTED UNITS

- Sampling schemes may be *without replacement* ('WOR' - no element can be selected more than once in the same sample) or *with replacement* ('WR' - an element may appear multiple times in the one sample).
- For example, if we catch fish, measure them, and immediately return them to the water before continuing with the sample, this is a WR design, because we might end up catching and measuring the same fish more than once. However, if we do not return the fish to the water (e.g. if we eat the fish), this becomes a WOR design.

SYSTEMATIC SAMPLING

- Systematic sampling relies on arranging the target population according to some ordering scheme and then selecting elements at regular intervals through that ordered list.
- Systematic sampling involves a random start and then proceeds with the selection of every *k*th element from then onwards. In this case, *k*=(population size/sample size).
- It is important that the starting point is not automatically the first in the list, but is instead randomly chosen from within the first to the *k*th element in the list.
- A simple example would be to select every 10th name from the telephone directory (an 'every 10th' sample, also referred to as 'sampling with a skip of 10').

SYSTEMATIC SAMPLING……

As described above, systematic sampling is an EPS method, because all elements have the same probability of selection (in the example given, one in ten). It is *not* 'simple random sampling' because different subsets of the same size have different selection probabilities - e.g. the set {4,14,24,...,994} has a one-in-ten probability of selection, but the set {4,13,24,34,...} has zero probability of selection.

* SYSTEMATIC SAMPLING……

* **ADVANTAGES:**
* Sample easy to select
* Suitable sampling frame can be identified easily
* Sample evenly spread over entire reference population
* **DISADVANTAGES:**
* Sample may be biased if hidden periodicity in population coincides with that of selection.
* Difficult to assess precision of estimate from one survey.

STRATIFIED SAMPLING

Where population embraces a number of distinct categories, the frame can be organized into separate "strata." Each stratum is then sampled as an independent sub-population, out of which individual elements can be randomly selected.

- Every unit in a stratum has same chance of being selected.
- Using same sampling fraction for all strata ensures proportionate representation in the sample.
- Adequate representation of minority subgroups of interest can be ensured by stratification & varying sampling fraction between strata as required.

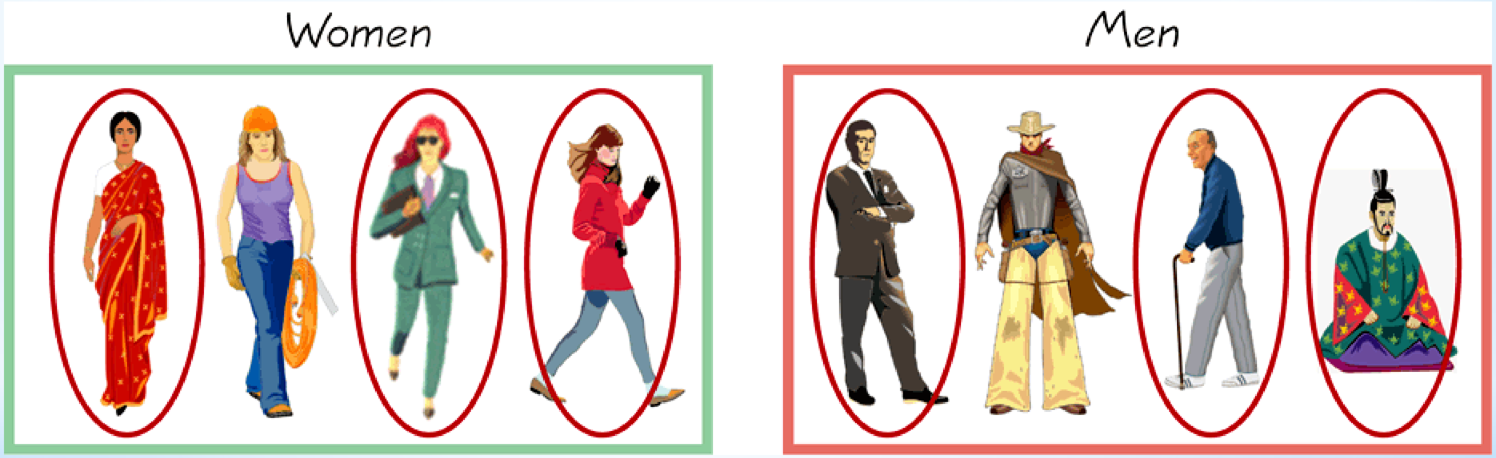
* STRATIFIED SAMPLING……

* Finally, since each stratum is treated as an independent population, different sampling approaches can be applied to different strata.

* **Drawbacks** to using stratified sampling.
* First, sampling frame of entire population has to be prepared separately for each stratum
* Second, when examining multiple criteria, stratifying variables may be related to some, but not to others, further complicating the design, and potentially reducing the utility of the strata.
* Finally, in some cases (such as designs with a large number of strata, or those with a specified minimum sample size per group), stratified sampling can potentially require a larger sample than would other methods

* STRATIFIED SAMPLING…….

Draw a sample from each stratum



POSTSTRATIFICATION

- Stratification is sometimes introduced after the sampling phase in a process called "poststratification".
- This approach is typically implemented due to a lack of prior knowledge of an appropriate stratifying variable or when the experimenter lacks the necessary information to create a stratifying variable during the sampling phase. Although the method is susceptible to the pitfalls of post hoc approaches, it can provide several benefits in the right situation. Implementation usually follows a simple random sample. In addition to allowing for stratification on an ancillary variable, poststratification can be used to implement weighting, which can improve the precision of a sample's estimates.

* Choice-based sampling is one of the stratified sampling strategies. In this, data are stratified on the target and a sample is taken from each strata so that the rare target class will be more represented in the sample. The model is then built on this biased sample. The effects of the input variables on the target are often estimated with more precision with the choice-based sample even when a smaller overall sample size is taken, compared to a random sample. The results usually must be adjusted to correct for the oversampling.

OVERSAMPLING

CLUSTER SAMPLING

* Cluster sampling is an example of 'two-stage sampling' .
* First stage a sample of areas is chosen;
* Second stage a sample of respondents *within* those areas is selected.
* Population divided into clusters of homogeneous units, usually based on geographical contiguity.
* Sampling units are groups rather than individuals.
* A sample of such clusters is then selected.
* All units from the selected clusters are studied.

* Advantages :
* Cuts down on the cost of preparing a sampling frame.
* This can reduce travel and other administrative costs.
* Disadvantages: sampling error is higher for a simple random sample of same size.
* Often used to evaluate vaccination coverage in EPI

CLUSTER SAMPLING…….

*CLUSTER SAMPLING…….

- **Identification of clusters**
  - List all cities, towns, villages & wards of cities with their population falling in target area under study.
  - Calculate cumulative population & divide by 30, this gives sampling interval.
  - Select a random no. less than or equal to sampling interval having same no. of digits. This forms 1^{st} cluster.
  - Random no.+ sampling interval = population of 2^{nd} cluster.
  - Second cluster + sampling interval = 4^{th} cluster.
  - Last or 30^{th} cluster = 29^{th} cluster + sampling interval

*CLUSTER SAMPLING…….

Two types of cluster sampling methods.

One-stage sampling. All of the elements within selected clusters are included in the sample.

Two-stage sampling. A subset of elements within selected clusters are randomly selected for inclusion in the sample.

*CLUSTER SAMPLING…….

	Freq	c f	cluster
I	2000	2000	1
II	3000	5000	2
III	1500	6500	
IV	4000	10500	3
V	5000	15500	4, 5
VI	2500	18000	6
VII	2000	20000	7
VIII	3000	23000	8
IX	3500	26500	9
X	4500	31000	10
XI	4000	35000	11, 12
XII	4000	39000	13
XIII	3500	44000	14,15
XIV	2000	46000	
XV	3000	49000	16
XVI	3500	52500	17
XVII	4000	56500	18,19
XVIII	4500	61000	20
XIX	4000	65000	21,22
XX	4000	69000	23
XXI	2000	71000	24
XXII	2000	73000	
XXIII	3000	76000	25
XXIV	3000	79000	26
XXV	5000	84000	27,28
XXVI	2000	86000	29
XXVII	1000	87000	
XXVIII	1000	88000	
XXIX	1000	89000	30
XXX	1000	90000	

- 90000/30 = 3000 sampling interval

* Difference Between Strata and Clusters

* Although strata and clusters are both non-overlapping subsets of the population, they differ in several ways.
* All strata are represented in the sample; but only a subset of clusters are in the sample.
* With stratified sampling, the best survey results occur when elements within strata are internally homogeneous. However, with cluster sampling, the best results occur when elements within clusters are internally heterogeneous

* MULTISTAGE SAMPLING

* Complex form of cluster sampling in which two or more levels of units are embedded one in the other.

* First stage, random number of districts chosen in all states.

* Followed by random number of talukas, villages.

* Then third stage units will be houses.

* All ultimate units (houses, for instance) selected at last step are surveyed.

MULTISTAGE SAMPLING…….

* This technique, is essentially the process of taking random samples of preceding random samples.
* Not as effective as true random sampling, but probably solves more of the problems inherent to random sampling.
* An effective strategy because it banks on multiple randomizations. As such, extremely useful.
* Multistage sampling used frequently when a complete list of all members of the population not exists and is inappropriate.
* Moreover, by avoiding the use of all sample units in all selected clusters, multistage sampling avoids the large, and perhaps unnecessary, costs associated with traditional cluster sampling.

* MULTI PHASE SAMPLING

* Part of the information collected from whole sample & part from subsample.

* In Tb survey MT in all cases – Phase I
* X –Ray chest in MT +ve cases – Phase II
* Sputum examination in X – Ray +ve cases - Phase III

* Survey by such procedure is less costly, less laborious & more purposeful

* MATCHED RANDOM SAMPLING

A method of assigning participants to groups in which pairs of participants are first matched on some characteristic and then individually assigned randomly to groups.

* The Procedure for Matched random sampling can be briefed with the following contexts,
* Two samples in which the members are clearly paired, or are matched explicitly by the researcher. For example, IQ measurements or pairs of identical twins.
* Those samples in which the same attribute, or variable, is measured twice on each subject, under different circumstances. Commonly called repeated measures.
* Examples include the times of a group of athletes for 1500m before and after a week of special training; the milk yields of cows before and after being fed a particular diet.

QUOTA SAMPLING

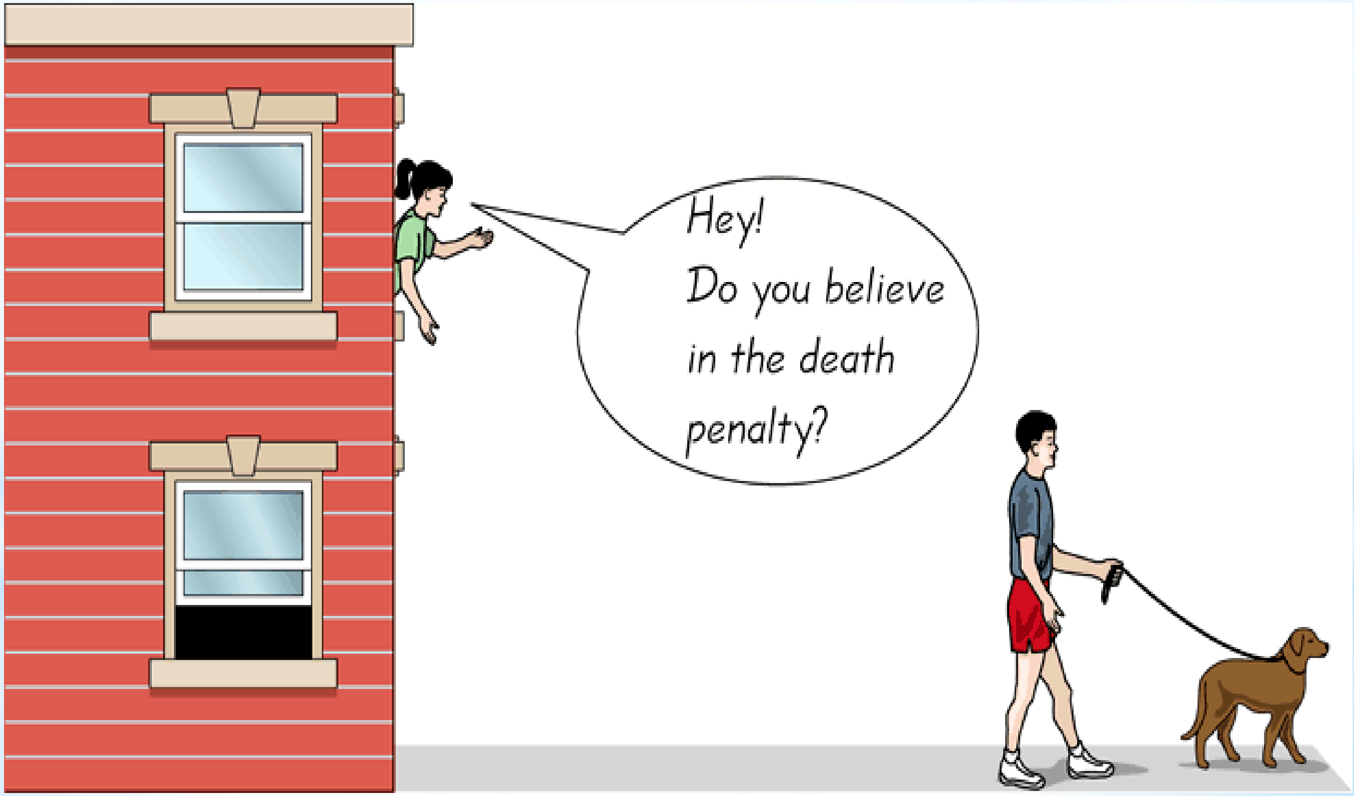
- The population is first segmented into mutually exclusive sub-groups, just as in stratified sampling.
- Then judgment used to select subjects or units from each segment based on a specified proportion.
- For example, an interviewer may be told to sample 200 females and 300 males between the age of 45 and 60.
- It is this second step which makes the technique one of non-probability sampling.
- In quota sampling the selection of the sample is non-random.
- For example interviewers might be tempted to interview those who look most helpful. The problem is that these samples may be biased because not everyone gets a chance of selection. This random element is its greatest weakness and quota versus probability has been a matter of controversy for many years

* CONVENIENCE SAMPLING

* Sometimes known as **grab** or **opportunity sampling or accidental or haphazard sampling**.
* A type of nonprobability sampling which involves the sample being drawn from that part of the population which is close to hand. That is, readily available and convenient.
* The researcher using such a sample cannot scientifically make generalizations about the total population from this sample because it would not be representative enough.
* For example, if the interviewer was to conduct a survey at a shopping center early in the morning on a given day, the people that he/she could interview would be limited to those given there at that given time, which would not represent the views of other members of society in such an area, if the survey was to be conducted at different times of day and several times per week.
* This type of sampling is most useful for pilot testing.
* In social science research, snowball sampling is a similar technique, where existing study subjects are used to recruit more subjects into the sample.

* CONVENIENCE SAMPLING…….

* Use results that are easy to get



101

* - The researcher chooses the sample based on who they think would be appropriate for the study. This is used primarily when there is a limited number of people that have expertise in the area being researched

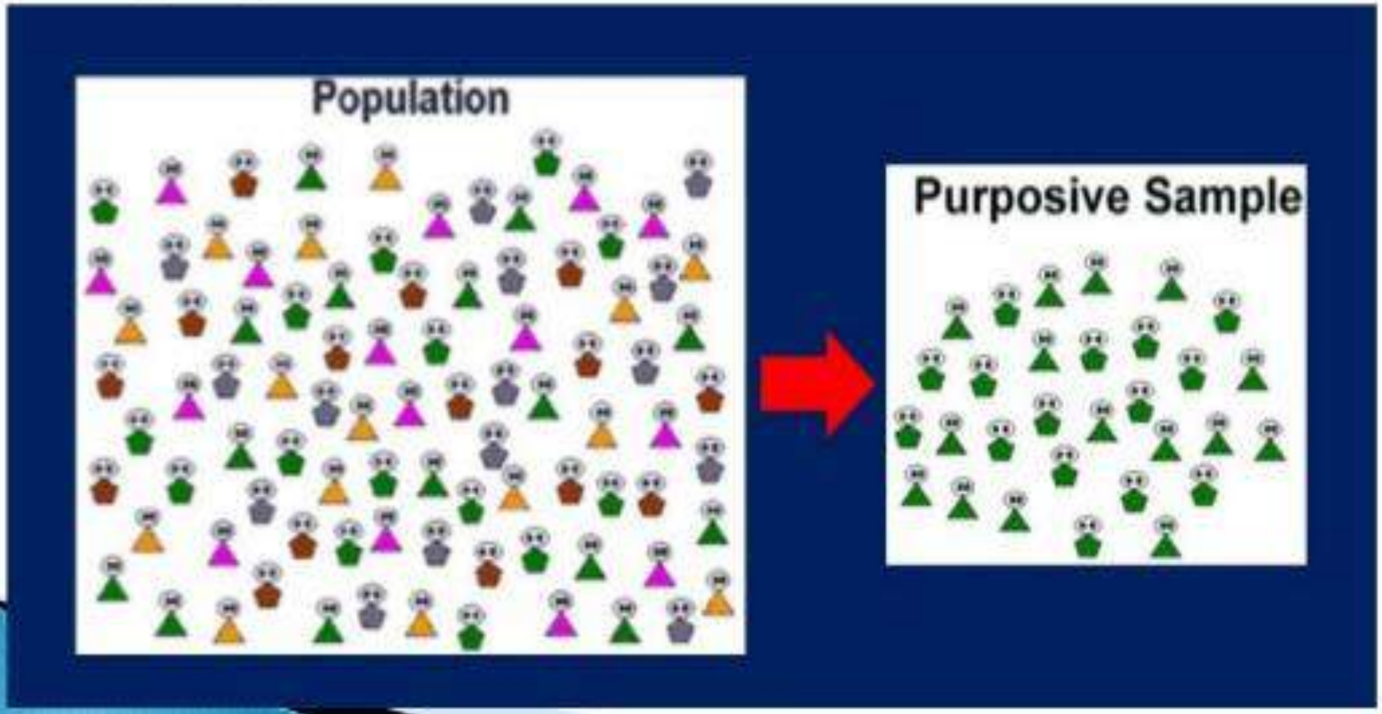
Judgmental sampling or Purposive sampling

* PANEL SAMPLING

* Method of first selecting a group of participants through a random sampling method and then asking that group for the same information again several times over a period of time.
* Therefore, each participant is given same survey or interview at two or more time points; each period of data collection called a "wave".
* This sampling methodology often chosen for large scale or nation-wide studies in order to gauge changes in the population with regard to any number of variables from chronic illness to job stress to weekly food expenditures.
* Panel sampling can also be used to inform researchers about within-person health changes due to age or help explain changes in continuous dependent variables such as spousal interaction.
* There have been several proposed methods of analyzing panel sample data, including growth curves.

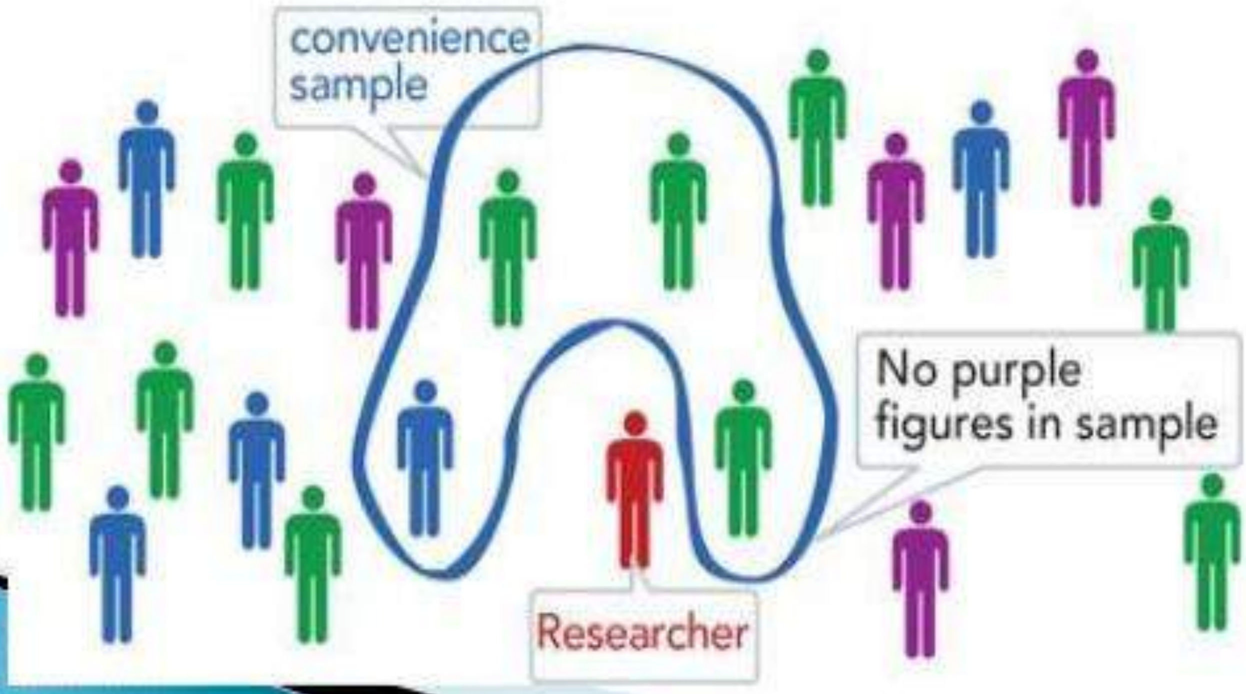
Non Probability Sampling

1. **Purposive Sampling: R**esearcher selects a "typical group" of individuals who might represent the larger population and then collects data from this group. Also known as **Judgmental Sampling.**



2. **Convenience Sampling :** Obtaining units or members who are most conveniently available. It consists of units which are obtained because cases are readily available.

Researcher determines the required sample size and then simply collects data on that number of individuals who are available easily.



3. **<u>Quota Sampling:</u>** The selection of the sample is made by the researcher, who decides the quotas for selecting sample from specified sub groups of the population.

- For example, an interviewer might be **needed data from 40 adults and 20 adolescents in order to study students' television viewing habits.**

<u>Selection will be</u>

- **20** Adult men and 20 adult women
- 10 adolescent girls and 10 adolescent boys

4. Snowball Sampling:

- **In snowball sampling,** the researcher **Identifying and selecting available respondents who meet the criteria** for inclusion.
- After the data have been collected from the subject, **the researcher asks for a referral of other individuals**, who would also meet the criteria and represent the population of concern.

- **chain sampling**, **chain-referral**, **sampling referral sampling**

* NON-PROBABILITY SAMPLING TECHNIQUES

Sampling

- Measuring a small portion of something and then making a general statement about the whole thing.

- Process of selecting a number of units for a study in such a way that the units represent the larger group from which they are selected.

General Types of Sampling

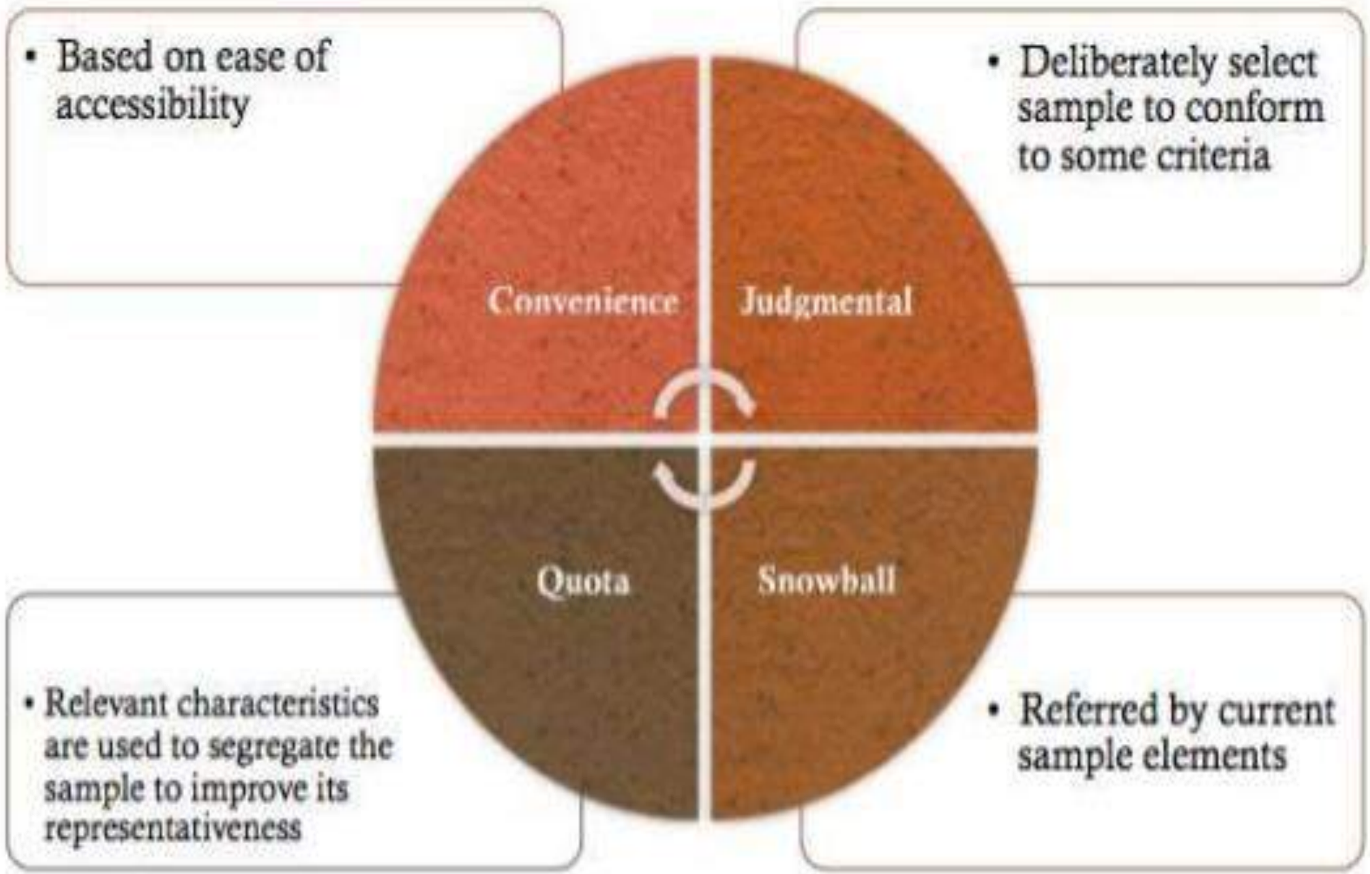
1. **Probability sampling**

2. **Non-probability sampling**

Non-probability sampling

- Unequal chance of being included in the sample (non-random)
- Non random or non - probability sampling refers to the sampling process in which, the samples are selected for a specific purpose with a pre-determined basis of selection.
- The sample is not a proportion of the population and there is no system in selecting the sample. The selection depends upon the situation.
- No assurance is given that each item has a chance of being included as a sample
- There is an assumption that there is an even distribution of characteristics within the population, believing that any sample would be representative.

Non-Probability Methods



DATA COLLECTION

INTRODUCTION

- Various methods of collecting data are employed by social scientists. Here we will discuss the varied dimensions relevant to : data generation and attempt to arrive at an abstract typology involving stimuli, responses and setting for data collection. The task of data collection begins after a research problem has been defined and research design /plan chalked out.

TYPES OF DATA

1) PRIMARY DATA : Are those which are collected a **fresh** and for the **first time** and thus happen to be **original in character** and known as Primary data.

2) SECONDARY DATA : Are those which have been **collected by someone else** and which have **already been passed** through the statistical process are known as Secondary data.

COLLECTION OF PRIMARY DATA

- There are several methods of collecting primary data, particularly in surveys and descriptive researches. In descriptive research, we obtain primary data either through observation or through direct communication with respondents in one form or another or through personal interviews.

COLLECTION OF SECONDARY DATA

- These are already available i.e. they refer to the data which have **already been collected and analyzed by someone else.**
- Secondary data may either be published or unpublished data. Researcher must be very careful in using secondary data, because the data available may be sometimes unsuitable.

Methods of data Collection :Primary Data

- 1) OBSERVATION METHOD : Observation method is a method under which data from the field is collected with the help of observation by the observer or by personally going to the field.
- In the words of P.V. Young, "Observation may be defined as systematic viewing, coupled with consideration of seen phenomenon."

ADVANTAGES:

- Subjective bias eliminated (No bias info)
- Information researcher gets is Current information
- Independent to respondent's variable (as in interview and may be bias)
- DISADVANTAGES :
- It is expensive method (time requires more)
- Limited information
- Unforeseen factors may interfere with observational task
- Respondents opinion can not be recorded on certain subject

TYPES OF OBSERVATION

Structured and Unstructured Observation

- When observation is done by characterizing style of recording the observed information, standardized conditions of observation , definition of the units to be observed , selection of pertinent data of observation then it is structured observation
- When observation is done without any thought before observation then it is unstructured observation

Participant & Non Participant Observation

- When the Observer is member of the group which he is observing then it is Participant Observation
- In participant observation Researcher can record natural behavior of group , Researcher can verify the truth of statements given by informants in the context of questionnaire , Difficult to collect information can obtain through this method but in this researcher may loose objectivity of research due emotional feelings. Prob. of control in observation isn't solved.

Non Participant Observation

- When observer is observing people without giving any information to them then it is non participant observation

Controlled & Uncontrolled Observation

- When the observation takes place in natural condition i.e. uncontrolled observation. It is done to get spontaneous picture of life and persons
- When observation takes place according to definite pre arranged plans , with experimental procedure then it is controlled observation generally done in laboratory under controlled condition.

INTERVIEW METHOD

- This method of collecting data involves presentation or oral-verbal stimuli and reply in terms of oral-verbal responses.
- Interview Method This is Oral Verbal communication . Where interviewer asks questions(which are aimed to get information required for study) to respondent

There are different type of interviews as follows :

PERSONAL INTERVIEWS :

The interviewer asks questions generally in a face to face contact to the other person or persons.

Types of Personal Interview

Personal Interview

- Predetermined questions
- Standardized techniques of recording
- Interviewer follows rigid procedure laid down i.e. asking questions in form & order prescribed
- Time required for such interview is less than non structured interview
- Not necessary of skill or specific knowledge
- Analysis of data becomes easier Becoz information is collected in prescribed manner

Structured Interview

- Flexibility in asking questions
- No Predetermined questions
- No Standardized techniques of recording
- Interviewer has freedom to ask , omit , add questions in any manner
- Ask questions without following sequence
- Deep knowledge & skill required
- Analysis of data is difficult

Merits of Personal Interview

- Information at greater depth
- Flexibility of restructuring the Questionnaire
- Interviewer by his skill can come over resistance
- Non Response generally low
- Samples can controlled more effectively
- Personal information can be obtained

- Interviewer can collect supplementary information about respondent's personal characteristics and environment which has value in interpreting results

De Merits Of Interview

- Expensive method
- Respondent may give bias information
- Some Executive people are not approachable so data collected may be inadequate
- Takes more time when samples are more
- Systematic errors may be occurred
- Supervisors has to do complex work of selecting ,training and supervising the field staff.

TELEPHONIC INTERVIEWS

- Contacting samples on telephone
- Uncommon method may be used in developed regions

MERITS

- Flexible compare to mailing method
- Faster than other methods
- Cheaper than personal interview method
- Callbacks are simple and economical also
- High response than mailing method.
- when it is not possible to contact the respondent directly, then interview is conducted through – Telephone.

- Replies can be recorded without embarrassment to respondents
- Interviewer can explain requirements more easily
- No field staff is required
- Wider distribution of sample is possible

DEMERITS

- Little time is given to respondents
- Survey is restricted to respondents who have telephones
- Not suitable for intensive survey where comprehensive answers are required
- Bias information may be more
- Very difficult to make questionnaire because it should short and to the point

- structured interviews : in this case, a set of pre-decided questions are there.
- unstructured interviews : in this case, we don't follow a system of pre-determined questions.
- focused interviews : attention is focused on the given experience of the respondent and its possible effects.
- clinical interviews : concerned with broad underlying feelings or motivations or with the course of individual's life experience, rather than with the effects of the specific experience, as in the case of focused interview.

- group interviews : a group of 6 to 8 individuals is interviewed.
- qualitative and quantitative interviews : divided on the basis of subject matter i.e. whether qualitative or quantitative.
- individual interviews : interviewer meets a single person and interviews him.
- selection interviews : done for the selection of people for certain jobs.
- depth interviews : it deliberately aims to elicit unconscious as well as other types of material relating especially to personality dynamics and motivations.

QUESTIONNAIRE METHOD

- This method of data collection is quite popular, particularly in case of big enquiries. The questionnaire is mailed to respondents who are expected to read and understand the questions and write down the reply in the space meant for the purpose in the questionnaire itself. The respondents have to answer the questions on their own.
- Questionnaire Method Questionnaire is sent to persons with request to answer the questions and return the questionnaire Questions are printed in definite order , mailed to samples who are expected to read that questions understand the questions and write the answers in provided space .

Merits of Questionnaire

- Merits of Questionnaire Low cost even the geographical area is large to cover Answers are in respondents word so free from bias Adequate time to think for answers Non approachable respondents may be conveniently contacted Large samples can be used so results are more reliable

Demerits of Questionnaire

- Demerits of Questionnaire Low rate of return of duly filled questionnaire Can be used when respondent is educated and co operative It is inflexible Omission of some questions Difficult to know the expected respondent have filled the form or it is filled by some one else Slowest method of data collection

Main Aspects of Questionnaire

- Main Aspects of Questionnaire General Form Structured Questionnaire Alternatives or yes no type questions are asked Easy to interpret the data but unuseful for the survey which is aimed to probe for attitudes, and reasons for certain actions Unstructured Questionnaire open ended questions

- Respondents gives answers in his own words On the basis of the pre test researcher can decide about which type of questionnaire should be used Question Sequence Question sequence should be clear and smoothly moving (relation of one question to another should readily apparent First question important for creating interest in respondents mind

- Question which gives stress on memory or of a personal character and wealth should be avoided as opening questions Easier question should be at the start of the questionnaire General to specific questions should be the sequence of questions Question Formulation and Wording Question should easily understood Question should be simple and concrete.

- Closed questions are easy to handle but this is like fixing the answers in people's mouth. So depending upon problem for which survey is going on both close ended and open ended question may be asked in Questionnaire. Words having ambiguous meaning should be avoided, catch words ,words with emotional connotations , danger words should be avoided

Essentials of Good Questionnaire

- Essentials of Good Questionnaire Should Short & simple Questions should arranged in logical sequence (From Easy to difficult one) Technical terms should avoided Some control questions which indicate reliability of the respondent (To Know consumption first expenditure and then weight or qty of that material)
- Questions affecting the sentiments of the respondents should avoided Adequate space for answers should be provided in questionnaire Provision for uncertainty (do not know, No preference) Directions regarding the filling of questionnaire should be given Physical Appearance - - Quality of paper, color

HOW TO CONSTRUCT A QUESTIONNAIRE

Researcher should note the following with regard to these three main aspects of a questionnaire:

- General form
- Question Sequence
- Determine the type the Questions :
  - A) Direct Question
  - B) Indirect Question
  - C) Open Form Questionnaire
  - D) Closed Form Questionnaire
  - E) Dichotomous Questions
  - F) Multiple Choice Questions (MCQ)

SCHEDULE METHOD

- It is one of the important methods for the study of social problems.
- Schedules Like Questionnaires but it filled by enumerator . Enumerators are specially appointed for filling questionnaire Enumerators explain the aim and objective to respondent and fill the answers in provided space .
- In the words of Thomas Carson Macormic, "The schedule is nothing more than a list of questions which it seems necessary to test the hypothesis ."

Questionnaire V/S Schedule

Questionnaire

- Q generally sent through mail and no further assistance from sender
- Q is cheaper method
- Non Response is high

Schedule

- Schedule is filled by the enumerator or research worker
- Costly requires field workers
- Non response is low

Questionnaire

- In questionnaire it is not confirmed that expected respondent have filled the answers
- Very slow method
- No Personal contact
- Q can be used only when respondent is educated and co operative

Schedule

- In Schedule identity of person is known
- Information is collected well in time
- Direct personal contact
- Info can collected from illiterates also

- Wider distribution of sample is possible
- Incomplete and wrong information is more
- Depends on quality of questionnaire
- Physical appearance of questionnaire should attractive
- Observation method can not use

- Difficulty for wider area
- Relatively more correct and complete
- Depends on Honesty and competence of enumerator
- Not necessary in Schedule method
- It is possible to use observation at the time of filling schedule by enumerator

Other Methods Of Data Collection

- **Warranty Cards** Post card size cards sent to customers and feedback collected through asking questions on that card
- **Distributor or Store Audits** Audits are done by distributor or manufacturer's salesperson. Observation or copying information about inventory in retail shops. Useful method for knowing market share ,market size , effect of in store promotion.

- **Pantry Audits** From the observation of pantry of customer to know purchase habit of the people (which product , of what brand etc.) Questions may be asked at the time of audit
- **Consumer Panels** When pantry audit is done at regular basis, Daily record of consumption of certain customers. Or repeatedly interviewed at the specific periods to know their consumption.
- **Transitory consumer panels** – for limited time Continuing Consumer panel For indefinite period

- **Use of Mechanical Device** Eye Cameras to record eyes focus on certain sketch
- Psycho galvanometer to measure body excitement to visual stimulus
- Motion Picture camera to record movement of body at the time of purchase
- Audiometer concerned to TV . Useful to know Channel, program preference of people

- **Depth Interview** To discover the underlying motives or desires of samples . To explore needs , feelings of respondents. Skill is required , indirect question or projective techniques are used to know behavior of the respondent.
- **Content Analysis** analyzing contents of documentary material as news paper , books , magazines about certain characteristics to identify and count

CASE STUDY METHOD

- It is essentially an intensive investigation of the particular unit under consideration. Its important characteristics are as follows :

a) the researcher can take one single social unit or more of such units for his study purpose.

b) the selected unit is studied intensively i.e. it is studied in minute details.

SURVEY METHOD

- One of the common methods of diagnosing and solving of social problems is that of undertaking surveys.
- Festinger and Kat of the opinion that, "Many research problems require systematic collection of data from population through the use of personal interviews or other data gathering devices".

PANEL METHOD

In this method, data is collected from the same sample respondents at the some interval either by mail or by personal interview. This is used for studies on :

- 1) Expenditure Pattern
- 2) Consumer Behaviour
- 3) Effectiveness of Advertising
- 4) Voting Behaviour and so on

Secondary Data

Sources of data

- Publications of Central, state , local government
- Technical and trade journals
- Books, Magazines, Newspaper
- Reports & publications of industry ,bank, stock exchange
- Reports by research scholars, Universities, economist
- Public Records

Factors to be considered before using secondary data

- Reliability of data – Who, when , which methods, at what time etc.
- Suitability of data – Object ,scope, and nature of original inquiry should be studied, as if the study was with different objective then that data is not suitable for current study
- Adequacy of data– Level of accuracy,
- Area differences then data is not adequate for study

Selection of proper Method for collection of Data

- Nature ,Scope and object of inquiry
- Availability of Funds
- Time Factor
- Precision Required

TYPES OF DATA

- Quantitative data

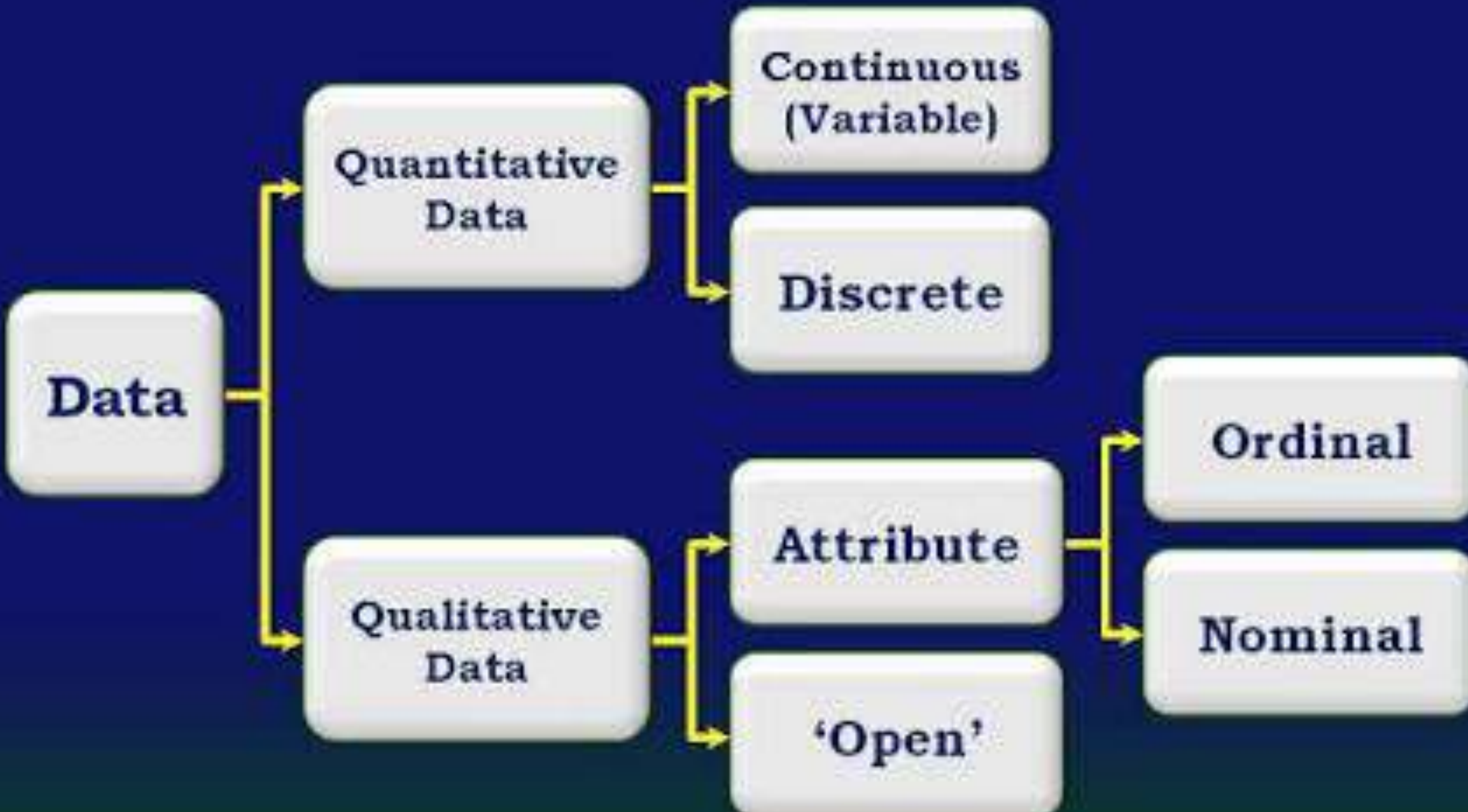
  measurements use scale with equal intervals
  examples include mass (g), length (cm), volume (mL), temperature (°C or K)

- Qualitative data

  non-standard scales with unequal intervals or discrete categories
  examples include gender, choice, color scales

SG-3

Types of Data



www.Velaction.com

Types of quantitative data

- Nominal (categorical) data:
  - Categories, e.g. gender, occupation.
  - Analysis type: Show frequency.
- Ordinal (ranked) data:
  - Numbers allocated to a quantitative scale, e.g. examination results.
  - Can be used to code answers to questionnaires (e.g. 1="Strongly disagree").
  - Ranking does not provide information about intervals (i.e. difference among ranks).
- Interval data:
  - Includes information about the interval between numbers, e.g. year number.
- Ratio data:
  - Interval data with a true zero point, e.g. people's age.

Qual Vs Quan: Basic differences

	Qualitative	Quantitative
Purpose	To describe a situation, gain insight to particular practice...	To measure magnitude-How widespread is a practice...
Format	No pre-determined response categories	Pre-determined response categories, standard measures
Data	In-depth explanatory data from a small sample	Wide breadth of data from large statistically representative sample
Analysis	Draws out patterns from concepts and insights	Tests hypotheses, uses data to support conclusion
Result	Illustrative explanation & individual responses	Numerical aggregation in summaries, responses are clustered
Sampling	Theoretical	Statistical

Variables

- Quantitative Variable
  - A variable that is counted or measured on a numerical scale
  - Can be continuous or discrete (always a whole number).
- Qualitative Variable
  - A non-numerical variable that can be classified into categories, but can't be measured on a numerical scale.
  - Can be nominal or ordinal

SCALE

Background

- The "levels of measurement" is an expression which typically refers to the theory of scale types developed by the psychologist Stanley Smith Stevens.
- Stevens proposed his theory in a 1946 article titled "On the theory of scales of measurement".
- In this article Stevens claimed that all measurement in science was conducted using four different types of numerical scales which he called "nominal", "ordinal", "interval" and "ratio".

The Theory Of Scale Types

Stevens (1946, 1951) proposed that measurements can be classified into four different types of scales. These were:

- Nominal
- Ordinal
- Interval
- Ratio

Nominal Scale

- A **categorical variable**, also called a **nominal variable**, is for mutual exclusive, but not ordered, categories.
- Nominal scales are mere codes assigned to objects as labels, they are not measurements.
- Not a measure of quantity. Measures identity and difference. People either belong to a group or they do not.
- Sometimes numbers are used to designate category membership.

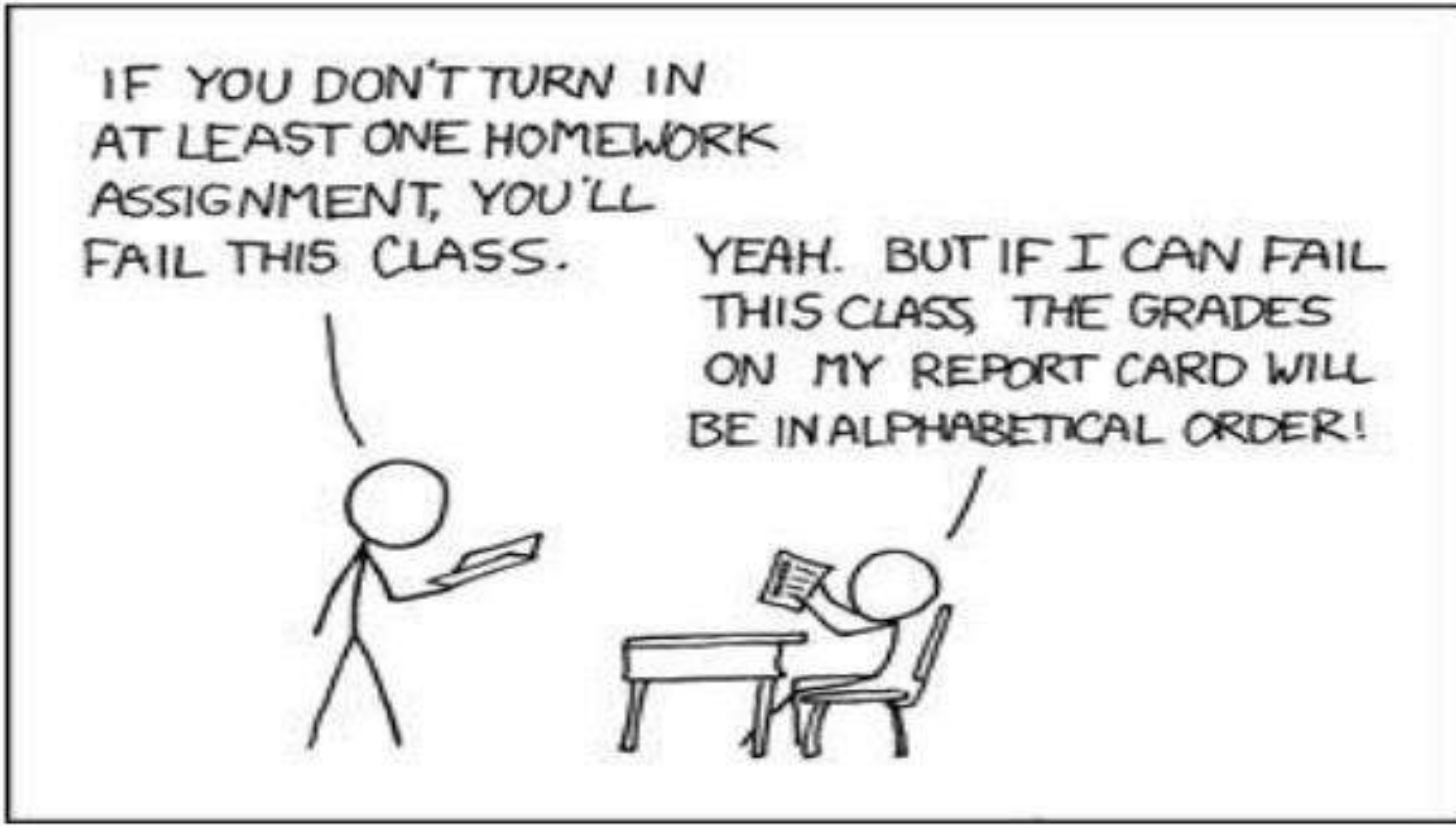
Examples

- Eye color: blue, brown, green, etc.
- Biological sex (male or female)
- Democrat, republican, green, libertarian, etc.
- Married, single, divorced, widowed
- Country of Origin
- 1 = United States 3 = Canada
- 2 = Mexico 4 = Other

(Here, the numbers do not have numeric implications; they are simply convenient labels)

What Statistic Can I Apply?

OK to compute....	Nominal
Frequency Distribution and mode	Yes
Median And Percentiles.	No
Add Or Subtract.	No
Mean, Standard Deviation, Standard Error Of The Mean.	No
Ratio, Or Coefficient Of Variation.	No
Chi-square	Yes



Ordinal Scale

Ordinal Scale

- This scale has the ability to rank the individual attributes of to items in same group but unit of measurement is not available in this scale, like student A is taller than student B but their actual heights are not available.
- Designates an ordering: greater than, less than.
- Does not assume that the intervals between numbers are equal.

Examples

- Rank your food preference where 1 = favorite food and 4 = least favorite:

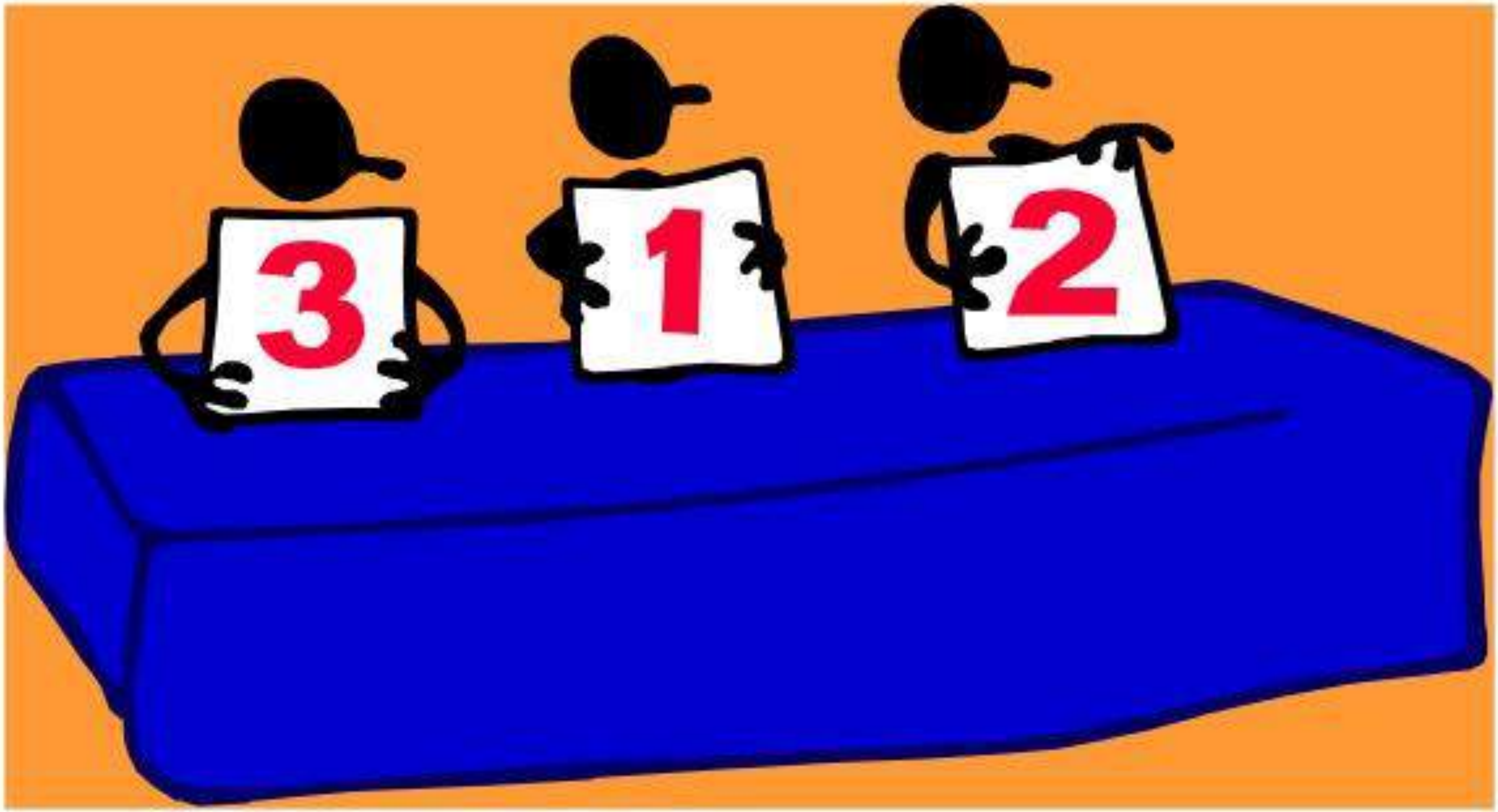
  ____ sushi ____ chocolate

  ____ hamburger ____ papaya

- Final position of horses in a thoroughbred race is an ordinal variable. The horses finish first, second, third, fourth, and so on. The difference between first and second is not necessarily equivalent to the difference between second and third, or between third and fourth.

What Statistic Can I Apply?

OK To Compute....	Ordinal
Frequency Distribution.	Yes
Median And Percentiles.	Yes
Add Or Subtract.	No
Mean, Standard Deviation, Standard Error Of The Mean.	No
Ratio, Or Coefficient Of Variation.	No



Interval Scale

Interval Scale

- Classifies data into groups or categories
- Determines the preferences between items
- Zero point on the internal scale is arbitrary zero, it is not the true zero point
- Designates an equal-interval ordering.
- The difference in temperature between 20 degrees f and 25 degrees f is the same as the difference between 76 degrees f and 81 degrees f.

Examples

- Temperature in Fahrenheit is interval.
- Celsius temperature is an interval variable. It is meaningful to say that 25 degrees Celsius is 3 degrees hotter than 22 degrees Celsius, and that 17 degrees Celsius is the same amount hotter (3 degrees) than 14 degrees Celsius. Notice, however, that 0 degrees Celsius does not have a natural meaning. That is, 0 degrees Celsius does not mean the absence of heat!
- Common IQ tests are *assumed* to use an interval metric.

Examples

Likert scale: How do you feel about Stats?

1 = I'm totally dreading this class!

2 = I'd rather not take this class.

3 = I feel neutral about this class.

4 = I'm interested in this class.

5 = I'm SO excited to take this class!

What Statistic Can I Apply?

OK To Compute....	Interval
Frequency Distribution.	Yes
Median And Percentiles.	Yes
Add Or Subtract.	Yes
Mean, Standard Deviation, Correlation, Regression, Analysis Of Variance	Yes
Ratio, Or Coefficient Of Variation.	No

Ratio Scale

- This is the highest level of measurement and has the properties of an interval scale; coupled with fixed origin or zero point.
- It clearly defines the magnitude or value of difference between two individual items or intervals in same group.

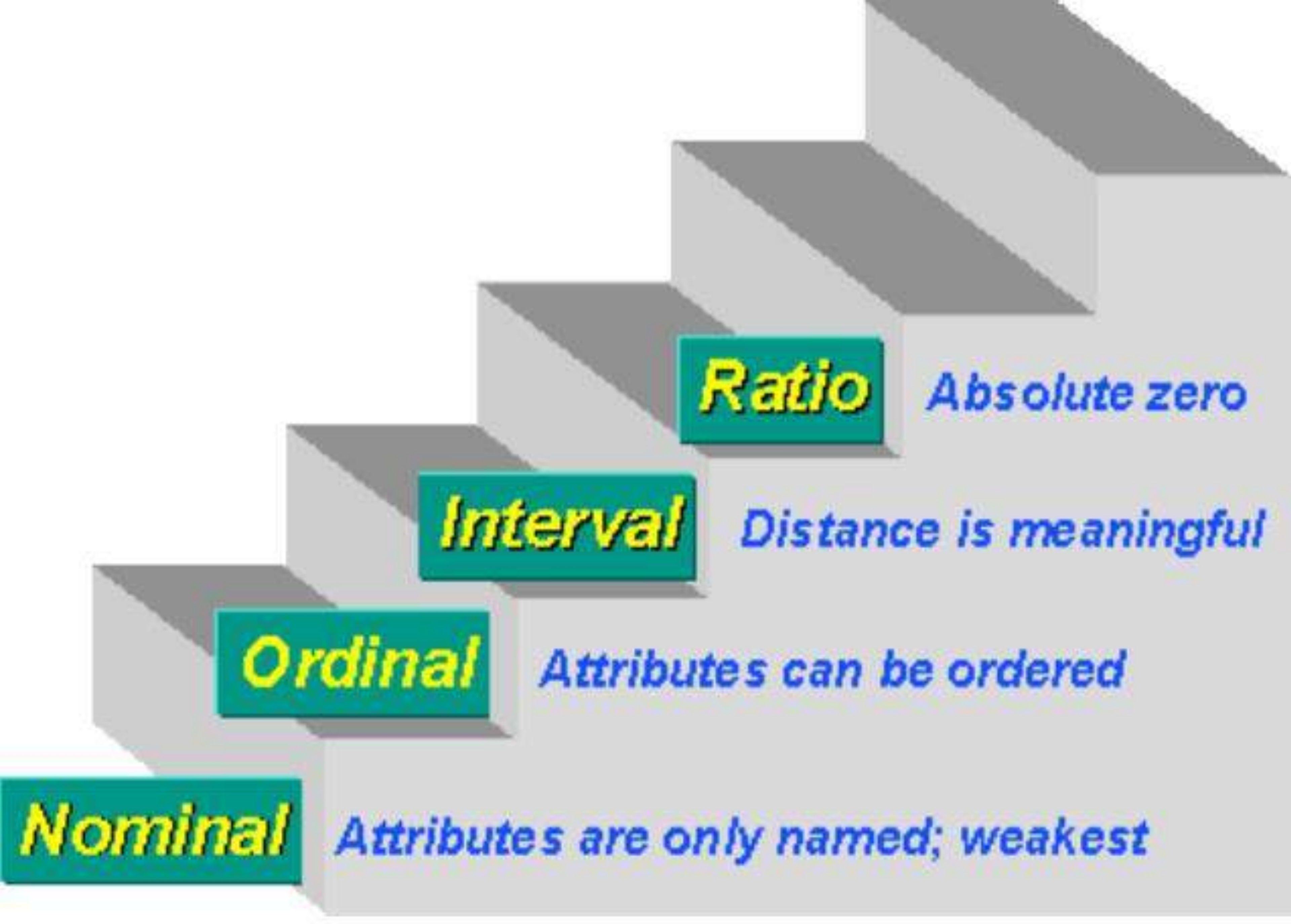
Examples

- Temperature in Kelvin (zero is the absence of heat. Can't get colder).
- Measurements of heights of students in this class (zero means complete lack of height).
- Someone 6 ft tall is twice as tall as someone 3 feet tall.
- Heart beats per minute has a very natural zero point. Zero means no heart beats.

What Statistic Can I Apply?

OK To Compute....	Ratio
Frequency Distribution.	Yes
Median And Percentiles.	Yes
Add Or Subtract.	Yes
Mean, Standard Deviation, Correlation, Regression, Analysis Of Variance	Yes
Ratio, Or Coefficient Of Variation.	Yes

Putting It Together

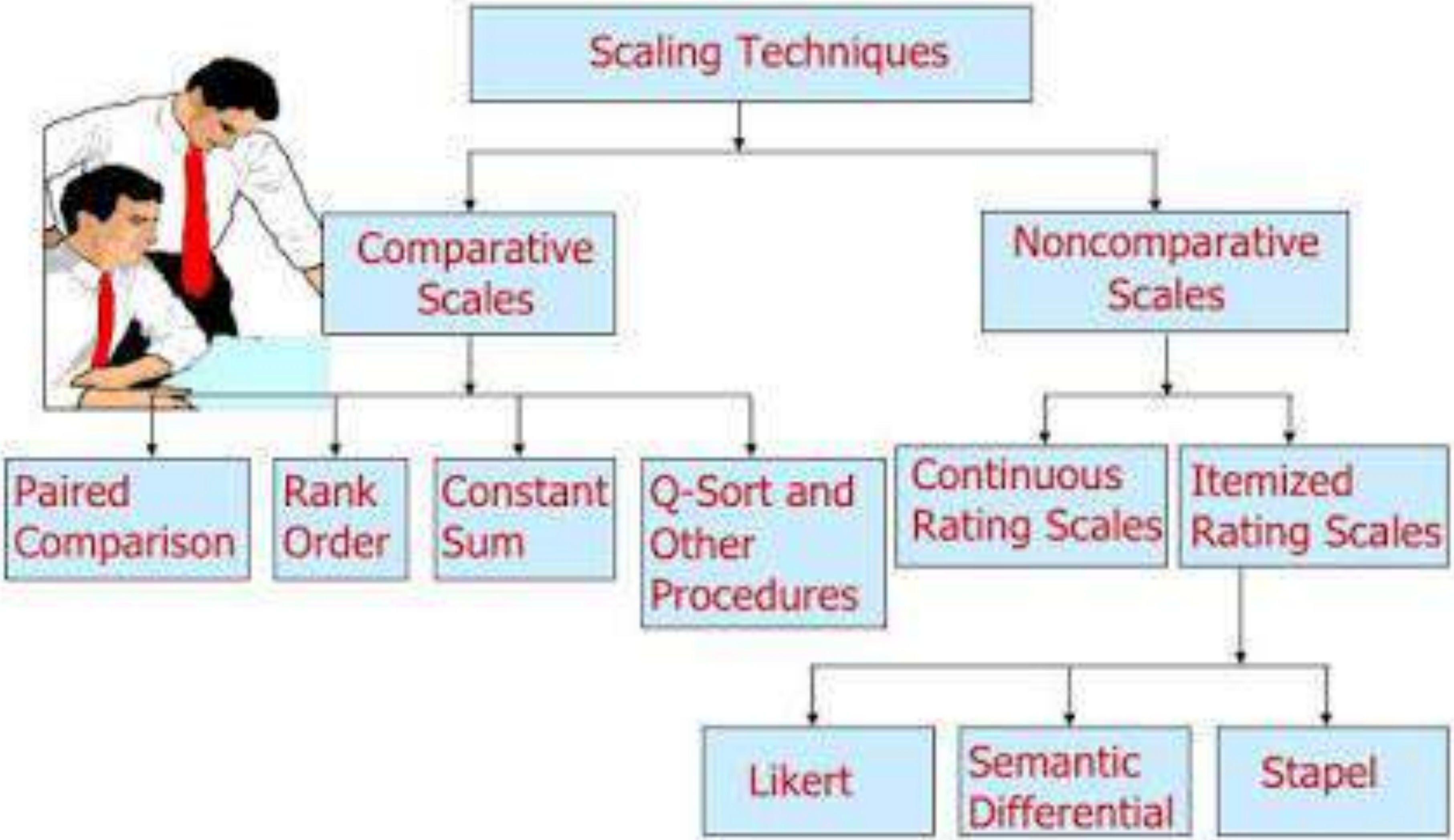

Ratio
Absolute zero
Interval
Distance is meaningful
Ordinal
Attributes can be ordered
Nominal
Attributes are only named; weakest

Summary of Levels of Measurement

Level of measurement	Put data in categories	Arrange data in order	Subtract data values	Determine if one data value is a multiple of another
Nominal	Yes	No	No	No
Ordinal	Yes	Yes	No	No
Interval	Yes	Yes	Yes	No
Ratio	Yes	Yes	Yes	Yes

A Classification of Scaling Techniques

Figure 8.2



Primary Scales of Measurement

Scale Fig. 8.1

Nominal Numbers Assigned to Runners

Finish

Ordinal Rank Order of Winners

Finish

Third place	Second place	First place

Interval Performance Rating on a 0 to 10 Scale

8.2	9.1	9.6

Ratio Time to Finish in Seconds

15.2	14.1	13.4

Primary Scales of Measurement

Table 8.1

Scale	Basic Characteristics	Common Examples	Marketing Examples	Permissible Statistics Descriptive	Inferential
Nominal	Numbers identify & classify objects	Social Security nos., numbering of football players	Brand nos., store types	Percentages, mode	Chi-square, binomial test
Ordinal	Nos. indicate the relative positions of objects but not the magnitude of differences between them	Quality rankings, rankings of teams in a tournament	Preference rankings, market position, social class	Percentile, median	Rank-order correlation, Friedman ANOVA
Interval	Differences between objects	Temperature (Fahrenheit)	Attitudes, opinions, index	Range, mean, standard	Product-moment
Ratio	Zero point is fixed, ratios of scale values can be compared	Length, weight	Age, sales, income, costs	Geometric mean, harmonic mean	Coefficient of variation

A Comparison of Scaling Techniques

- **Comparative scales** involve the direct comparison of stimulus objects. Comparative scale data must be interpreted in relative terms and have only ordinal or rank order properties.

- In **noncomparative scales**, each object is scaled independently of the others in the stimulus set. The resulting data are generally assumed to be interval or ratio scaled.

Relative Advantages of Comparative Scales

- Small differences between stimulus objects can be detected.
- Same known reference points for all respondents.
- Easily understood and can be applied.
- Involve fewer theoretical assumptions.
- Tend to reduce halo or carryover effects from one judgment to another.

Obtaining Shampoo Preferences Using Paired Comparisons

Fig. 8.3

Instructions: We are going to present you with ten pairs of shampoo brands. For each pair, please indicate which one of the two brands of shampoo you would prefer for personal use.

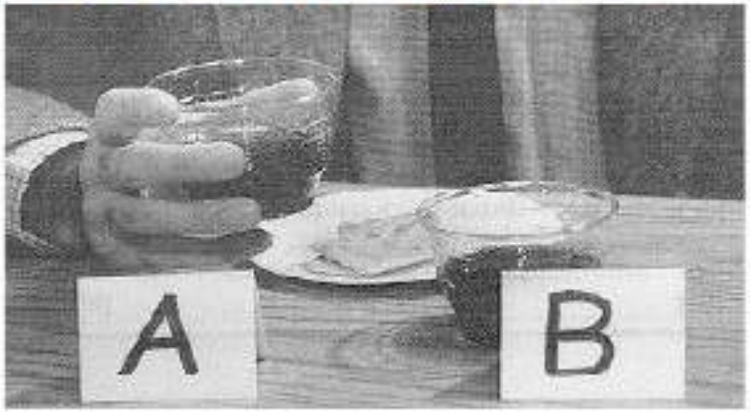
Recording Form:

	Jhirmack	Finesse	Vidal Sassoon	Head & Shoulders	Pert
Jhirmack		0	0	1	0
Finesse	1[a]		0	1	0
Vidal Sassoon	1	1		1	1
Head & Shoulders	0	0	0		0
Pert	1	1	0	1	
Number of Times Preferred[b]	3	2	0	4	1

[a]A 1 in a particular box means that the brand in that column was preferred over the brand in the corresponding row. A 0 means that the row brand was preferred over the column brand. [b]The number of times a brand was preferred is obtained by summing the 1s in each column.

Paired Comparison Selling

The most common method of taste testing is paired comparison. The consumer is asked to sample two different products and select the one with the most appealing taste. The test is done in private and a minimum of 1,000 responses is considered an adequate sample. A blind taste test for a soft drink, where imagery, self-perception and brand reputation are very important factors in the consumer's purchasing decision, may not be a good indicator of performance in the marketplace. The introduction of New Coke illustrates this point. New Coke was heavily favored in blind paired comparison taste tests, but its introduction was less than successful, because image plays a major role in the purchase of Coke.



A paired comparison taste test

Comparative Scaling Techniques
Rank Order Scaling

- Respondents are presented with several objects simultaneously and asked to order or rank them according to some criterion.
- It is possible that the respondent may dislike the brand ranked 1 in an absolute sense.
- Furthermore, rank order scaling also results in ordinal data.
- Only (n - 1) scaling decisions need be made in rank order scaling.

Preference for Toothpaste Brands Using Rank Order Scaling

Fig. 8.4

Instructions: Rank the various brands of toothpaste in order of preference. Begin by picking out the one brand that you like most and assign it a number 1. Then find the second most preferred brand and assign it a number 2. Continue this procedure until you have ranked all the brands of toothpaste in order of preference. The least preferred brand should be assigned a rank of 10.

No two brands should receive the same rank number.

The criterion of preference is entirely up to you. There is no right or wrong answer. Just try to be consistent.

Preference for Toothpaste Brands Using Rank Order Scaling

Fig. 8.4

Instructions: Rank the various brands of toothpaste in order of preference. Begin by picking out the one brand that you like most and assign it a number 1. Then find the second most preferred brand and assign it a number 2. Continue this procedure until you have ranked all the brands of toothpaste in order of preference. The least preferred brand should be assigned a rank of 10.

No two brands should receive the same rank number.

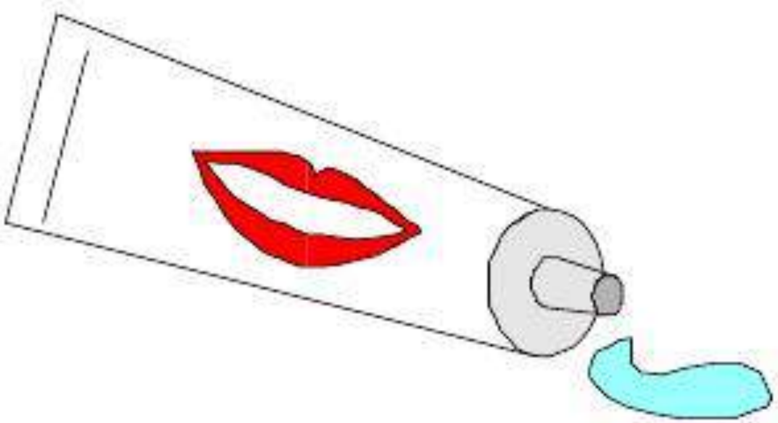
The criterion of preference is entirely up to you. There is no right or wrong answer. Just try to be consistent.

Preference for Toothpaste Brands Using Rank Order Scaling

Fig. 8.4 cont.

Form

Brand	Rank Order
1. Crest	________
2. Colgate	________
3. Aim	________
4. Gleem	________
5. Sensodyne	________
6. Ultra Brite	________
7. Close Up	________
8. Pepsodent	________
9. Plus White	________
10. Stripe	________



Comparative Scaling Techniques
Constant Sum Scaling

- Respondents allocate a constant sum of units, such as 100 points to attributes of a product to reflect their importance.
- If an attribute is unimportant, the respondent assigns it zero points.
- If an attribute is twice as important as some other attribute, it receives twice as many points.
- The sum of all the points is 100. Hence, the name of the scale.

Importance of Bathing Soap Attributes Using a Constant Sum Scale

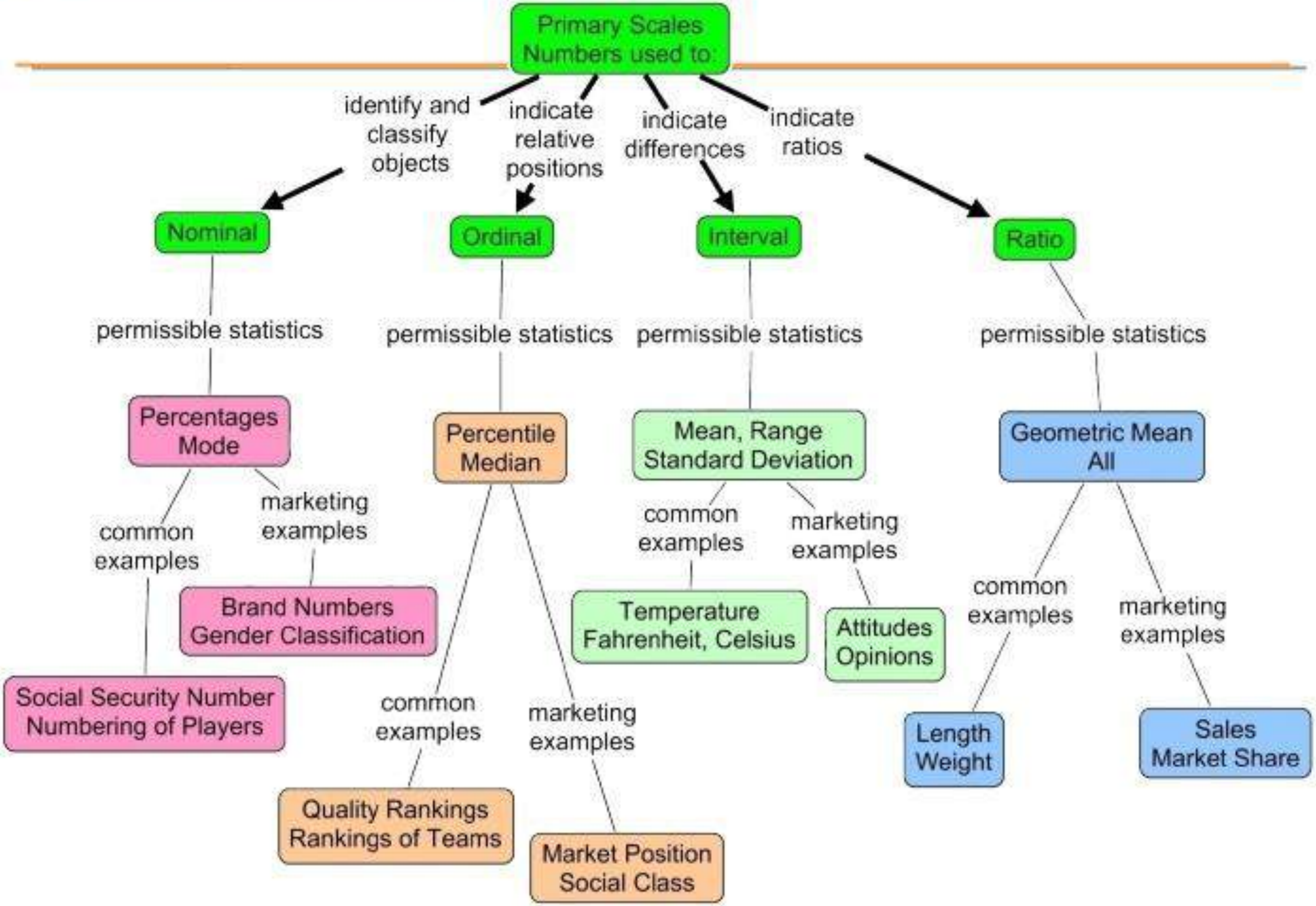
Fig. 8.5 cont.

Form

Average Responses of Three Segments

Attribute	Segment I	Segment II	Segment III
1. Mildness	8	2	4
2. Lather	2	4	17
3. Shrinkage	3	9	7
4. Price	53	17	9
5. Fragrance	9	0	19
6. Packaging	7	5	9
7. Moisturizing	5	3	20
8. Cleaning Power	13	60	15
Sum	100	100	100

Figure 8.7 A Concept Map for Primary Scales



Noncomparative Scaling Techniques

- Respondents evaluate only one object at a time, and for this reason **non-comparative scales** are often referred to as monadic scales.

- Non-comparative techniques consist of **continuous** and **itemized rating scales**.

Continuous Rating Scale

Respondents rate the objects by placing a mark at the appropriate position on a line that runs from one extreme of the criterion variable to the other.

The form of the continuous scale may vary considerably.

How would you rate Wal Mart as a department store?

Version 1

Probably the worst - - - - - - - -I - Probably the best

Version 2

Probably the worst - - - - - - - -I --Probably the best

0 10 20 30 40 50 60 70 80 90 100

Version 3

Very bad Neither good nor bad Very good

Probably the worst - - - - - - - -I ---Probably the best

0 10 20 30 40 50 60 70 80 90 100

Itemized Rating Scales

- The respondents are provided with a scale that has a number or brief description associated with each category.
- The categories are ordered in terms of scale position, and the respondents are required to select the specified category that best describes the object being rated.
- The commonly used itemized rating scales are the Likert, semantic differential, and Stapel scales.

Likert Scale

The **Likert scale** requires the respondents to indicate a degree of agreement or disagreement with each of a series of statements about the stimulus objects.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. Sears sells high-quality merchandise.	1	2X	3	4	5
2. Sears has poor in-store service.	1	2X	3	4	5
3. I like to shop at Sears.	1	2	3X	4	5

- The analysis can be conducted on an item-by-item basis (profile analysis), or a total (summated) score can be calculated.

- When arriving at a total score, the categories assigned to the negative statements by the respondents should be scored by reversing the scale.

Semantic Differential Scale

The **semantic differential** is a seven-point rating scale with end points associated with bipolar labels that have semantic meaning.

SEARS IS:

Powerful --:--:--:--:-X-:--:--: Weak

Unreliable --:--:--:--:--:-X-:--: Reliable

Modern --:--:--:--:--:--:-X-: Old-fashioned

- The negative adjective or phrase sometimes appears at the left side of the scale and sometimes at the right.
- This controls the tendency of some respondents, particularly those with very positive or very negative attitudes, to mark the right- or left-hand sides without reading the labels.
- Individual items on a semantic differential scale may be scored on either a -3 to +3 or a 1 to 7 scale.

A Semantic Differential Scale for Measuring Self-Concepts, Person Concepts, and Product Concepts

1) Rugged	:---:---:---:---:---:---:---:	Delicate
2) Excitable	:---:---:---:---:---:---:---:	Calm
3) Uncomfortable	:---:---:---:---:---:---:---:	Comfortable
4) Dominating	:---:---:---:---:---:---:---:	Submissive
5) Thrifty	:---:---:---:---:---:---:---:	Indulgent
6) Pleasant	:---:---:---:---:---:---:---:	Unpleasant
7) Contemporary	:---:---:---:---:---:---:---:	Obsolete
8) Organized	:---:---:---:---:---:---:---:	Unorganized
9) Rational	:---:---:---:---:---:---:---:	Emotional
10) Youthful	:---:---:---:---:---:---:---:	Mature
11) Formal	:---:---:---:---:---:---:---:	Informal

Stapel Scale

The **Stapel scale** is a unipolar rating scale with ten categories numbered from -5 to +5, without a neutral point (zero). This scale is usually presented vertically.

SEARS

+5	+5
+4	+4
+3	+3
+2	+2X
+1	+1
HIGH QUALITY	POOR SERVICE
-1	-1
-2	-2
-3	-3
-4X	-4
-5	-5

The data obtained by using a Stapel scale can be analyzed in the same way as semantic differential data.

Basic Noncomparative Scales

Table 9.1

Scale	Basic Characteristics	Examples	Advantages	Disadvantages
Continuous Rating Scale	Place a mark on a continuous line	Reaction to TV commercials	Easy to construct	Scoring can be cumbersome unless computerized
Itemized Rating Scales				
Likert Scale	Degrees of agreement on a 1 (strongly disagree) to 5 (strongly agree) scale	Measurement of attitudes	Easy to construct, administer, and understand	More time-consuming
Semantic Differential	Seven- point scale with bipolar labels	Brand, product, and company images	Versatile	Controversy as to whether the data are interval
Stapel Scale	Unipolar ten - point scale, - 5 to +5, without a neutral point (zero)	Measurement of attitudes and images	Easy to construct, administer over telephone	Confusing and difficult to apply

Summary of Itemized Scale Decisions

Table 9.2

1) Number of categories	Although there is no single, optimal number, traditional guidelines suggest that there should be between five and nine categories
2) Balanced vs. unbalanced	In general, the scale should be balanced to obtain objective data
3) Odd/even no. of categories	If a neutral or indifferent scale response is possible for at least some respondents, an odd number of categories should be used
4) Forced vs. non-forced	In situations where the respondents are expected to have no opinion, the accuracy of the data may be improved by a non-forced scale
5) Verbal description	An argument can be made for labeling all or many scale categories. The category descriptions should be located as close to the response categories as possible
6) Physical form	A number of options should be tried and the best selected

Balanced and Unbalanced Scales

Fig. 9.1

Balanced Scale		Unbalanced Scale	
Jovan Musk for Men is:		Jovan Musk for Men is:	
Extremely good	______	Extremely good	______
Very good	______	Very good	______
Good	______	Good	______
Bad	______	Somewhat good	______
Very bad	______	Bad	______
Extremely bad	______	Very bad	______

Rating Scale Configurations

Fig. 9.2

Cheer detergent is:

1) Very harsh --- --- --- --- --- --- --- Very gentle

2) Very harsh 1 2 3 4 5 6 7 Very gentle

3) . Very harsh
.
.
. Neither harsh nor gentle
.
.
. Very gentle

4) ____ Very harsh ____ Harsh ____ Somewhat harsh ____ Neither harsh nor gentle ____ Somewhat gentle ____ Gentle ____ Very gentle

-3 -2 -1 0 +1 +2 +3

5)

Very harsh Neither harsh nor gentle Very gentle

TOPIC: CORRELATION

C.SATHYALAKSHMI,
ASSOCIATE PROFESSOR OF
HOMESCIENCE COLLEGE

CORRELATION

DEFINITION

- The variables are said to be correlated if the changes in one variable results in a corresponding change in the other variable. That is, when two variables move together we say they are correlated.
- Boddington states that " whenever some definite connection exists between the two or more groups, classes or series or data there is said to be correlation".
- Bowely defines correlation as, " when two quantities are so related that the fluctuations in one are in sympathy with the fluctuations of the other, that an increase or decrease of the one is found in connection with the increase or decrease of the other and greater the magnitude of change in the other, the quantities are said to be correlated"

- According to A. M Tuttle, “correlation is an analysis of the association between two or more variables.

- Simply, correlation may be defined as the degree of relationship between two variables.

- “Correlation analysis” the purpose of which is the determination of degree of relationship between the variables

- The method of correlation is developed by FRANCIS GALTON in 1885.

TYPES OF CORRELATION

The different types of correlation are

- Positive and Negative correlation
- Linear and Non-linear correlation
- Simple, Multiple and Partial correlation.

- Positive Correlation

  When the values of two variables move same direction, correlation is said to be positive

  ie; an increase in the value of one variable results into an increase in the other variable also or if decrease in the value of one variable results into a decrease in the other variable also correlation is said to be positive.

  Eg. Temperature and volume

- Negative correlation

  When the values of two variables move opposite direction, correlation is said to be negative.

  ie; an increase in the value of one variable results into an decrease in the other variable also or if decrease in the value of one variable results into a increase in the other variable also correlation is said to be positive.

  Eg. Pressure and volume

- Linear Correlation

  When the amount of change in one variable leads to a constant ratio of change in the other variable, correlation is said to be linear.

- Non linear Correlation

  Correlation is said to be non linear (curve linear) when the amount of change in one variable is not in constant ratio to the change in the other variable.

- Simple correlation

    In the study of relationship between the variables, if there are only two variables, the correlation is said to be simple.

    When one variable is related to a number of others, the correlation is not simple.

- Multiple correlation

    In the study of multiple correlation we measure the degree of association between one variable on one side and all the other variable together on the other side.

- Partial correlation

    In partial correlation we study the relationship of one variable with one of the other variables presuming that the other variable remains constant.

Degree of correlation

The degree or the intensity of the relationship between two variables can be ascertained by finding the value of coefficient of correlation. The degree of correlation can be classified into

- Perfect correlation

When the change in the two variables is such that with an increase in the value of one, the value of the other increases in a fixed proportion, correlation is said to be perfect. The perfect correlation may be positive or negative. Coefficient of correlation is +1 for perfect positive correlation and it is –1 for perfect negative correlation.

- No correlation

If the changes in the value of one variable are in association with the changes in the value ot other variablether will be no correlation. When there is no correlation the coefficient of correlation is zero.

- Limited degree of correlation

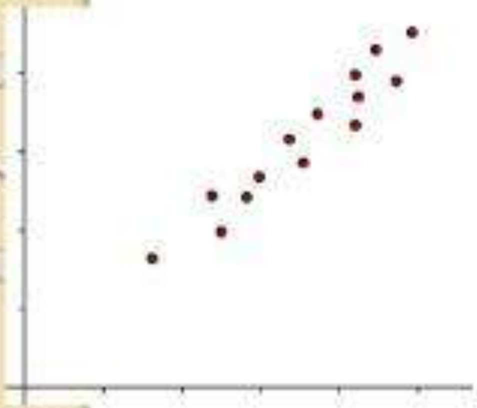
In between no correlation and perfect correlation there may be limited degree of correlation. It may also be positive or negative. Limited degree of correlation may be termed as high, moderate or low. For limited degree of correlation the coefficient of correlation lies between 0 and 1 numerically.

METHODS FOR STUDYING CORRELATION

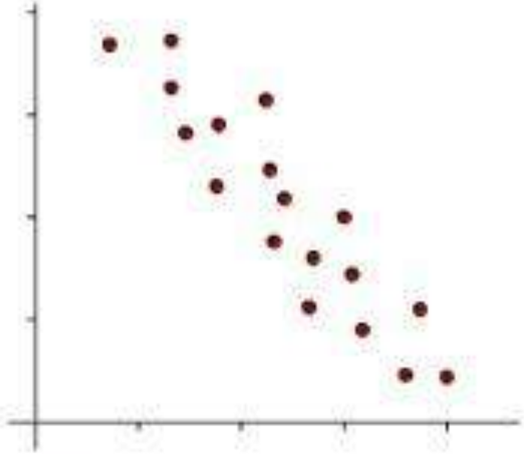
Correlation between two variables can be measured by both graphic and algebraic method. **Scatter diagram** and **correlation graph** are the two important graphic methods while **coefficient of correlation** is an algebraic method used for measuring correlation.

a) Scatter diagram

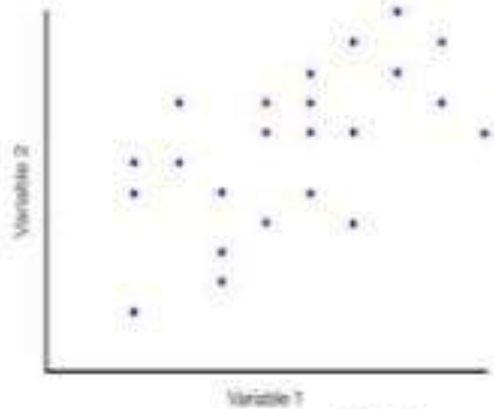
This is a graphical method of studying the correlation between two variables. One of the variable is shown on the X- axis and the other on the Y-axis. Each pair of values is plotted on the graph by means of a dot mark. After all the items are plotted we get as many dots on the graph paper as the number of points. If these points show some trend either upward or downward, the two variables are said to be correlated. If the point do not show any trend, the two variables are not correlated.



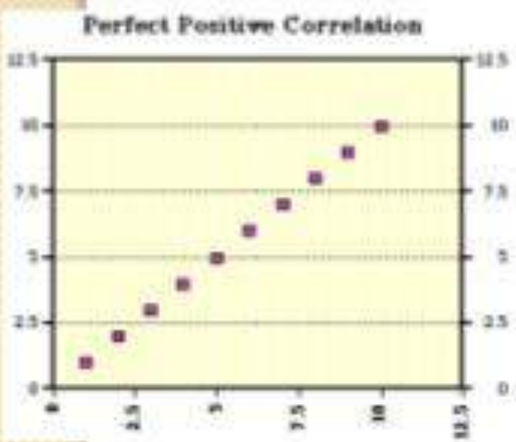
positive correlation



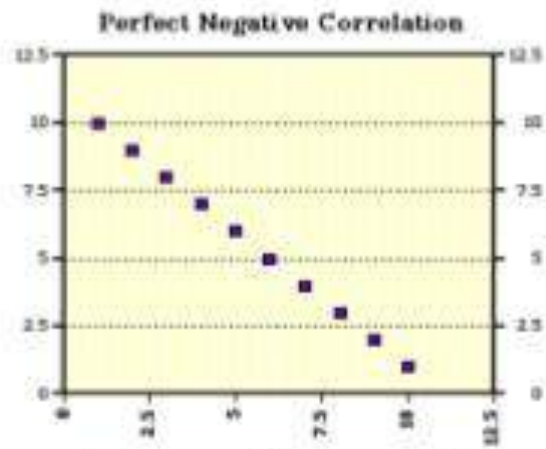
Negative correlation



zero correlation



perfect positive correlation



perfect negative correlation

b) Correlation Graph

Under this method, separate curves are drawn for the X variable and Y variable on the same graph paper. The values of the variable are taken as ordinates of the points plotted. From the direction and closeness of the two curves we can infer whether the variables are related. If both the curves are move in the same direction(upward or downward), correlation is said to be positive. If the curves are moving in the opposite direction correlation is said to be negative.

c) Coefficient of correlation

- Coefficient of correlation is an algebraic method of measuring correlation.
- Under this method, we measure correlation by finding a value known as the coefficient of correlation using an appropriate formula.
- Coefficient of correlation is a numerical value. It shows the degree or the extent of correlation between two variables.

- Coefficient of correlation is a pure number lying between -1 and +1.
- When the correlation is negative, it lies between -1 and 0.
- When the correlation is positive, it lies between 0 and 1.
- When the correlation of coefficient is zero, it indicates that there is no correlation between the variables.
- When the correlation coefficient is 1,there is perfect correlation.
- Between no correlation and perfect correlation there are varying degree of correlation.

Coefficient of correlation can be computed by applying the methods given below

- Karl Pearson's method
- Spearman's method
- Concurrent deviation method

PROPERTIES OF COEFFICIENT OF CORRELATION

1. Correlation coefficient has a well defined formula
2. Correlation coefficient is a pure number and is independent of its units of measurement.
3. It lies between -1 and +1.
4. Correlation coefficient does not change with reference to change of origin or change of scale.
5. Correlation of coefficient between x and y is same as that between y and x.

IMPORTANCE OF CORRELATION

- Correlation helps to study the association between two variables.
- Coefficient of correlation is vital for all kinds of research work.
- It helps in establishing Validity or Reliability of an evaluation tool.
- It helps to ascertain the traits and capacities of pupils while giving guidance or counselling.
- Correlation analysis helps to estimate the future values.

? What would be your interpretation if the correlation coefficient equal to

1) $r = 0$

Ans : There is no correlation between the variables

2) $r = -1$

Ans: negative perfect correlation

3) $r = 0.2$

Ans: low positive correlation

4) $r = 0.9$

Ans: high positive correlation

5) $r = -0.3$

Ans: low negative correlation

6) $r = -0.8$

Ans: High negative correlation

CHECKLIST AND RATING SCALE

OBSERVATION CHECKLIST

- It is the most commonly used instrument for performance evaluation.
- A checklist enables the observer to note to only whether a trait is present or not.
- If consist of a listening of steps, activities or behaviours the observer records when an incident occurs.
- The observer has to judge a certain behaviour has taken place.

DEFINITION OF CHECKLIST

- A checklist is a simple instrument consisting prepared list of expected items of performance or attributes which are checked by an evaluator for the presence or absence.

- Checklist are constructed by breaking a performance and quality of a product, which specifies the presence or absence of an attribute or trait which is then 'checked' by the traitor or observer.

CHARACTERISTICS OF CHECKLIST

- Observe one respondent at one time
- Clearly specify the characteristics of the behaviour to the observer
- Use only carefully prepared checklists to avoid more complex traits.
- The observer should be trained how to observe, how to record the observed behaviour.
- Use checklist only when you are interested in calculating in particular characteristics

Procedures Testing Checklist - (Sample Checklist)

Sr.	Check Point / Defect Statement	Check Mark (√) the Appropriate Column	
		Yes	N/A
01	Whether start-up procedures have been validated?		
02	Whether query procedures have been validated?		
03	Whether file-mounting procedures have been validated?		
04	Whether updating procedures have been validated?		
05	Whether backup procedures have been validated?		
06	Whether off-site storage procedures have been validated?		
07	Whether recovery procedures have been validated?		
08	Whether terminal operating procedures have been validated?		
09	Whether procedures needed to operate the terminal when the main computer is down have been validated?		
10	Whether procedures needed to capture data when the terminals are down have been validated?		

CONSTRUCTION OF A CHECKLIST

- Express each item in clear, simple language.
- The type of checklist to be used in particular assessment/ evaluation
- The list of items in the checklist may be continuous or divided into groups of related items.
- These lists of the items are formulated on the basis of the judgement of experts and each item is evaluated with respect to the number of favourable and unfavourable responses.

- Avoid negative statement whenever possible.
- Avoid lifting statements verbatim from the text.
- Ensure that each item has a clear response: yes or no, true or false.
- Review the items independently.
- Checklists must have a quality of completeness and comprehensiveness.

ADVANTAGES OF CHECKLIST

- A checklist allows inter individual comparisons.
- They provide a simple method to record observations.
- They are adaptable to subject matter areas.
- It is useful in evaluating learning activities expected to be performed.
- They are helpful in evaluating procedure work.
- Properly prepared checklists allow the observer to constrains the direct attention.
- Checklist have objectively to evaluate the characteristics.
- Decreases the chances of errors in observation

DISADVANTAGES OF CHECKLIST

- Checklist don't indicate quality of performance so the usefulness of checklists is limited.
- Only a limited component of overall clinical performance can be evaluated.
- Only the presence or absence of an attribute, behaviour or performance or parameter may be assured.
- It has limited use in qualitative observations.
- Checklist are not easy to prepare.

RATING SCALE

RATING SCALE

- Rating is the term used to express opinion or judgement regarding some performance of a person, object, situation and character.
- The rating scale involves qualitative description of a limited number of aspects of a thing or traits of a person.

DEFINITION OF RATING SCALE

- Rating scale refers to a scale with a set of opinion, which describes varying degree of the dimensions of an attitude or a phenomenon being observed.

- Rating scale is a device by which judgements may be qualified or an opinion concerning a trait can be systematized. -Suresh. K. Sharma.

- A rating scale is a tool in which one person simply checks off another person's level of performance.
- It could be a 3- point, 5- point or 7- point rating scale.
- Rating scalers measure how much or how well something happened, where generally quantitative and qualitative terms are used to judge the performance.
- A wide variety of attributes may be assessed by using rating scales.

TYPES OF RATING SCALES

- **A)Graphic rating scale:**
- **B)Descriptive rating scale**
- **C)Numerical rating scale**
- **D)Comparative rating scale**

TYPES OF RATING SCALES

- **Graphic rating scale:**
- In this scale, the performance is printed horizontally at various points from lowest to highest. It includes the numerical point on the scales. It is anchored by two extremes presented to the respondents for the evaluation of a concept or object.

Graphic Rating Scale: Performance is assessed along one or more continua with specified intervals.

Example: A supervisor of a nurse answers the question"How would you rate the quality of care this nurse provides to patients?"

B)Descriptive rating scale

- This type of rating scale does not use numbers but divides the assessment into series of verbal phrases to indicate the level of performance.

- In which descriptive phrases or terms assigned to each trait. The rater enters the appropriate phrase after each name to indicate judgement of the person
- .A B C D E

A	B	C	D	E
Excellent	Good	Average	Below average	poor

Numerical rating scale

- It divides the evaluation criteria into a fixed number of point and defines only numbers except at the extremes. In these scalers, each statement is generally assigned a numerical score ranging from 1- 10 or even more.
- E.g.: - pain assessment numerical scale

Comparative rating scale

- In this type of rating scale, the rater makes a judgement about a person's attributes by a comparing with another similar person.

Please divide 100 points between the following attributes of compact disc players according to the relative importance of each attribute to you.

Sound Quality	______
Physical Size	______
Brand Name	______
Durability	______
	100 points

CHARACTERISTICS OF A RATING SCALE

- They are value judgements of the attributes of one person by another person.
- These scales are most commonly used tools to carry out structured observations.
- They are generally developed to make quantitative judgements about qualitative attributes.
- They provide more flexibility to judge the level of performance or presence of attributes among subjects.

- **GUIFORD [1954] CHARACTERISTICS.**
- **Clarity:** it must be constructed using short, concise statements in simple and unambiguous language.
- **Relevance:** the statement should be relevant to the phenomenon and should be exactly in accordance with the variables understudy.
- **Variety:** monotony of the statements must be avoided and variety and difference statements must be ensured.
- **Objectivity:** it must be objective in a nature so that it is convenient for the rater to judge the attributes or performances of the subjects under study.
- **Uniqueness:** each statement constructed must be unique in itself so that the attributers can be judged appropriately.

ADVANTAGES OF A RATING SCALE

- Rating scale are easy to administer and score the measured attributes.
- They have a wide range of application in nursing educational; evaluation.
- Graphic rating scale in easier to make and less time- consuming.
- Rating can be easy used to a large group.
- They are also used for quantitative methods.
- Assessment of interest and attitudes and personal characteristics.
- They are used to evaluate performance, skills and product outcomes.
- Rating scales are adaptable and flexible assessment instruments

DISADVANTAGES OF RATING SCALES

- It is difficult or dangerous to fix up rating about aspects of an individual.
- Misuse can result in a decrease in objectivity.
- There are chances of subjective evaluation thus the scales may become unscientific and un reliable.

BIBLIOGRAPHY

- SURESH K. SHARMA, REENA SHARMA "COMMUNICATION AND EDUCATIONAL TECHNOLOGY" 2ND EDITION, ELSEVIER PUBLICATION, PAGE NO: 376-380.
- B. T BASAVANTHAPPA, "NURSING EDUCATION", 1ST EDITION, 2003, JAYPEE PUBLICATIONS, PAGE NO: 508-514.
- R. SUDHA, "NURSING EDUCATION, PRINCIPLES AND CONCEPTS" 2013, JAYPEE PUBLICATIONS, PAGE NO:181-184.
- NEELAM KUMARI, "A TEXT BOOK OF COMMUNICATION AND EDUCATIONAL TECHNOLOGY," 2014, PEEVEE BOOKS PRIVATE LIMITED PAGE NO: 279-282.
- BARBARA CHERRY, SUSAN.R. JACOB, "CONTEMPORARY NURSING ISSUES, TRENDS AND MANAGEMENT," 5TH EDITION, MOSBY ELSEVIER PUBLICATIONS, PAGE NO: 503-504.
- PAMELA EVANS SMITH, MARILEELEBON, "SKILL CHECKLIST TO ACCOMPANY TAYLORS CLINICAL NURSING SKILLS A NURSING PROCESS APPROACH, LIPPINCOTT WILLIAMS AND WILKINS PUBLICATIONS.
- D. ELAKUVANABHASKARA RAJ, NIMA BHASKER, "A TEXT BOOK OF NURSING EDUCATION" EMMESS PUBLICATIONS, PAGE NO: 282-283.

- **NET REFERENCE**
- www.wikipedia.org
- www.kkhsource.in
- www.eduhk.hk
- https://content.wisestep.com
- https://www.slideshare.net

- **JOURNEL REFERENCE**
- www.nejm.org
- www.Bcg.perpectives
- www.geopolitica.hu/en
- Sgo.sagepub.com
- https://www.ncbi.nlm.nih.gov

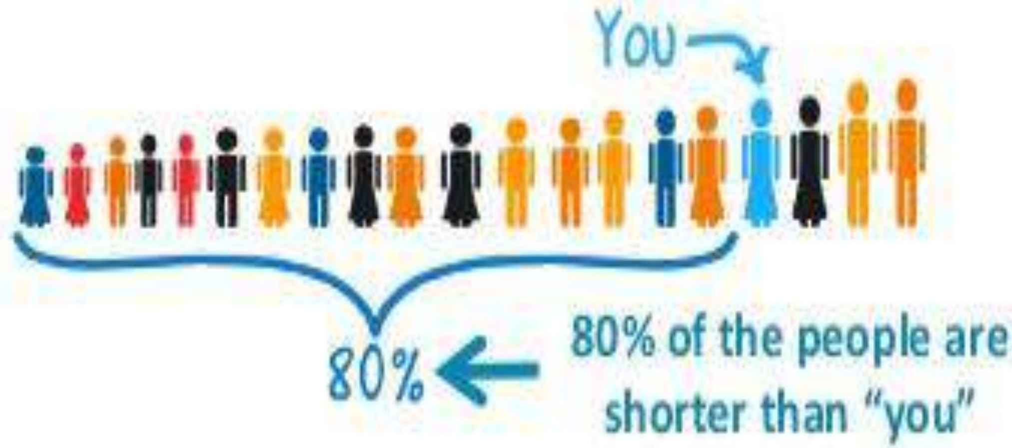
PERCENTILES

Percentiles are positional measures used mainly in <u>educational</u> and <u>health-related</u> fields to indicate the position of an individual in a group.

For example, the graphs and tables show the percentiles for various measures such as test scores, height or weight.

Percentiles (denoted P_x) divide a set of data into 100 equal parts.

They are used as positional measures to indicate what percent of the data set have a value <u>less than a specified value</u>.



Percentiles <u>are not the same</u> as percentages.

For example, if a student gets 72 correct answers out of 100 on a test they earn 72%.

If a score of 72 correct answers corresponds to the 64th percentile, then they did better than 64% of the students in the class, but still received a score of 72%.

Special Cases

Some frequently used percentiles have specific names:

- The median is the value that is in the halfway position of a data set. Median = P_{50}
- The quartiles are the values that are in the quarter positions of a data set.
  First Quartile = $Q_1 = P_{25}$
  Third Quartile = $Q_2 = P_{75}$
- The deciles are the values that are in the positions that divided the data into 10 pieces.
  $D_1 = P_{10}$
  $D_2 = P_{20}$
  ...
  $D_{10} = P_{100}$

Formula

Provided data is ordered least to greatest, the **position number of the percentiles** can be calculated using the formula:

$$\text{Position number of } \mathrm{P}_x = \frac{n+1}{100}(x)$$

With percentiles, there are two items of interest:

1. the position of the percentile (found using the formula), and

2. the value of the percentile (for example, P50 = 139, means the 50th percentile has a value

Calculating Percentiles

The following data lists the number of calories in 22 manufactures vanilla-flavoured ice cream bars.

111	131	147	151	151	179
182	190	197	201	209	234
286	294	295	310	319	337
353	377	377	439		

Given the data above determine:

a) the median

b) the first quartile

c) the third quartiles

d) the fourth decile

Calculating Percentiles

a) the median

$$\text{Position number of } P_{50} = \frac{22+1}{100}(50)$$
$$= 11.5$$

111	131	147	151	151	179
182	190	197	201	209	234
286	294	295	310	319	337
353	377	377	439		

11^{th} & 12^{th} positions

$$P_{50} = 11^{th} \text{ value} + 0.5\left(12^{th} \text{ value} - 11^{th} \text{ value}\right)$$
$$= 209 + 0.5(234 - 209)$$
$$= 221.5$$

Calculating Percentiles

b) the first quartile

$$\text{Position number of } P_{25} = \frac{22+1}{100}(25)$$
$$= 5.75$$

5th & 6th positions

111	131	147	151	151	179
182	190	197	201	209	234
286	294	295	310	319	337
353	377	377	439		

$$P_{25} = 5^{\text{th}} \text{ value} + 0.75\left(6^{\text{th}} \text{ value} - 5^{\text{th}} \text{ value}\right)$$
$$= 151 + 0.75(179 - 151)$$
$$= 172$$

Calculating Percentiles

c) the third quartile

$$\text{Position number of } P_{75} = \frac{22+1}{100}(75)$$
$$= 17.25$$

111	131	147	151	151	179
182	190	197	201	209	234
286	294	295	310	319	337
353	377	377	439		

17th & 18th positions

$$P_{75} = 17^{\text{th}} \text{ value} + 0.25\left(18^{\text{th}} \text{ value} - 17^{\text{th}} \text{ value}\right)$$
$$= 319 + 0.25(337 - 319)$$
$$= 323.5$$

Calculating Percentiles

d) the fourth decile

$$\text{Position number of } P_{40} = \frac{22+1}{100}(40)$$
$$= 9.2$$

111	131	147	151	151	179
182	190	197	201	209	234
286	294	295	310	319	337
353	377	377	439		

9th & 10th positions

$$P_{40} = 9^{\text{th}} \text{ value} + 0.2\left(10^{\text{th}} \text{ value} - 9^{\text{th}} \text{ value}\right)$$
$$= 197 + 0.2\left(201 - 197\right)$$
$$= 197.8$$

ANalysis Of VAriance

Contents

- Introduction – Various statistical tests
- What is ANOVA?
- One way ANOVA
- Two way ANOVA
- MANOVA (Multivariate ANalysis Of VAriance)
- ANOVA with repeated measures
- Other related tests
- References

Summary Table of Statistical Tests

Level of Measurement	Sample Characteristics					Correlation
	1 Sample	2 Sample		K Sample (i.e., >2)		
		Independent	Dependent	Independent	Dependent	
Categorical or Nominal	X^2 or bi-nominal	X^2	Macnarmar's X^2	X^2	Cochran's Q	
Rank or Ordinal	X^2	Mann Whitney U	Wilcoxin Matched Pairs Signed Ranks	Kruskal Wallis H	Friedman's ANOVA	Spearman's rho
Parametric (Interval & Ratio)	z test or t test	t test between groups	t test within groups	1 way ANOVA between groups	1 way ANOVA (within or repeated measure)	Pearson's r
		Factorial (2 way) ANOVA				

What is ANOVA

- ✓ Statistical technique specially designed to test whether the **means** of more than 2 quantitative populations are equal.
- ✓ Developed by Sir Ronald A. Fisher in 1920's.

EXAMPLE: Study conducted among men of age group 18-25 year in community to assess effect of SES on BMI

Lower SES	Middle SES	Higher SES
18,17,18,19,19	22,25,24,26,24,21	25,26,24,28,29
N1= 5	N2= 6	N3= 5
Mean=18.2	Mean= 23.6	Mean=26.4

ANOVA

One way ANOVA	Two way ANOVA	Three way ANOVA
Effect of SES on BMI	Effect of age & SES on BMI	Effect of age, SES, Diet on BMI

ANOVA with repeated measures - comparing >=3 group means where the participants are same in each group. E.g.

Group of subjects is measured more than twice, generally over time, such as patients weighed at baseline and every month after a weight loss program

One Way ANOVA

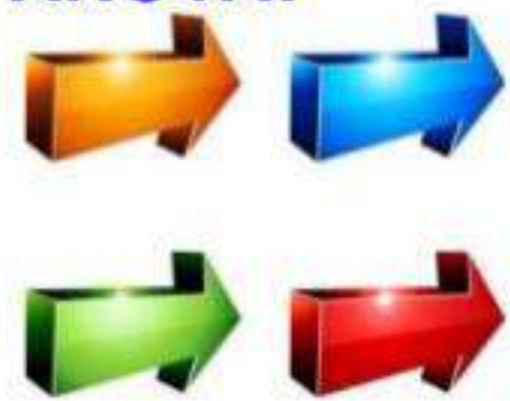
Data required

One way ANOVA or single factor ANOVA:

- Determines means of

≥ 3 independent groups

significantly different from one another.



- Only 1 independent variable (factor/grouping variable) with ≥3 levels
- Grouping variable- nominal
- Outcome variable- interval or ratio

Post hoc tests help determine where difference exist

Assumptions

1) **Normality:** The values
normally distributed.

Skewness
Kurtosis
Kolmogorov-Smirnov
Shapiro-Wilk test
Box-and-whiskers plots
Histogram

2) **Homogeneity of vari**
within each group should be equal for all groups.

3) **Independence of error:** The error(variation of each value around its own group mean) should be independent for each value.

Steps

1. State null & alternative hypotheses

2. State Alpha
3. Calculate degrees of Freedom
4. State decision rule

5. Calculate test statistic
- Calculate variance between samples
- Calculate variance within the samples
- Calculate F statistic
- If F is significant, perform *post hoc* test

6. State Results & conclusion

1. State null & alternative hypotheses

$$H_0 = \mu_1 = \mu_2 ... = \mu_i$$

H0 : all sample means are equal

$$H_a = \text{not all of the } \mu_i \text{ are equal}$$

At least one sample has different mean

2. State Alpha i.e 0.05

3. Calculate degrees of Freedom

K-1 & n-1

k= No of Samples,

n= Total No of observations

4. State decision rule

If calculated value of F >table value of F, reject Ho

5. Calculate test statistic

Calculating variance between samples

1. Calculate the **mean** of each sample.
2. Calculate the **Grand average**
3. Take the difference between means of various samples & grand average.
4. Square these deviations & obtain total which will give sum of squares between samples **(SSC)**
5. Divide the total obtained in step 4 by the degrees of freedom to calculate the mean sum of square between samples **(MSC)**.

Calculating Variance within Samples

1. Calculate **mean** value of each sample
2. Take the deviations of the various items in a sample from the mean values of the respective samples.
3. Square these deviations & obtain total which gives the sum of square within the samples **(SSE)**
4. Divide the total obtained in 3rd step by the degrees of freedom to calculate the mean sum of squares within samples **(MSE).**

The mean sum of squares

Calculation of **MSC**-
Mean sum of **S**quares
between samples

$$MSC = \frac{SSC}{k-1}$$

Calculation of **MSE**
Mean Sum Of
Squares within
samples

$$MSE = \frac{SSE}{n-k}$$

k= No of Samples, n= Total No of observations

Calculation of F statistic

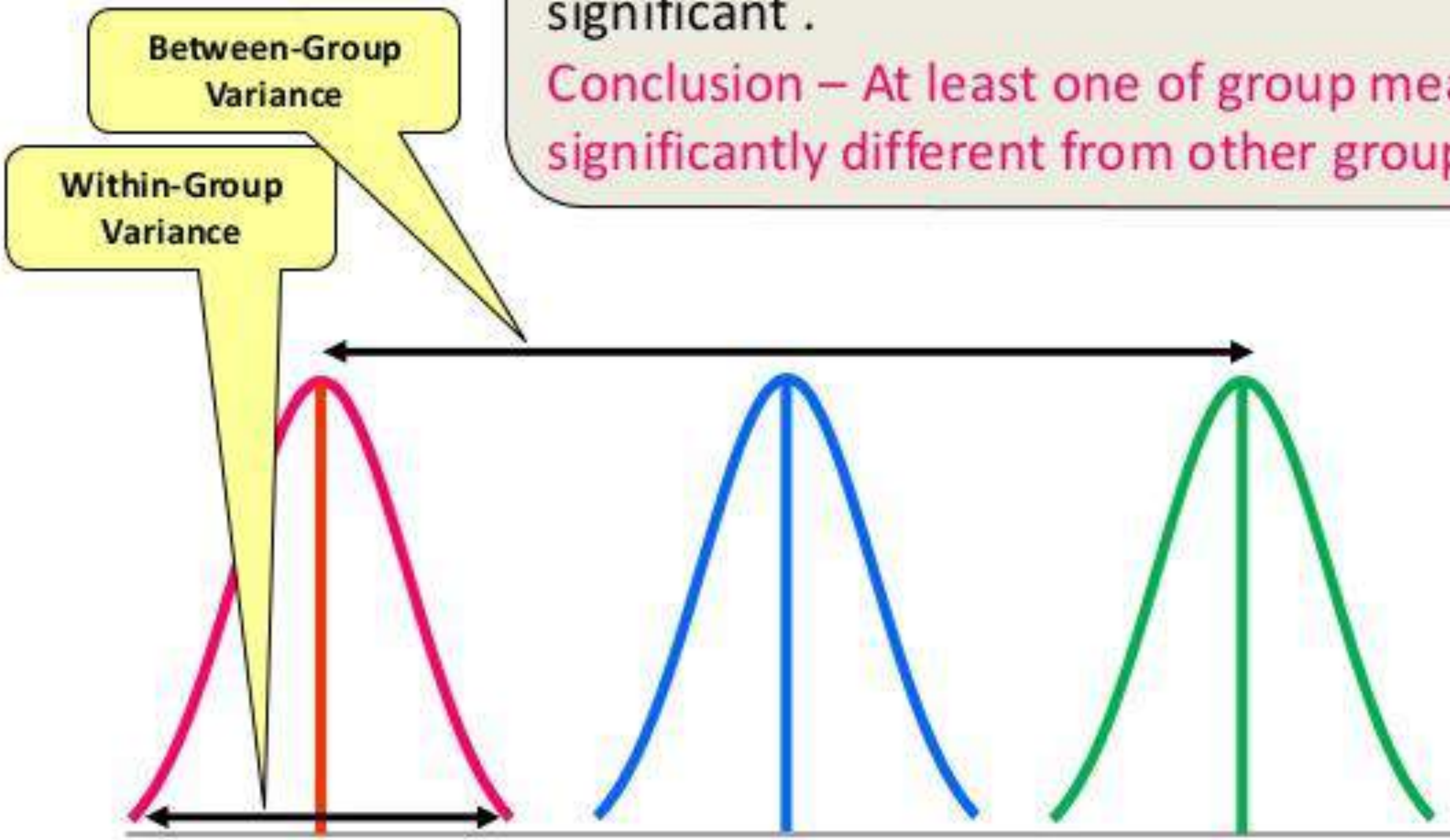
$$F = \frac{Variability\ between\ groups}{Variability\ within\ groups}$$

$$\text{F- statistic} = \frac{MSC}{MSE}$$

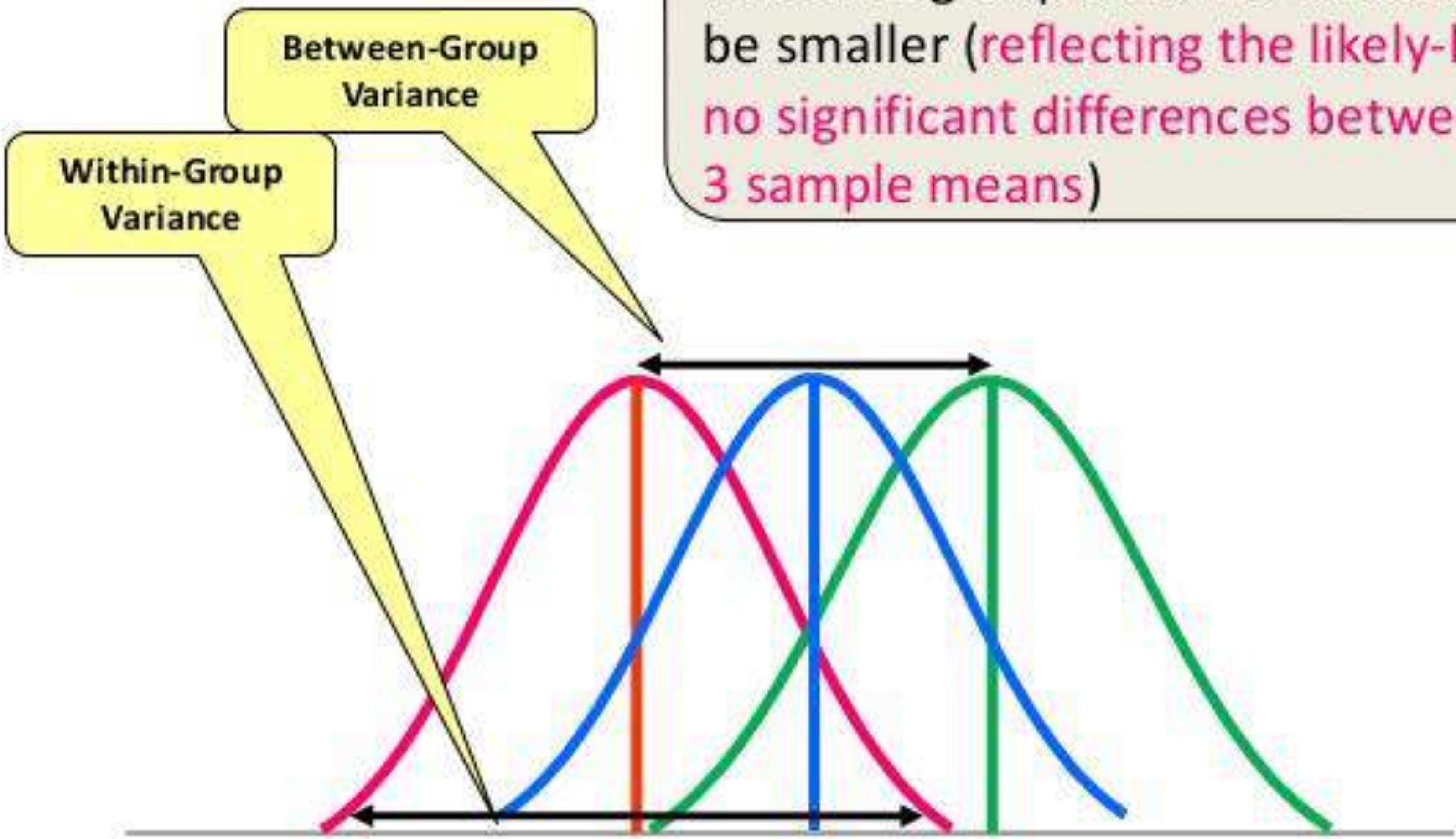
Compare the F-statistic value with F(critical) value which is obtained by looking for it in F distribution tables against degrees of freedom. The calculated value of F > table value H0 is rejected

Between-group variance is large relative to the within-group variance, so F statistic will be larger & > critical value, therefore statistically significant .

Conclusion – At least one of group means is significantly different from other group means



Within-group variance is larger, and the between-group variance smaller, so F will be smaller (reflecting the likely-hood of no significant differences between these 3 sample means)



Post-hoc Tests

- Used to determine which mean or group of means is/are significantly different from the others (**significant *F***)
- Depending upon research design & research question:

✓ **Bonferroni** (more powerful)

Only some pairs of sample means are to be tested

Desired alpha level is divided by no. of comparisons

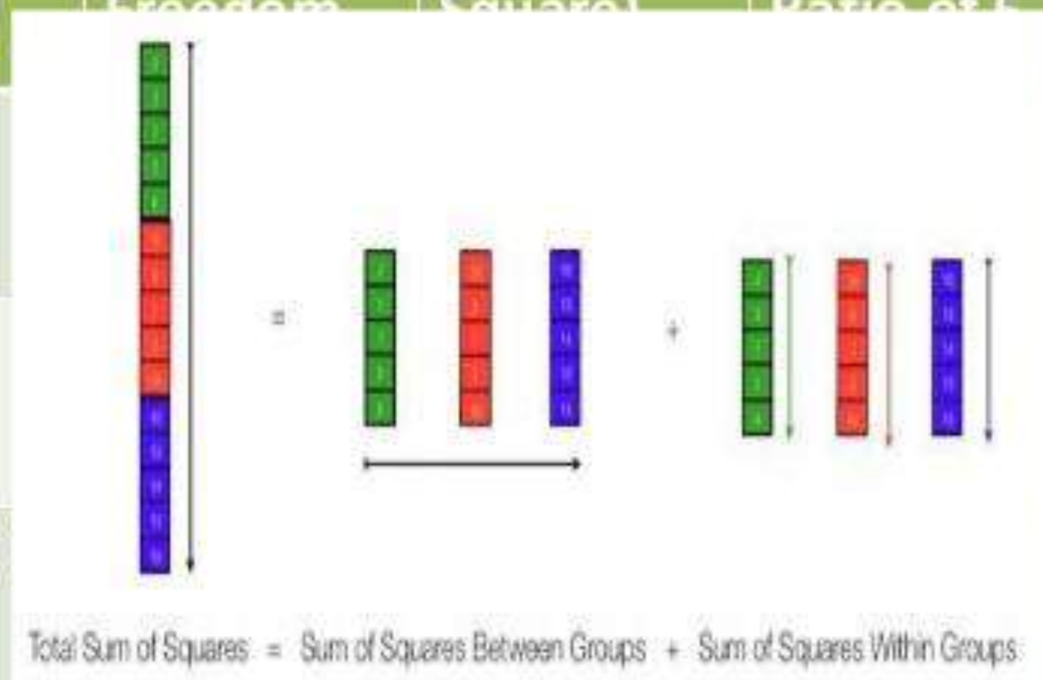
✓ **Tukey's HSD Procedure**

when all pairs of sample means are to be tested

✓ **Scheffe's Procedure** (when sample sizes are unequal)

One way ANOVA: Table

Source of Variation	SS (Sum of Squares)	Degrees of Freedom	MS (Mean Square)	Variance Ratio of F
Between Samples	SSC			
Within Samples	SSE			
Total	SS(Total)			



Example- one way ANOVA

Example: 3 samples obtained from normal populations with equal variances. Test the hypothesis that sample means are equal

8	7	12
10	5	9
7	10	13
14	9	12
11	9	14

1.Null hypothesis –

No significant difference in the means of 3 samples

2. State Alpha i.e 0.05

3. Calculate degrees of Freedom

k-1 & n-k = 2 & 12

4. State decision rule

Table value of F at 5% level of significance for d.f 2 & 12 is 3.88

The calculated value of F > 3.88 ,H0 will be rejected

5. Calculate test statistic

X1	X2	X3
8	7	12
10	5	9
7	10	13
14	9	12
11	9	14
Total 50 M1= 10	40 M2 = 8	60 M3 = 12

$$\text{Grand average} = \frac{10+8+12}{3} = 10$$

Variance BETWEEN samples (M1=10, M2=8,M3=12)

Sum of squares between samples (SSC) =

n1 (**M1** – Grand avg)^2 + n2 (**M2**– Grand avg)^2 + n3(**M3**– Grand avg)^2

5 (10 - 10)^2 + 5 (8 - 10)^2 + 5 (12 - 10)^2 = 40

Calculation of **M**ean sum of **S**quares between samples (**MSC**)

$$MSC = \frac{SSC}{k-1} = \frac{40}{2} = 20$$

k= No of Samples, n= Total No of observations

Variance WITH IN samples (M1=10, M2=8,M3=12)

X1	$(X1-M1)^2$	X2	$(X2-M2)^2$	X3	$(X3-M3)^2$
8	4	7	1	12	0
10	0	5	9	9	9
7	9	10	4	13	1
14	16	9	1	12	0
11	1	9	1	14	4
	30		16		14

Sum of squares within samples (SSE) = 30 + 16 +14 = 60

Calculation of Mean Sum Of Squares within samples (**MSE)**

$$MSE = \frac{SSE}{n-k} = \frac{60}{12} = 5$$

Calculation of ratio F

$$F = \frac{Variability\ between\ groups}{Variability\ within\ groups}$$

$$\text{F- statistic} = \frac{MSC}{MSE} = 20/5 \quad =4$$

The Table value of F at 5% level of significance for d.f 2 & 12 is 3.88
The calculated value of F > table value
H0 is rejected. Hence there is significant difference in sample means

Short cut method -

X1	$(X1)^2$	X2	$(X2)^2$	X3	$(X3)^2$
8	64	7	49	12	144
10	100	5	25	9	81
7	49	10	100	13	169
14	196	9	81	12	144
11	121	9	81	14	196
Total 50	530	40	336	60	734

Total sum of all observations = 50 + 40 + 60 = 150

Correction factor = $T^2 / N = (150)^2 / 15 = 22500/15 = 1500$

Total sum of squares= 530+ 336+ 734 – 1500= 100

Sum of square b/w samples=$(50)^2/5 + (40)^2/5 + (60)^2/5 - 1500 = 40$

Sum of squares within samples= 100-40= 60

Example with SPSS

Example:

Do people with private health insurance visit their Physicians more frequently than people with no insurance or other types of insurance ?

N=86

- **Type of insurance** - 1.No insurance
  2.Private insurance
  3. TRICARE
- No. of visits to their Physicians(dependent variable)

Violations of Assumptions

Normality

Choose the non-parametric **Kruskal-Wallis H Test** which does not require the assumption of normality.

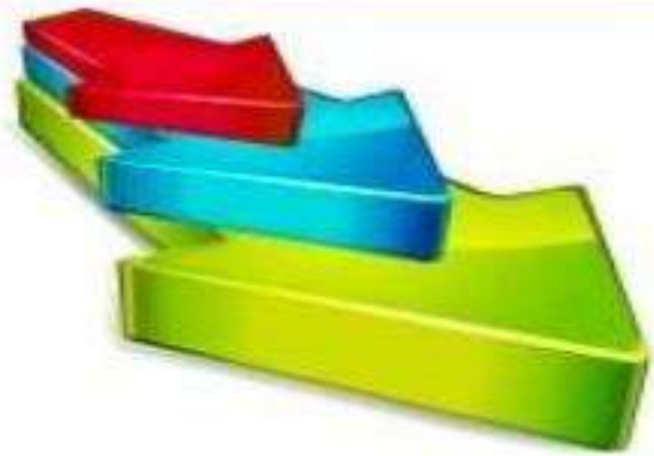
Homogeneity of variances

- ✓ Welch test or
- ✓ Brown and Forsythe test or Kruskal-Wallis H Test

Two Way ANOVA

Data required

- When 2 independent variables (Nominal/categorical) have an effect on one dependent variable (ordinal or ratio measurement scale)



- Compares relative influences on Dependent Variable
- Examine interactions between independent variables
- Just as we had Sums of Squares and Mean Squares in **One-way ANOVA**, we have the same in **Two-way ANOVA.**

Two way ANOVA

Include tests of three null hypotheses:

1) Means of observations grouped by one factor are same;

2) Means of observations grouped by the other factor are the same; and

3) There is no interaction between the two factors. The interaction test tells whether the effects of one factor depend on the other factor

Example-

we have test score of boys & girls in age group of 10 yr,11yr & 12 yr. If we want to study the effect of gender & age on score.

Two independent factors- Gender, Age

Dependent factor - Test score

Ho -Gender will have no significant effect on student score

Ha -

Ho - Age will have no significant effect on student score

Ha -

Ho – Gender & age interaction will have no significant effect on student score

Ha -

Two-way ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-ratio	*P-value*
Factor A	$r-1$	SS_A	MS_A	$F_A = MS_A / MS_E$	Tail area
Factor B	$c-1$	SS_B	MS_B	$F_B = MS_B / MS_E$	Tail area
Interaction	$(r-1)(c-1)$	SS_{AB}	MS_{AB}	$F_{AB} = MS_{AB} / MS_E$	Tail area
Error (within)	$rc(n-1)$	SS_E	MS_E		
Total	$rcn-1$	SS_T			

Example with SPSS

Example:

Do people with private health insurance visit their Physicians more frequently than people with no insurance or other types of insurance ?

N=86

- **Type of insurance** - 1.No insurance
  2.Private insurance
  3. TRICARE
- No. of visits to their Physicians(dependent variable)

Gender
0-M
1-F

MANOVA

Multivariate ANalysis Of VAriance

Data Required

- MANOVA is used to test the significance of the effects of one or more IVs on **two or more DVs**.

- It can be viewed as an extension of ANOVA with the key difference that we are dealing with **many dependent variables** (not a single DV as in the case of ANOVA)

- **Dependent Variables (at least 2)**
  - Interval /or ratio measurement scale
  - May be correlated
  - Multivariate normality
  - Homogeneity of variance

- **Independent Variables (at least 1)**
  - Nominal measurement scale
  - Each independent variable should be independent of each other

- Combination of dependent variables is called **"joint distribution"**

- MANOVA gives answer to question

" Is joint distribution of 2 or more DVs significantly related to one or more factors?"

- The result of a MANOVA simply tells us that a difference exists (or not) across groups.
- It does not tell us which treatment(s) differ or what is contributing to the differences.
- For such information, we need to run ANOVAs with post hoc tests.

Various tests used-

- ✓ **Wilk's Lambda**
  - ➢ Widely used; good balance between power and assumptions
- ✓ **Pillai's Trace**
  - ➢ Useful when sample sizes are small, cell sizes are unequal, or covariances are not homogeneous
- ✓ **Hotelling's (Lawley-Hotelling) Trace**
  - ➢ Useful when examining differences between two groups

Example with SPSS

Example:

Do people with private health insurance visit their Physicians more frequently than people with no insurance or other types of insurance ?

N=50

Gender(0-M,1-F)

- **Type of insurance** - 1.No insurance
  2.Private insurance
  3. TRICARE

Satisfaction with facility provided

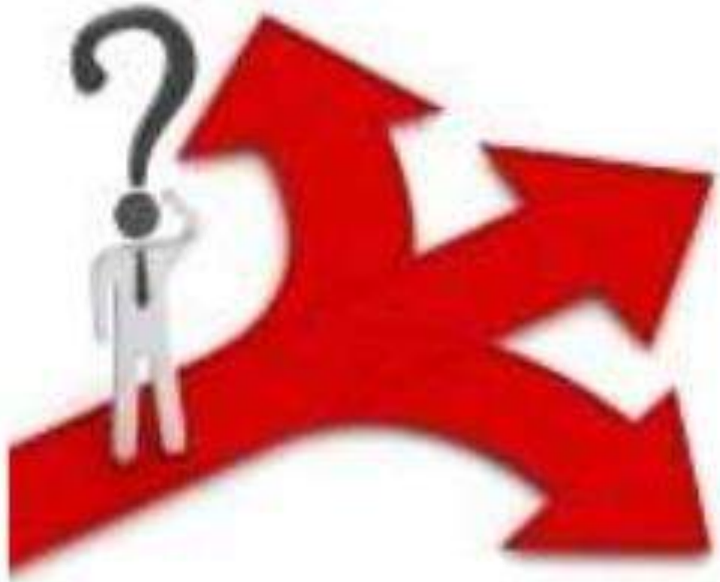
- No. of visits to their Physicians(dependent variable)

Research question

1. Do men & women differ significantly from each other in their satisfaction with health care provider & no. of visits they made to a doctor

2. Do 3 insurance groups differ significantly from each other in their satisfaction with health care provider & no. of visits they made to a doctor

3. Is there any interaction b/w gender & insurance status in relation to satisfaction with health care provider & no. of visits they made to a doctor



ANOVA with repeated measures

ANOVA with Repeated Measures

- Determines whether means of 3 or more measures from same person or matched controls are similar or different.

- Measures DV for various levels of one or more IVs

- Used when we repeatedly measure the same subjects multiple times

Assumptions

- Dependent variable is **interval /ratio** (**continuous**)
- Dependent variable is **approximately normally distributed**.
- One independent variable where participants are tested on the same dependent variable at least 2 times.
- **Sphericity-** condition where variances of the differences between all combinations of related groups (levels) are equal.

Sphericity violation

- Sphericity can be like homogeneity of variances in a between-subjects ANOVA.
- The violation of sphericity is serious for the Repeated Measures ANOVA, with violation causing the test to have an increase in the Type I error rate).
- **Mauchly's Test of Sphericity** tests the assumption of sphericity.

Sphericity violation

- The corrections employed to combat violation of the assumption of sphericity are:
  - ✓ **Lower-bound estimate**,
  - ✓ **Greenhouse-Geisser correction** and
  - ✓ **Huynh-Feldt correction**.

- The corrections are applied to the degrees of freedom (df) such that a valid critical F-value can be obtained.

Steps ANOVA

1.Define null & alternative hypotheses

2. State Alpha
3. Calculate degrees of Freedom
4. State decision rule

5. Calculate test statistic
- Calculate variance between samples
- Calculate variance within the samples
- Calculate ratio F
- If F is significant, perform *post hoc* test

6. State Results & conclusion

Calculate Degrees of Freedom for

- D.f between samples = K-1
- D.f within samples = n- k
- D.f subjects=r -1
- D.f error= d.f within- d.f subjects
- D.f total = n-1

State decision rule

If calculated value of F >table value of F, reject Ho

Calculate test statistic (f= MS bw/ MS error)

	SS	DF	MS	F
Between				
Within -subjects - error				
Total				

State Results & conclusion

Example with SPSS

Example-

Researcher wants to observe the effect of medication on free T 3 levels before, after 6 week, after 12 week. Level of free T 3 obtained through blood samples. Are there any differences between 3 conditions using alpha 0.05?

- ✓ **Independent Variable- time 1, time 2, time 3**
- ✓ **Dependent Variable- Free T3 level**

Other related tests-

ANCOVA (Analysis of Covariance)

Additional assumptions-

- Covariate should be continuous variable
- Covariate & dependent variable must show a linear relationship & must be similar in each group

MANCOVA (Multivariate analysis of covariance)

One or more continuous covariates present

References

- Methods in Biostatistics *by* BK Mahajan
- Statistical Methods *by* SP Gupta
- Basic & Clinical Biostatistics *by* Dawson and Beth
- Munro's statistical methods for health care research


Thank You!

“T” TEST

Overview

- Background
- Different versions of t-test
- Main usage of t-test
- t-test v/s z-test
- Assumptions of t-test
- Examples

Background

- Introduced in 1908 by William Sealy Gosset.
- Gosset published his mathematical work under the pseudonym "Student".
- Definition of t test: " It's a method of testing hypothesis about the mean of small sample drawn from a normally distributed population when the standard deviation for the sample is unknown."

Assumptions of t-Test

- Dependent variables are interval or ratio.
- The population from which samples are drawn is normally distributed.
- Samples are randomly selected.
- The groups have equal variance (Homogeneity of variance).
- The t-statistic is robust (it is reasonably reliable even if assumptions are not fully met.)

Applications of t test

- The calculation of a confidence interval for a sample mean.
- To test whether a sample mean is different from a hypothesized value.
- To compare mean of two samples.
- To compare two sample means by group.

Types of “t” test

- Single sample t test – we have only 1 group; want to test against a hypothetical mean.

- Independent samples t test – we have 2 means, 2 groups; no relation between groups, Eg: When we want to compare the mean of T/T group with Placebo group.

- Paired t test – It consists of samples of matched pairs of similar units or one group of units tested twice. Eg: Difference of mean pre & post drug intervention.

One Sample t-test

- It is used in measuring whether a sample value significantly differs from a hypothesized value.
- For example, a research scholar might hypothesize that on an average it takes 3 minutes for people to drink a standard cup of coffee.
- He conducts an experiment and measures how long it takes his subjects to drink a standard cup of coffee.
- The one sample t-test measures whether the mean amount of time it took the experimental group to complete the task varies significantly from the hypothesized 3 minutes value.

Equation for a one-sample t-test

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

where

t = the t statistic

$\bar{x}$ = the mean of the sample

μ = the comparison mean

s = the sample standard deviation

n = the sample size

Example

- 10 individuals had taken an exam and we want to test whether their scores, all together, are significantly different from the score of 100.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{107.8 - 100}{5.35/\sqrt{10}} = 4.61$$

- We need to calculate the degrees of freedom.
- Here, the degrees of freedom is simply the sample size minus one.
- Therefore, Degrees of freedom = n – 1 = 10 – 1 = 9
- Now, we will refer to a t table to determine the critical t value for 9 degrees of freedom at the .05 level of significance.
- Looking at a t table, this value is 2.26 .
- Since our calculated t value of 4.61 is greater than the critical t value of 2.26, we can say that the scores of our sample of 10 individuals differ significantly from the score of 100.

T table

df	.05	.01	.001
1	12.706	63.657	636.619
2	4.303	9.925	31.598
3	3.182	5.841	12.924
4	2.776	4.604	8.610
5	2.571	4.032	6.869
6	2.447	3.707	5.959
7	2.365	3.499	5.408
8	2.306	3.355	5.041
9	2.262	3.250	4.781
10	2.228	3.169	4.587

Independent t test

- The independent sample t-test consists of tests that compare mean value(s) of continuous-level (interval or ratio data), in a normally distributed data.
- The independent sample t-test compares two means.
- The independent samples t-test is also called unpaired t-test/ two sample t test.
- It is the t-test to be used when two separate independent and identically distributed variables are measured.
- Eg: 1. Comparision of quality of life improved for patients who took drug Valporate as opposed to patients who took drug Levetiracetam in myoclonic seizures.
- 2.Comparasion of mean cholesterol levels in treatment group with placebo group after administration of test drug.

Assumptions

- A random sample of each population is used.
- The random samples are each made up of independent observation.
- Each sample is independent of one another.
- The population distribution of each population must be nearly normal, or the size of the sample is large.

Independent t test

To test the null hypothesis that the two population means, $\mu 1$ and $\mu 2$, are equal:

- 1. Calculate the difference between the two sample means, $\bar{x}1 - \bar{x}2$.
- 2. Calculate the pooled standard deviation: sp
- 3. Calculate the standard error of the difference between the means:
- 4. Calculate the T-statistic, which is given by $T = \bar{x}1 - \bar{x}2 / SE(\bar{x}1 - \bar{x}2)$
- This statistic follows a t-distribution with $n1 + n2 - 2$ degrees of freedom.
- 5. Use tables of the t-distribution to compare your value for T to the $t_{n1+n2-2}$ distribution. This will give the p-value for the unpaired t-test.

Equation for the independent samples t-test

- The independent-Samples t-test procedure compares means for two groups of cases.

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left[\frac{SS_1 + SS_2}{n_1 + n_2 - 2}\right]\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

Here,

$\overline{X}_1$ and $\overline{X}_2$ are the means of the two different groups

$n_1 = n$ of Group 1

$n_2 = n$ of Group 2

SS = sum of squares

Example

- Suppose we have to compare the mean value of two groups, one with 7 subjects and the other with 5 subjects .
- These were their scores:

Case	Group 1	Group 2
1	78	87
2	82	92
3	87	86
4	65	95
5	75	73
6	82	
7	71	

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{SS_1 + SS_2}{n_1 + n_2 - 2}\right]\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} = \frac{77.14 - 86.60}{\sqrt{\left[\frac{334.86 + 285.20}{7 + 5 - 2}\right]\left[\frac{1}{7} + \frac{1}{5}\right]}}$$

$$= \frac{-9.46}{\sqrt{\left(\frac{620.06}{10}\right)\left(\frac{12}{35}\right)}} = \frac{-9.46}{\sqrt{21.26}} = -0.44$$

For an independent or between subjects' t test: df = n1+ n2 - 2

- Now, take the absolute value of this, which is 0.44.
- Now, for the .05 probability level with 10 degrees of freedom, we see from the table that the critical t score is 2.228 for a two-tailed test.
- Since the calculated t score is lower than the critical t score, the results are not significant at the .05 probability level.

T table

df	.05	.01	.001
1	12.706	63.657	636.619
2	4.303	9.925	31.598
3	3.182	5.841	12.924
4	2.776	4.604	8.610
5	2.571	4.032	6.869
6	2.447	3.707	5.959
7	2.365	3.499	5.408
8	2.306	3.355	5.041
9	2.262	3.250	4.781
10	2.228	3.169	4.587

Paired t test

- A paired t-test is used to compare two population means where you have two samples in which observations in one sample can be paired with observations in the other sample.
- A comparison of two different methods of measurement or two different treatments where the measurements/treatments are applied to the same subjects.
- Eg: 1.pre-test/post-test samples in which a factor is measured before and after an intervention,
- 2.Cross-over trials in which individuals are randomized to two treatments and then the same individuals are crossed-over to the alternative treatment,
- 3.Matched samples, in which individuals are matched on personal characteristics such as age and sex,

Paired t test

- Suppose a sample of "n" subjects were given an antihypertensive drug we want to check blood pressure before and after treatment . We want to find out the effectiveness of the treatment by comparing mean pre & post t/t.
- To test the null hypothesis that the true mean difference is zero, the procedure is as follows:

1. Calculate the difference ($d_i = y_i - x_i$) between the two observations on each pair.

2. Calculate the mean difference, d.

3. Calculate the standard error of the mean differences. $S.E = S.D/\sqrt{n}$

4. Calculate the t-statistic, which is given by $T = d/S.E$, Under the null hypothesis, this statistic follows a t-distribution with $n - 1$ degrees of freedom.

5. Use tables of the t-distribution to compare your value for T to the t_{n-1} distribution. This will give the p-value for the paired t-test.

Z-tests

- A z-test is a statistical test used to determine whether two population means are different when the variances are known and the sample size is large. The test statistic is assumed to have a normal distribution with a known S.D.
- The z-test is a hypothesis test in which the z-statistic follows a normal distribution.
- The z-test is best used for greater than 30 samples because, under the central limit theorem, as the number of samples gets larger, the samples are considered to be approximately normally distributed. When conducting a z-test, the null and alternative hypothesis, alpha and z-score should be stated.
- Next, the test statistic should be calculated, and the results and conclusion stated.
- z-test is used to test hypotheses about means for large samples ($N>100$) with a known variance,We use t-test when the sample size is small ($N < 100$) and the population variance is unknown.
- <u>Ex:</u> Comparing the prevelance of disease in men versus women.

Sampling Statistics and Introduction to Statistical Package for Social Sciences (SPSS)

Introduction

- SPSS stands for *Statistical Package for the Social Sciences*
- SPSS Incorporated is a leading worldwide provider of predictive analytics software and solutions.
- First version of SPSS was released in 1968, after being developed by Norman H. Nie, Dale H. Bent and C. Hadlai Hull..
- The company announced on July 28, 2009 that it was being acquired by IBM for US$1.2 billion.

SPSS is now owned by IBM.

Between 2009-10 the primer Vender of SPSS was called PASW (Predictive Analytics SoftWare) .

IBM SPSS Statistics 21.0 - Released on August 2012 Latest version

SPSS®

AN IBM® COMPANY

Company Logo

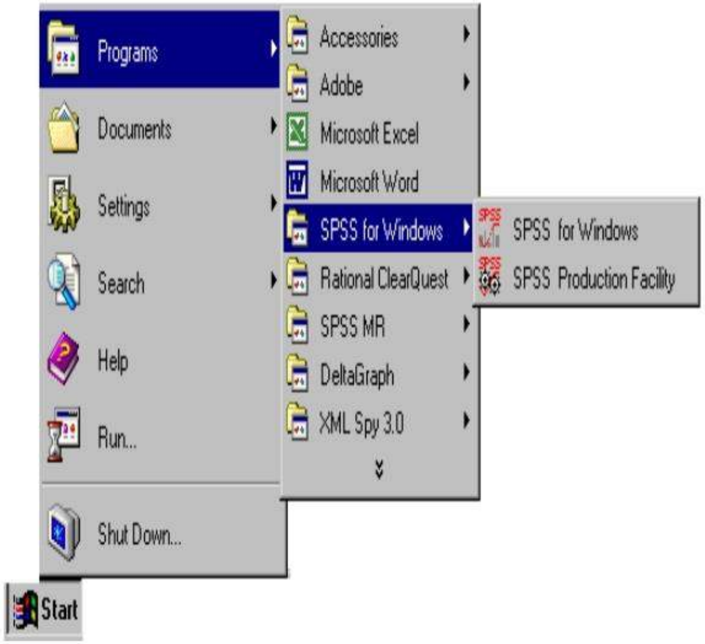
What is it Used For

- With SPSS we can analyze data in three basic ways:

- ✓ Describe data using descriptive statistics example frequency, mean, minimum and maximum.
- ✓ Examine Relationships between variables example correlation, regression, factor analysis etc.
- ✓ Compare groups to determine if there are significant difference between these groups example t-test, ANOVA etc.

How to open SPSS

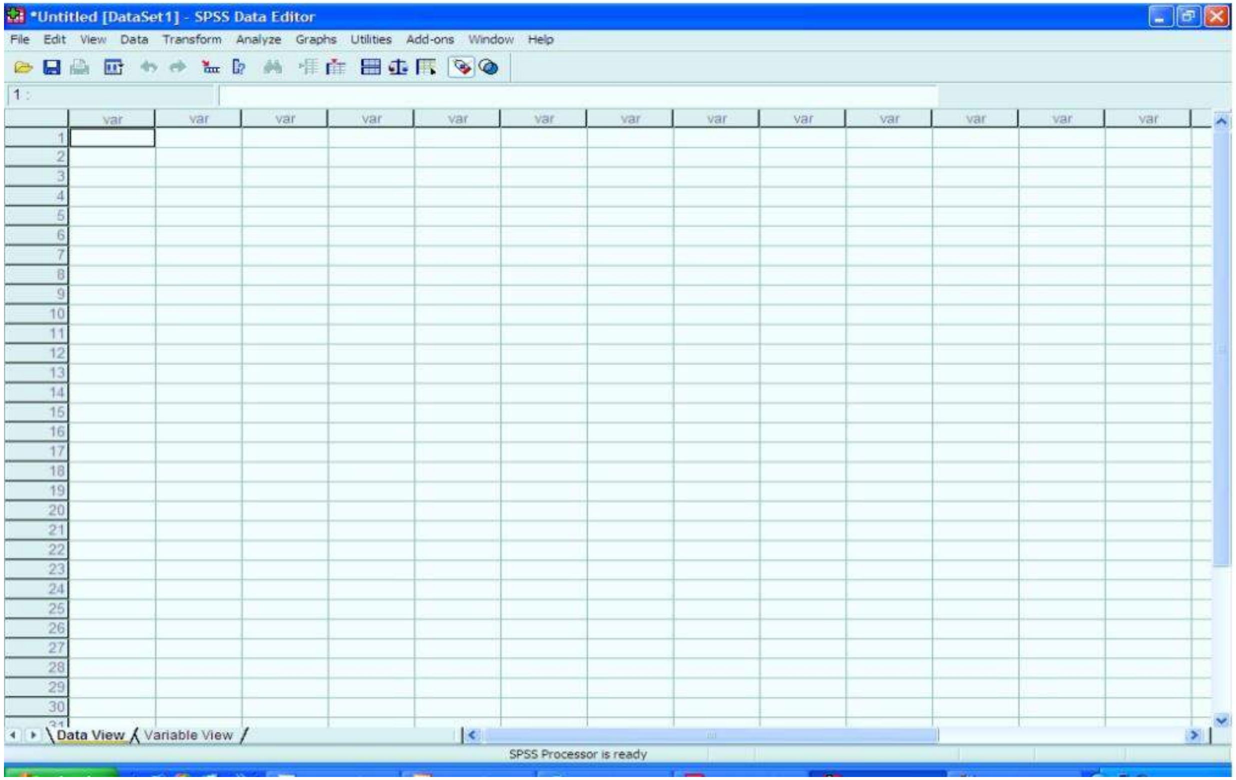
- Go to START
- Click on PROGRAMS
- Click on SPSS INC
- Click on SPSS 16.0



Basic Structure of SPSS

- There are two different windows in SPSS
- 1st – **<u>Data Editor Window</u>**
  In data editor we can create variables, enter data and carry out statistical functions.
- 2nd – **<u>Output Viewer Window</u>**
  Output window shows what results are produced by analyzing the functions.

DATA EDITOR WINDOW



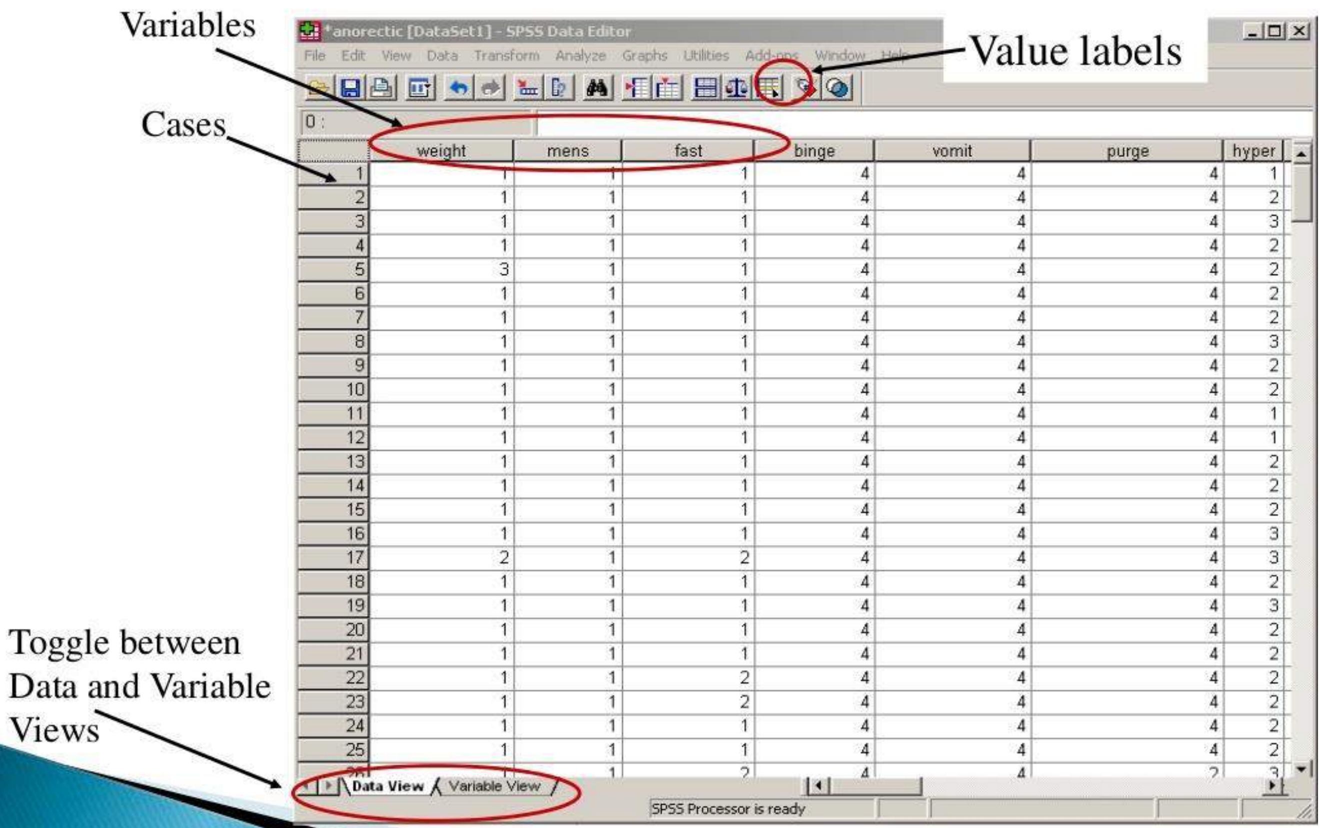
Data Editor Window Consists Two TABs

- Data View
- Variable View

Data View

- Data view is used to enter data and view data.
- In data view:
  - Rows represent individual cases. It can be state, company, business etc.
  - Columns represent particular variable in your data file.

DATA VIEW

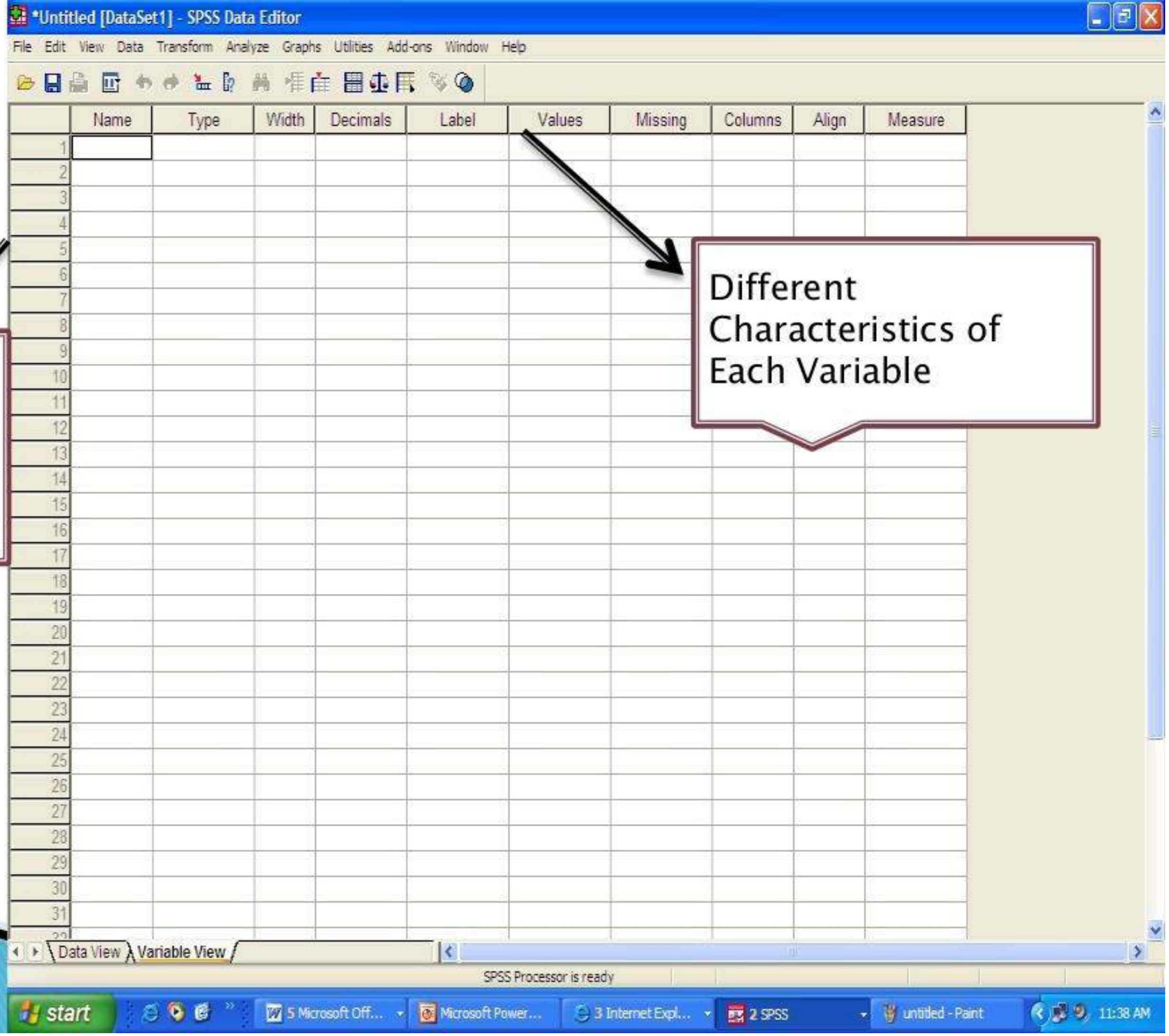

Variables
Cases
Value labels
Toggle between Data and Variable Views
*anorectic [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help
weight
mens
fast
binge
vomit
purge
hyper
Data View
Variable View
SPSS Processor is ready

Variable View

- Variable View is used to create and define various variables.
- In Variable View:
  - Row represent individual variable or define the variable
  - Column represent the specific characteristic of variable like Name, Type, Width, Decimals, Label, Missing, Align, Measure etc.

Row show individual Variable

Different Characteristics of Each Variable

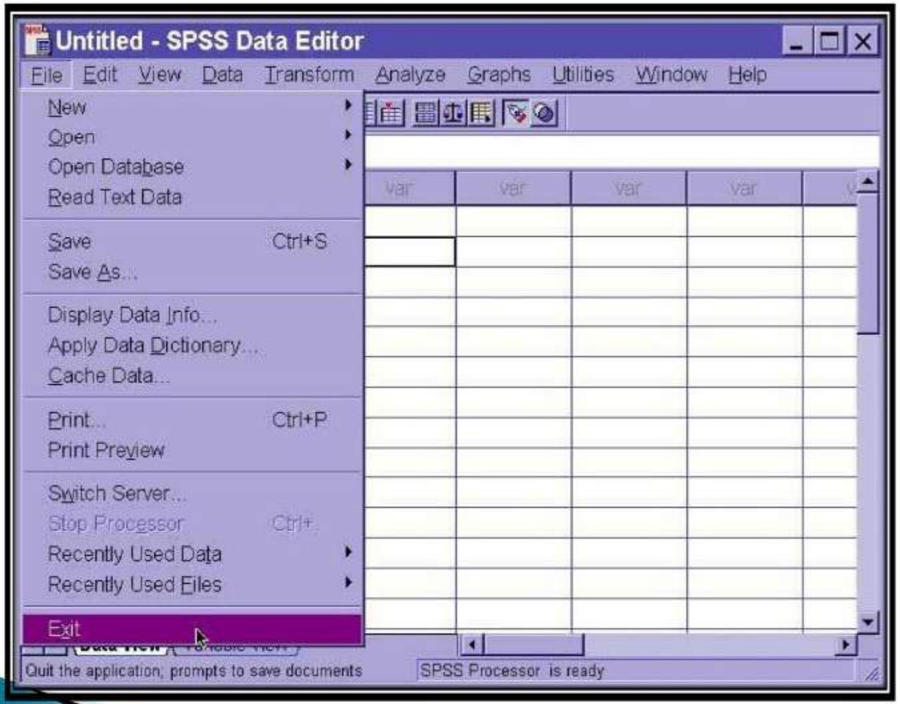


MENU BAR IN DATA EDITOR OF SPSS

FILE MENU

- The file menu SPSS contains standard option like other programs.
- File menu allows creating new files, open existing files, save files, read text data, print, print preview, exit from SPSS and other basic activities

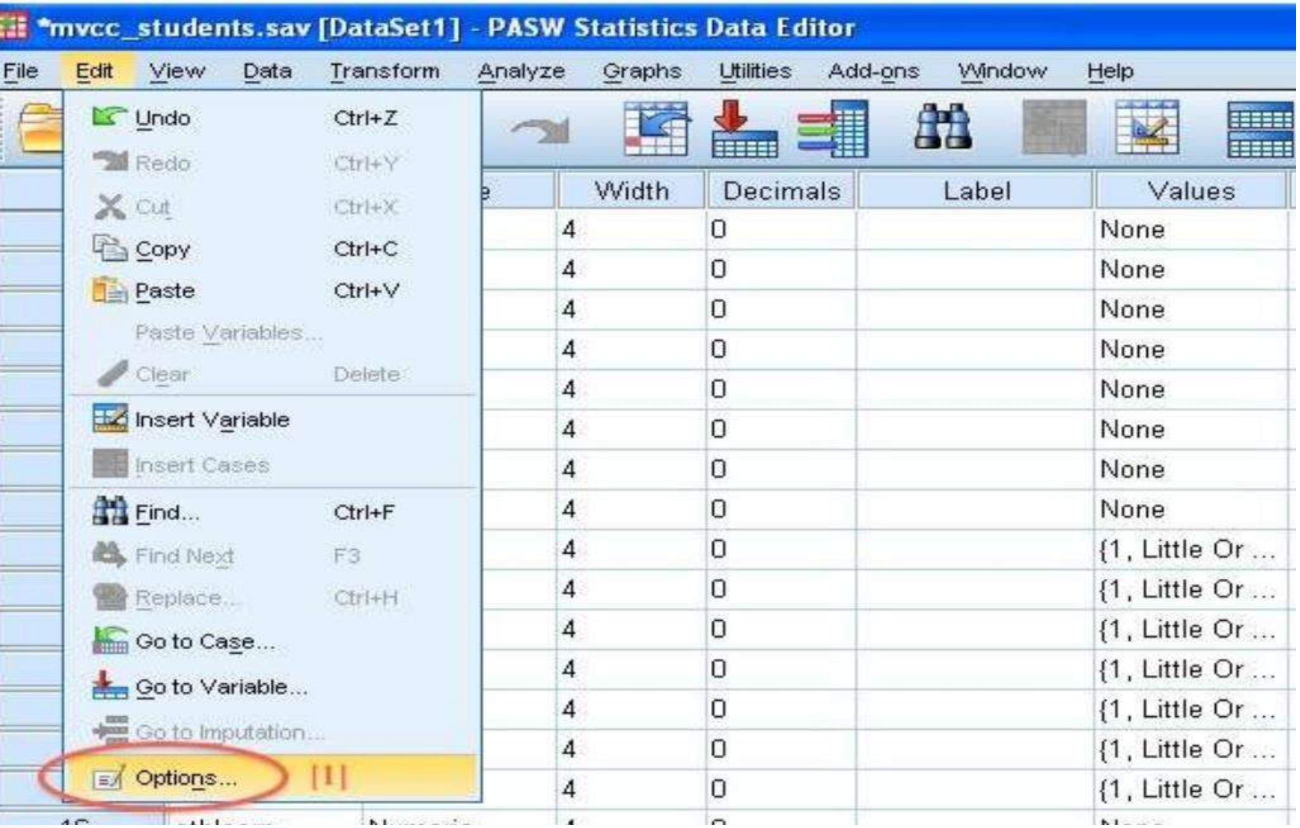
File Menu


Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
New
Open
Open Database
Read Text Data
Save Ctrl+S
Save As...
Display Data Info...
Apply Data Dictionary...
Cache Data...
Print... Ctrl+P
Print Preview
Switch Server...
Stop Processor Ctrl+.
Recently Used Data
Recently Used Files
Exit
var
Quit the application; prompts to save documents
SPSS Processor is ready

Edit Menu

- The edit menu allows the standard functions like to cut, copy, paste, edit, redo and undo.
- It has some other functions like Insert variable/case, Go to variable/case and Edit SPSS preference by OPTION

Edit Menu


*mvcc_students.sav [DataSet1] - PASW Statistics Data Editor
File
Edit
View
Data
Transform
Analyze
Graphs
Utilities
Add-ons
Window
Help
Undo Ctrl+Z
Redo Ctrl+Y
Cut Ctrl+X
Copy Ctrl+C
Paste Ctrl+V
Paste Variables...
Clear Delete
Insert Variable
Insert Cases
Find... Ctrl+F
Find Next F3
Replace... Ctrl+H
Go to Case...
Go to Variable...
Go to Imputation...
Options...
Width
Decimals
Label
Values

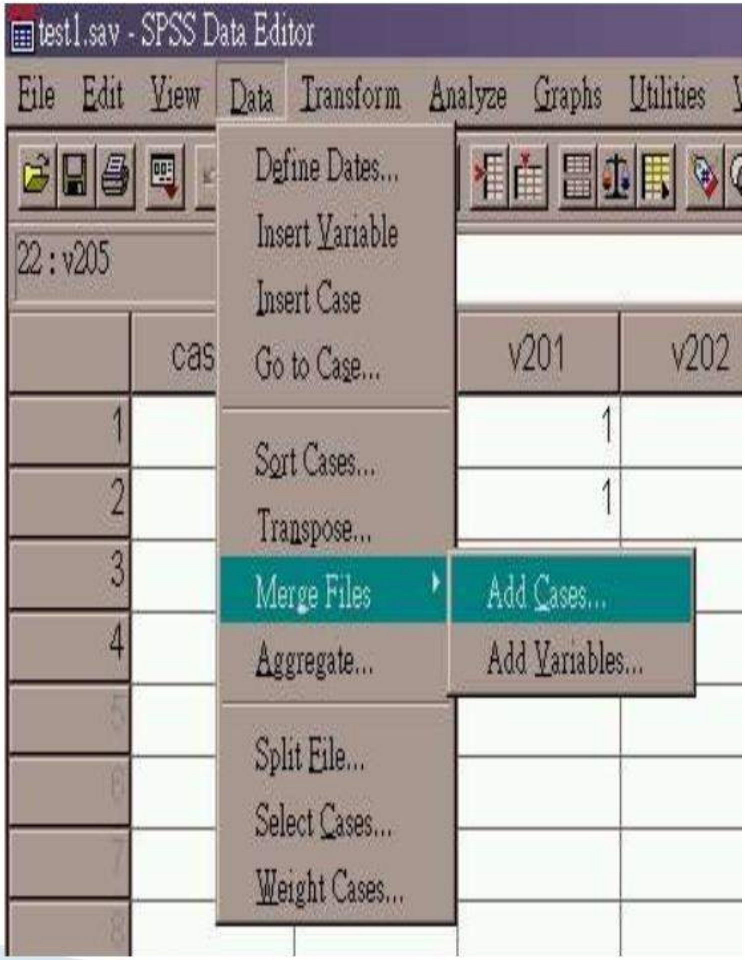
View Menu

- The view menu allow us to activate/deactivate the Status bar and other toolbars.

- We can change fonts, gridlines, value labels and see variables.

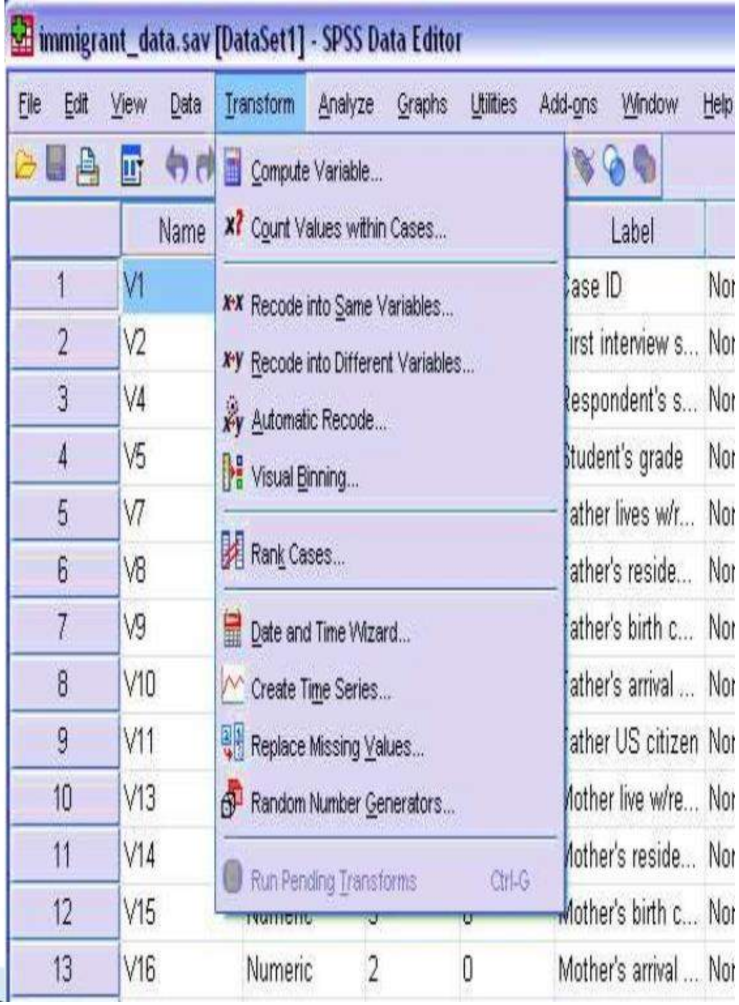
Data Menu

- ❖ The data menu allows us to define variable properties, dates, sorting cases, sorting variables.
- ❖ We can merge files, split a file, select cases and weight cases.



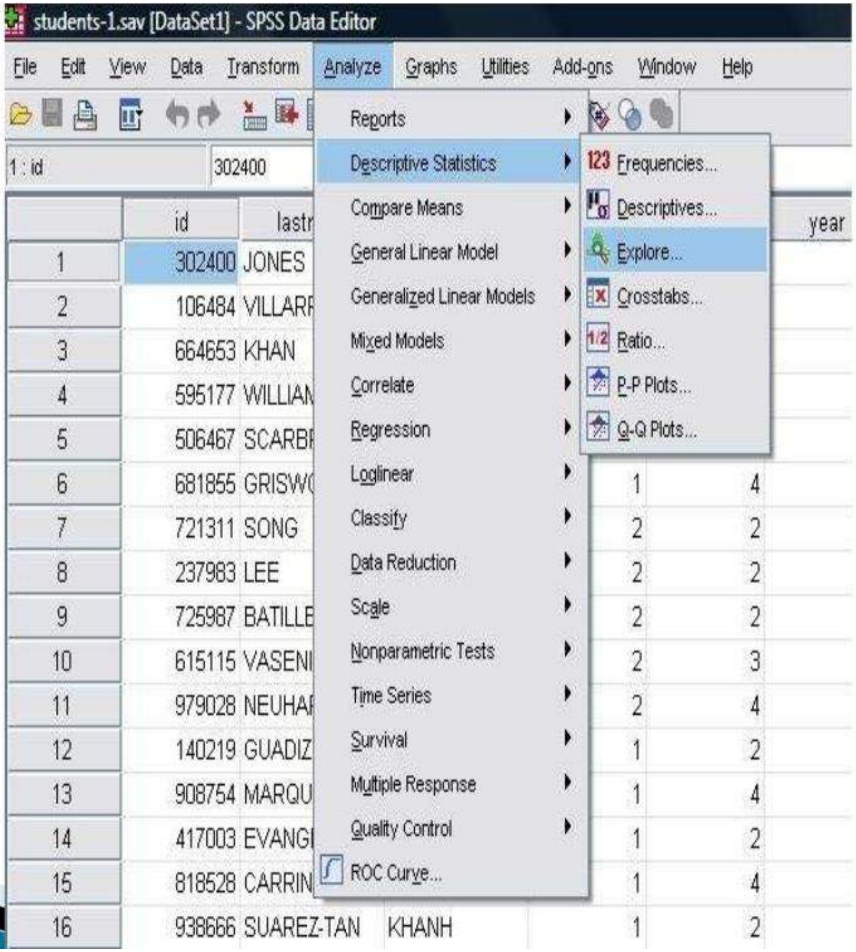
Transform Menu

- The transform menu allows creating new variables using compute variable.
- Helps in changing variables in the data file through recode, rank cases etc.
- Create time series, date & time wizard, replace missing numbers etc.



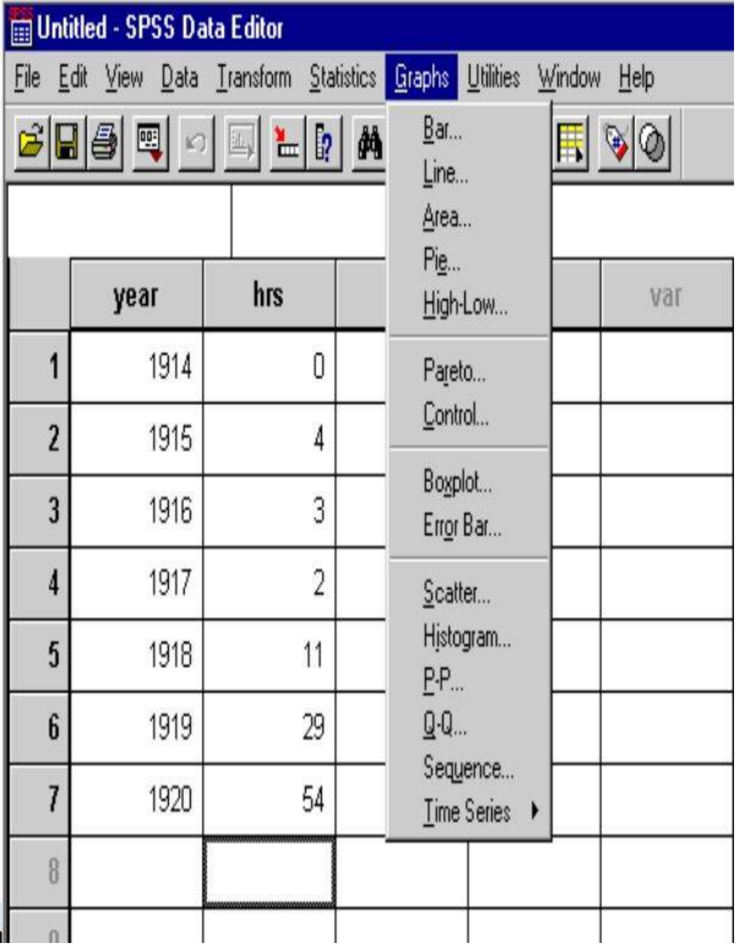
Analyze Menu

The analyze menu allows the analysis of data with help of various statistical tools & techniques.



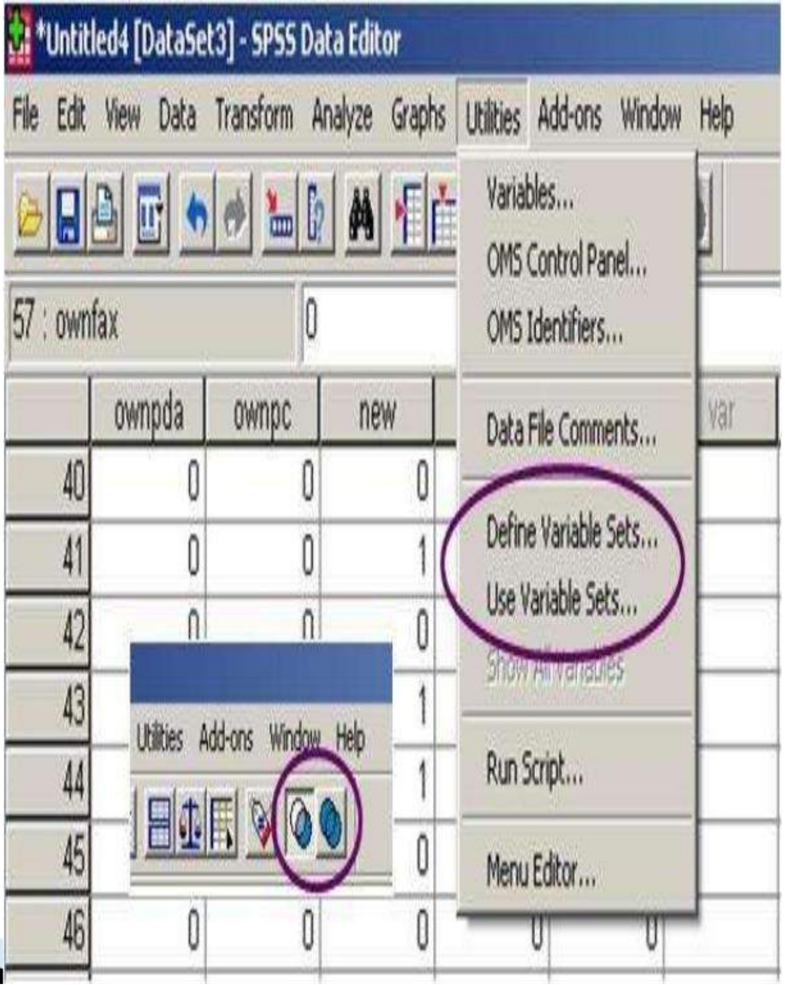
Graph Menu

- The graphs menu allows us to create bar chart, line chart, area chart, pie chart, histogram, scatter plots along with many other variations.
- This option can be used when we want to produce graph without any analysis of data.



Utilities Menu

The utilities menu helps in using various other utilities, which may be not required in normal usage.



Add-Ons Menu

- **Add-ons** are programs that can be added to the base SPSS package which provides a list of various features requisite for special requirements and are mostly meant for advanced level users.
- It include options like forecasting, advanced statistics, decision trees and many more options.

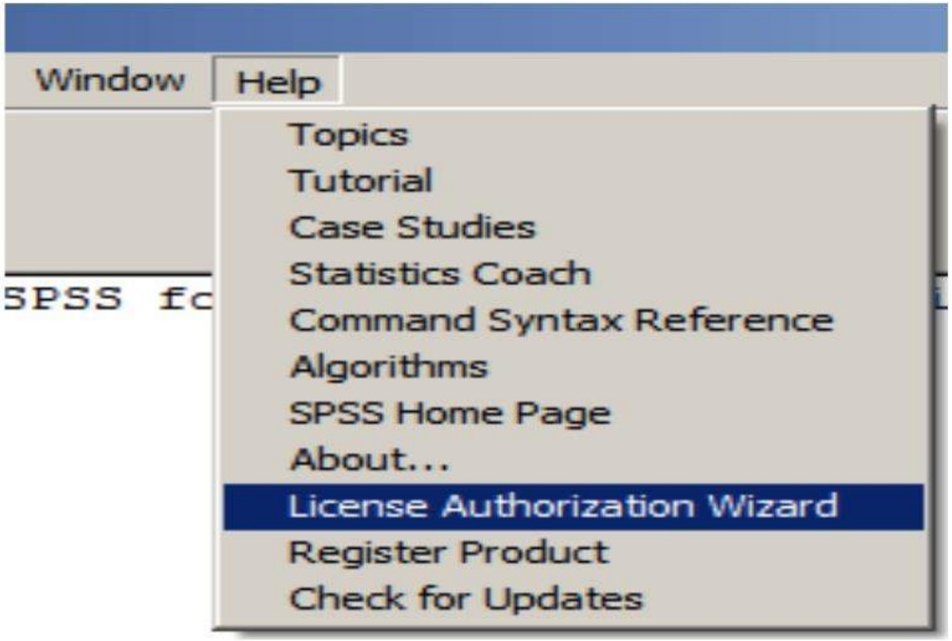
Window Menu

- **Window** can be used to select which window you want to view (i.e., Data Editor, Output Viewer, or Syntax). Since we have a data file and an output file open, let's try this.
- Select **Window/Data Editor**. Then select **Window/SPSS Viewer**.

Help Menu

- **Help** has many useful options including a link to the SPSS homepage, a statistics coach, and a syntax guide. Using **topics**, you can use the index option to type in any key word and get a list of options, or you can view the categories and subcategories available under contents. This is an excellent tool and can be used to troubleshoot most problems.

Help Menu



Window
Help
Topics
Tutorial
Case Studies
Statistics Coach
Command Syntax Reference
Algorithms
SPSS Home Page
About...
License Authorization Wizard
Register Product
Check for Updates

Opening a Data File

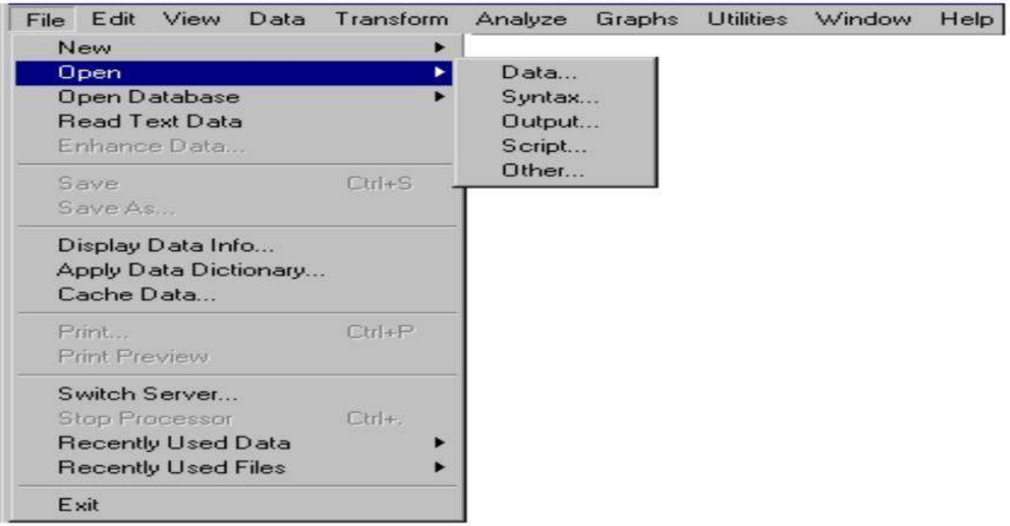
- Before you can analyze data, you need some data to analyze.

To open a data file:

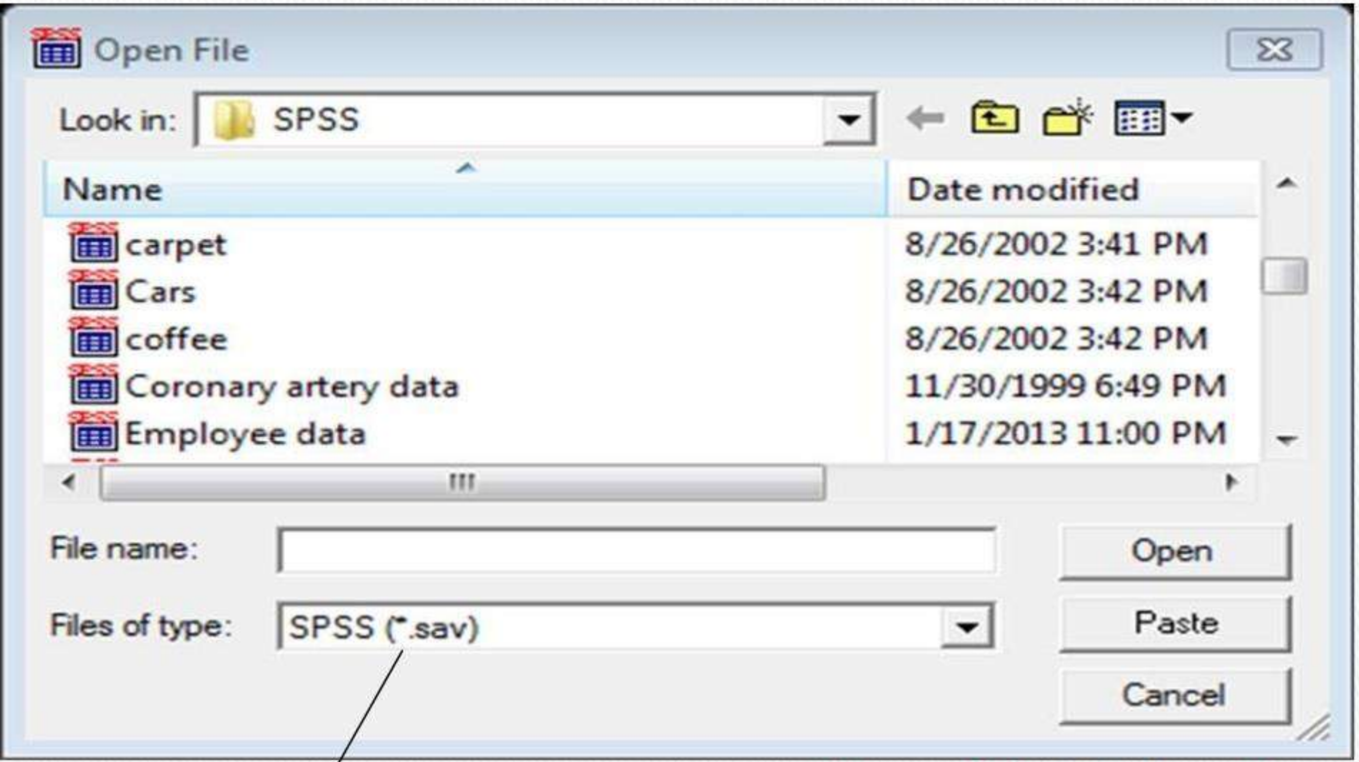
Step 1: From the menus choose: File

Open

Data...

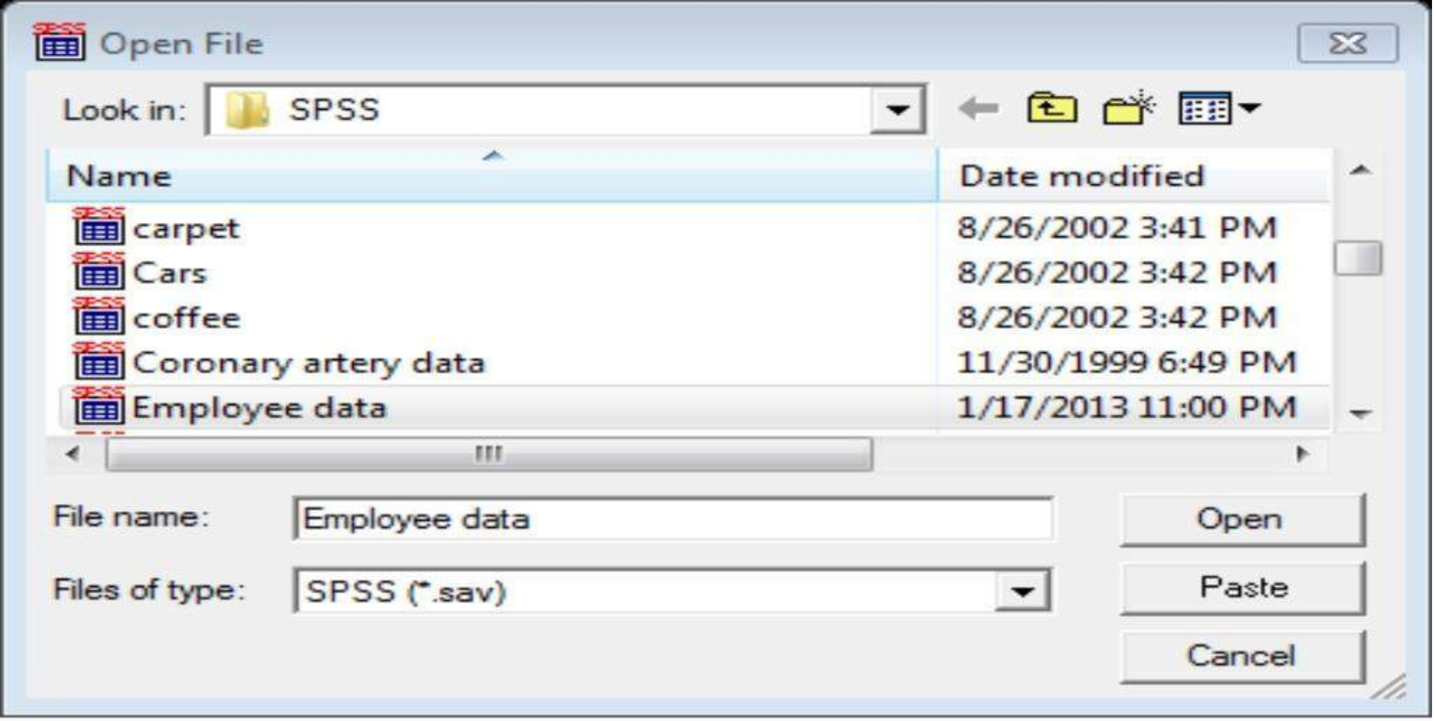


Now in next step it opens the OPEN FILE DIALOG BOX



Open File Dialog Box

By default, SPSS-format data files (.*sav* extension) are displayed



Now we select the Employee data file to open from the Open File Dialog Box

Another option to open data file

- Alternatively, you can use the Open File button on the toolbar.

Open File toolbar button

This opens the Open File dialog box.

Employee data - SPSS Data Editor

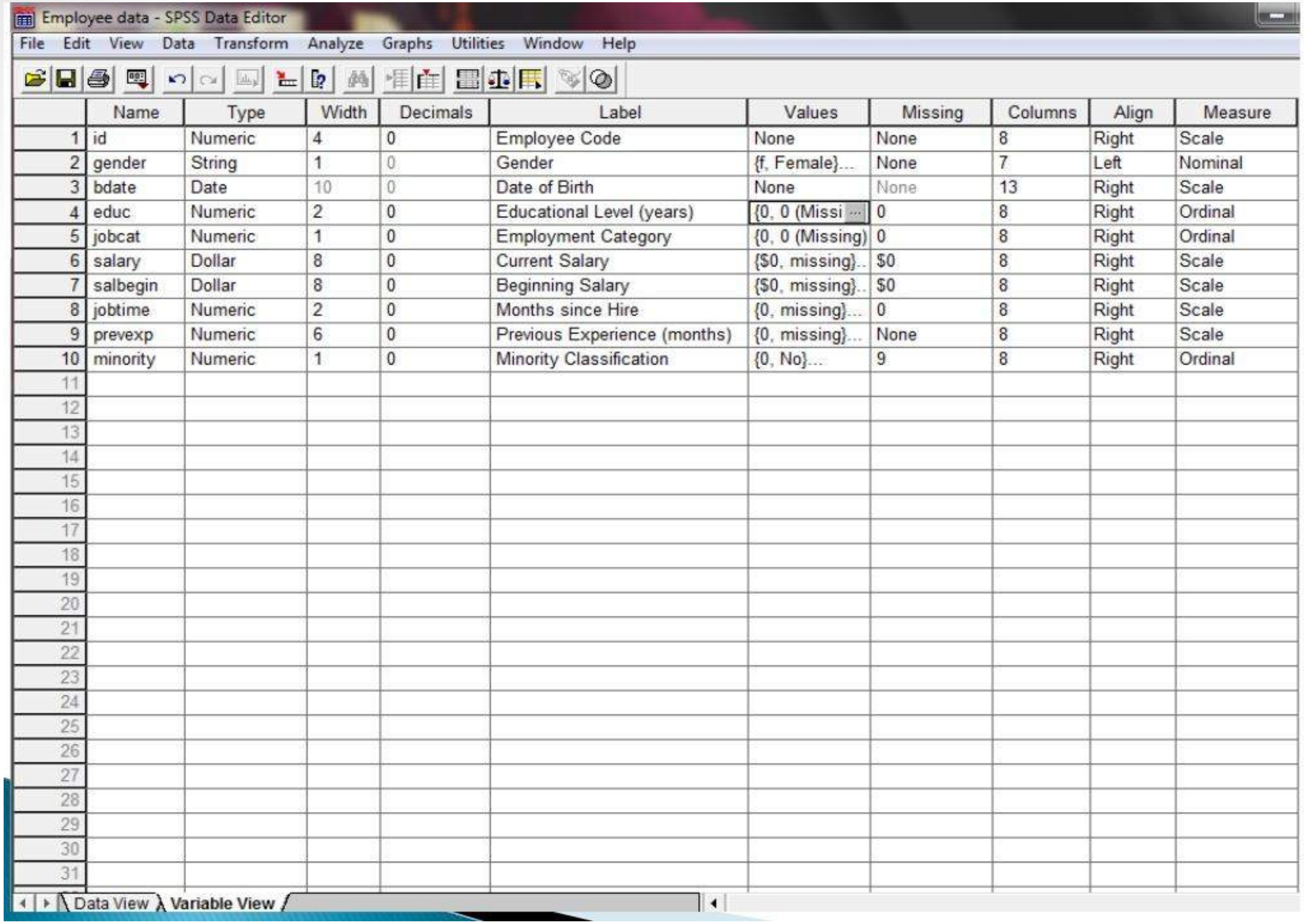
File Edit View Data Transform Analyze Graphs Utilities Window Help

1 : id 1

	id	gender	bdate	educ	jobcat	salary	salbegin	jobtime	prevexp	minority
1	1	m	02/03/1952	15	3	$57,000	$27,000	98	144	0
2	2	m	05/23/1958	16	1	$40,200	$18,750	98	36	0
3	3	f	07/26/1929	12	1	$21,450	$12,000	98	381	0
4	4	f	04/15/1947	8	1	$21,900	$13,200	98	190	0
5	5	m	02/09/1955	15	1	$45,000	$21,000	98	138	0
6	6	m	08/22/1958	15	1	$32,100	$13,500	98	67	0
7	7	m	04/26/1956	15	1	$36,000	$18,750	98	114	0
8	8	f	05/06/1966	12	1	$21,900	$9,750	98	0	0
9	9	f	01/23/1946	15	1	$27,900	$12,750	98	115	0
10	10	f	02/13/1946	12	1	$24,000	$13,500	98	244	0
11	11	f	02/07/1950	16	1	$30,300	$16,500	98	143	0
12	12	m	01/11/1966	8	1	$28,350	$12,000	98	26	1
13	13	m	07/17/1960	15	1	$27,750	$14,250	98	34	1
14	14	f	02/26/1949	15	1	$35,100	$16,800	98	137	1
15	15	m	08/29/1962	12	1	$27,300	$13,500	97	66	0
16	16	m	11/17/1964	12	1	$40,800	$15,000	97	24	0
17	17	m	07/18/1962	15	1	$46,000	$14,250	97	48	0
18	18	m	03/20/1956	16	3	$103,750	$27,510	97	70	0
19	19	m	08/19/1962	12	1	$42,300	$14,250	97	103	0
20	20	f	01/23/1940	12	1	$26,250	$11,550	97	48	0
21	21	f	02/19/1963	16	1	$38,850	$15,000	97	17	0
22	22	m	09/24/1940	12	1	$21,750	$12,750	97	315	1
23	23	f	03/15/1965	15	1	$24,000	$11,100	97	75	1
24	24	f	03/27/1933	12	1	$16,950	$9,000	97	124	1
25	25	f	07/01/1942	15	1	$21,150	$9,000	97	171	1
26	26	m	11/08/1966	15	1	$31,050	$12,600	96	14	0
27	27	m	03/19/1954	19	3	$60,375	$27,480	96	96	0
28	28	m	04/11/1963	15	1	$32,550	$14,250	96	43	0
29	29	m	01/28/1944	19	3	$135,000	$79,980	96	199	0
30	30	m	09/17/1961	15	1	$31,200	$14,250	96	54	0

Data View / Variable View

SPSS Processor is ready



Employee data - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	id	Numeric	4	0	Employee Code	None	None	8	Right	Scale
2	gender	String	1	0	Gender	{f, Female}...	None	7	Left	Nominal
3	bdate	Date	10	0	Date of Birth	None	None	13	Right	Scale
4	educ	Numeric	2	0	Educational Level (years)	{0, 0 (Missi...	0	8	Right	Ordinal
5	jobcat	Numeric	1	0	Employment Category	{0, 0 (Missing)	0	8	Right	Ordinal
6	salary	Dollar	8	0	Current Salary	{$0, missing}..	$0	8	Right	Scale
7	salbegin	Dollar	8	0	Beginning Salary	{$0, missing}..	$0	8	Right	Scale
8	jobtime	Numeric	2	0	Months since Hire	{0, missing}...	0	8	Right	Scale
9	prevexp	Numeric	6	0	Previous Experience (months)	{0, missing}...	None	8	Right	Scale
10	minority	Numeric	1	0	Minority Classification	{0, No}...	9	8	Right	Ordinal
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										

Data View Variable View

Employee data - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	id	Numeric	4	0	Employee Code	None	None	8	Right	Scale
2	gender	String	1	0	Gender	{f, Female}...	None	7	Left	Nominal
3	bdate	Date	10	0	Date of Birth	None	None	13	Right	Scale
4	educ	Numeric	2	0	Educational Level (years)	{0, 0 (Missi ...	0	8	Right	Ordinal
5	jobcat	Numeric	1	0	Employment Category	{0, 0 (Missing)	0	8	Right	Ordinal
6	salary	Dollar	8	0	Current Salary	{$0, missing}..	$0	8	Right	Scale
7	salbegin	Dollar	8	0	Beginning Salary	{$0, missing}..	$0	8	Right	Scale
8	jobtime	Numeric	2	0	Months since Hire	{0, missing}...	0	8	Right	Scale
9	prevexp	Numeric	6	0	Previous Experience (months)	{0, missing}...	None	8	Right	Scale
10	minority	Numeric	1	0	Minority Classification	{0, No}...	9	8	Right	Ordinal
11										

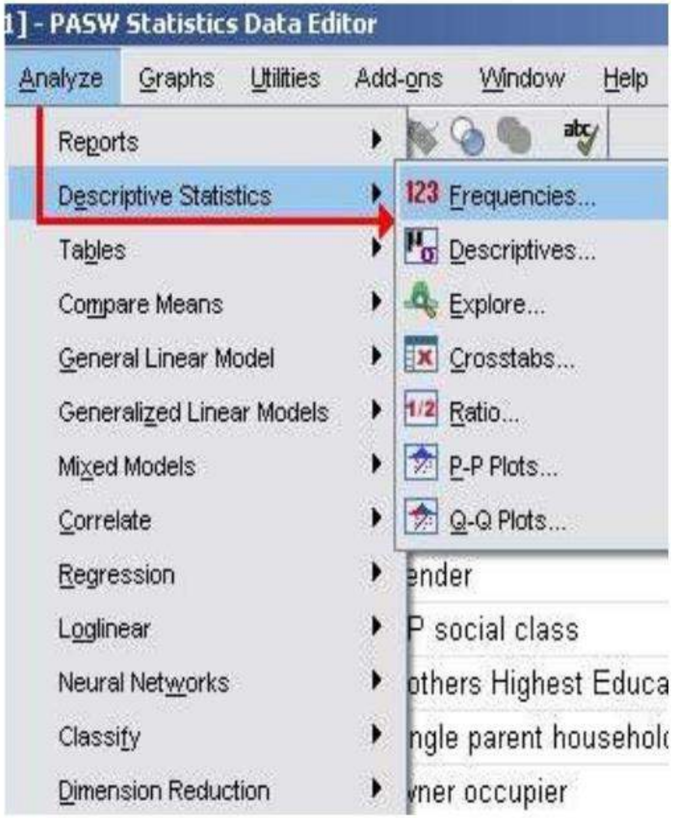
Running an Analysis

- The Analyze menu contains a list of general reporting and statistical analysis categories. Most of the categories are followed by an arrow, which indicates that there are several analysis procedures available within the category; they will appear on a submenu when the category is selected.
- We'll start with a simple frequency table (table of counts).

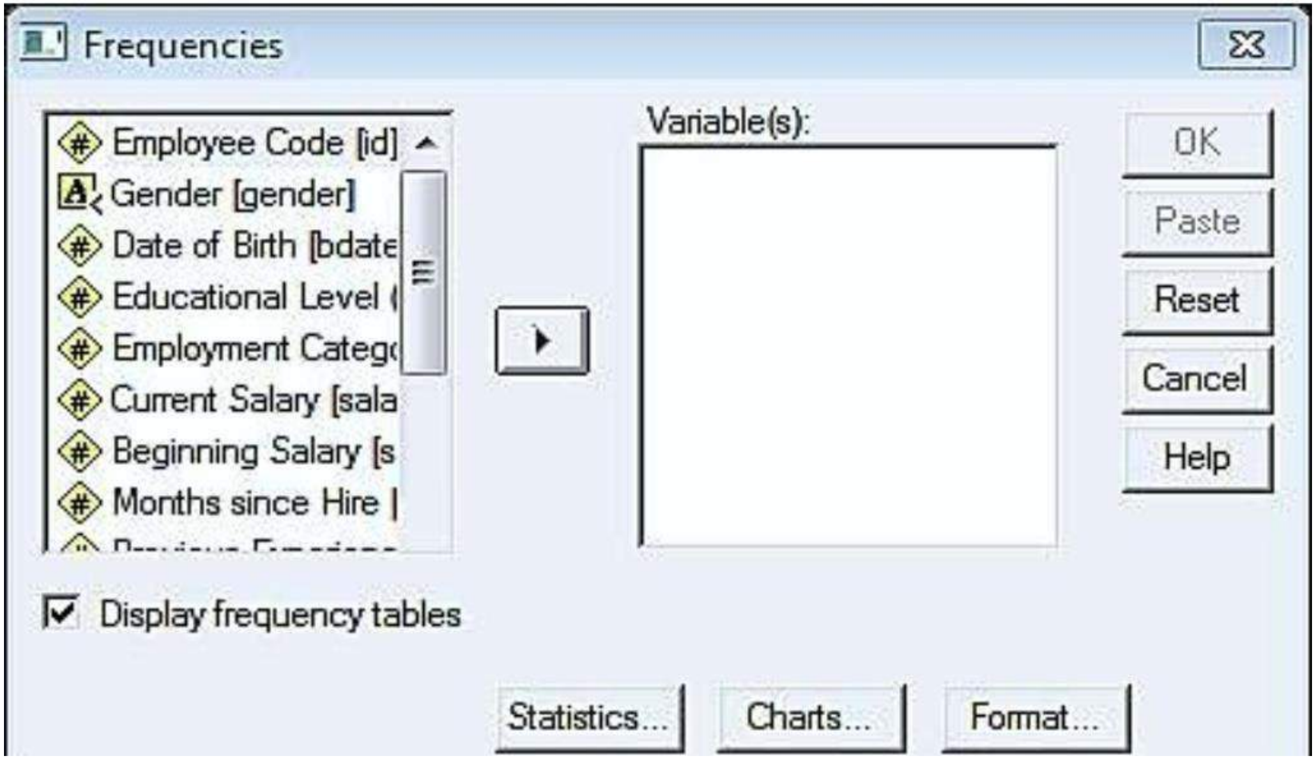
Steps for frequency table

- From the menus choose:

Step 1: Select Analyze
Step 2: Select Descriptive
Statistics
Step 3: Select Frequencies
Now……..

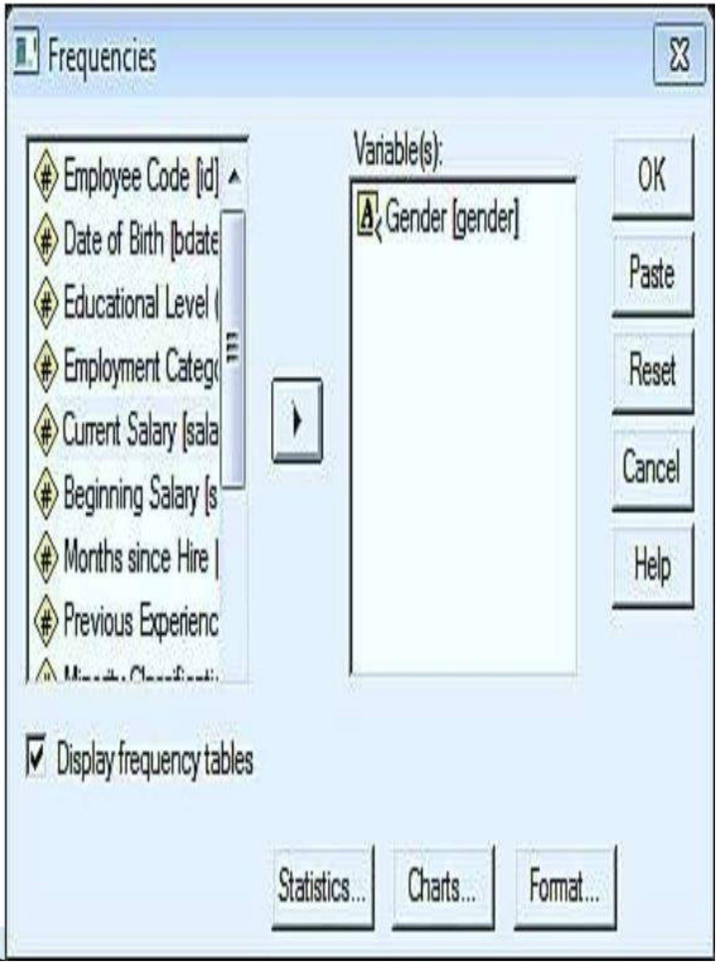


Now……. this open the Frequencies Dialog Box

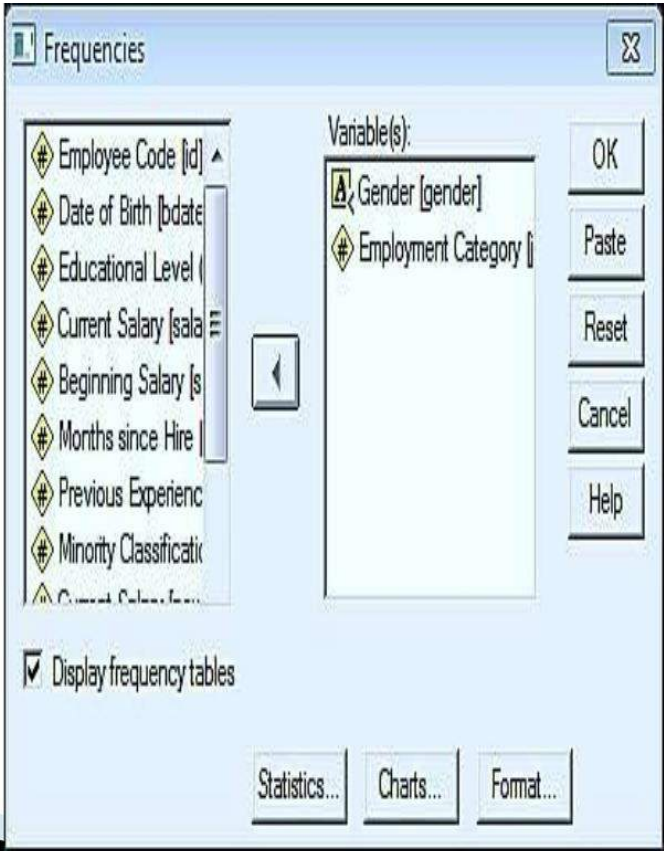


- In the dialog box, you choose the variables you want to analyze from the source list on the left and move them into the Variable(s) list on the right.

- Click *Gender [gender]* in the source variable list, and then click the right-arrow button to move the variable into the target Variable(s) list.



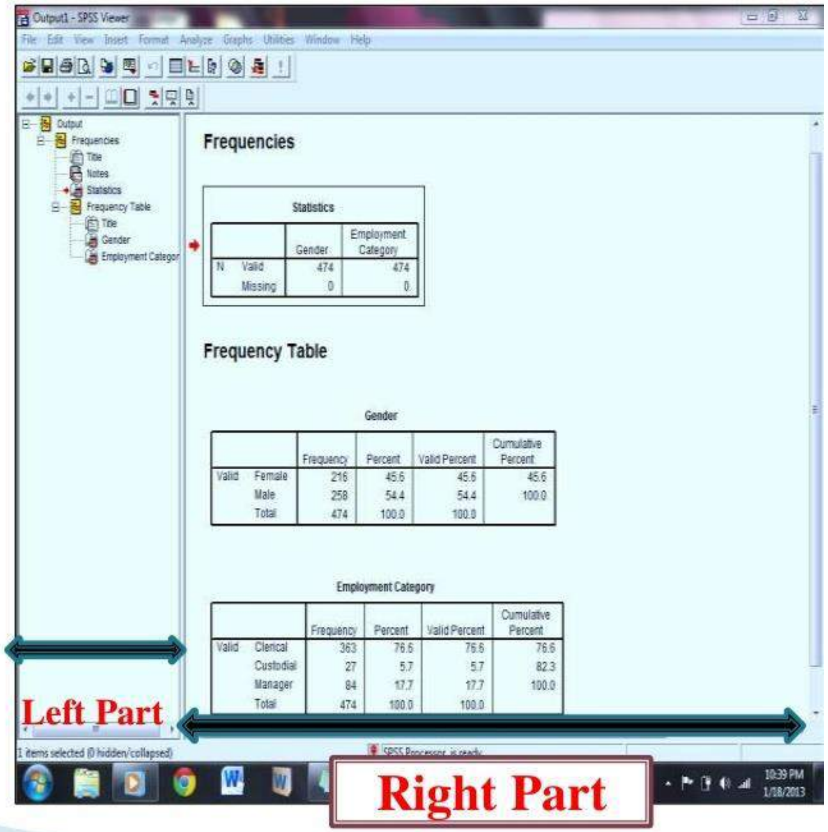
- Click *Employment category [jobcat]* in the source list, and then click the right arrow button again.
- Click OK to run the procedure.



Viewing Results in Output Window

Output window shows the result that are produced by analyzing the functions.

- ➢ The output window is broken in two parts:
  - ➢ Left part: Outline of Output Pane
  - ➢ Right part: Result of Analysis



- Output
  - Frequencies
    - Title
    - Notes
    - Statistics
    - Frequency Table
      - Title
      - Gender
      - Employment Categor

Frequencies

Statistics

		Gender	Employment Category
N	Valid	474	474
	Missing	0	0

Frequency Table

Gender

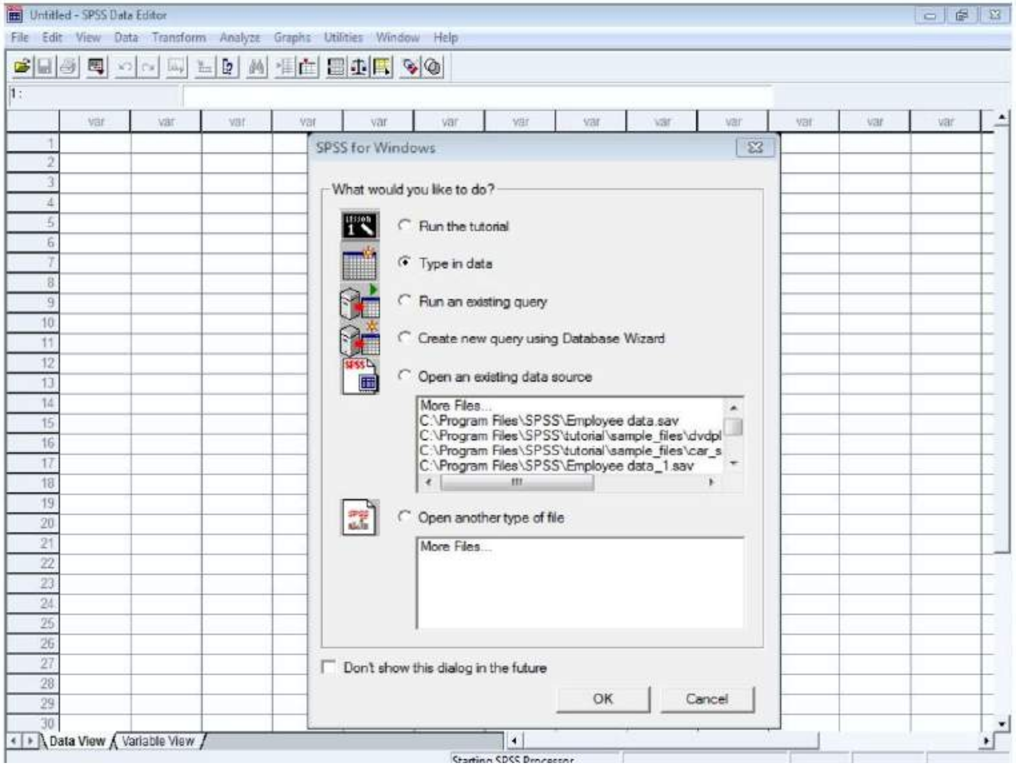
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	216	45.6	45.6	45.6
	Male	258	54.4	54.4	100.0
	Total	474	100.0	100.0	

Employment Category

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Clerical	363	76.6	76.6	76.6
	Custodial	27	5.7	5.7	82.3
	Manager	84	17.7	17.7	100.0
	Total	474	100.0	100.0	

How to Input Data Manually

- First open SPSS through above procedure and in the window click Type in data
- Click OK


Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
SPSS for Windows
What would you like to do?
Run the tutorial
Type in data
Run an existing query
Create new query using Database Wizard
Open an existing data source
More Files...
C:\Program Files\SPSS\Employee data.sav
C:\Program Files\SPSS\tutorial\sample_files\dvdpl
C:\Program Files\SPSS\tutorial\sample_files\car_s
C:\Program Files\SPSS\Employee data_1.sav
Open another type of file
More Files...
Don't show this dialog in the future
OK
Cancel
Data View
Variable View

Example: A close ended questionnaire is developed to know the choice for educational courses

- Information required related to variables: Unique ID, Name, Age, Sex, Educational Qualification, Educational Course Preference.
- First Define Variables

Variable view

- Name
- Type (Numeric)
- Width
- Decimal
- Label
- Values (= the codes of the answers)
- Measure (= Level of Measurement Nominal, Ordinal, Scale)

Filled Questionnaire…How to code it

- Unique ID: 1
- Name: Ajay
- Age: 1

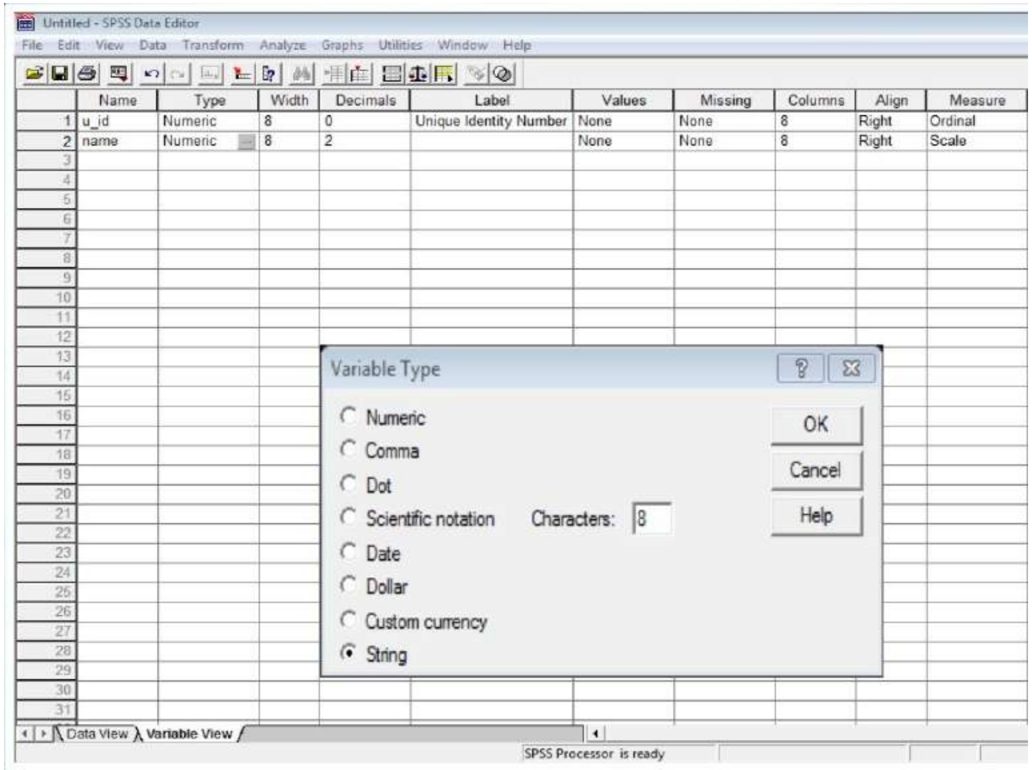
  1 = "<20"; 2 = "20-30"; 3 = "30-40"; 4 = ">40"
- Sex (Categorical Data): 1

  1= "Male" ; 2 = "Female"
- Educational Qualification: 2

  0 = "Illiterate"; 1= "Matric"; 2= "Sr. Secondary";

  3= "Graduate"; 4= "Post Graduate"
- Educational Course Preference: 2

  1= "B. Ed."; 2= "C.S."; 3= "C.A."; 4= "ICWA"; 5= "LLB"

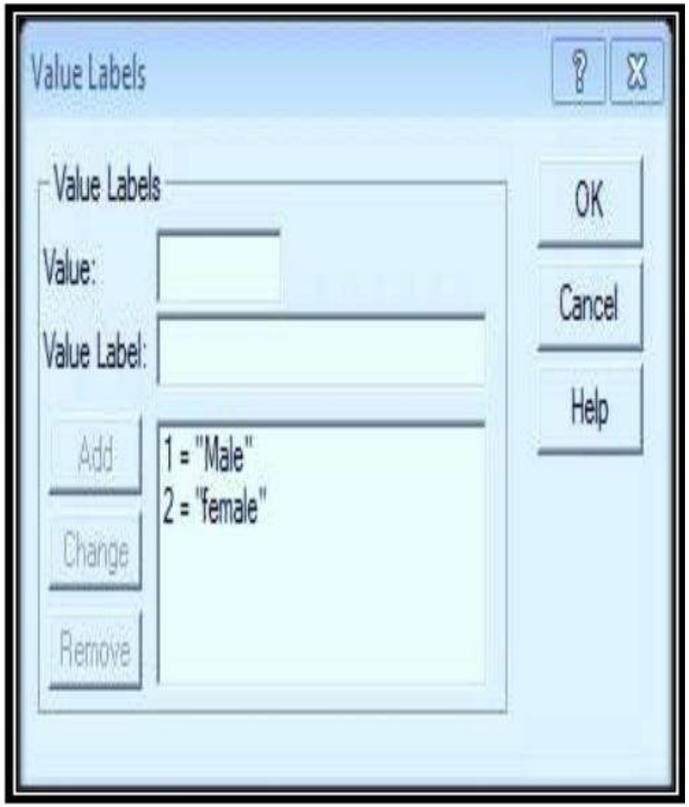

Untitled - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure
1 u_id Numeric 8 0 Unique Identity Number None None 8 Right Ordinal
2 name Numeric 8 2 None None 8 Right Scale
Variable Type
Numeric
Comma
Dot
Scientific notation
Characters: 8
Date
Dollar
Custom currency
String
OK
Cancel
Help
Data View
Variable View
SPSS Processor is ready

Variables

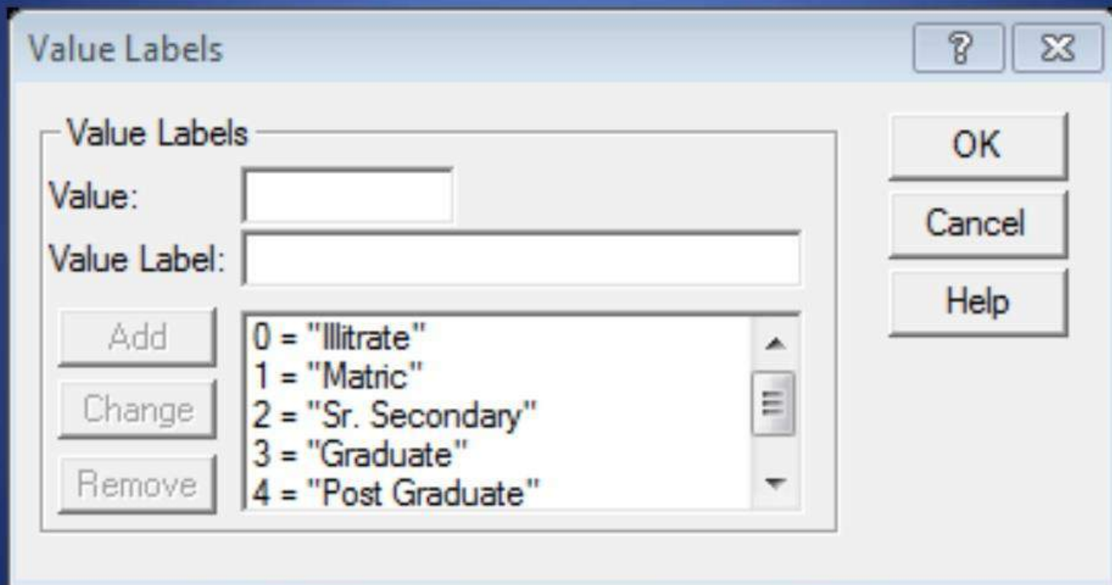
Variable Name	Type	Label	Value Labels (Code)
u_id	Numeric	Unique Identification Number	None
name	String	Name	none
age	Numeric	Age	1 = “<20” 2 = “20-30” 3 = “30-40” 4 = “>40”
sex	Numeric	Gender	1 = “Male” 2 = “Female”
edu_qua	Numeric	Educational Qualification	0 = “Illiterate” 1= “Matric” 2= “Sr. Secondary” 3= “Graduate” 4= “Post Graduate”
edu_cprf	Numeric	Educational Course Preference	1= “B. Ed.” 2= “C.S.” 3= “C.A.” 4= “ICWA” 5= “LLB”

How to Value Labels

- Click on the cell Value label.
- Some dotted lines will appear ahead of “None”
- Click those lines, a new box will pop up i.e. Value Labels Dialog Box as on right.
- Type Value Code in Value Box and type value label
- Click ADD.
- Then OK



Another Example of Value Label for Variable : Educational Qualification



ques_eduprf - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

17 :

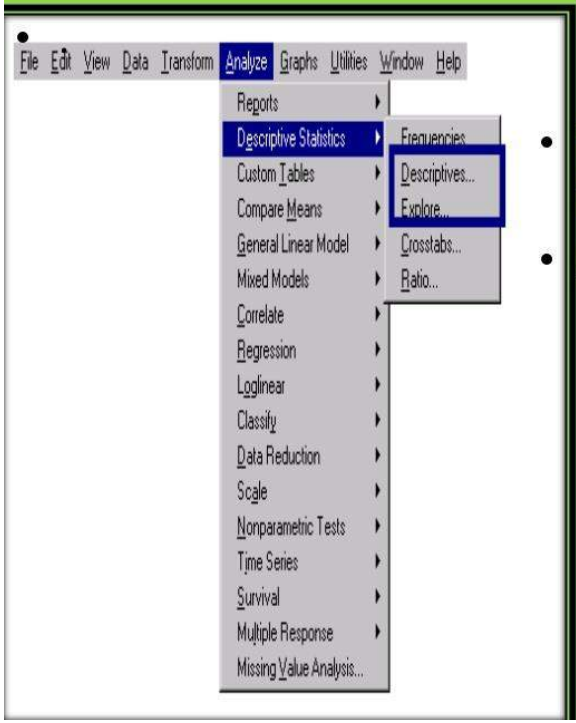
	u_id	name	age	sex	edu_qua	edu_cprf	var
1	1	Ajay	1	1	2	2	
2	2	Sahil	2	1	3	3	
3	3	Ankit	3	1	3	4	
4	4	Ashu	2	1	2	5	
5	5	Rahul	1	1	1	1	
6	6	Hitesh	3	1	4	2	
7	7	himanshu	3	1	4	3	
8	8	Sapna	4	2	4	4	
9	9	Aarti	3	2	2	1	
10	10	Leena	3	2	3	2	
11	11	Shikha	4	2	4	2	
12	12	Seema	2	2	1	2	
13	13	Sarita	1	2	2	3	
14	14	Rinki	3	2	3	3	
15	15	Ritu	2	2	4	4	
16	16	Rashmi	3	2	3	4	
17	17	Dimple	4	2	2	4	
18	18	Diksha	1	2	1	5	
19	19	Dhruv	2	1	1	5	
20	20	Rakesh	1	1	1	2	
21	21	Piyush	2	1	2	2	
22	22	Puneet	2	1	3	2	
23	23	Sachine	3	1	4	2	
24	24	Yuvraj	4	1	4	3	
25	25	Rohit	2	1	4	3	
26	26	Mohit	1	1	3	2	
27	27	Mukesh	1	1	3	3	
28	28	Mohan	2	1	3	4	
29	29	Sunaina	2	2	3	4	
30	30	Divya	3	2	4	2	

Data View / Variable View

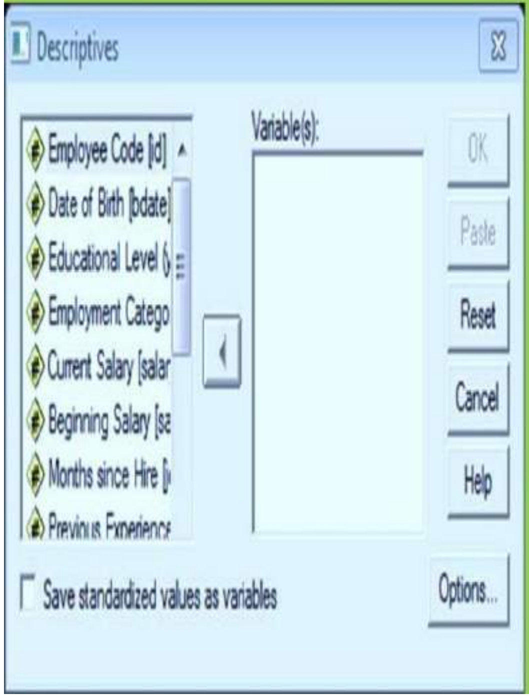
Descriptive Analysis

- The Descriptive procedure displays univariate summary statistics for several variables in a single table and calculates standardized values (z scores). Variables can be ordered by the size of their means (in ascending or descending order), alphabetically, or by the order in which you select the variables (the default).

Procedure of Descriptive Analysis

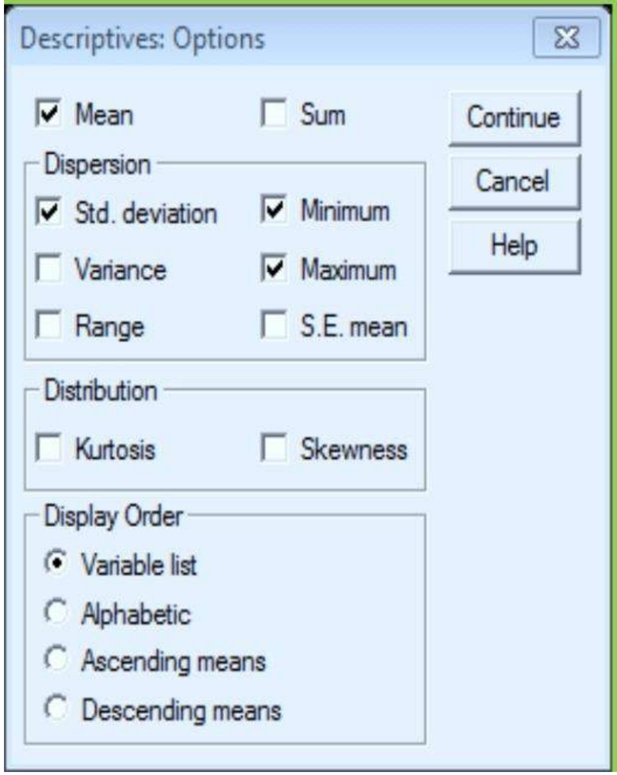


- To run a Descriptive analysis, from the menus choose:
- Analyze
  Descriptive Statistics
  Descriptives...



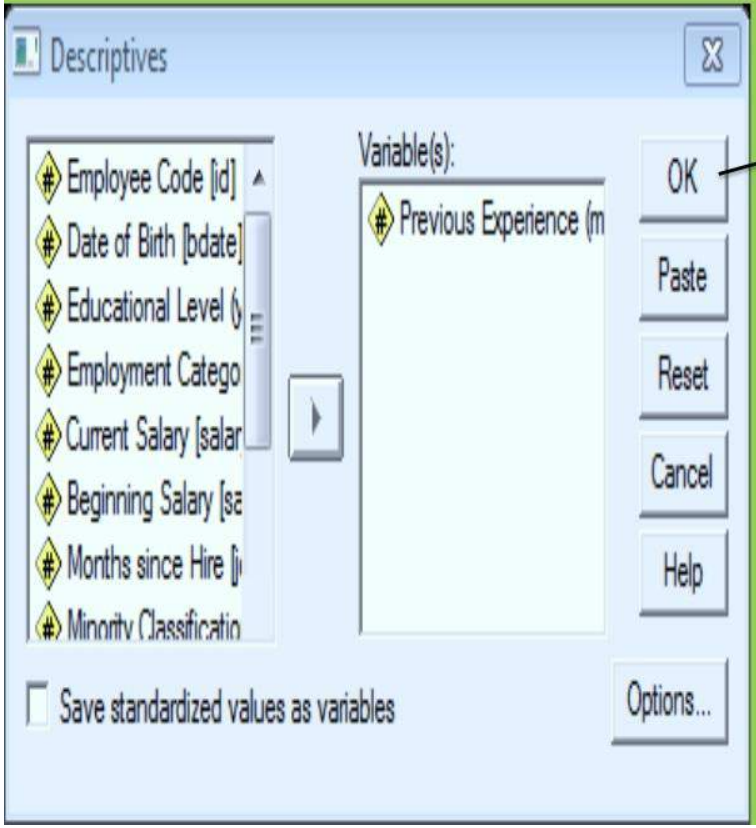
- Now the dialog box appears.
- Now Select the variable Previous Experience in the left side box by clicking the right arrow button between the two boxes.
- The variable will shift to the right side in Variable(s) list.

Procedure of Descriptive Analysis



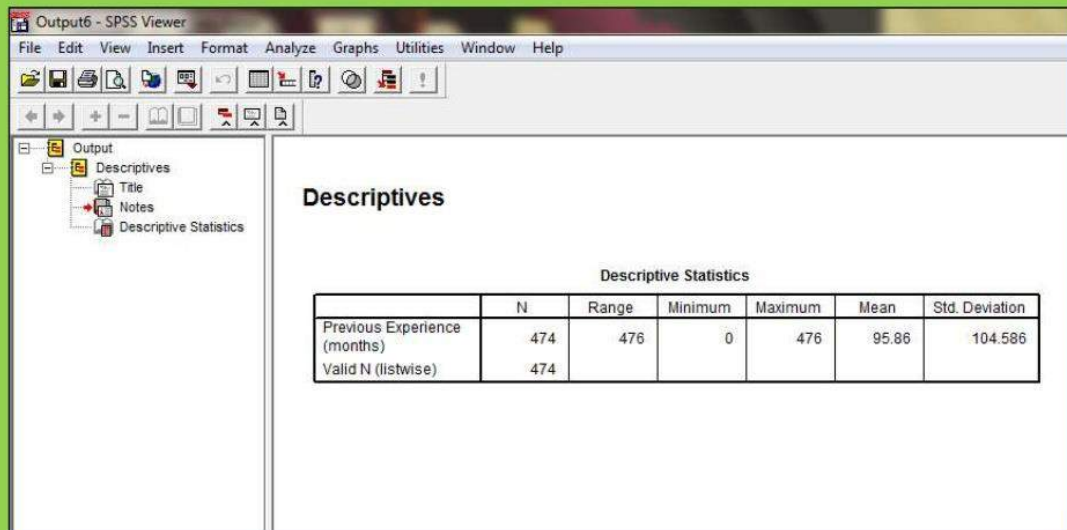
- Select Options from the Descriptive Option dialog box.
- For example: Mean, Minimum, Maximum & Standard Deviation.
- Now CLICK Continue.

Procedure of Descriptive Analysis



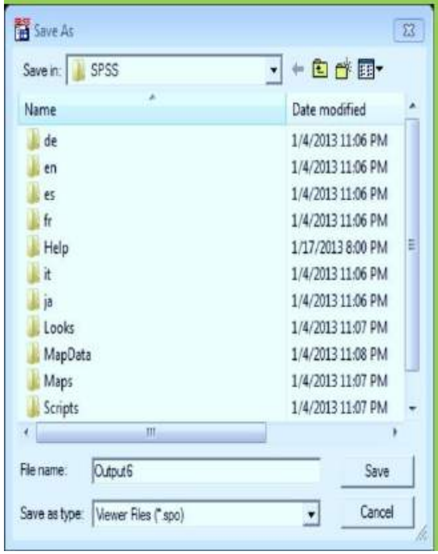
- *Click OK tab.*

Output of Descriptive Analysis in Output Window



	N	Range	Minimum	Maximum	Mean	Std. Deviation
Previous Experience (months)	474	476	0	476	95.86	104.586
Valid N (listwise)	474					

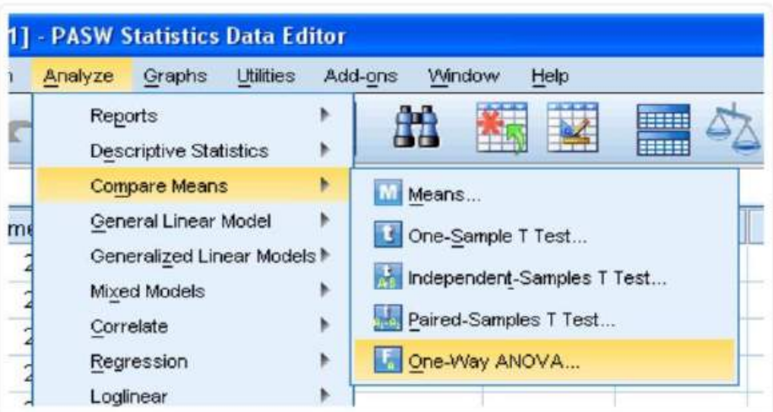
How to Save the Output……..



- Now Click on the FILE MENU
- SELECT the option Save the dialog box appears.
- Click the Tab Save.

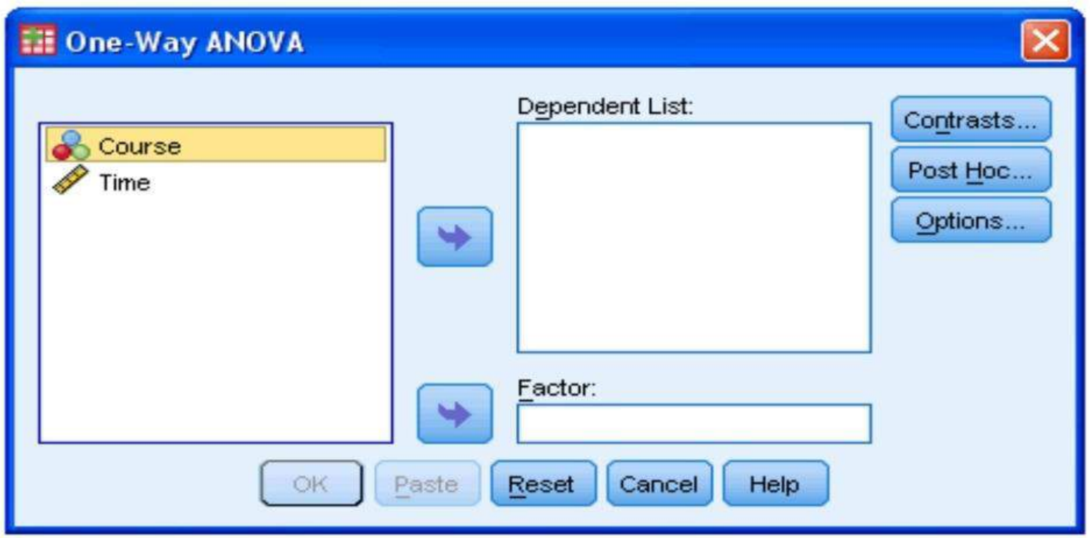
Procedure of ONE-WAY ANOVA

- Click **Analyze** > **Compare Means** > **One-Way ANOVA...** on the top menu as shown below.



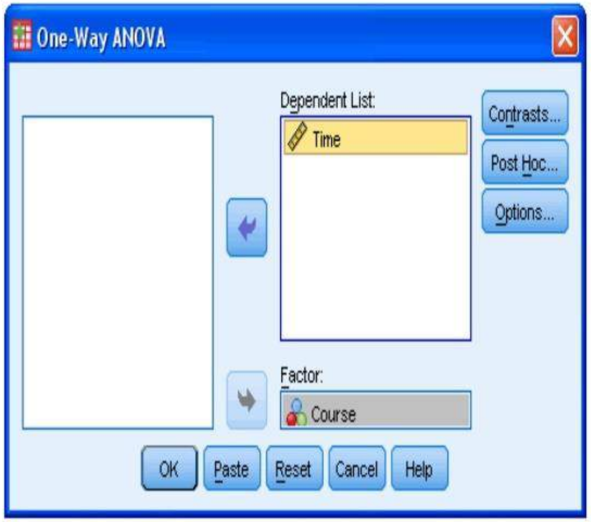
Published with written permission from SPSS Inc., an IBM Company.

Now you will be presented with the following screen:



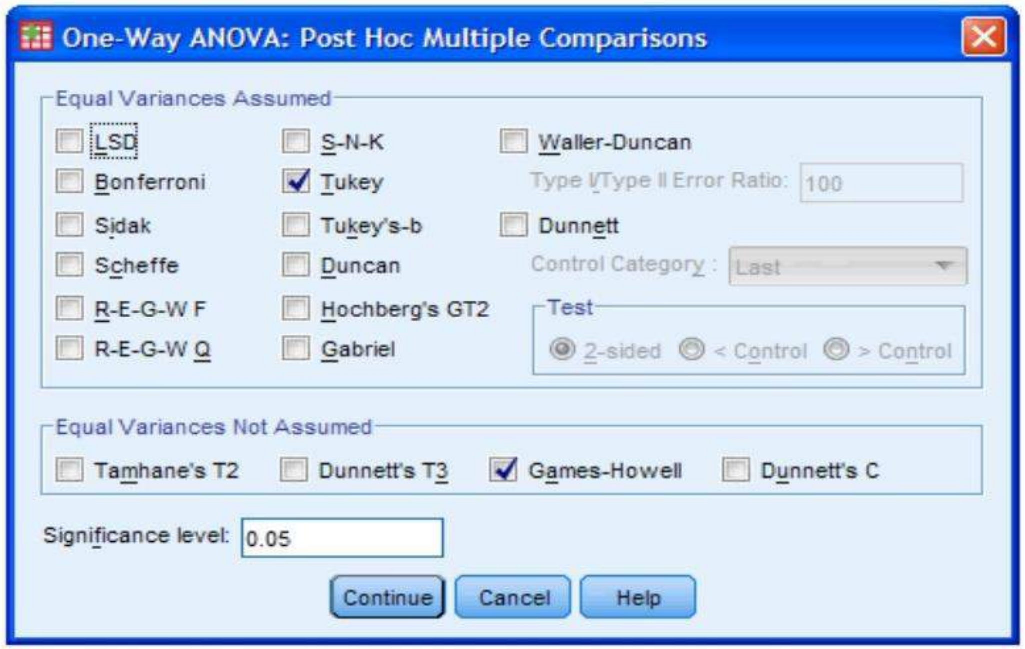
Published with written permission from SPSS Inc., an IBM Company.

Drag-and-drop (or use the buttons) to transfer the dependent variable (Time) into the Dependent List: box and the independent variable (Course) into the Factor: box as indicted in the diagram below:



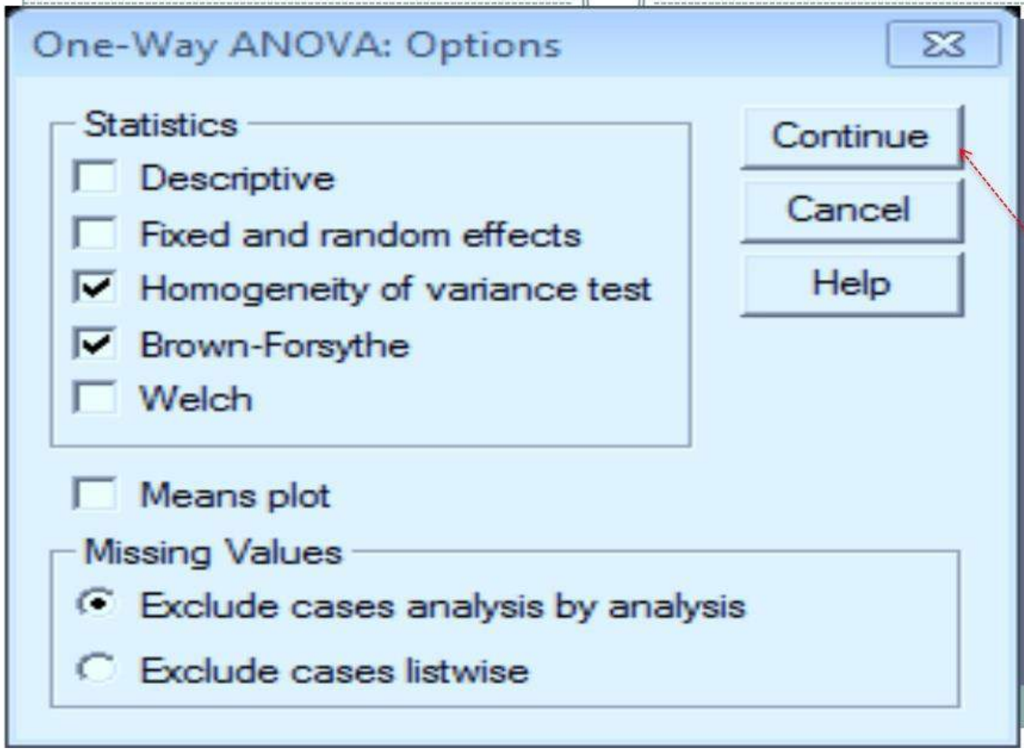
Published with written permission from SPSS Inc., an IBM Company.

Click the Post Hoc... button. Tick the Tukey checkbox as shown below:



Click the Continue button.

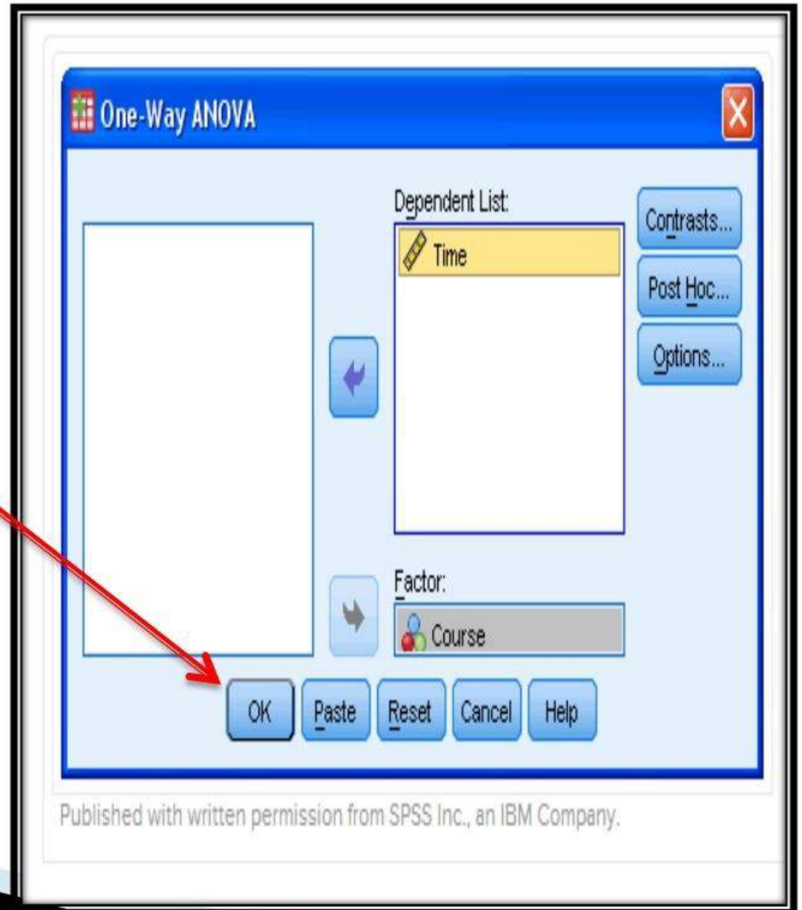
Analysis of One-Way ANOVA……



Then :
Click the Continu e Button

Now Click the 

Button



Published with written permission from SPSS Inc., an IBM Company.

Output of ANOVA in SPSS

Descriptives

Time

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Beginner	10	27.2000	3.04777	.96379	25.0198	29.3802	22.00	33.00
Intermediate	10	23.6000	3.30656	1.04563	21.2346	25.9654	18.00	29.00
Advanced	10	23.4000	3.23866	1.02415	21.0832	25.7168	18.00	29.00
Total	30	24.7333	3.56161	.65026	23.4034	26.0633	18.00	33.00

ANOVA

Time

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	91.467	2	45.733	4.467	.021
Within Groups	276.400	27	10.237		
Total	367.867	29			

Output of ANOVA in SPSS

Multiple Comparisons

Dependent Variable:Time

	(I) Course	(J) Course	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	Beginner	Intermediate	3.60000*	1.43088	.046	.0523	7.1477
		Advanced	3.80000*	1.43088	.034	.2523	7.3477
	Intermediate	Beginner	-3.60000*	1.43088	.046	-7.1477	-.0523
		Advanced	.20000	1.43088	.989	-3.3477	3.7477
	Advanced	Beginner	-3.80000*	1.43088	.034	-7.3477	-.2523
		Intermediate	-.20000	1.43088	.989	-3.7477	3.3477
Games-Howell	Beginner	Intermediate	3.60000	1.42205	.052	-.0314	7.2314
		Advanced	3.80000*	1.40633	.037	.2096	7.3904
	Intermediate	Beginner	-3.60000	1.42205	.052	-7.2314	.0314
		Advanced	.20000	1.46363	.990	-3.5356	3.9356
	Advanced	Beginner	-3.80000*	1.40633	.037	-7.3904	-.2096
		Intermediate	-.20000	1.46363	.990	-3.9356	3.5356

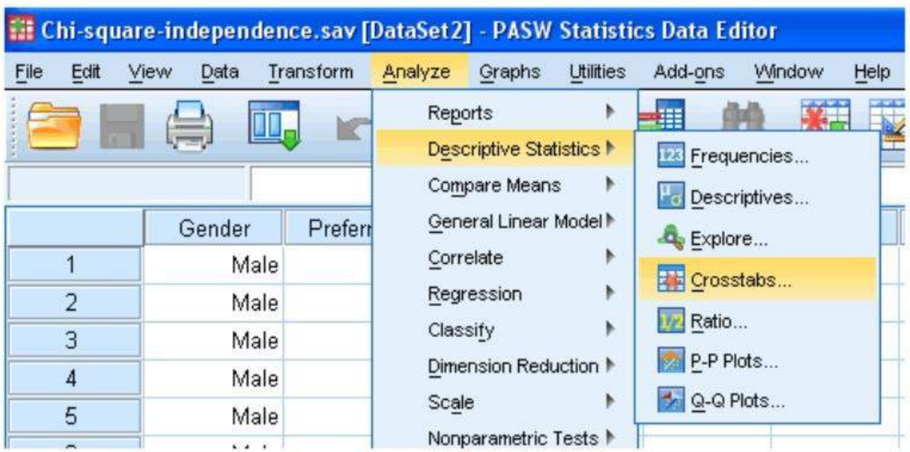
*. The mean difference is significant at the 0.05 level.

*CHI - SQUARE

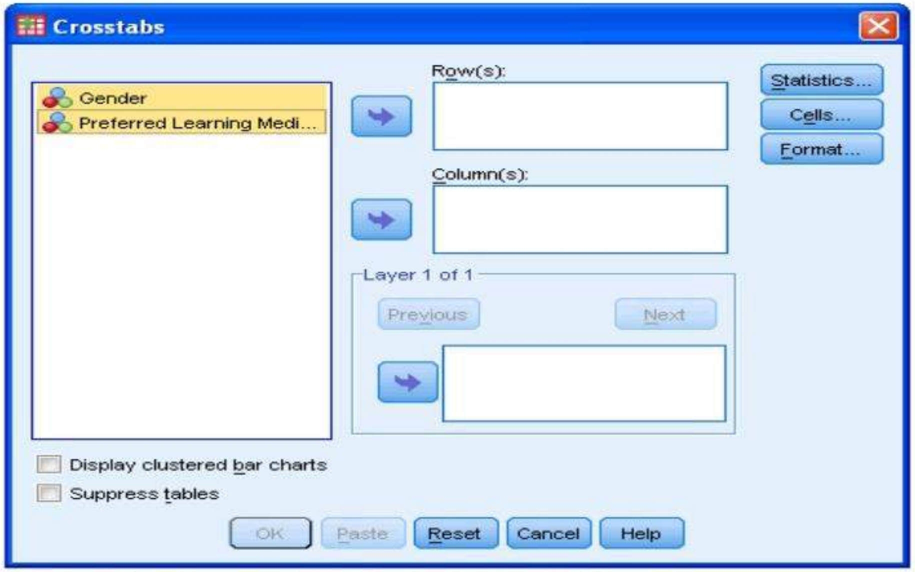
- The chi-square test for independence, also called Pearson's chi-square test or the chi-square test of association, is used to discover if there is a relationship between two categorical variables.
- **Assumption #1:** Your **two variables** should be measured at an **ordinal** or **nominal level** (i.e., **categorical** data).
- **Assumption #2:** Your two variable should consist of **two or more categorical**, **independent groups**. Example independent variables that meet this criterion include gender (2 groups: Males and Females), ethnicity (e.g., 3 groups: Caucasian, African American and Hispanic), physical activity level (e.g., 4 groups: sedentary, low, moderate and high), profession (e.g., 5 groups: surgeon, doctor, nurse, dentist, therapist), and so forth.

Analysis of Chi – Square….

Click **Analyze** > **Descriptives Statistics** > **Crosstabs...** on the top menu, as shown below:



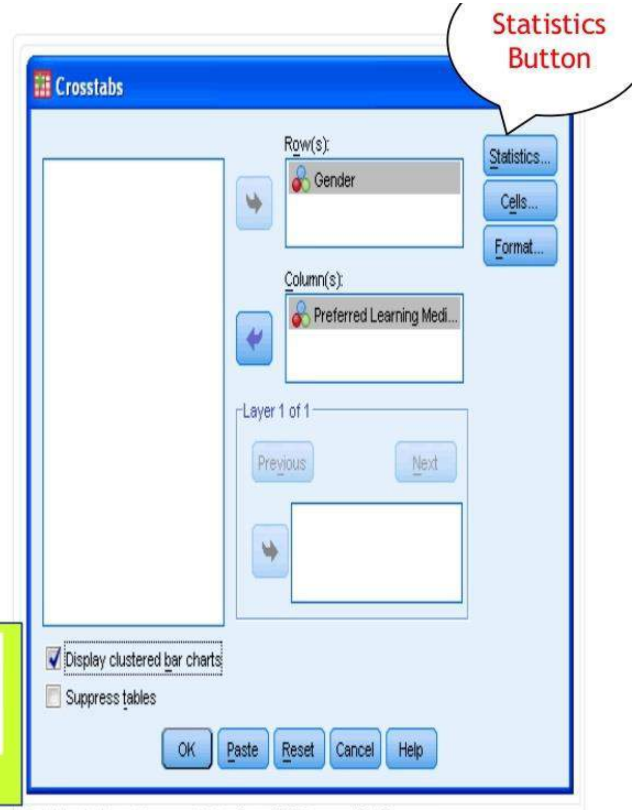
Published with written permission from SPSS Inc., an IBM Company.



Published with written permission from SPSS Inc., an IBM Company.

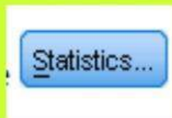
*Now you will be presented with the following screen:

* **Now transfer one variable into the Row i.e. Gender and the other variable to Column i.e. Preferred Learning Medium**

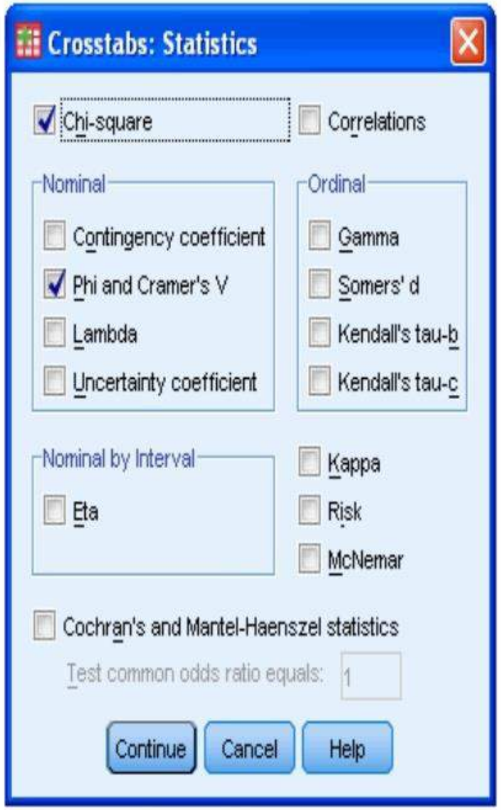


Published with written permission from SPSS Inc., an IBM Company.

Now Click Button



- Now Click the option Chi-square & Phi and Cramer's V as show in the left.
- Click the Continue Button.
- Then Click the Cells.. Button.

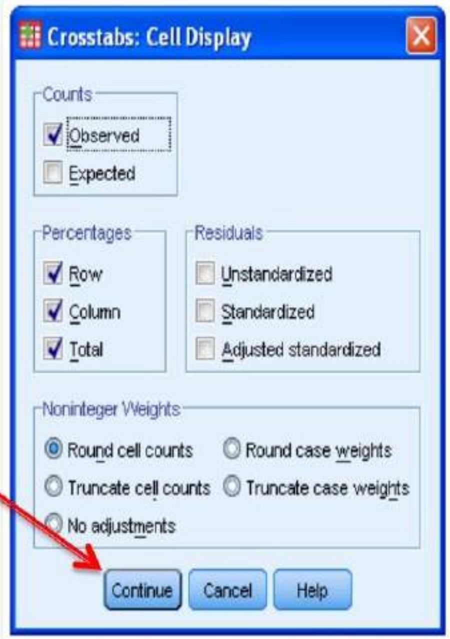


Published with written permission from SPSS Inc., an IBM Company.

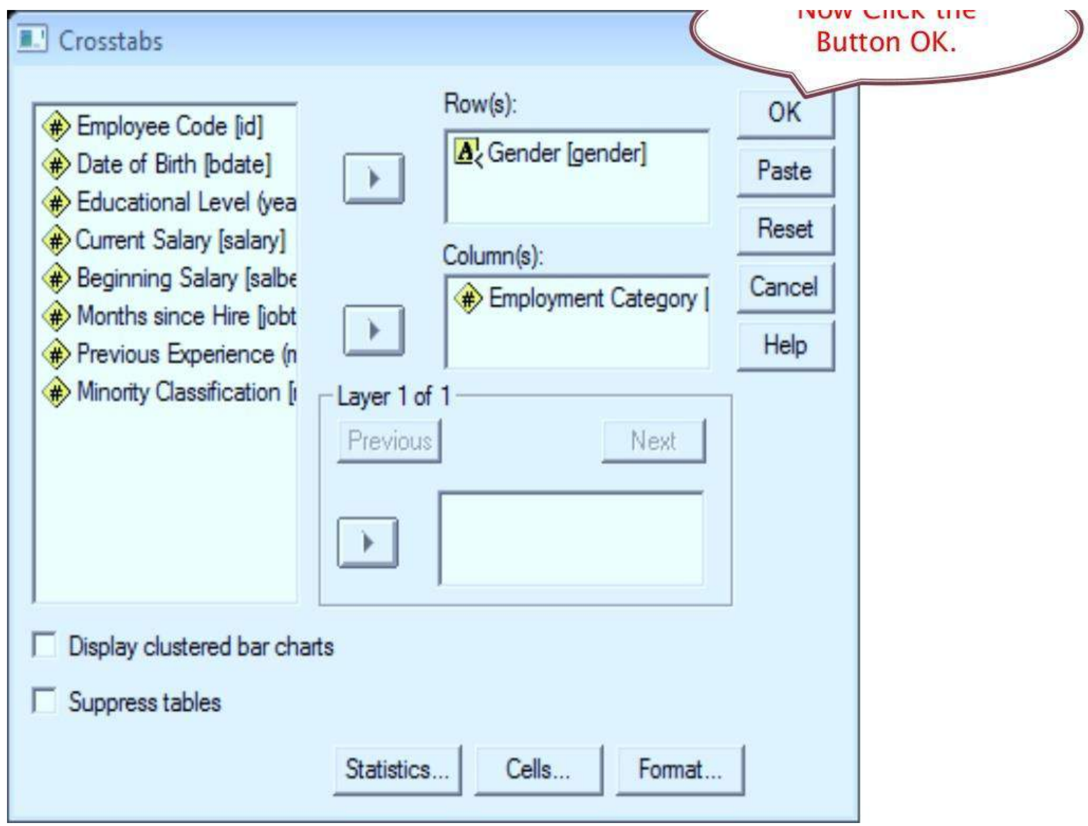
Select the ✓ Observed

✓ Row, ✓ Column and ✓ Total from the -Percentages-

Now Click the Continue Button



Published with written permission from SPSS Inc., an IBM Company.


Crosstabs
Button OK.
Row(s):
Gender [gender]
Column(s):
Employment Category [
Employee Code [id]
Date of Birth [bdate]
Educational Level (yea
Current Salary [salary]
Beginning Salary [salbe
Months since Hire [jobt
Previous Experience (n
Minority Classification [
Layer 1 of 1
Previous
Next
OK
Paste
Reset
Cancel
Help
Display clustered bar charts
Suppress tables
Statistics...
Cells...
Format...

*Output of Chi Square…..

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender * Employment Category	474	100.0%	0	.0%	474	100.0%

Gender * Employment Category Crosstabulation

			Employment Category			
			Clerical	Custodial	Manager	Total
Gender	Female	Count	206	0	10	216
		% within Gender	95.4%	.0%	4.6%	100.0%
		% within Employment Category	56.7%	.0%	11.9%	45.6%
		% of Total	43.5%	.0%	2.1%	45.6%
	Male	Count	157	27	74	258
		% within Gender	60.9%	10.5%	28.7%	100.0%
		% within Employment Category	43.3%	100.0%	88.1%	54.4%
		% of Total	33.1%	5.7%	15.6%	54.4%
Total		Count	363	27	84	474
		% within Gender	76.6%	5.7%	17.7%	100.0%
		% within Employment Category	100.0%	100.0%	100.0%	100.0%
		% of Total	76.6%	5.7%	17.7%	100.0%

Output of Chi Square…..

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	79.277[a]	2	.000
Likelihood Ratio	95.463	2	.000
N of Valid Cases	474		

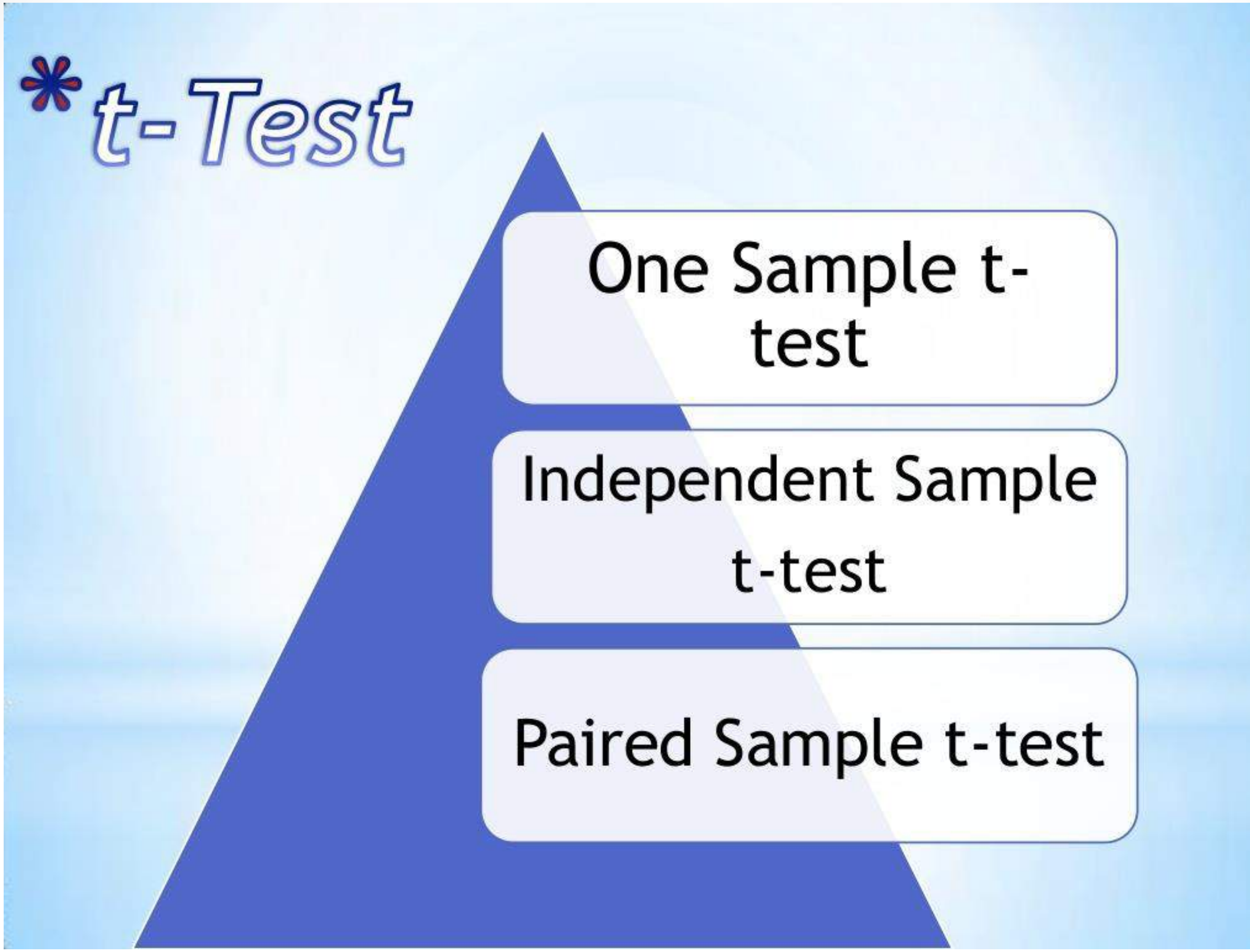
a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 12.30.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.409	.000
	Cramer's V	.409	.000
N of Valid Cases		474	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.


*t-Test
One Sample t-test
Independent Sample t-test
Paired Sample t-test

*t-Test

- A one sample t-test is used when you want to know if there is a significant difference between a sample mean and a test value. The test value can be the known mean from a population or some other value used to compare the sample mean against.
- An independent-samples t-test is used when you want to compare the mean scores on some continuous variable for two different groups of subjects.
- The Paired Samples t test compares the means of two variables. It computes the difference between the two variables for each case, and tests to see if the difference is significant or not.