Research Methodology
— A Perspective

Samarth Borker
Goa College of Engg.
@SGGSIE&T.

What is Research?

- Research in common parlance refers to a search for knowledge.

Defn: It is a scientific and systematic search for pertinent information on a specific topic.

Dictionary: It is a careful investigation or inquiry specially through search for new facts in any branch of knowledge.

When the unknown confronts us, more and more our inquisitiveness makes us probe and attain understanding of the unknown. This inquisitiveness is the mother of all knowledge and method, which one employs for obtaining the knowledge of whatever the unknown, can be termed as Research

Prof. Clifford Woody:

Research comprises defining & redefining problems, formulating hypothesis or suggested solutions; collecting, organising, and evaluating data; making deductions and research conclusions; and at last carefully testing the conclusions to determine whether they fit the formulating hypothesis.

D. Slesinger & M. Stephenson

Research is the manipulation of the things, concepts or symbols for the purpose of generalising to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or the practise of an art.

In short;

Research, is an ORIGINAL contribution to the existing stock of knowledge making for its advancement.

To, sum it up :—
Research refers to the systematic method consisting of enunciating the problem, formulating a hypothesis, collecting the facts or data, analyzing the facts and reaching certain conclusions either in the form of solution(s) towards the concerned problem or in certain generalisations for the theoretical formulation.

Samarth Borkar

Objectives of Research

1. Exploratory or formulative Research studies.
   To gain familiarity with a phenomenon or to achieve new insights into it.
2. Descriptive Research studies
   To potray accurately the characteristics of a particular individual, situation or a group.
3. Diagonastic Research
   To determine the frequency with which something occurs or with which it is associated with something else
4. Hypothesis Testing Research
   To test a hypothesis of a causal relationship between variables.

Types of Research

i) Descriptive v/s Analytical.
ii) Applied v/s Fundamental.
iii) Quantitative v/s Qualitative.
iv) Conceptual v/s Empirical.
v) Some other types of Research.
   - one time research or longitudinal research
   - field setting or laboratory research
   - Clinical or diagnostic research
   - Historical research
   - Conclusion oriented & decision oriented
   - Exploratory Research.

i] Descriptive Research (DR) v/s Analytical R (AR)

DR includes surveys & fact-finding enquiries of different kinds. The major purpose of DR is to describe the state of affairs as it exists at present. for example frequency of shopping, preferences of people or similar data. It includes attempts by researchers to discover causes even when they cannot control the variables.

In Analytical research, the researcher has to use facts or information already available and analyze these to make a critical evaluation of the material.

ii) Applied v/s fundamental Research

Applied research aims at finding a solution for an immediate problem facing a society or an industrial/business organisation, whereas fundamental research is mainly concerned with generalisations and with the formulation of a theory.

Central aim of applied research is to discover a solution for some pressing practical problems, whereas basic research is directed towards finding information that has a broad base of application and thus, adds to the already existing organized body of scientific knowledge.

iii) Quantitative v/s Qualitative

Quantitative research is based on the quantitative measurements of some characteristics. It is applicable to phenomenon that can be expressed in terms of quantities.

Qualitative research, is concerned with qualitative phenomenon relating to or involving quality or kind. For eg "Motivation Research".

Qualitative practice is relatively a difficult job and therefore, while doing such research one should seek guidance from experimental psychologists

iv) Conceptual v/s Empirical:

Conceptual research is related to some abstract idea(s) or theory. It is generally used by philosophers and thinkers to develop new concepts or to re-interprete the existing ones.

Empirical research, relies on experience or observation alone, often without due regard for system & theory.

It is data based research, coming up with conclusions which are capable of being verified by observation or experiment.

Evidence gathered through experiments, or empirical studies are considered to be the most powerfull support possible for testing a given hypothesis.

Some other types of Research

i) **One-time Research**
- Confined to single time period.

ii) **Longitudinal Research**
- Research is carried out over several time periods.

iii) **field setting Research**
iv) **laboratory research**
v) **Simulation research**

} depending upon Place of work & environment.

vi) **Clinical or diagnotic Research**
- Studies usually go deep into the causes, using very small samples & very deep probing data gathering devices.

vii) **Exploratory Research**
- Concerned with development of hypothesis rather than their testing.

viii) **Formalized Research**
- Given substantial structure, researcher has to test specific hypothesis.

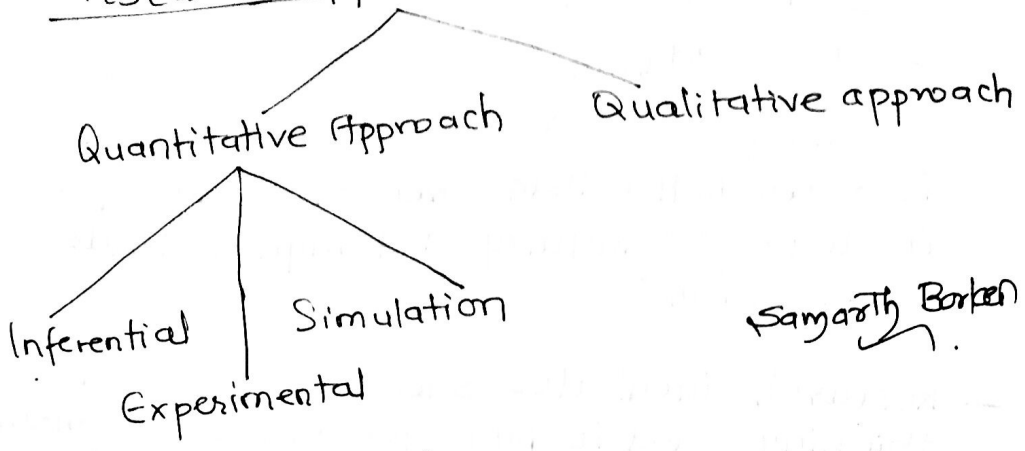
ix) **Conclusion Oriented**
- Researcher is free to pick up any problem redesign the enquiry & conceptualize as he wish

x) **Decision Oriented Research**

Depends upon need of decision maker & researcher is not free to embark upon research according to his own inclination.

ex. Operations Research.

Research Approaches



Inferential Approach:

- Is to form a database to infere characteristics or relationship of population ex Survey Research.

Experimental Approach:

- In this case, using greater control, some variables are manipulated to observe their effect on other variables

Simulation Approach

- It involves construction of an artificial environment within which relevant information and data can be generated.

Qualitative Approach

- It is concerned with subjective assessment of attitudes, opinions & behavior.
- Research in such a situation is a function of researchers insights & impressions.

Techniques used are.

a) Focus group interviews.
b) Projective techniques.
c) depth interviews.

Significance of Research

Prof Hudson Maxim :

"All progress is born of inquiry. Doubt is often better than overconfidence, for it leads to inquiry and inquiry leads to invention."

- Research inculcates scientific & inductive thinking and it promotes the development of logical habits of thinking & organization.
- The role of research in several fields of applied economics, whether related to business or to the economy as a whole, has greatly increased in modern times.
- Research provides the basis for nearly all government policies in our economic system.
- Research has its special significance in solving various operational & planning problems of buisness & Industry.
- Research is equally important for social scientists in studying social relationship and in seeking answers to various social problems.

Research in social science is concerned with.

1) The development of a body of principles that help in understanding human interactions.
2) Practical guidance in solving immediate problems of human relations

Significance of Research can also be understood keeping in view the following points

a) To those candidates who are to write Master's or Ph.D thesis, research may mean a careerism or a way to attain a high position/post in the social structure. (QIP).

b) To professionals in research methodology, research may mean a source of livelihood.

c) To philosophers & thinkers, research may mean the outlet for new ideas & insights.

d) To literary men & women, research may mean the development of new styles and creative work; and.

e) To analyst and intellectuals, research may mean the development of new theories.

In short;

It is a formal sort of training which enables one to understand the new developments in one's field in a better way.

Research Methods v/s Research Methodology

Research methods may be defined as all those methods/techniques that are used for conduction of Research.

Thus, it refers to the methods the researchers use in performing research operations.

In other words, all those methods which are used by the researcher during the course of studying his research problem are termed as research methods.

Classification

i) Methods concerned with collection of data.
ii) Statistical techniques.
iii) Techniques used for evaluation of accuracy & results.

What is the difference between Research methods & research techniques?

- Research technique refer to behaviour and instruments we use in performing research operations such as : i) making observations
  ii) recording data
  iii) Techniques of preprocessing

Research Methods refer to the behaviour and instruments used in selecting and constructing research techniques.

ex: the difference between methods & techniques of data collection.

Research Methodology

Sampath Barker

It is a systematic way to solve the research problem. In this we study, the various steps that are adopted by researcher in studying his research problem along with the logic behind them.

It is necessary for the researcher to know not only the research methods/techniques but also the methodology.

Researchers also need to understand the assumptions underlying various techniques and they need to know the criteria by which they can decide that certain techniques and procedures will be applicable to certain problems and others will not.

Research Methodology has many dimensions and research methods to constitute a part of it.

So, when we talk of Research Methodology, we not only talk of the research methods but also consider the logic behind the methods we use in the context of our research study and explain why we are using a particular method or technique, and why we are not using others, so that research results are capable of being evaluated either by the researcher himself or by others.

It raises questions like:-

i] Why research study has been undertaken?
ii] How research problem has been designed?
iii] How & why, hypothesis has been formulated?
iv] What data has been collected?
v] Why particular method & technique adopted?

Research & Scientific Method

In Research; we study the nature, reasons and the consequences of a set of circumstances which are controlled experimentally or observed as they appear.

Scientific Method refers to the philosophy common to all research methods and techniques, although they may vary considerably from one science to another.

Scientific method is a persuit of thruth as determined by logical considerations.

All this is done through experimentation & survey investigations which constitutes the integral parts of scientific method.

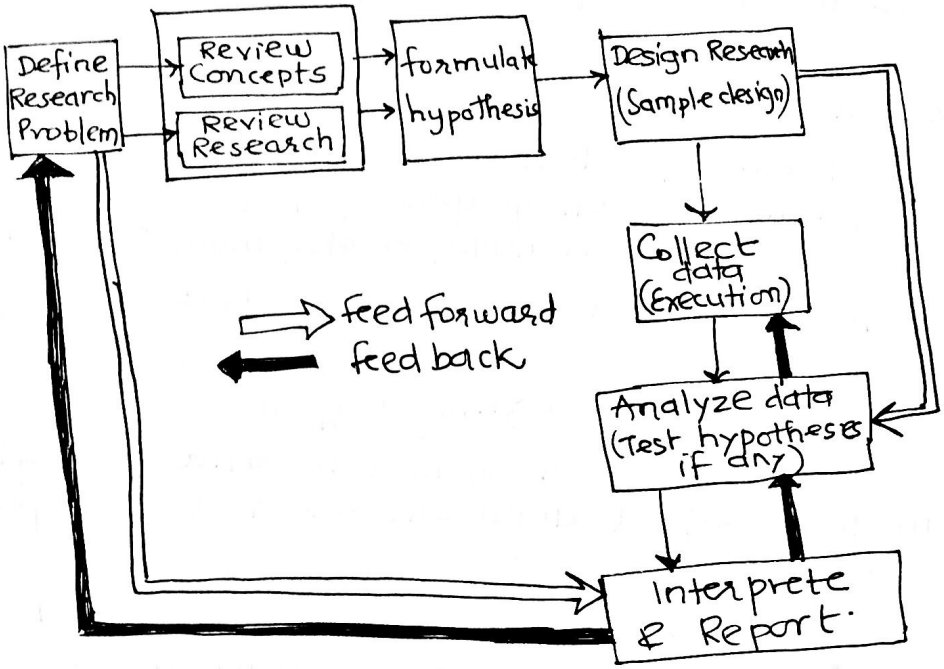
Experimentation is done to test hypotheses and to discover new relationships if any, among variables.

The scientific method based on certain postulates are stated as under

1. It relies on empirical evidence
2. It uses relevant concepts.
3. It only commited to objective considerations.
4. Making only adequate & correct statements.
5. It results into probabilistic predictions.
6. Well defined methodology & are for use in testing the conclusions through replication.
7. It aims at formulating general axioms or what can be termed as scientific theories.

Research Process

fig below illustrates a research process.



Procedural guideline regarding the research process.

i] formulating the research problem
ii] Extensive literature survey
iii] developing the hypothesis.
iv] Preparing the research design.
v] Determining sample design.
vi] Collecting the data
vii] Execution of the project
viii] Analysis of data
ix] hypothesis testing.
x] Generalizations & Interpretation
xi] Preparation of the report or presentation of results.

Samarth Borken

A brief discussion of above states steps.

i) Formulating the research problem

Initially the problem is stated in broad category. This formulation into specific research problem, constitutes first step of scientific enquiry.

ii) Extensive literature survey

- Abstracting, indexing journals, published or unpublished bibliographies, Academic journals, conference proceedings, books, monlogues etc.
- Google scholar is the best approach to begin.

iii) Development of working hypotheses

Working hypothesis is a tentative assumption made in order to draw out and tests 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 hypothesis is to guide the researcher by delimiting the area of research and to keep him on the right track

It sharpens his thinking and focusses his attention on the more important facets of the problem.

It also indicates the type of data required and the type of methods of data analysis to be used.

Approach

i) Discussion with collegue, experts etc
ii) Examination of data & Records
iii) Review of similar studies
iv) Exploratory personal investigation.

iv] Preparing the research design

The preparation of research design facilitates research to be as efficient as possible yielding maximal information.

i] Exploration
ii] Description
iii] Diagnosis, and
iv] Experimentation.

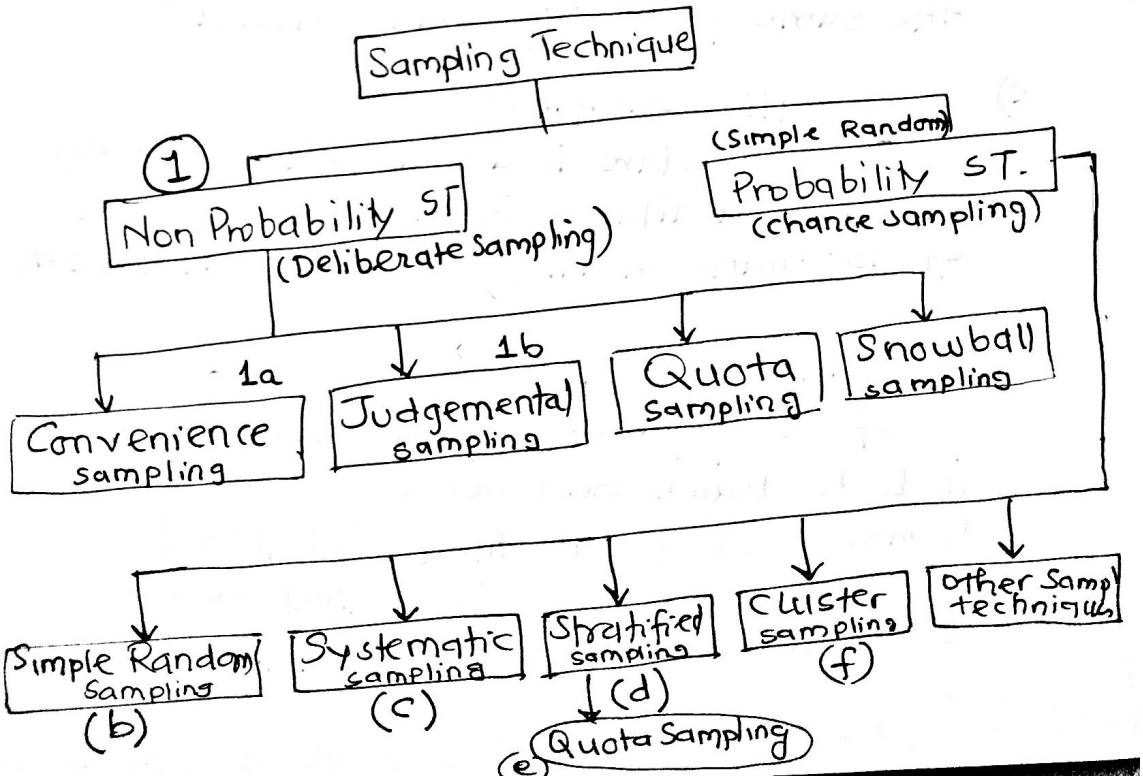
Samarth Borker
GOA COLLEGE OF ENGG

The preparation of research design, involves usually the consideration of the following

a] the means of obtaining the information.
b] the availability of skills of researcher & staff
c] Organized Explanation
d] time available for research
e] Cost factor relating to research.

v] Determining the sample design.

Classification of Sampling Techniques



a) Deliberate Sampling

- It involves purposive or deliberate selection of particular units of the universe for constituting a sample which represents the universe.

Convenience Sampling

If the elements are selected for inclusion in the sample based on the ease of access.

Judgement sampling

Researchers judgement is used for selecting items which he considers as representative of the population.

b) Simple random sampling

(chance or Probability sampling)

- In this each & every item in the population has an equal chance of inclusion in the sample and each one of the possible samples, incase of finite universe, has the same probability being selected.

c) Systematic sampling

This procedure is useful when sampling frame is available in the form of a list. eg 15th name on a list, or every 10th house etc

d) Stratified Sampling

If the population from which a sample is to be drawn does not constitute a homogeneous group, then a stratified sampling technique is applied so as to obtain a representative sample

e] Quota Sampling

In this interviewers are simply given quota to be filled from different strata, the actual selection of items for sample being left to the interviewers judgement.

f) Cluster sampling and Area sampling

Cluster sampling involves grouping the population and then selecting the groups or the clusters rather than individual elements for inclusion in the sample.

Area Sampling. If geographical area seems to be big one, the big area is divided into overlapping clusters, then a number of these smaller areas are randomly selected, and all units in these small areas are included in the sample

g) Multi-Stage Sampling

eg. Extension of cluster sampling

Country → state → districts → towns → families

If random sampling is applied at all stages, it is then termed as multi-stage random sampling

h) Sequential Sampling

In this case ultimate size of the sample is not fixed in advance but is determined according to mathematical decisions on the basis of information yielded as survey progresses.

Summary

The sample design to be used must be decided by the researcher taking into consideration the nature of the inquiry and other related factors.

vi] Collecting the data

Primary data can be collected either through
a) Experiment, or
b) survey

Samarth Booken

If the researcher conducts an experiment, he observes some quantitative measurements, or the data, with the help of which he examines the thruth contained in his hypothesis.

In case of survey data can be collected by any one or more of the foll ways.

i] By Observation.
- without interviewing the respondents
- Information depends only on current affair.
- Limited Information
- Expensive method

ii] Through Personal Interview
- Rigid procedure
- set of pre-conceived questionaire

iii] Through Telephone interviews
- Used in industrial surveys in developed regions.
- Limited survey time at disposal

iv] By mailing of questionaires.
- Most extensively used in economic and business surveys.
- Pilot study is needed to be carried out beforehand.

v] Through schedules
- Enumerators are appointed & given training
- depends on capability of enumerators.
- Occasional field checks necessary to ensure sincere work

vii] Execution of the Project

If the execution of the project proceeds on correct lines, the data to be completed or collected would be adequate & dependable

viii] Analysis of data

The analysis of data requires a number of closely related operations such as the establishment of categories, the application of these categories to raw data through coding, tabulation and then drawing statistical inferences.

Coding

- The categories of data is transformed into symbols that may be tabulated and counted.

Editing

- It is the procedure that improves the quality of the data for coding. With coding the stage is ready for tabulation.

Tabulation

- It is a part of the technical procedure wherein the classified data is put in the form of tables.

Analysis work after tabulation is generally based on the computation of various percentages, coefficients etc by applying various well defined statistical formulae.

In brief, the researcher can analyze the collected data with the help of various statistical measures.

ix] Hypothesis-Testing

The hypothesis, is tested through tests like chi square test, t-test, f-test, depending upon the nature and object of research inquiry. Hypothesis testing will result in either accepting the hypothesis or rejecting it.

If the researcher had no hypothesis to begin with, generalizations established on the basis of data may be stated as hypothesis to be tested by subsequent researchers in times to come.

x] Generalization & Interpretation

If hypothesis is tested & upheld several times, it may be possible for the researcher to arrive at generalisation, ie to build a theory.

The process of interpretation may quite often trigger off new questions which in turn may lead to further researches.

xi] Preparation of the Report or the thesis

A] The layout of the report should be as follows;
- the preliminary pages
- the main text, and
- the end matter.

In its preliminary pages, the report should carry title and date followed by acknowledgement and foreword.

Followed by Table of contents, List of tables, list of figures, instructions, if any, given in report

The Main text of the Report should have foll.

a] Introduction

It should contain a clear statement of the objective of the research and an explanation of the methodology adopted in accomplishing the research. The scope of the study along with various limitations should be as well stated in this part

b] Summary of findings

After introduction there would appear a statement of findings and recommendations in non-technical language. If the findings are extensive, they should be summarised.

c] Main Report

The main body of the report should be presented in logical sequence and broken down into readily identifiable sections.

d] Conclusion

Towards the end of the main text, researcher should again put down the results of his research clearly & precisely.

At the end of the report appendicies should be enlisted in respect of all technical data. Bibliography, i.e. list of books, journals, reports etc

B]. In report avoid terms like, 'it seems', 'there may be'

C] Charts & illustrations to be used only if they present information more clearly & forcibly.

D] Calculated 'confidence limits' must be mentioned and the various constraints experienced in conducting research operations, may as well be stated.

Criteria of Good Research

a] Purpose should be clearly defined and common concepts to be used.

b] Detailed sufficient description of research procedure for further advancement

c] Procedural design of the research to yield results that are as objective as possible

d] Should report any flaws with full frankness.

e] Methods of analysis should be appropriate Validity & reliability should be checked thoroughly.

f] Conclusions should be confined to justifiable data.

g] Greater confidence in research is warranted if the researcher is experienced has a good reputation in research and is a person of integrity.

Qualities of Good Research —

a] Good research is systematic

b] Good research is logical.

c] Good research is empirical.

d] Good research is replicable

Samarth Borkar
Goa college of Engg

Elements of Research Proposal

The title page provides the first impression, for your audience of your proposal. Your title should be complete and it should provide the focus of your investigation.

1. Your title should serve as a mini-abstract of your investigation and should put the most important words first.
2. Word choice and syntax should be precise

The introduction is very important. If your introduction gets your audience's attention, they will stay with you throughout your proposal. It establishes the nature, context & scope of your project.

It includes

1. Clear statement of the problem
2. Purpose of the study
3. Definitions
4. Significance of the Investigation
5. Literature Review
6. Questions or hypothesis.
7. Cite Relevant references.

The introduction should be at a level that makes it easy to understand for readers with a general science background.

In detail;

1. Clear statement of the problem

It is just one sentence. In it, you are looking for something wrong, that needs close attention, or something where existing methods no longer seem to be working

2. Purpose of the study

The goal of this study is to

... overcome the difficulty with...
.... discover what...
... Understand the causes or effect of...
... refine our ~~our~~ current understanding of...
... Provide a new interpretation of...
... understand what makes _ succesful or unsuccessful

3. Definitions

This section gives the definition of important terms and concepts that are usually stated in the objectives, hypothesis, and research questions.

4. Significance of the Investigation

It may include the meaning of the research work to you personally and should include how your research benefits or impacts others in part or whole.

Discuss what people or groups of people might benefit from reading your research. Show how this project is significant to developing a body of knowledge.

5. Literature Review

The literature review develops broad ideas of what is already known in field, and what questions are still unanswered. This process will assit you in further narrowing the problem for investigation, and will highlight any theories that may exist to support developing hypothesis.

6. Questions or hypothesis

These are testable explanations that are proposed before the methodology of a project is conducted, but after the researcher has had an opportunity to develop background knowledge (lit. review).

Hypothesis

It represents a declarative statement, a sentence instead of question, of the cause effect relationship between two or more variables. Make a clear and careful distinction between the dependent and the independent variables and be certain that they are clear to the reader.

Hypothesis can be created as four kinds of statements.

1. Literary Null
   - a "no difference" forms in terms of theoretical constructs

   eg "There is no relationship between support services and academic persistence of non-traditional aged college women"

2. Operational Null
   - a "no difference" forms in terms of the operation required to test the hypothesis.

   eg. "There is no relationship between the number of hours non-traditional aged college women use the students union and their persistence at the college after their freshman year."

3. Literary Alternative

- A form that states the hypothesis you will accept if the null hypothesis is rejected, stated in terms of theoretical constructs

eg. "~~There is no relationship between~~
"The more that non traditional aged women use support services, the more they will persist academically."

4. Operational Alternative

- Similar to the literary alternative except that the operations are specified.

eg. "The more that non-traditional aged college women use the student union, the more they will persist at the college after their freshman year".

Finally, it should be realistic and feasible, and be formulated with time and resource constraint in mind.

2nd example:

1. There is no difference in school achievement for high & low self regulated students.
2. There is no difference between the mean grade point averages achieved by students in the upper & lower quartiles of the distribution of the self regulatory Inventory
3. High self regulated students will achieve more in their classes than low self regulated students.
4. Students in the upper quartile of the self regulated inventory distribution achieve significantly higher grade points averages than do students in lower quartile.

Ethnographic Research

It develops in depth analytical descriptions of current systems, processes & phenomenon, and/or understandings of the shared beliefs and practices of a particular group or culture. This type of design collects extensive narrative data (non-numerical data) based on many variables over an extended period of time in a natural setting within a specific context.

The background, the development, current conditions and environmental interaction of one or more individuals, groups, communities, businesses or institutions is observed, recorded and analyzed for patterns in relation to internal & external influences. It is a complete description of present phenomenon

One specific form of ethnographic research is called a **Case Study**
It is a detailed examination of a single group, individual, situation or site

& Meta-Analysis is another specific form. It is a statistical method which accumulates experimental & correlational results across independent studies.

It is analysis of analyses.

Examples of Ethnographic Research

1. A case study of parental involvement at a specific magnet school.
2. A multi case study of children of drug addicts who excels despite early childhoods in poor environments.
3. A psychological case study with extensive notes based on observations of and interviews with immigrant workers.
4. A study of primate behaviour in the wild measuring the amount of time an animal engaged in a specific behaviour.

How data can be collected?

Samarth Borker
GOA COLLEGE OF ENGG

- Research tools

1. Interviews.
2. Focus groups.
3. Questionaire / Survey / Checklists.
4. Observation
5. Case studies.
6. Documentation Review/Analysis.

1. Interview

Purpose: To fully understand someones impressions, attributes or experiences or learn more about their answers to questionaires (generally qualitative)

Preview DATA COLLECTION

Primary Data: It is the data which is collected a fresh & for the first time. and thus happen to be original in character.

Secondary Data: are those which have already been collected by someone else and which have already been passed through the statistical process.

Experiment v/s Surveys

We collect primary data during the course of doing experiment in an experimental research. An expt refers to an investigation in which a factor or variable under test is isolated and its effects measured.

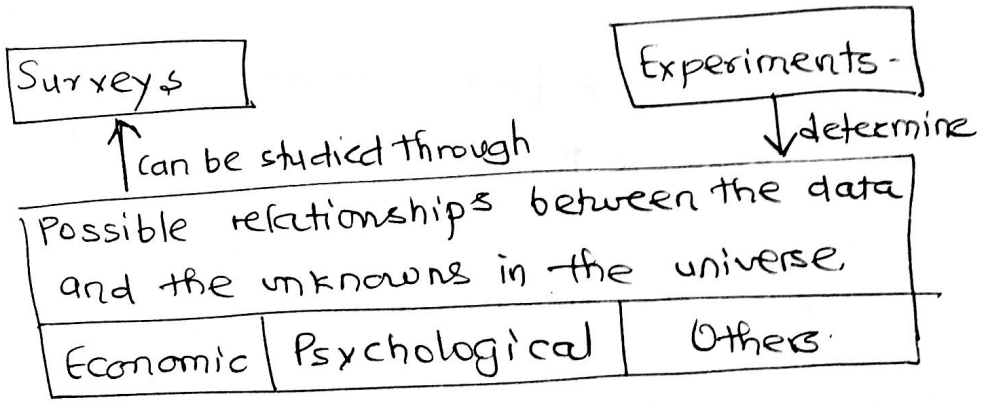
Surveys

In case, we do research of the descriptive type and perform surveys, whether sample survey or census survey, then we can obtain primary data either through observation or through direct communication with respondent in one form or another or through personal interviews.

Survey refers to the method of securing information concerning a phenomenon, under study from all or a selected number of respondents of the concerned universe.

In a survey, the investigator examines those phenomenon which exists in the universe independent of his action

The diff between experiment & survey can be depicted through, as under;



Difference betn survey & experiments

1. Surveys are conducted in case of descriptive research studies where as experiments are a part of experimental research studies.

2. Survey type research have large samples as responses generally happens to be 20% especially in mailed questionaire

   Experimental studies generally need small samples.

3. Surveys are concerned with
   - describing
   - Recording
   - Analysing &
   - Interpreting conditions.
   - No scope for manipulating variables.

Experimental research provides a systematic & logical method for answering the question. "what will happen if this is done when certain variables, are carefully controlled or manipulated?"

4. Surveys are appropriate for
   - social & behavioural science

Experiments
   - physical & natural science

5. Survey → field research
   Experiments → laboratory research

6. Surveys are concerned with hypothesis formulation and testing the analysis of the relationship between non-manipulated variables.

Experimentation provides a method of hypothesis testing.
After experimenters define a problem, they propose a hypothesis.

7. Surveys
   - Sample surveys
   - Social surveys
   - Economic surveys
   - public opinion surveys.

   Method of data collection
   - Observation
   - Interview
   - Questionaire / opioniomaive.
   - Case study. (projective technique)

   But incase of experiments, data is collected from several readings of experiments.

8. Survey research
   - rigid design
   - No bias
   - Maximize reliability
   - Accurate information
   - No scope for assumption

   whereas research design in case of experimental studies, apart reducing bias and ensuring reliability, must permit drawing inferences about causality.

9. Survey, relationship between data and unknowns can be studied.

   where as experments needs to establish such relationship.

10. Surveys
    - understanding & controlling variables

    Experiments
    - Causal analysis is considered relatively more important in experiments.

Page intentionally kept blank

Chi-Square Test
(parametric test)

Samarth Borkar
Goa College of Engg

In testing of hypothesis, all the tests of significance are based on certain parameters and their estimation. These tests were applied under certain assumptions about the population, like normal distribution. These tests can be categorised as parametric Tests.

Chi-Square distribution is used to obtain confidence interval estimate of unknown population variance.

Chi-Square distribution

- Based on mean, variance or proportion
- Non-parametric tests
  - i) test of diff of more than 2 parameters
  - ii) test of independent of attributes.
  - iii) test of goodness of fit.

Caution in using Chi-Square Test

This test is to be applied only when the individual observations of sample are independent which means that the occurance of one individual observation (event) has no effect upon the occurance of any other observation (event) in the sample under consideration.

Possible reasons for improper application:

i) Neglect of frequencies of non-occurance
ii) failure to equalize the sum of observed and the sum of expected frequencies.
iii) wrong determination of degrees of freedom
iv) wrong computation, and the like.

Ex 1.

A die is thrown 132 times with the following results.

Number turned up	1	2	3	4	5	6
frequency	16	20	25	14	29	28

Is the die unbiased?

Soln: let us take the hypothesis that die is unbiased. If that is so, the probability of obtaining any one of the six numbers is $1/6$ and as such the expected frequency of any one number coming upward is $132 \times \frac{1}{6} = 22$

Now we can write the observed frequencies along with expected frequencies. and work out the value of χ^2 as follows (shown in Table →)

Note: The test statistic is

$$\chi_c^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

where O_i = Observed freq of i^{th} class

E_i = Expected freq of i^{th} class

= Total freq × Prob. of that class.

Probability of each class is obtained using the specified theory or distribution.

i) Total sample size should be large (more than 50)

ii) Each cell in the contigency table has expected frequency of at least five

iii) $\sum O_i = \sum E_i$

35

No turned up	Observed frequency	Expected frequency	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	16	22	-6	36	36/22
2	20	22	-2	4	4/22
3	25	22	3	9	9/22
4	14	22	-8	64	64/22
5	29	22	7	49	49/22
6	28	22	6	36	36/22

$\therefore \Sigma [(O_i - E_i)^2 / E_i] = 9$

hence the calculated value of $\chi^2 = 9$

∵ the degrees of freedom in the given problem is $(n-1) = 6-1 = 5$

The table value of χ^2 for 5 degrees of freedom at 5 per cent level of significance is 11.071

Refer Pg 428, Table 3 ~~three~~ D.O.F ·05

5 → 11.070

Comparing calculated and table values of χ^2, we find that the calculated value is less than the table value and as such could be arisen due to fluctuation of sampling.

The results thus support the hypothesis and it can be concluded that die is unbiased.

Ex 2 Find the value of χ^2 for the following info^n

Class	A	B	C	D	E
Observed freq	8	29	44	15	4
Expected (T) freq	7	24	38	24	7

Soln Since some of the frequencies less than 10, we have to re-group the given data as follows and then will work out the value of χ^2

Class	Observed freq O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2/E_i$
A & B	$(8+29)=37$	31	6	36/31
C	44	38	6	36/38
D & E	$(15+4)$ 19	31	-12	144/31

$$\therefore \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 6.76 \text{ app.}$$

Ex 3. The table given below shows the data obtained during outbreak of small pox.

	Attacked	Not attacked	Total.
Vaccinated	31	469	500
Not vaccinated	185	1315	1500
Total	216	1784	2000

Test the effectiveness of vaccination in preventing the attack from smallpox. Test your results with the help of χ^2 at 5% level of significance.

Soln let us take the hypothesis that vaccination is not effective in preventing the attack from smallpox. i.e. vaccination and attack are independent.

On the basis of this hypothesis, the expected frequency corresponding to the number of persons vaccinated and attacked would be

$$\text{Expectation of } (AB) = \frac{(A) \times (B)}{N}$$

when A represents vaccination & B, attack.

$(A) = 500$

$(B) = 216$

$N = 2000$

$$\text{Expectation } (AB) = \frac{500 \times 216}{2000} = 54.$$

Now using the expectation of (AB), we can write the table of expected values as follows.

	Attacked : B	Not attacked : b	Total.
Vaccinated (A)	(AB) = 54	(Ab) = 446	500
not vaccinated (a)	(aB) = 162	(ab) = 1338	1500
total	216	1784	2000

Calculation of Chi-Square

Group	O_{ij}	E_{ij}	$O_{ij}-E_{ij}$	$(O_{ij}-E_{ij})^2$	$(O_{ij}-E_{ij})^2/E_{ij}$
AB	31	54	−23	529	529/54 = 9.796
Ab	469	446	+23	529	529/446 = 1.186
aB	158	162	+23	529	529/162 = 3.265
ab	1315	1338	−23	529	529/1338 = 0.395

$$\chi = \sum \frac{(O_{ij}-E_{ij})^2}{E_{ij}} = 14.642$$

∴ Degrees of freedom in this case

$= (r-1)(c-1) = (2-1)(2-1) = 1$

The table value of χ^2 for 1 degree of freedom at 5 per cent level of significance is 3.841.

The calculated value of χ^2 is much higher than this table & hence the result of the experiment does not support the hypothesis

We can thus, conclude that vaccination is effective in preventing the attack from small pox.

Samarth Borkar
GOA COLLEGE OF ENGG

Ex 4. Two research workers classified some people in income groups on the basis of sampling studies. Their results are as follows:

Investigators	Income group			Total
	Poor	Middle	Rich	
A	160	30	10	200
B	140	120	40	300
Total	300	150	50	500

Show that the sampling technique of at least one research worker is objective.

Sol: Let us take the hypothesis that the sampling techniques adopted by research workers are similar (i.e. there is no difference between the techniques adopted by research workers). This being so, the expectation of A investigator classifying the people in.

i] Poor income group $= \frac{200 \times 300}{500} = 120$

ii] Middle " " $= \frac{200 \times 150}{500} = 60$

iii] Rich " " $= \frac{200 \times 50}{500} = 20$

Similarly, the expectation of B investigator's classifying the people in

i] Poor income group $= \frac{300 \times 300}{500} = 180$

ii] Middle " " $= \frac{300 \times 150}{500} = 90$

iii] Rich " " $= \frac{300 \times 50}{500} = 30$.

We can now calculate value of χ^2 as follows.

Ex-4. Two research workers classified some people in income groups on the basis of sampling studies. Their results are as follows.

Investigators	Income group			Total.
	Poor	Middle	Rich	
A	160	30	10	200
B	140	120	40	300
Total	300	150	50	500

Show that the sampling technique of at-least one research worker is defective.

Soln let us take the hypothesis that the sampling techniques adopted by research workers are similar (ie There is no difference between the techniques adopted by research workers).

This being so, the expectation of A investigator classifying the people in.

i) Poor Income group $= \frac{200 \times 300}{500} = 120$

ii) Middle " " $= \frac{200 \times 150}{500} = 60$

iii) Rich " " $= \frac{200 \times 50}{500} = 20$

Similarly, the expectation of B investigator classifying the people in

i) Poor Income group $= \frac{300 \times 300}{500} = 180$

ii) Middle " " $= \frac{300 \times 150}{500} = 90$

iii) Rich " " $= \frac{300 \times 50}{500} = 30.$

We can now calculate value of χ^2 as follows:

Groups.	O_{ij}	E_{ij}	$O_{ij}-E_{ij}$	$(O_{ij}-E_{ij})^2/E_{ij}$
Invesitagutor A				
i) Poor	100	120	40	13.33
ii) Middle	30	60	-30	15.00
iii) Rich	10	20	-10	5.00
Investigator B				
i) Poor	140	180	-40	8.88
ii) Middle	120	90	30	10.00
iii) Rich	40	30	10	3.33

Hence, $\chi^2 = \Sigma \frac{(O_{ij}-E_{ij})^2}{E_{ij}} = 55.54$

∴ Degrees of freedom

$= (c-1) \times (r-1)$

$= (3-1)(2-1) = 2$

from table for 5% significance = 5.991.

Calculated value is much larger than the table value, which means that the calculated value cannot be said to have arisen just because of chance.

It is significant.

Hence the hypothesis does not hold good.

This means that the sampling techniques adopted by two research investigators differ and are not similar.

Naturally, then the technique of one must be superior than that of other.

Ex 5. Eight coins were tossed 256 times and the following results were obtained.

Number of heads	0	1	2	3	4	5	6	7	8
frequency	02	06	30	52	67	56	32	10	01

Are the coins biased? Use χ^2 test.

Soln: Let us take the hypothesis that the coins are not biased. If that is so, the probability of any one coin falling with head upward is 1/2 and with tail upward is 1/2 and it remains the same, whatever the number of throws may be.

In such case, the expected value of getting 0, 1, 2 ... heads in a single throw of eight coins will be worked out as follows.

Events of Number of heads.	Expected frequencies.
0	$8C_0 \left[\frac{1}{2}\right]^0 \left[\frac{1}{2}\right]^8 \times 256 = 1$
1	$8C_1 \left[\frac{1}{2}\right]^1 \left[\frac{1}{2}\right]^7 \times 256 = 8$
2	$8C_2 \left[\frac{1}{2}\right]^2 \left[\frac{1}{2}\right]^6 \times 256 = 28$
3	$8C_3 \left[\frac{1}{2}\right]^3 \left[\frac{1}{2}\right]^5 \times 256 = 56$
4	$8C_4 \left[\frac{1}{2}\right]^4 \left[\frac{1}{2}\right]^4 \times 256 = 70$
5	$8C_5 \left[\frac{1}{2}\right]^5 \left[\frac{1}{2}\right]^3 \times 256 = 56$
6	$8C_6 \left[\frac{1}{2}\right]^6 \left[\frac{1}{2}\right]^2 \times 256 = 28$
7	$8C_7 \left[\frac{1}{2}\right]^7 \left[\frac{1}{2}\right]^1 \times 256 = 8$
8	$8C_8 \left[\frac{1}{2}\right]^8 \left[\frac{1}{2}\right]^0 \times 256 = 1$

Note:

Using binomial principle $(p+q)^n$ where $p=q=\frac{1}{2}$, $n=8$

$\left[{}^nC_r \cdot p^r \cdot q^{n-r} \times 256\right]$

The value of χ^2 can be worked out as follows.

No. of heads	Observed freq O_i	Expected freq E_i	$O_i - E_i$	$(O_i - E_i)^2 / E_i$
0	2	1	1	1
1	6	8	-2	0.5
2	30	28	2	0.14
3	52	56	-4	0.29
4	67	70	-3	0.13
5	56	56	0	0
6	32	28	4	0.57
7	10	8	2	0.50
8	1	1	0	0

$$\therefore \chi^2 = \sum \left(\frac{(O_i - E_i)^2}{E_i} \right) = 3.13 .$$

$\because$ Degrees of freedom $= (n-1) = (9-1) = 8$ (15.507)

At 5% level of significance from table $= 15.507$

The calculated value is much less than the this table value & hence it is insignificant and can be ascribed due to fluctuations of sampling.

The result, thus supports the hypothesis and we may say that the coins are not biased.

Ex 6

Following are the number of aircraft accidents during the various days of the week. Test at 5% of significance, if the accidents are uniformly distributed over the week.

Days	Mon	Tues	wed	Thurs	fri	Sat	Sun
No of accid.	14	16	08	12	11	09	14

Soln we wish to test:

H_0: Fit is good (Data follows discrete uniform distribution)

H_1: Fit is not good

In case of (discrete) uniform distribution, the probability of each class is the same, which is $\frac{1}{7}$

The test statistics is

$$\chi_c^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

								Total
O_i	14	16	8	12	11	9	14	84
E_i	12	12	12	12	12	12	12	84
$\frac{(O_i - E_i)^2}{E_i}$	0.33	1.33	1.33	0	0.083	0.75	0.3	41

D.of f for $7-1=6$ at 5% is 12.592

Table value is larger than the computed value of 4.1667. Therefore at 5% level of significance we cannot reject the null hypothesis.

Thus there is no statistical evidence that the accidents are not uniformly distributed over the week.

Ex.7

A survey of 320 families with 5 children each revealed the following distribution.

No of boys	5	4	3	2	1	0
No of girls	0	1	2	3	4	5
No of familis	14	56	110	88	40	12

Is this result consistent with the hypothesis that the male & female births are equally probable?

Soln: We wish to test

H0: Male & female births are equally probable
H1: Male & female births are not equally probable

Success: A randomly chosen ~~girl~~ child is girl.
X = No of girl children in a fly with 5 children.

Under H0: $X \sim B(5, \frac{1}{2})$

$P(X=0) = {}^5C_0 \left[\frac{1}{2}\right]^0 \left(1-\frac{1}{2}\right)^{5-0} = {}^5C_0 \left(\frac{1}{2}\right)^5 = 0.03125$

$P(X=1) = {}^5C_1 \left[\frac{1}{2}\right]^5 = 0.15625$

$P(X=2) = {}^5C_2 \left[\frac{1}{2}\right]^5 = 0.31250$

$P(X=3) = {}^5C_3 \left[\frac{1}{2}\right]^5 = 0.31250$

$P(X=4) = {}^5C_4 \left[\frac{1}{2}\right]^5 = 0.15625$

$P(X=5) = {}^5C_5 \left[\frac{1}{2}\right]^5 = 0.03125$

The test statistic is

$$\chi^2_c = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

Therefore;

O_i	E_i	$\frac{(O_i-E_i)^2}{E_i}$
14	$320 \times P(x=0) = 10$	1.6
56	$320 \times P(x=1) = 50$	0.72
110	$320 \times P(x=2) = 100$	1
88	$320 \times P(x=3) = 100$	1.44
40	$320 \times P(x=4) = 50$	2
12	$320 \times P(x=5) = 10$	0.4
$N = 320$	Total 320	$\chi_c^2 = 7.16$

The test statics follows chi-square distribution with 6-1=5 df @5%, critical value is 11.070 which is larger than the computed value 7.16.

Therefore at 5% level of significance, we cannot reject the null hypothesis,

Thus there is no statistical evidence that male & female births are not equally probable.

Samarth Borkar
Goa college of Engg

Ex 8.

Find the value of χ^2 for the foll. info^n.

Class	A	B	C	D	E	F	G	H
obs. freq	8	29	44	15	4	44	29	08
Exp. freq	7	24	38	24	7	38	24	07

Soln: Since some of the frequencies are less than 10

Hint: Regroup & solve, as shown in example 2.

Ex 9: Genetic theory states that children having one parent of blood type A and other of blood type B will always be one of the three types, A, AB, B. and their proportion of three types will be on average be as 1 : 2 : 1.

A report states that out of 300 children having one A parent and B parent, 30 percent found to be type A, 45% type AB & remainder type B.

Test the hypothesis by χ^2 test.

Soln: The observed frequencies of types A, AB & B is given in the question are 90, 135 & 75. resp.

The expected frequencies of type, A, AB & B (as per the genetic theory) should have been 75, 150 and 75 resp.

We now calculate the value of χ^2 as follows.

Type	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2 / E_i$
A	90	75	15	225	3
AB	135	150	−15	225	1.5
B	75	75	0	0	0

$$\therefore \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = 3 + 1.5 + 0 = 4.5$$

$$\therefore df = (n-1) = 3-1 = 2$$

Table value of χ^2 of 2df @5% is 5.991.

The calculated value of χ^2 is 4.5 which is less than the table value and hence can be ascribed to have taken place because of chance.

This supports the theoretical hypothesis of the genetic theory that on average type A, AB & B, stand in the proportion of 1:2:1.

Ex 10. (Repeat / Similar structure problems)

An experiment was conducted to test the efficacy of chloromycetin in checking typhoid. In a certain hospital chloromycetin was given to 285 out of the 392 patients suffering from typhoid. The number of typhoid cases were as follows.
(Table value χ^2 @5% → 3.841)

	Typhoid	No typhoid	Total
Chloromycetin	35	250	285
No chloromycetin	50	57	107
Total.	85	307	392

Soln

Chloromycetin : Typhoid $= \frac{285 \times 85}{392} = 61.79$

No-Typhoid $= \frac{285 \times 307}{392} = 223$

No-chloromycetin Typhoid $= \frac{107 \times 85}{392} = 23$

No-typhoid $= \frac{107 \times 307}{392} = 83$

	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
Chloromycetin Th	35	61.8	−26.8	718.24	11.62
NoTh	250	223	26.8	"	3.22
No chloromycetin Th	50	23	26.8	"	30.96
No Th	57	83	−26.8	"	8.57

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$= 54.37.$$

Comment :

Ex 11. On the basis of information given below about the treatment of 200 patients suffering from a disease state whether the new treatment is comparatively superior to the conventional treatment.

Treatment	No. of Patients. Favourable Response	No Response
New	60	20
Conventional	70	50

Given: χ^2 for one degree of freedom @5% = 3.84.

Soln New: favor. Response $= \frac{130 \times 80}{200} = 52$

No response $= \frac{70 \times 80}{200} = 28$

Conventional: FR $= \frac{130 \times 120}{200} = 78$

NR $= \frac{70 \times 120}{200} = 42$

Treatment	Oi	Ei	Oi - Ei	$(Oi-Ei)^2$	$(Oi-Ei)^2/Ei$
New FR	60	52	8	64	1.23
NR	20	28	-8	64	2.29
Convent FR	70	78	-8	64	0.82
NR	50	42	8	64	1.52
					5.86.

$$\chi^2_c = \sum \frac{(Oi-Ei)^2}{Ei} = 5.85.$$

$$\chi^2_{Table} = 3.84.$$

Comment :

Samarth Borkar
Goa college of Engg

Ex:
Prob 12

200 digits were chosen at random from a set of tables. The freq of digits were

Digit	0	1	2	3	4	5	6	7	8	9
freq	18	19	23	21	16	25	22	20	21	15

Calculate χ^2

Soln:

digit	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
0	18	20	-2	4	0.2
1	19	20	-1	1	0.05
2	23	20	3	9	0.45
3	21	20	1	1	0.05
4	16	20	-4	16	0.8
5	25	20	5	25	1.25
6	22	20	2	4	0.2
7	20	20	0	0	0
8	21	20	1	1	0.05
9	15	20	-5	25	1.25
					4.3

$$\chi^2_{\text{cal}} = \sum \frac{(O_i - E_i)^2}{E_i} = 4.3$$

~~χ^2_{table}~~ =

Collection of Primary Data

i) Observation method.

ii) Interview method
   a) Personal Interviews.
   b) Telephone Interviews

iii) Collecting data through questionaires.
   1. General form
   2. Question sequence
   3. Question formulating & wording

iv) Collection of data through schedules.

We shall also study;
1. Difference between questionaire & schedule
2. Guidelines for constructing Quest/Schedule

Collection of secondary Data

- Reliability of data
- Suitability of data
- Adequacy of data

We shall also study:

1. Selection of appropriate method for data collection
   a) Nature scope & object of enquiry
   b) Availability of funds
   c) Time factor
   d) Precision required

2. Case Study Method & characteristics.

i) Observation method

Under the observation method, the info is sought by way of investigators own direct information/observation without asking from the respondent.

eg study of consumer behaviour

Advantage

a) Subjective bias is eliminated, if observation is done accurately.
b) Information obtained under this method relates only to what is currently happening
c) Method is independent of respondents willingness to respond & cooperation

limitations

a) Expensive method
b) very limited information
c) some people are rarely accesible for direct observation.

Objective

a) what should be observed?
b) How observation should be recorded?
c) how accuracy ensured?

Types

a) structural observation
b) Unstructural observation
c) Participant observation
d) non-participant observation.

Merits

a) Record natural behaviour
b) disinterested fashion
c) truth of statement verification.

ii) Interview Method

The interview method of collecting data involves presentation of oral-verbal stimuli and reply in terms of oral-verbal responses. This method can be used through personal interviews and if possible through telephone interviews.

a) Personal Interviews

objective

- Suitable for intensive investigations.
- Structured interviews
  - predetermined questions and of highly standardized techniques of recording
- Focussed Interviews
  - freedom to explore reasons
  - power to respondent
- Clinical Interview
  - broad underlying feelings
  - Interviewers discretion
- Non directive Interview
  - The interviewer's function is simply to encourage the respondents to talk about the given topic

Merits of Interview

1. More info in greater depth
2. Interviews still play vital role
3. greater flexibility as restructuring can be done
4. Can be applied to recording verbal answers.
5. Personal information can also be obtained.
6. non-response generally becomes Low.
7. Total control in hands with interviewer
8. Spontaneous reaction on account of off guard.
9. language of the interview can be adopted.
10. Supplementary information regarding respondents personal characteristics and environment can be obtained.

Weaknesses

1. Expensive on account of large geographical spread.
2. Headache of supervision & bias.
3. High income respondents not easily available
4. Time consuming
5. Respondent may provide imaginery information, just to make the interview interesting.
6. Organization required for selecting, training and supervising the field staff is more complex.
7. May introduce systematic errors.
8. Proper rapport with respondent needed.

b) Telephone Interview

1. more flexible compared to mailing method.
2. faster
3. cheaper than personal interview method
4. Recall is easy, callback
5. Higher rate of response
6. Replies can be recorded.
7. Better access to respondent
8. No field staff required

Demerits

1. little time is given to respondent
2. Telephone facility may be restricted
3. Cost factor ↑, with extensive geo. coverage
4. Not suitable for intensive surveys.
5. Possibility of interviewer bias is more
6. Questions has to be short & to the point

Points to be considered

1. Advance planning
2. Informal & friendly approach
3. Proper rapport
4. Ability to listen with understanding, respect
5. free flowing interview
6. Resonable uniformity in respect of all Salient points in the study.

iii) Collection of Data through Questionaire

Merits

1) low cost
2) free from interviewers bias
3) Respondent get adequate time
4) Respondent not easily approachable, can also be reached conveniently.
5) Results more reliable

Demerits

1) low rate of return.
2) respondent needs to be educated and cooperating
3) Control over questionair is lost
4) Inbuilt inflexibility
5) Interpretation of ommision is difficult
6) difficult to know if respondent is fake or not.
7) This method likely to be slowest of all.

Main aspect of Questionaire

Researcher should note the following with regard to three main aspects of questionaire.

1. General form
2. Question sequence
3. Question formulation & wording

1. General form

- type 'Yes' or 'No' (closed)
- free response (open)
- alternative questions.
- highly structured questionaire is one, where respondent's own words are held to the minimum.

2. Question Sequence

- Meaning relation of one question to another should be readily apparent to the respondent, with the questions that are easiest to answer being put in the beginning

Following type of questions should be avoided
a) which puts strain on memory
b) Questions of personal character
c) Questions related to personal wealth.

3. Question formulating & wording

- Question must be very clear
- Should be easily understood
- should be simple
- Should be concrete
- should be realistic

Essentials of good questionaire

- size of the questionaire should be kept minimum.
- Questions should proceed in logical sequence
- Personal & intimate questions at end.
- terms with vague interpretation to be avoided.
- Question may be dichotomous (Y or N)
- Questions affecting sentiments to be avoided.
- Adequate space should be provided in questionaire to help editing & tabulation.
- Provision for uncertainity indication (do not know)
- Brief description about filling should be provided
- Physical appearance of questionaire also counts.
- Can use internet.

iv] Collection of Data through Schedules

(1) This method of data collection is very much like the collection of data through questionaire, with a little difference, which lies in the fact that schedules (performa containing a set of questions) are being filled in by the enumerators who are specially appointed for the purpose.

(2) These enumerators along with schedules go to respondents, put them the questions and record replies in the space meant for same

(3) Enumerators may explain aims & objectives of investigation and also remove the difficulties which may respondent feel.

(4) These enumerators or assisting respondents should be carefully selected.

(5) Should be intelligent & possess capability of cross-examination

Merit

1. Useful in extensive enquiries
2. fairly reliable results.

Demerit

1. Expensive

Example

Population census all over the world is conducted through this method.

Samarth Borkar
Goa College of Engg

Difference between Questionnaire & Schedule

	Questionnaire	Schedule
1.	Sent with specific covering letter without any further assistance	1. filled out by Research worker or enumerator.
2.	cheap & economical way of data collection No field staff required	2. Money is spent on preparing schedules.
3.	Non-response is usually high	3. Very low
4.	Not clear as to who actually replies.	4. Identity of respondent is known.
5.	Very slow	5. collected well in time
6.	Personal contact generally not possible	6. Direct personal contact is established.
7.	Used only when respondent is literate & cooperative	7. can be used even with illiterates.
8	Wider and more representative distn is possible	8. Difficulty in sending enumerators.
9.	Risk of collecting incomplete & wrong information	9. More accurate
10.	Success lies on questionaire itself	10. Depends on honesty and competancy of enumerator
11.	Physical appearance of questionaire must be good	11. Not so
12.	Observation method cannot be employed.	12. Observation method can also be employed.

Guidelines for constructing Questionnaire/Sche

1. Researcher must be clear about research proble
2. Open or close, & logical aspect should be dec
3. Previous drafts can be referred to
4. Re-examine & Revise
5. Pilot study is encouraged
6. Simple straight forward directions to be include

Some other methods of data collection

1. Warranty Cards.
2. Distributor or store audits
3. Pantry audit
4. Consumer panels.
5. Use of mechanical devices.

6 Projective Techniques

i. Word association tests.
ii. Sentence completion test
iii. Story completion test
iv. Verbal project tests
v. Pictorial Techniques.
vi. TAT - Thematic Appreception Test.
vii Rosenzweig Test
viii Rorschach test
ix Holtzman inkbolt test
x Tomkins Horn picture arrangement test
xi Play Techniques.
xii Quizes, tests & examination
xiii Sociometry.

7. Depth Interviews
8. Content Analysis.

Case Study Method

The case study method is very popular form of qualitative analysis & involves a careful complete observation of a social unit, be that unit a person, a family, an institution, a cultural group, or even the entire community.

- It is a method of depth rather than breadth.

Characteristics

1. Researcher can take one or more units.
2. Selected unit is studied minutely
3. All the complex factors are studied
4. Every possible effort is made to collect info ex criminal study – & how to reform also.
5. Study of mutual inter relationship of causal factors
6. Behaviour pattern is studied
7. It results in fruitful hypotheses.

Evolution & Scope

Credit goes to 1) Frederic Le Play.
2) Dr William Healy

Scope: Anthropologists, historians novelists & dramatists.

Assumptions

1) Uniform basic human nature
2) Natural history of unit needs to be studied.
3) Comprehensive study.

Major phases involved

1) Recognition & determination
2) Collection of data, Examination & history
3) Diagonosis & identification of causal factors
4) Application of remedial measures.
5) follow up programme to determine the effectiveness of the treatment applied.

Advantages

a) Exhaustive study
b) real & enlightened record of Personal history
c) Natural history of social unit
d) Enable generalized knowledge to get richer
e) Intensive study of social units.
f) helps in constructing questionnaire
g) Researcher can use other methods like depth interview, questionnaire, documents, study report of individuals, letters under case study method.
h) mode of organizing data
i) Can understand past social nature of units.
j) Increases researchers analysing ability and skill.
k) are indespensable for therapeautic and administrative purpose

Limitations (according to Read Bain)

1) Case studies are seldom comparable.
2) Real info is often not collected
3) Danger of false generalization
4) Consumes more time & expenditure
5) Case data often vitiated
6) Based on several assumptions
7) Limited sphere
8) Response of investigator is important.

Depth Interviews

① These interviews are designed to discover underlying motives and desires.

② Often used in motivational research

③ These are held to explore needs, desires, & feelings of respondents.

④ They aim to elicit unconscious especially to personality dynamics.

⑤ These interviews require great skill on the part of interviewer and at the same time involve considerable time

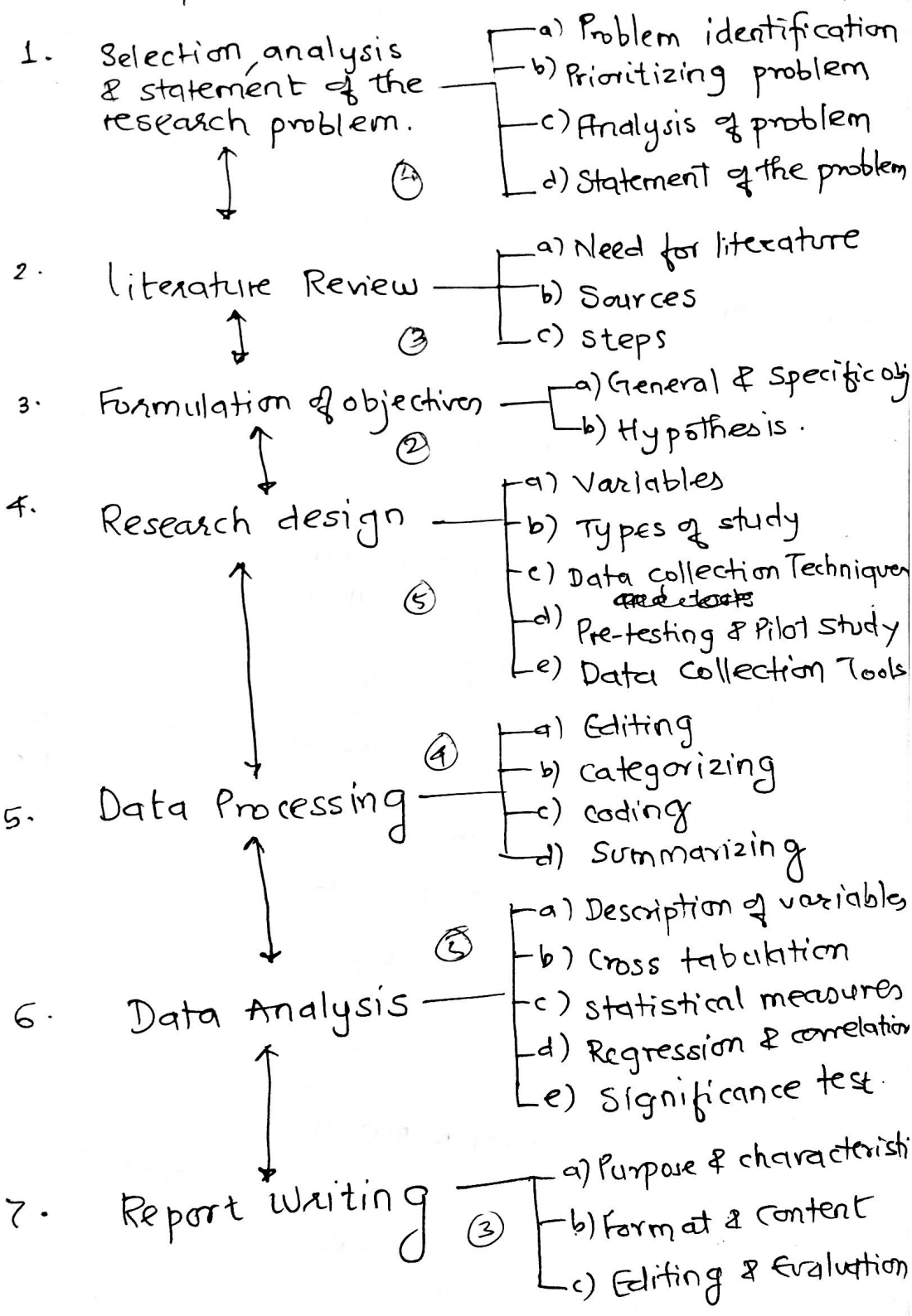
⑥ Unless the researcher is specialized in training, depth interviewing should not be attempted.

⑦ It may be projective or non-projective in nature.

The diff lies in the nature of question asked. Indirect questions on seemingly irrelevant subject provide info that can be related to the informants behaviour.

Samarth Borkar
Goa college of Engg

Revision



Steps & Elements in Research process

Hypothesis Testing for mean

In testing of mean, we test on the basis of random if the mean of a population mean is the same as hypothesized value.

Case 1:

Standard Normal Distribution Test.

↳ Condition: i) Population S.D. σ is known.

Null Hypothesis $H_0: \mu = \mu_0$

Alternate hypothesis $H_1: \mu \neq \mu_0$

where μ is the population mean

μ_0 is the population mean under null hypothesis.

Case 1 **If σ is known**

In this case the population should be normally distributed or sample should be large enough (in practice $n \geq 30$).

The test statistics is

$$Z_c = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

where

$\bar{x}$ = sample mean

σ = known population standard deviation

n = sample size

Conclusion : If $Z_{calculated} > Z_{tabulated}$ (@ reqd level of significance)

then null hypothesis is rejected (H_0) and alternate hypothesis is accepted (H_1)

* If H_0 is rejected then there is significant difference between sample mean & population mean.
* and sample may not belong to population with respect to character under consideration

Example

The average daily production of a cement factory over several years were found out to be 46 tonnes. with a S.D of 6 tonnes.
On inspection for 10 days, the daily production was recorded as 38, 56, 44, 40, 36, 43, 51, 49, 37, 26
Test whether there is any significant difference between sample production and average production recorded over several years at 5% level of significance

Soln $\rightarrow$ $H_0 = \mu = \mu_0 = 46$

$$\bar{x} = \frac{38+56+44+40+36+43+51+49+37+26}{10}$$

$$= \frac{420}{10} = 42$$

$\bar{x} = 42$

$\sigma = 6$

$$Z = \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} = \frac{|42-46|}{6/\sqrt{10}} = 2.108 = 2.11$$

Here $Z_{calculated} > Z$ tabulated @ 5% LOS

Concln $\therefore$ H_0 is rejected & H_1 is accepted. necessary steps hv to be taken for restoring prodn.

Case 2] Population standard deviation σ is not known

Condition i] σ is not known

ii] Size of sample is large ($n > 30$)

(readers pls. verify) but "Kothari" says it should be small.

$H : \mu = \mu_0$

$H_1 : \mu \neq \mu_0$

μ is the population mean

μ_0 is the population mean under null hypothesis

The test statistic is

$$T_c = \frac{|\bar{x} - \mu|}{S_1/\sqrt{n}}$$

where $S_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ for type I data

& $S_1^2 = \frac{1}{N} \Sigma f_i (x_i - \bar{x})^2$ for type II & III

where S_1^2 = sample standard deviation.

$\bar{x}$ = sample mean

n = size of sample

Conclusion: $Z_{cal} > Z_{tab}$, then H_0 is rejected & H_1 is accepted

Prob The distribution of urea production on different days in a fertilizer factory is given as follows. The daily production of urea over several years was found to be 25 tonnes. Test whether the sample daily production is significantly different from average production at 1% level of significance

Production	10-15	15-20	20-25	25-30
Days	8	10	11	2

Soln

Production	Days (f_i)	x_i	$f_i x_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
10-15	8	12.5	100	-6.13	37.57	300.61
15-20	10	17.5	175	-1.13	1.28	12.77
20-25	11	22.5	247.5	3.87	14.98	164.74
25-30	2	27.5	55	8.87	78.68	157.35
Total			577.5			635.47

$$\bar{x} = \frac{f_i x_i}{f_i} = \frac{577.5}{31} = 18.63$$

$$s_1^2 = \sum \frac{f_i (x_i - \bar{x})^2}{N} = \frac{635.47}{31} = 20.50$$

$$s_1 = 4.527.$$

$$T_c = \frac{|\bar{x} - M|}{s_1/\sqrt{n}} = \frac{|18.63 - 25|}{4.527/\sqrt{31}} = 7.834.$$

Z calculated is 7.834

Z table @ 1% LOS is 2.58 ?

$Z_{cal} > Z_{tab}$

Thus H_0 is Rejected & H_1 is accepted.

Conclusion

Hence the average sample daily production is significantly different from average production of a factory.

t- Test (Students t-distribution)

* If the sample size, n becomes large, This distribution tends to standard normal distribution with mean equal to zero and variance equal to one

* If the sample size is small $n \leq 30$ and population S.D. σ is not known then the t distribution is used instead of std normal distribution and the test based on t-distribution are called as t-tests

Conditions i) Population S.D. σ is not known
ii) Size of the sample is small ($n \leq 30$)

$H_0 : \mu = \mu_0$
$H_1 : \mu \neq \mu_0$

formula: $$t = \frac{|\bar{x} - \mu|}{\sqrt{\frac{s^2}{n}}}$$

where $$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

The confidence limits for population mean is given as.

$$\bar{X} \pm t_{(n-1)} \times \sqrt{\frac{s^2}{n}}$$

The population mean, μ lies betn lower & upper limit, where $t_{(n-1)}$ is the tabulated value of t-distribution with $(n-1)$ dof.

Conclusion: if $t_{cal} > t_{tab}$ with $(n-1)$ dof then H_0 is rejected & H_1 is accepted.

Note: Population mean, & the sample does not belong to the population with respect to character under consideration.

Prob The average sales per day of cakes prepared by bakery was 5.6 based on several years data. The sample sales for 10 days were recorded as 4.8, 5.2, 4.9, 5.7, 4.2, 5.1, 5.0, 4.8, 5.4, 5.3 Test the significant diff between sample sales & recorded sales at 5% LOS and also find the confidence limits for the population mean.

Soln

Sr No:	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	4.8	−0.24	0.0576
2	5.2	0.16	0.0256
3	4.9	−0.14	0.0196
4	5.7	0.66	0.4356
5	4.2	−0.84	0.7056
6	5.1	0.06	0.0036
7	5.0	−0.04	0.0016
8	4.8	−0.24	0.0576
9	5.4	0.36	0.1296
10	5.3	0.26	0.0676
Total	50.4		1.504

$$\bar{x} = \frac{x_i}{n} = \frac{50.4}{10} = 5.04$$

$$s^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2$$

$$= \frac{1}{10-1} \times 1.504$$

$$s^2 = 0.1671$$

$$s = 0.4088$$

$$t = \frac{|\bar{x} - \mu|}{\sqrt{s^2/n}} = \frac{|5.04 - 5.6|}{\sqrt{0.1671/10}} = 4.33.$$

Samarth Bankar Goa College of Engg

The confidence limit for population mean is

$$\bar{x} - t_{n-1} \times \sqrt{\frac{s^2}{n}} = 5.04 - 2.262 \times \sqrt{\frac{0.1671}{10}} = 4.7875$$

$$\bar{x} + t_{n-1} \times \sqrt{\frac{s^2}{n}} = 5.04 + 2.262 \times \sqrt{\frac{0.1671}{10}} = 5.3341$$

Conclusion $t_{cal} > t_{tab}$ for 5% LOS

→ The avg sample sales are significantly lower than average sales over the years.

→ Also it can be seen that average sales over the years do not lie between confidence limit of 4.75 & 5.33

Quartile Deviation

It is based on the lower quartile Q_1 and the upper quartile Q_3. The difference $Q_3 - Q_1$ is called the inter quartile range. The difference $Q_3 - Q_1$ divided by 2 is called semi-inter-quartile range or the quartile deviation. Thus,

$$Q.D = \frac{Q_3 - Q_1}{2}$$

The Q.D. is slightly better measure of absolute dispersion than the range.

Coefficient of Quartile deviation

$$= \frac{(Q_3 - Q_1)/2}{(Q_3 + Q_1)/2} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Prob 1 The wheat production in kgs of 20 acres is given as 1120, 1240, 1320, 1040, 1080, 1200, 1440, 1360, 1680, 1730, 1785, 1342, 1960, 1880, 1755, 1720, 1600, 1470, 1750 & 1885.
Find QD & Coeff of QD

Soln – Step 1 Arrange all the values in ascending order.

Q_1 = Value of $\left(\frac{n+1}{4}\right)^{th}$ item
= value of $\left(\frac{20+1}{4}\right)$
= Value of $(5.25)^{th}$ item
= 5^{th} item + 0.25 (6^{th} item – 5^{th} item)
= 1240 + 0.25(1320 – 1240)
Q_1 = 1260

Q_3 = value of $3\left(\frac{n+1}{4}\right)$
= value of 15.75^{th} item
= 15^{th} item + 0.75 (16^{th} – 15^{th})
= 1750 + 0.75(1755 – 1750)
= 1753.75

$$QD = \frac{Q_3 - Q_1}{2} = \frac{1753.75 - 1260}{2} = 246.875$$

$$\text{Coeff of } QD = \frac{Q_3 - Q_1}{Q_3 + 1} = 0.164$$

Ex-2

Calculate the quartile deviation and coefficient of quartile deviation from the data given below

Maximum load (Short-tons)	Number of cables
9.3 – 9.7	2
9.8 – 10.2	5
10.3 – 10.7	12
10.8 – 11.2	17
11.3 – 11.7	14
11.8 – 12.2	6
12.3 – 12.7	3
12.8 – 13.2	1

Soln;

Max load	Cables	class boundaries	Cumulative frequency
9.3 – 9.7	2	9.25 – 9.75	2
9.8 – 10.2	5	9.75 – 10.25	2+5=7
10.3 – 10.7	12	10.25 – 10.75	19
10.8 – 11.2	17	10.75 – 11.25	36
11.3 – 11.7	14	11.25 – 11.75	50
11.8 – 12.2	6	11.75 – 12.25	56
12.3 – 12.7	3	12.25 – 12.75	59
12.8 – 13.2	1	12.75 – 13.25	60

Q_1 = value of $\left(\frac{n}{4}\right)^{th}$ item = value of $\frac{60}{4}$ = 15^{th} item

Q_1 lies between 10.25 – 10.75.

$\therefore Q_1 = L + \frac{h}{f}\left(\frac{n}{4} - c\right)$ where $L = 10.25$, $h = 0.5$, $f = 12$, $\frac{n}{4} = 15$, $c = 7$

$= 10.58$

Q_3 = value of $\frac{3n}{4} = \frac{3 \times 60}{4} = 45^{th}$ item

Q_3 lies between 11.25 – 11.75

$\therefore Q_3 = L + \frac{h}{f}\left(\frac{3n}{4} - c\right)$ where $L = 11.25$, $h = 0.5$, $f = 14$, $\frac{3n}{4} = 45$, $c = 36$

$\therefore Q_1 = 11.25 + \frac{0.5}{14}(45 - 36)$

$= 11.57$

$QD = \frac{Q_3 - Q_1}{2} = \frac{11.57 - 10.58}{2} = 0.495$

Coeff of $QD = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{11.57 - 10.58}{11.57 + 10.58} = 0.045$

Standard deviation

It is defined as the positive square root of the mean of the square deviations taken from arithmetic mean of data

- for the sample data the standard deviation is denoted as S and is defined as.

$$S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

- for the frequency distribution the formula becomes

$$S = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}}$$

Actual Mean Method

for ungrouped data

$$S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

for grouped data

$$S = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}}$$

Assumed Mean method

a) Short cut Method

for grouped data

$$S = \sqrt{\frac{\Sigma D^2}{n} - \left(\frac{\Sigma D}{n}\right)^2}$$

for ungrouped data

$$S = \sqrt{\frac{\Sigma f D^2}{\Sigma f} - \left(\frac{\Sigma f D}{\Sigma f}\right)^2}$$

where $D = X - A$, A is any assumed mean other than zero

If $A = 0$, then $D = X$

Ex Calculate the standard deviation from the distribution of marks.

Marks	No of students
1-3	40
3-5	30
5-7	20
7-9	10

Soln: Method I Actual mean method

Marks	f	x	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
1-3	40	2	80	4	160
3-5	30	4	120	0	0
5-7	20	6	120	4	80
7-9	10	8	80	16	160
	100				

$$\bar{x} = \frac{\Sigma f x}{\Sigma f} = \frac{400}{100} = 4$$

$$S = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}} = \sqrt{\frac{400}{100}} = 2 \text{ marks}.$$

Method-II

Taking assumed mean as A = 2

Marks	f	x	D = x-2	fD	fD^2
1-3	40	2	0	0	0
3-5	30	4	2	60	120
5-7	20	6	4	80	320
7-9	10	8	6	60	160
	100			200	800

$$S = \sqrt{\frac{\Sigma fD^2}{\Sigma f} - \left(\frac{\Sigma fD}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{800}{100} - \left(\frac{200}{100}\right)^2}$$

$$= \sqrt{8-4}$$

$$= 2$$

Samarth Bakan
Goa College of Engg

Approaches to Research
Deductive and Inductive

Deductive reasoning works from the more general to more specific. It is a 'top-down' approach. We might begin with thinking up a theory, about our topic of interest. We then narrow that down into more specific hypotheses, that we can test. We narrow down even further when we collect observations to address the hypotheses. This ultimately leads us to test the hypotheses with specific data.. a confirmation (or not) of our original theories.

Inductive Approach

Inductive reasoning works the other way, moving from specific observations to broader generalizations and theories. This is a 'bottoms-up' approach.

In inductive reasoning we begin with.

1) Specific Observations & measures.
2) begin to detect patterns & regularities.
3) formulate some tentative hypotheses that we can explore, and finally end up developing some general conclusions or theories.

Sample Chapter Organization of study

Chapter 1: Introduces problem statement
2: Review of literature
3: Methodology & procedures.
4: Analysis of data and presentation of results.
5: - Summary & discussion of the researcher's findings
- Implications for practice
- recommendation for future research.

Validity issues in research

25/26-11

Internal Validity

1) History effect ex Halloween
2) Maturation
3) Testing
4) Instrumentation ex caliberation changes
5) statistical regression
6) Differential selection
7) Experimental Morality
8) Selection maturation interaction

External Validity

1) Interactive or reactive effects of testing
2) Interaction & selection biases.
3) Reactive effects of experimental arrangement
4) multiple treatment interference.

what is Research Proposal?

1. A RP is the presentation of an idea that you wish to persue.
2. It is intended to convince funding agency/ RDC that you have a worthwhile research project and that yai have the competance and the work plan to complete it.
3. A good RP assumes that you have already thought about your project and have devoted some time and efforts in gathering information reading and organizing your thoughts
4. A high quality proposal not only promises success for the project, but also impresses RDC about your potential as researcher.

Samarth Borkar
GOA COLLEGE OF ENGG

Two Mean, Proportion & Variance

Mean

Z test

both samples are independent & σ_x^2 & σ_y^2 are known

$$Z_c = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

t-test

both samples are independent but σ_x^2 & σ_y^2 unknown

$$T_c = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{j=1}^{m}(y_j - \bar{y})^2\right]$$

$$d.f = m + n - 2$$

both samples are related

i) Obtain diff

$$d_i = x_i - d_i$$

ii) Mean diff

$$\bar{d} = \sum_{i=1}^{n} d_i$$

iii) Sampled std deviation Sd

$$Sd = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2}$$

iv) T_c

$$T_c = \frac{\bar{d}}{Sd/\sqrt{n}}$$

$$df = n - 1$$

Proportion:

$$Z_c = \frac{P_1 - P_2}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ where $\hat{\pi} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$

Variance:

$$F_c = \frac{S_1^2(x)}{S_2^2(y)} = \frac{\text{Sample variance of 1st sample}}{\text{" " 2nd "}}$$

Testing of hypothesis

H_0 = null hypothesis
H_1 = alternative hypothesis

Type I and Type II error

Decision	Actual situation: H_0 true	H_0 false (H_1 true)
Accept H_0	No Error prob = $1-\alpha$	Type II error prob = β
Reject H_0	Type I Error Prob = α	No Error Prob = $1-\beta$

where α is level of significance
ie probability of type I error

– Two tailed and one tailed test

i) $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_0$
is a two tailed test

ii) $H_0 : \mu = \mu_0$ against $H_0 : \mu > \mu_0$

OR

$H_0 : \mu \leq \mu_0$ against $H_1 : \mu > \mu_0$
is a right tailed test

iii) $H_0 : \mu = \mu_0$ against $H_0 : \mu < \mu_0$

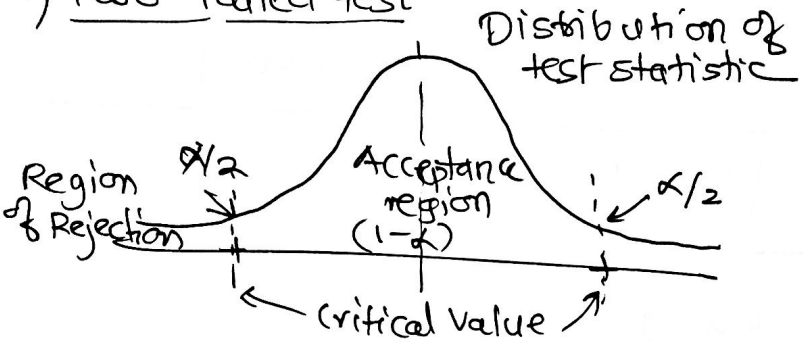
OR

$H_0 : \mu \geq \mu_0$ against $H_1 : \mu < \mu_0$
is a left tailed test.

Critical Value & Decision Rule

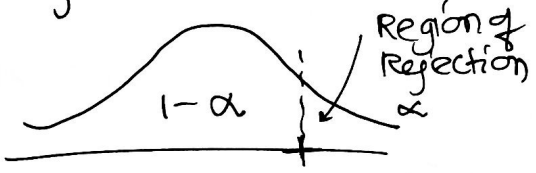
Critical value devides the area under the probability curve of distribution of test statistic into two regions, critical (or rejection) region and acceptance region

1) Two tailed test

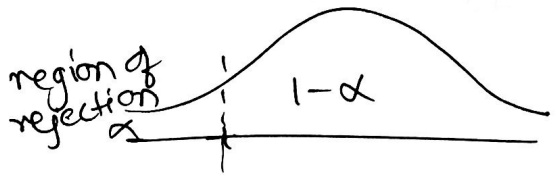


when the statistical outcome falls into Critical region Ho is rejected. when it falls in acceptance region, it is accepted.

ii) Right tailed test



iii) left tailed test



Procedure for hypothesis testing

1) Setting up the hypothesis
2) Selecting a significance level
3) Test statistic
4) Critical value
5) Decision.