Research Methodology UE18CS400SG
Unit 2

Aditeya Baral

February 2022

1 Research Design

1. Research Design is the preparation of the design of the research project
2. It constitutes the blueprint for the collection, measurement and analysis of data
3. Revolves around questions like what the study is about, why is it being made, where will it be carried out, the type of data that will be required etc

1.1 Need for Research Design

1. Facilitate the smooth sailing of the various research operations
2. Make research efficient (maximum information with minimal cost of money, effort, time)
3. For collection and analysis of data that will be required

Research Design stands for advance planning of methods to be adopted for relevant data collection and techniques used for their analysis keeping in view the objective of research and availability of staff, time and money

1.2 Research Design Break Down

1. **Sampling Design** – method to select items for study
2. **Observational Design** – conditions under which observations are made
3. **Statistical Design** – how many items to observe? how to gather and analyse data?
4. **Operational Design** – techniques of implementing steps in sampling, observational and statistical design

1.3 Concepts Related to Research Design

1.3.1 Dependent and Independent Variables

A variable is a concept that can take on different quantitative values. Can be discrete or continuous.

A **dependent variable** depends on, or is a consequence of another variable. A variable *antecedent* to the dependent variable is termed an **independent variable**.

1.3.2 Extraneous Variable

Extraneous variables are independent variables unrelated to the purpose of the study but may affect dependent variables.

Any effect on a dependent variable due to an extraneous variable is termed an *experimental error*. The term **control** is used to refer to restraining experimental conditions to minimise effects of extraneous variables.

1.3.3 Confounded Relationship

When the dependent variable is not free from the effect of extraneous variables, the relationship between the dependent and independent variables is said to be confounded by the extraneous variables.

1.3.4 Research Hypothesis

Testing of a prediction or hypothetical relationship using scientific methods.

It is a predictive statement relating one or more independent and dependent variables (must contain atleast one of each)

1.3.5 Experiment and Non-Experimental Hypothesis Testing Research

While testing the Research Hypothesis,

1. **Experimental** – independent variable is manipulated
   (a) Experiment under usual conditions – **control group**
   (b) Experiment under special conditions – **experimental group**

   A study can include both control as well as both experimental groups.
2. **Non-experimental** – independent variable ***not*** manipulated

1.3.6 Treatments

The different conditions under which experimental and control groups are put are known as *treatments*. Example, different types of techniques applied in a study are each considered a treatment.

1.3.7 Experiment

An *experiment* is the examination of the truth of a statistical hypothesis related to a research problem.

1. **Absolute Experiment** – determine impact or outcome of a study
2. **Comparative Experiment** – compare outcomes between studies

1.3.8 Experimental Units

Pre-determined plots or blocks where different treatments are used are called *experimental units*. These units must be selected and defined carefully.

1.4 Basic Principles of Experimental Design

Fisher's Principles of Experimental Designs state the following

1. **Principle of Replication**
   - Repeat experiment multiple times
   - Each treatment is applied in many experimental units
   - Increases statistical accuracy of experiment
2. **Principle of Randomization**
   - Design or plan experiment such that variations by extraneous factors can be all combined as due to *chance*
   - Provides protection against extraneous factors by randomization
3. **Principle of Local Control**
   - Vary extraneous factor (or known source of variability) over a wide range such that variability due to it can be measured
   - Allows elimination of variability due to extraneous factors from experimental error
   - Two-way analysis of variance – treatments, extraneous factors and experimental error
   - Divide the field into *num_treatments* homogeneous parts (process known as *blocking*) where each block contains fixed extraneous factors. Measure value to check contribution to total variability using two-way variance analysis and then eliminate variability from extraneous factors from experimental error.

1.5 Experimental Design

Experimental Design refers to the framework or structure of an experiment.

1.5.1 Informal Experimental Design

1. **Before-and-after without Control**
   - Measure dependent variable, apply treatment, measure again
   - $Treatment\ Effect = phenomenon\ level_{after} - phenomenon\ level_{before}$
2. **After-only with Control**
   - Two groups – *test* and *control* with treatment added to test group
   - $Treatment\ Effect = phenomenon\ level_{test} - phenomenon\ level_{control}$
3. **Before-and-after with Control**
   - Two groups – *test* and *control* with treatment added to test group
   - Measure dependent variable in both groups, apply treatment to test group, measure both groups again
   - $Treatment\ Effect = (phenomenon\ level_{test,after} - phenomenon\ level_{test,before}) - (phenomenon\ level_{control,after} - phenomenon\ level_{control,before})$

Important – Time periods for measurement must remain constant.

1.5.2 Formal Experimental Design

1. **Completely Randomized**
   - Involves only principles of replication and randomization
   - Two approaches
     (a) **Two-group Simple Randomized** – sample from defined population and assign to experimental or control group (follows principle of randomization). **Disadvantage** – extraneous factors are not controlled.
     (b) **Random Replications** – each treatment is replicated a number of times to reduce effect of extraneous factors. Two populations (for study and to conduct experiments), sample from each, randomly assign to multiple experimental and control groups.
2. **Randomized Block**
   - All principles are applied
   - Subjects divided into groups (called blocks), keep extraneous factor fixed in each block to measure contribution to total variability
   - Each treatment appears same number of times in each block
   - Analyse using two-way variance analysis (two-way ANOVA technique)
3. **Latin Square**

- Frequently used in agricultural research
- An n x n table with n symbols, each symbol referring to a treatment such that each symbol appears once per row and column
- Used to control variation in two different directors (or two factors)
- Assign treatments randomly to combinations of the two factors keeping each treatment's occurrence as per point 2.

	Cow 1	Cow 4	Cow 3	Cow 2
Period 2	T1	T3	T2	T4
Period 4	T2	T1	T4	T3
Period 1	T4	T2	T3	T1
Period 3	T3	T4	T1	T2

- Refer section 1.1 and 2

4. **Factorial**
   - Used to vary more than one factor (independent variable) and find effect on dependent variable
   - Mostly used for social and economic phenomena
   - (a) Simple Factorial/Two-Factor Design – 2 factors
     (b) Complex Factorial/Multifactor Design – > 2 factors
   - If $n^{(i)}_{levels/treatments}$ corresponds to the number of levels or treatments for factor i, then the total number of cells in the design will be

$$Number\ of\ cells\ in\ table = \prod_{i=1}^{factors} n^{(i)}_{levels/treatments} \tag{1}$$

1.5.3 Features of a Good Research Design

It must consider the following factors

1. Means to obtain information
2. Availability and skills of researchers
3. Objective and nature of problem
4. Availability of time and money for research
5. Flexibility to consider different aspects
6. Maximum accuracy with minimum bias

2 Sampling Design

- A **population** is a large group from which individuals are selected to participate in a study
- A **sample** is a smaller collection of individuals taken from a population for study. The sample must be representative of the **target population** from which the individuals were selected.
- Taking entire population in study is called **census** (*impossible for cost reasons*)
- A **sampling frame** is a list of all elements or other units containing the elements of a population
- Sample Design
  - Plan for sampling
  - Technique for selecting sample
  - Sample design detected before and after sample collection

2.1 Steps in Sampling

1. **Objective** – research objective in proportion with manpower, money and time
2. **Population** – clearly defined
3. **Sampling Units and Frames** – select unit for sample and sample source list
4. **Sample Size** – optimize for efficiency, flexibility, reliability
5. **Parameters of Interest** – statistical constants like mean
6. **Data Collection** – relevant information only
7. **Non respondents** – practical difficulties lead to data not being collected, changes results
8. **Selecting Sampling Design** – select technique that yields least error
9. **Organize Field Work** – reliable, trained personnel with supervisory staff
10. **Pilot Survey** – research on small scale before field scale
11. **Budgetary Constraints** – practical cost consideration, affects sampling decisions like size and technique

2.2 Sampling and Non-Sampling Errors

Sampling Error	Non-Sampling Error
Due to inferences made on non-representative samples	Due to improper data collection and preparation
Present only in sample	Present in both census and sample
Precision is measured for a sample size and design, can be improved by increasing sampling size (but with cost)	Reduced by defining sampling unit, frame and population correctly

2.3 Sampling Techniques

1. **Probabilistic** – simple random, systematic, stratified, cluster
2. **Non-Probabilistic** – sequential, quota

2.3.1 Simple Random Sampling

- Probability based
- Randomly select without replacement
- Each individual has equal probability
- Becomes biased in large populations

2.3.2 Systematic Sampling

- Randomly select start point, then select every n^{th} individual
- List should not contain any hidden order
- Works well for large populations

2.3.3 Stratified Sampling

- Performed when sample is not homogeneous, but possible to form homogeneous groups of population
- Divide population into homogeneous groups (called *strata*) based on a factor that may influence dependent variable
- Perform *simple random sampling* on each stratum

2.3.4 Cluster Sampling

- Population divided into groups (if groups are geographic areas, called *Area Sampling*)
- Select samples from select groups

- **Advantages** – useful when population is spread over large geographic area, convenient, reduced cost
- **Disadvantage** – Less precise, representation issues likely

2.3.5 Multistage Sampling

- More than 1 sampling technique used
- Complex and rarely used, requires more effort, time and cost

2.3.6 Sequential Sampling

- Complex because size is not fixed
- Used for acceptance sampling
- A sequence of samples are taken from a lot
  - When a particular lot to be accepted/rejected on basis of a single sample, it is called *single sampling*. If two samples are used, it is called *double sampling*. **If multiple but undefined samples are used, it is called sequential sampling**.

2.3.7 Quota Sampling

- Divide population into groups (like stratified)
- Judgement used to select individuals related to study from each group

3 Data Collection

1. **Primary Data** – data being collected for the first time, original, fresh, new
2. **Secondary Data** – data already collected before, been used for analysis

3.1 Collection of Primary Data

3.1.1 Observation

- Related to behavioral sciences
- Information without asking respondents
- Methods
  - **Non-Scientific** – observe surroundings
  - **Scientific** – plan and record, checks performed, validity tested

- **Advantages** – no subject bias, current happenings, independent of respondents
- **Disadvantages** – expensive, limited information, unforeseen factors, people not always accessible
- **Terminologies**
  - **Structured Observation** – units, styles, standardised conditions, descriptive study
  - **Unstructured Observation** – exploratory study
  - **Participant Observation**
  - **Non-Participant/Disguised Observation**
  - **Controlled Observation**
  - **Non-Controlled Observation**

3.1.2 Interview

1. **Personal Interview**
   (a) Direct face-to-face questions asked
   (b) Types
      i. **Direct** – interview source and collect data
      ii. **Indirect** – interview 3rd party close to source or someone who has knowledge about the problem
      iii. **Structured** – structured data collection, predetermined question set and fixed order
      iv. **Unstructured** – flexible, no predetermined question set or order, interviewer given freedom to add/remove questions
      v. **Focused**
      vi. **Clinical**
      vii. **Non-Directive**
   (c) **Advantages** – more information, greater flexibility, easily obtained, low non-respondents, choice of respondent, less misinterpretation of questions
   (d) **Disadvantages** – expensive, time consuming, respondents not always approachable, bias due to interview presence, selection and training of staff required
   (e) **Prerequisites** – selection and training of interviewer, honesty, technical competence, practical experience, should not deviate from instructions
2. **Telephonic Interview**

(a) Collect data over telephone

(b) Not widely used (industry survey mainly)

(c) **Advantages** – flexible, fast, cheap, responses can be recorded, easy to call back, no field staff, higher number and wider range of respondents

(d) **Disadvantages** – little time to answer, less geographic coverage, short questions and point answers, more bias of interviewer

3.1.3 Questionnaire

1. For economic and business surveys
2. Conducted by private individuals, research workers, organizations, governments
3. Fixed number and order of questions (open-ended, MCQ, T/F) to be filled out and returned
4. **Advantages** – low cost, large geographic area, no interviewer bias, larger response time, can contact non-approachable respondents, more responses leads to more accurate results
5. **Disadvantages** – low return rate, educated and cooperative respondents needed, no flexibility (cannot make changes to questionnaire once sent), slowest, may have incomplete or ambiguous responses
6. **Pilot Study** needed to test questionnaire and make modifications based on study – add, remove, reword, rephrase etc

3.1.4 Schedule

1. A *schedule* is a set of questions (direct, open/close ended, tabular) which are asked by the interviewer (also called *enumerator*), who also fills out the respondent's answers
2. Schedules may be handed out to respondents to fill themselves
3. Enumerators are appointed, help respondents fill out answers, explain objectives of study, clarify doubts
4. **Advantages** – useful for illiterate respondents, less non-respondents, reliable data
5. **Disadvantages** – expensive, requires selection and training of enumerators, enumerator bias, respondent not anonymous

Questionnaire	Schedules
Respondent fills	Enumerator fills
Cheap, economical	More expensive – enumerators selection, training
Higher non-response	Lower non-response
Anonymous responses	Identity of respondent is known
Wider range of respondents	Smaller range covered by enumerators
No personal contact	Needs personal contact
Required literate and cooperative respondents	Respondents may be illiterate
Depends on respondent answers	Depends on honesty of enumerator

3.2 Collection of Secondary Data

Data already collected and analysed by someone else, available publicly for others to use.

3.2.1 Published Secondary Data

Publications from government bodies, journals, reports from organizations and academia, books, magazines, newspapers, websites, public records and historical documents.

3.2.2 Unpublished Secondary Data

Diaries, letters, biographies made available publicly

3.2.3 Checklist for Secondary Data Collection

1. **Reliability** – who collected, sources, methods used, time period
2. **Suitability** – data must suit the study
3. **Adequacy** – data must be adequate for study

3.3 Survey vs Experiment

Survey	Experiment
Descriptive Research	Experimental Research
Large data size	Small data size
No manipulation	Deliberate manipulation
Relationship between data and unknown studied through survey	Relationship between data and unknown determined
Casual analysis	Correlation analysis