Basic Research Design and Methods

Survey Research

Experimentation

Documentation/Historical Analysis

SURVEY

The social survey is conventionally associated with questionnaires and interviewing. Survey is best conceptualized as a research design

Sampling Theory: Every day we are involving in sampling. The notion behind sampling theory is that a small number of sampling units can tell you something about the total population.

Units: An important phase in the collection of data is the selection of the units to which the data relate. In social research-and particularly in sociology-by far the most common unit of analysis is the individual without forgetting the collective actors as unit of analysis.

Population: Population refers to a universe of units which can be people, people, organizations, sub-units within organizations and so on. A sample must be selected from a population. In order to enhance a sample is representative; the researcher will need to engage in probability or random sampling. Probability sampling means that selection bias may be largely eliminated.

Types of Probability Sampling

Simple Random Sample

The most basic form of probability sample is the simple random sample (SRS). With SRS each unit in the population has an equal probability of inclusion in the sample. Listing of the population, known as the sampling frame, is required. A non-probability sampling may ensue with unrepresentative sample to the population. By removing the human hand and relying on random numbers, the possibility of a systematic bias is reduced.

Stratified Random Sample

A further type of probability sample that is often used is the stratified random sample. The population is divided into strata. A simple random or systematic sample taken from the entire population might have lead to an accurate representation, but the stratified random sample can enhance the likelihood of accuracy

Multi-stage Cluster Sample

This form of sampling is likely to be of particular use when the researcher is interested a widely dispersed population. The chief advantage of this approach is, when the interview is to be conducted, interviewers will be more concentrated in particular areas, thereby saving much time and money on travel. A convenience sample occurs when the sample has not been driven from any form of probability sampling

Ways to Collect Data Using Survey Design

The mail questionnaire

Some advantages

- Minimum time, money and other costs
- Allow greater anonymity
- Reduce errors that may arise in the process of interviewing

Some disadvantages

- Loss of control over the survey
- Problem of comprehending complex issues by some respondents
- Inaccurate mailing list associated to mobility of people

Telephone Survey

Advantages

- Less expensive compared to face to face interview
- Interviewers have better control over the choice of respondents

- Help to access individuals who do not open their doors for interview for several reasons

Disadvantage

- The interviewer cannot see the respondent
- The interviewer misses facial expressions
- The quality of interviewers voice is important
- It must be kept short
- Exclusion of potential respondents who do not have telephones

Personal interviews

Advantages

- Gives to the interviews more control on the process than the mail or telephone surveys
- The interviewer can:
    - Read facial expressions and moods
    - Monitor environmental distractions
    - Observational data can be determined from personal interviews

Disadvantages

- Much more expensive than the two other approaches
- Sometimes loss of respondents' locations
- The safety of the interviewers is a concern

Experiment and Quazi-Experiment

There are two types of quantitative research deign: Experimental design and Non-experimental design.

Experimental deigns are sometimes known as the scientific method. The basis of the experimental method is the *experimen*t. Experiment is a test under controlled condition that is made to demonstrate a known truth or examine the validity of hypothesis. Experiments traditionally take place in laboratories, environments where all extraneous influences can be shut

out. In an experiment, the researcher manipulates the variable that is supposed to affect the outcome of the experiment.

Designing Experimental Study

1 Define Research Objectives

This step needs to e taken before deciding whether or not to do experimental research. It is important to spell out the objectives of the research. Research objectives need to be realistic; Need to limit ourselves to what is actually researchable and Need to be clear on what our population is.

2 Formulate Hypotheses

A research hypothesis can be defined as a tentative explanation that accounts for a set of facts and can be tested by further investigation. In experimental research, we traditionally look at two distinct types of hypotheses: The null hypothesis and the alternative hypothesis. In practice, most researchers test a null hypothesis of no difference

3 Set up the research design

The traditional experimental design is the pre-test post-test design. Subjects are placed into two groups: the experimental group and the control group. Both groups get the pre-test and a post-test. Another decision to be taken is whether or not to give the control group a placebo. Selecting a placebo can be hard in social science experiments. It is not as simple as giving patients a sugar pill selecting instruments. Prepare appropriate pre and post test measures. The measurements instrument first of all must measure what we want them to: this is known as validity.

The instrument must be reliable select appropriate levels to test the hypotheses. Think carefully about the right level of measurement at which to test the hypotheses. A good example of this is the medical paracetamol: Underdose and overdose.

Assign persons to groups

Always try to minimize the influence of external factors. Try to make sure the experimental and control group differ as little as possible at the start of the experiment. The best way to achieve

this is through randomization carryout the Experiment Carefully. In carrying out the experiment, make sure that extraneous factors are controlled as much as possible. It would be hard to conduct an experiment in an environment in which all kinds of other things are going on. Be sure that whatever outcome we find is a result of the treatment. We need also to control how the experiment is carried out. Do not introduce experimental bias: the effect of the experimenter on the experiment

Analyze the Data

Once the experiment is done and the post-test administered we have to analyze the results. Typically such as t-test and ANoVA are used.

Quazi-Experimental Designs

The main distinction between experimental and quazi-experimental designs lies in the allocation of persons to groups. Rather than randomly allocating we will have to choose a control group that is as similar to the experimental group as possible. We are not using random allocation, we call this control group the *comparison* group as it is not a pure control group. It is crucial to ensure that the experimental and comparison groups are as similar as possible. Quazi-experimental designs do have one clear advantage over pure experimental design.

a. Groups are studied a natural settings

Survey Research Planning

Probability and Non-Probability Sampling Techniques

Designing a Survey Study

The most popular quantitative research design in the social sciences is survey research. Designing survey research is not that simple. The phases in designing a survey study are similar to those in experimental research.

Step 1: Define research objectives

Survey research is well suited to descriptive study. Or where researchers want to look at relationships variables occurring in particular real life context. Once we define our objectives we can proceed to the research design which will be dependent on those objectives.

Step 2: Formulate a hypotheses

While in experimental designs it is common to develop and test hypotheses in all cases, this is not necessarily so in survey designs. Survey studies can be more wide-ranging complex than experiment studies. Not all survey studies test specific hypotheses. Some survey studies can be purely descriptive. Whether one wants specific hypotheses or conduct a more descriptive study will depend on your research question.

Step 3: Define what information you need

Define what information you need and decide exactly what your population is going to be. Sometimes it is possible to take the whole population as a sample. Sampling the whole population is known as census. It is important to note that we can only generalize to a population we have actually sampled from.

Step 4: Decide how to sample from the population

We need to have unbiased sample of the population. The sample should be a representative of the population we are studying. The best way is using probability sampling methods. Simple random method is the best well known. In practical terms most social science research do not use simple random sampling for a number of reasons, some good, and some less so. In some cases stratified random sampling is better than simple random sampling. Sometimes we want to be sure that different sub-groups are represented according to their presence in the population. In some instances cluster sampling is by far better than simple random. Still in other instances multi-stage sampling may be required. The problem with cluster and multi-stage sampling is that they are not purely random.

Step 5: Design your research instruments

Survey instruments include, a written questionnaire, a phone questionnaire, or online survey questionnaire. The quality of the data depends on the quality of instruments.

Step 6: Collect the data

This is the phase where problems can occur in survey studies. Non-response to the survey is one of the challenges. More response rates make the sample smaller and less statistical power to test our hypotheses.

Step 7: Analyze the data

The final stage is data analysis

Advantages and Disadvantages of Survey Research

Advantages

- Survey research is highly flexible
- Easy to generalize findings to the real world situation unlike experiments
- Efficient in terms of gathering large amounts of data at reasonably low cost and effort
- Easy to guarantee respondents' anonymity

Disadvantages

- Unable to help to control the environment
- Therefore less suited to answering questions of causality
- It is difficult to come to deeper understanding of processes and contextual differences through questionnaires, which are standardized and by their nature limited
- A combination of survey and qualitative methods can help here

Sampling

Why Sampling?

A major reason for studying samples rather than a whole group is that the whole group is sometimes so large to study.

Sampling terminologies

Population: all possible cases of what we are interested in studying

Sampling frame: Sampling frame is a listing of all the elements in a population. In experiments we talk about random assignments whereas, in surveys we are concerned about sampling.

Precision: The term refers to how close the sample estimate is likely to be to the value that would be obtained if the whole population had been included in the survey. The amount of sampling error a researcher wants to accept determines the level of precision

Sampling models

Survey study has two fundamental goals: Efficiency and Economy. Efficiency is the attempt to balance consideration of cost with those of precision

Non-Probability Sampling

Use of non-probability sampling

Non-probability sampling is useful when the researcher plans to see whether there is a relationship between independent and dependent variables with no interest to generalize the results beyond the sample. It is useful in some qualitative research where the goal is to understand the social process and meaning structure of a particular setting. When it is impossible to develop a sampling frame of a population.

Drawbacks

Not possible to talk about representativeness on the basis of non-probability sampling the techniques we use to estimate sample size is not applicable to non-probability sampling.

Examples of Non-probability Samples

Haphazard: Get any cases in any manner that is convenient

Quota: Get a preset number of cases that will reflect the diversity of the population

Purposive: Get all possible cases that fit particular criteria

Snowball sampling: Get cases using referrals from one or a few cases

Extreme case: Get cases that substantially differ from the dominant pattern – A special type of purposive sampling

Sequential: Get cases until there is no additional information or new characteristics - often used with other sampling methods

Theoretical: Get cases that will help reveal features that are theoretically important about a particular setting/topic

Purposes and Designs of Survey Research: Types of Variables and Instruments Design

Levels of Measurement

Levels of measurement are categories of variables. There are three levels measurement:

- Nominal/categorical
- Ordinal
- Continuous/interval/ratio

Nominal measures

Nominal variables are measured at the lower level. For nominal variables numbers only replace names (e.g. for the variable gender, 1 represents boys and 2 girls). The values cannot be placed in order (girls cannot be more than boys). Numbers given are merely a descriptor of that category

Ordinal variables

Ordinal variables do possess a natural ordering of categories. Example for the question I get good marks in English

- Code 4 agree strongly
- Code 3 agree
- Code 2 disagree

- Code 1 disagree strongly

What we cannot do is measure the distance between the scale points. We cannot assume that the distance between each scale points is exactly the same like it is measuring using a ruler

Continuous variables

Continuous variables are variables that serve as a ruler. Not only can we order categories but also the distance between each category is the same. Variables are measured in a continuous scale: temperature, weight and height are examples of variables.

Type	Description	Example
Nominal	A classification of objects	Ethnic group
Ordinal	The categories associated with a variable can be ranked-order	Commitment
Interval (a)	With true interval variables, can be ordered	Income, age
Interval (b)	Ordinal but which have large number of categories	Satisfaction

dichotomous	A variable that comprises only two categories	Gender, attendance

Measurement and analysis

Whether a variable is nominal, ordinal or continuous has significant consequence for what type of analyses we can do with it. The level of measurement can interpret the variable.

Do nominal and ordinal variables really constitute measurement?

Some researchers say nominal and ordinal variables are not real measures in natural sciences. To them real measures must be continuous and conform to mathematical measurement models.

Measures of central tendency

The mean, the Median and the Mode

When we talk about average the value we mean is about the mean. How can we calculate a mean? Can we calculate a mean for variables with nominal measures, example gender? What about variables that we measure at ordinal levels?

There are other measures of central tendency

- Median: Middle category of a distribution, More suitable to ordinal variables
- The mode: The most common value, More suitable to nominal variables

Measures of Spread

Range and Standardized Deviation

To give good descriptions of our variables, we need measure of spread. Range is the difference between the lowest and the highest values. An outlier may distort a statistic if we trust range as a

measure of spread. A measure that does use all the information we have is the standard deviation (SD). SD measures the cluster of values around the mean. SD is the square root of the variance. Variance = sum of square of deviations from the mean divided by number of observations - one

Validity and reliability of measures

Reliability

- The reliability of a measure refers to its consistency
- Two separate aspects
  - Internal
  - External
- External reliability is the degree of consistency of a measure over time
- It is administering a test at two occasions for a similar group and see the consistency of results
- Test-retest-reliability is related to external reliability
- Internal reliability is important in multiple item scales
- It should answer the question whether the items that make up the scale are internally consistent

Validity

- How far a measure really measures the concept that it purports to measure
  - Example how do we know that our measure of job satisfaction is really getting at job satisfaction not at something else

Scaling

Operationalizing complex concepts

Basic concepts in social sciences

    - Optimism
    - Depression
    - Racial prejudice
    - Authoritarianism
    - Religious devotion, social integration, social conflict
- How could these be translatable into empirical research?
- Complex concepts can be operationalized by indicators, e.g. for religion, religious devotion is an indicator
- The technique of scaling serves to achieve the objective of operationalizing
- Scaling replaces the concept into a coherent and organic indicators
- Therefore a scale is a coherent set of items that are regarded as indicator of a more general concept
- In the field of sociology and social psychology, the most common application of scaling is seen in attitude measurement, where the unit of analysis is the individual and the general concept is an attitude and the specific concepts are opinions
- A scale is made up of several items, and these items are question

Basic and advanced statistical tools, Statistical inferences

Properties of the Data

- To be properly used, some statistics require that the data to be analyzed have certain mathematical or other properties
- The level of measurement involves the mathematical properties required for particular statistical tests

- One important assumption of many statistical procedure is that the observation on the dependent variable are independent of one another
- Another important assumption for some statistics concerns the shape of the distribution of observations on a variable
- Some statistics requires normal distribution or symmetric distribution of observations

Audience

- Consider the audience for whom the data analysis is intended
- If the audience has limited statistical expertise, simple, easily understood statistic is preferred over a more complex statistic that might confuse them

Descriptive statistics

- Descriptive statistics are most commonly used in the social sciences and human services
- Descriptive statistics provide quantitative indicators of what is common or typical about a variable, how much diversity or difference there is in the variable, and how values on one variable are associated with values on one or more other variables

Measures of central tendency

- Measures of central tendency/average summarize distributions by identifying the typical or average value and are one of the most commonly used statistics
- The three most widely used measures of central tendency are:
    - The mode
    - Median
    - mean

- The mode is the least stable of the three measures of central tendency

- The modal value can be changed with minor additions
- With ordinal data the median is the most appropriate measure of central tendency
- The mean is only suitable for interval or ration data
- Because it takes the values of all scores the mean is less stable than the median

Measures of dispersion

- Like measures of central tendency, measures of dispersion describes and summarizes distributions
- Measures of dispersion indicates how dispersed or spread out the values are in a distribution
- The commonly
- The three commonest measures of dispersion are:
- Range
- Semi-interquartile range
- Standard deviation
- The range indicates the total spread of distribution
- Standard deviation indicates the average (mean) spread of the scores from the mean
- Standard deviation has important application in inferential statistics

Measures of association

- Measures of associations describe the nature of relationships between variables, particularly the strength of relationship or how closely variables are related

- The strongest relationship is a perfect one
- The positive relationship is one in which a change in value of the variable is in the same direction
- The negative relationship is one in which on variable increases while the other variable decreases
- Strong relationship near +1 or -1
- Weak relationship near to zero
- The overriding determinant in selecting a measure of association is the level of measurement of the data at hand

Measures of association and levels of measurement

Nominal data

- A good measure of association for dichotomous data is the phi coefficient
- Phi is a good measure for three reasons

    1. It is quite easy to compute
    2. It is mathematically associated to measures of association

3 Phi is a member of a group of measures of association that can be given what is called proportional reduction error (PRE) interpretation

- The PRE interpretation means that the measure shows how much the independent variable helps reduce error in predicting the values of the dependent variable

- To interpret Phi it is necessary to square it
- If phi is 0.11 is squared to become 0.01

- The latter value is treated as a percentage and is interpreted to mean that the independent variable reduced the error in predicting value of the dependent variable by 1 percent
- Because phi is suitable only for two dichotomous variable, a different measure of association must be used for nominal data with more categories

Ordinal data

- The most popular measure of association for fully ordered data is spearman's rho (rs)
- Rho facilitates comparisons of relationships
- With ordinal data the negative sign (-1) has meaning and indicates a negative relationship
- Rho also may be squared and given the PRE interpretation

Interval scale

- The most used measures of association for interval data is the correlation coefficient or Pearson's r
- Correlation refers to the linear relationship between two variables
- The correlation coefficient is a measure of the association between two numerical variables, usually denoted as *x* and *y*
- A coefficient of +1 is a perfect positive correlation between *x and y, while a coefficient* of −1 is a perfect negative correlation (sometimes referred to as an inverse correlation)

$$r = \frac{(x - \bar{x})(y - \bar{y})}{(x - \bar{x})2(y - \bar{y})2}$$

- The correlation coefficient is mathematically related to both phi and rho
- R varies between -1.00 and + 1.00 and may be squared and given the PRE interpretation
- When squared, this value is called the coefficient of determination

- If the correlation coefficient is squared it gives the proportion of the variance of one variable that is 'explained' by the other variable

Example

- If the correlation coefficient is squared it gives the proportion of the variance of one variable that is 'explained' by the other variable by *x*.
- But this should not be interpreted as meaning that 41 per cent of *y is caused by x*
- Pearson's r indicates the degree to which the relationship between two interval-level variables can be described by a straight line when plotted on a scattegram

Inferential statistics

- Findings from sample data are of little scientific value if they cannot be generalized beyond the members of the sample to the larger population from which the samples were drawn
- Statistics from which inferences are made about situations or social groupings that have not been observed directly
- In contrast to descriptive statistics, inferential statistics are used to make generalizations derived from estimates based on probability
- Descriptive statistics simply state what has been observed.
- Inferential statistics make inferences based on those observations
- Inferential statistics enable hypotheses to be tested and generalizations to be made
- These can be:

1 determining whether a sample came from a particular population

2 whether differences observed between groups are likely to be real differences, or are more likely to have occurred by chance

3 whether a relationship between two or more variables is significant

- Inferential statistics also enable us to make predictions about what is likely to happen on the basis of what has been observed
- Statistics such as chi-square, Student's *t and Pearson's product moment* correlation coefficient are all inferential
- While the result given by a random sample produces a reasonable estimate of what the population is like, it seldom produces exactly the same result as if the entire population had been surveyed
- Different samples drawn from the same population are likely to give slightly different results.
- These variations are known as sampling errors
- Assuming, that the figure for the population (the mean age, for example) is not known, and all one has is the sample result, then an inference is made about the mean age of the population based on the sample result, but one knows that it is unlikely to be exactly the same as the figure for the population
- So the population figure is going to be an estimate
- This estimate is based on probability
- Probability theory is therefore the basis for inferential statistics, and an awareness of the role played by probability is important