Study Guide

Applied Statistics in Gender Studies

Code No. 4646

Units 1 – 9

Department of Gender & Women Studies
Faculty of Social Sciences & Humanities
ALLAMA IQBAL OPEN UNIVERSITY
Code No. 4646

(All Rights Reserved with the Publisher)

Edition          First
Printing         2017
Quantity         1600
Price            Rs. 175/-
Printing Coordinator  Management Committee for P.P.U
Printer          Kashish Printers, Lahore.
Publishers       Allama Iqbal Open University, Islamabad.
ACKNOWLEDGMENT

The Gender & Women Studies, a relatively new discipline at post graduate level (MSc.) has been on the priority agenda of the Allama Iqbal Open University programme since the early 1990s. The Gender & Women Studies department launched its M.Sc programme since 2005. The present course is one of the compulsory courses of the programme.

We would also like to acknowledge the inputs of the members of the Committee of Courses. We would also like to acknowledge the work of all the writers whose works have been used as reference in allied material for this course.

Finally. I wish to acknowledge the course development co-coordinator of this course Ms. Atifa Nasir, who has worked very hard and diligently for the study guide to be in your hands. Last, but not the least, I would like to acknowledge all those who have contributed to the course in one way or the other.

Dr. Riffat Haque
Chairperson
Department of Gender & Women Studies
Course Team members

Dr. Riffat Haque
Ms. Mamoona Fayaz
Ms. Atifa Durrani
Ms. Najeeba Batool
Dr. Anila Kamal
Mr. Abdul Wadood
Ms. Atifa Durrani

Chairperson
Member/Unit Writer
Member/Unit Writer
Member/Unit Writer
Reviewer
Editor
Course Development Coordinator
and Course Coordinator
<table>
<thead>
<tr>
<th>Unit</th>
<th>Title</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Inferential statistics</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>Statistical Hypothesis Testing</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>Descriptive Method of Data Analysis</td>
<td>51</td>
</tr>
<tr>
<td>4</td>
<td>Data Collection and Data Analysis</td>
<td>51</td>
</tr>
<tr>
<td>5</td>
<td>Sampling</td>
<td>101</td>
</tr>
<tr>
<td>6</td>
<td>Research tools for Measuring Behaviours</td>
<td>130</td>
</tr>
<tr>
<td>7</td>
<td>Historical and Descriptive Research</td>
<td>164</td>
</tr>
<tr>
<td>8</td>
<td>Experimental Designs</td>
<td>189</td>
</tr>
<tr>
<td>9</td>
<td>Use of Personal Computer in Data Analysis</td>
<td>211</td>
</tr>
</tbody>
</table>
INTRODUCTION OF THE COURSE

Basic Statistics is a form of mathematics that allows you to properly collect, analyzes, interpret and present data in an easy to understand format. It is a basic skill needed to ensure proper understanding of many of the more centralized statistic fields. Statistics is the science of collection, analysis, interpretation or explanation, and presentation of data. It has wide usage in the field of research. In fact all the data collection and interpretation techniques used in Research are part of statistics. It makes use of descriptive statistics for collection of data and inferential statistics for drawing inferences from this set of data. Whatever type of research we undertake (for example, naturalistic observation, case study, surveys or experiments) our efforts can usually generate a considerable amount of data: numbers that represent our research findings and provide the basis for our conclusions. Statistical analyses are the methods most often used to summarize and interpret data. Very often it is the fear of using statistics that prevents us from fully realizing the potential of our data. In reality, statistics can be a straightforward and enjoyable process!

As you know that statistics is a very vast discipline, however, in this course we have introduced to you some of very basic concepts which you must be familiar with and may be using them in gender studies research.

Why Study Statistics?

Hopefully, the discussion above has helped you to understand a little better about statistics. However, you may still be wondering "Why do I need to learn statistics? There are five major reasons to study statistics:

The first reason is to be able to effectively conduct research. Without the use of statistics it would be very difficult to make decisions based on the data collected from a research project. Statistics, however, provides us with a tool to make an educated decision. We will be able to decide about the possibilities which may likely to be true. We may base this decision on our knowledge of probability and inferential statistics.

A point about research should be made. It is extremely important for a researcher to know what statistics they want to use before they collect their data.
Otherwise data might be collected that is un-interpretable. Unfortunately, when this happens it results in a loss of data, time, and money.

The second reason to study statistics is to be able to read journals. Most technical journals you may read contain some form of statistics. Usually, you will find them in something called the results section. Without an understanding of statistics, the information contained in this section will be meaningless. An understanding of basic statistics will provide you with the fundamental skills necessary to read and evaluate most results sections. The ability to extract meaning from journal articles and the ability to critically evaluate research from a statistical perspective are fundamental skills that will enhance your knowledge and understanding in related coursework.

The third reason is to further develop critical and analytic thinking skills. The study of statistics may serve to enhance and further develop these skills. To do well in statistics one must develop and use formal logical thinking abilities that are both high level and creative.

The fourth reason to have a working knowledge of statistics is to know when you need to hire a statistician. Conducting research is time consuming and expensive. If you are in over your statistical head, it does not make sense to risk an entire project by attempting to compute the data analyses yourself. It is very easy to compute incomplete or inappropriate statistical analysis of one's data. In other words, you want to be able to make sure that your statistician is on the right track.

**Why statistics is needed in Gender Studies research?**

In many developing countries today, much of the rural sector, especially women, live in poverty. Despite the fact that sustainable agricultural development aims at balancing greater productivity and better yields with natural resource conservation, enhanced incomes, job creation and improved levels of food and nutritional security, many development programmes and policies have actually exacerbated poverty or done nothing to improve local standards of living, especially those of women.

Development plans are formulated primarily in terms of economic criteria,
while social and human parameters are seen mostly as justifications for economic decisions. When the human factor is given as much importance as the economic aspects, planning exercises become very complex; introducing a gender perspective complicates the issue even more. Planners rarely see the relevance of the gender perspective, partly because they lack reliable, impartial data on the type and extent of men's and women's separate contributions.

In a world in which economic value is reckoned in purely monetary terms, women's work, which is often unpaid, is not considered to be productive work. So, although women are the pillars of subsistence economies and pivotal to food security, their activities tend to be excluded from economic accounts. Agricultural statistics therefore tend to under-represent, or even omit, variables that are essential to a clear understanding of rural sector activities and rural development. This severely limits planners' grasp of the real situation in rural economies which, in turn, constrains their potential to act.

Until a few years ago, the demand for specific data and indicators incorporating a gender perspective was limited to advocates of the rights of women and disadvantaged groups. Nowadays, the user audience has expanded to include decision-makers at every level and in every area of social and economic development. There is greater general awareness of the need for a gender perspective in development policy formulation, and of the corresponding need for pertinent statistics. At the same time, as reliable data become available, they help to promote and justify change and to dissipate doubts and skepticism with respect to the relevance of innovative approaches such as the gender perspective.

In short, statistics incorporating a gender perspective are now essential for:

- Advocates of gender equity, who want them to boost awareness of their concerns;
- Planners, who want them for economic and social policy formulation, implementation and monitoring;
- Development experts, who want to review and analyze gender aspects and interactions;
- International, government and non-governmental organizations (NGOs),
who use them in project and programme design, implementation and evaluation;

- The general public, who wants them for a better understanding of society.

- The concepts and methods proposed and adopted by the many countries who seek a more faithful mirror of reality have begun to bear fruit, and new strategies have been developed to improve data presentation and dissemination, incorporating a gender perspective into statistics production.

Keeping these facts, needs and requirements in mind this course has been developed to cater the basic statistics for gender studies students.

Atifa Durrani

Course Development Coordinator / Course Coordinator
HOW TO STUDY

The study material for this course comprises of this Study Guide and a textbook *Exploring Research*. Each unit requires one week’s study. If you spend two units daily to study your course you can complete the course in eighteen weeks. In mid of the study period a workshop will also be held which is an effort to help you to prepare for examinations and meet peer group and listen to the subject experts and exchange knowledge.

Please do not confine yourself to the materials, which are being supplied by the university. To enhance knowledge at postgraduate level the students are expected to extensively use library and internet.

Tutors Guidance

In distance learning system basically the students have to study on their own. However, if there is a viable group of 10-15 students the university does appoint a part time or a correspondence tutor. Part them tutors hold tutorial meeting in study centers established by the university. The students are required to regularly attend these fortnightly meetings. Other wise you are assigned a correspondence tutor who not only checks your assignment but you are encouraged to be in contact with the tutors for guidance regarding the course as is convenient for both of you. The regional office as well as your tutor will inform you about the appointment of the tutor.

Assessment and evaluation

According to university system your performance in the course will be evaluated through tow modes that are:

- Home Assignment
- Final Examination

You will be required to do two assignments for this course. The assignments are spread over course units and according to the schedule provided
Introducing Inferential Statistics
<table>
<thead>
<tr>
<th>Contents</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>15</td>
</tr>
<tr>
<td>Objectives</td>
<td>15</td>
</tr>
<tr>
<td>Inferential Statistics</td>
<td>15</td>
</tr>
<tr>
<td>Basic principles of Inferential Statistics</td>
<td>17</td>
</tr>
<tr>
<td>Central Limit theorem</td>
<td>18</td>
</tr>
<tr>
<td>What are tests for significance what is</td>
<td>19</td>
</tr>
<tr>
<td>statistical significance (p-value)</td>
<td></td>
</tr>
<tr>
<td>Type and Type II error</td>
<td>23</td>
</tr>
<tr>
<td>Self Assessment Questions</td>
<td>26</td>
</tr>
</tbody>
</table>
1.1 INTRODUCTION

Quantitative research in social sciences aims to test theories about the nature of the world in general (or some part of it) based on samples of “subjects” taken from the world (or some part of it). When we perform research on the effect of TV violence on children’s aggression, our intent is to create theories that apply to all children who watch TV, or perhaps to all children in cultures similar to our own who watch TV. We of course cannot study all children, but we can perform research on samples of children that, hopefully, will generalize back to the populations from which the samples were taken. Inferential statistics is the mathematics and logic of how this generalization from sample to population can be made. The fundamental question is: can we infer the population’s characteristics from the sample’s characteristics? Inferential statistics focuses on making statements about the population.

This chapter discusses some of the basic concepts in inferential statistics which includes central limit theorem and test of significance and their importance to statistical research.

1.2 OBJECTIVES

After reading the unit you will be able to:

1. Describe inferential statistics and its importance in research
2. Discuss the basic principle of inferential statistics
3. Define central limit theorem
4. Highlight their role of test of significance in statistics

INFERENTIAL STATISTICS

Inferential statistics are statistics which are used to make inferential a population. These statistics rely on the use of a random
performance of 1000 students that completed an examination. Of these, 500 students are girls and 500 students are boys. The 1000 students represent our "population". Whilst we are interested in the performance of all 1000 students, girls and boys, it may be impractical to examine the marks of all of these students because of the time and cost required to collate all of their marks. Instead, we can choose to examine a "sample" of these students and then use the results to make generalizations about the performance of all 1000 students. For the purpose of our example, we may choose a sample size of 200 students. Since we are looking to compare boys and girls, we may randomly select 100 girls and 100 boys in our sample. We could then use this, for example, to see if there are any statistically significant differences in the mean mark between boys and girls, even though we have not measured all 1000 students.

In other words, inferential statistics are used to make inferences about a population from a sample in order to generalize (make assumptions about this wider population) and/or make predictions about the future.

Analyzing statistics can be challenging, yet inferential statistics are used every day to make sweeping generalizations about populations which may shape public policy and other issues. Researchers who work with inferential statistics try to keep their methods and practices transparent and as rigorous as possible to ensure the integrity of their results.

1.3.1 Basic Principle of Inferential Statistics

At a certain basic level, all inferential statistics procedures are the same: they seek to determine if the observed (sample) characteristics are sufficiently deviant from the null hypothesis to justify rejecting it.

The ingredients for making this calculation are the same for all statistical procedures:

1. The size of the observed difference(s)
2. The variability in the sample
3. The sample size. If instead we are looking for relationships (e.g., correlation analysis), the size of the relationship replaces the size of the
difference.

Therefore, researchers seek to:

1. Find large differences (or relationships)
2. Hold down unwanted variability
3. Obtain large samples.

(Good research must also worry about internal and external validity and the ultimate goal: building theories.)

1.4 CENTRAL LIMIT THEOREM

A very important and useful concept in statistics is the Central Limit Theorem. There are essentially three things we want to learn about any distribution: 1) The location of its center; 2) its width, 3) and how it is distributed. The central limit theorem helps us approximate all three.

The central limit theorem is one of the most remarkable results of the theory of probability. In its simplest form, the theorem states that the sum of a large number of independent observations from the same distribution has, under certain general conditions, an approximate normal distribution. Moreover, the approximation steadily improves as the number of observations increases. The theorem is considered the heart of probability theory, although a better name would be normal convergence theorem.

For example, suppose an ordinary coin is tossed 100 times and the number of heads is counted. This is equivalent to scoring 1 for a head and 0 for a tail and computing the total score. Thus, the total number of heads is the sum of 100 independent, identically distributed random variables. By the central limit theorem, the distribution of the total number of heads will be, to a very high degree of approximation, normal. This illustrated graphically by repeating this experiment many times. The results of this experiment are displayed in a diagram. The percentage computed over the number of experiments is arranged along the vertical axis, and the total score or the number of heads is arranged along the
horizontal axis. After a large number of repetitions a curve appears that looks like the normal curve.

It has been empirically observed that various natural phenomena, such as the heights of individuals, follow approximately a normal distribution. A suggested explanation is that these phenomena are sums of a large number of independent random effects and hence are approximately normally distributed by the central limit theorem.

Reading: For more clarification read the following:

Exploring Research: pp184-188

1.5 WHAT ARE TESTS FOR SIGNIFICANCE

Statistical significance is a mathematical tool used to determine whether the outcome of an experiment is the result of a relationship between specific factors or due to chance. Statistical significance is commonly used in the medical field to test drugs and vaccines and to determine causal factors of disease. Statistical significance is also used in the fields of psychology, environmental biology, and any other discipline that conducts research through experimentation.

Statistics are the mathematical calculations of numeric sets or populations that are manipulated to produce a probability of the occurrence of an event. Statistics use a numeric sample and apply that number to an entire population. For the sake of example, we might say that 80% of all Americans drive a car. It would be difficult to question every American about whether or not they drive a car, so a random number of people would be questioned and then the data would be statistically analyzed and generalized to account for everyone.

Statistical significance means that there is a good chance that we are right in finding that a relationship exists between two variables. But statistical significance is not the same as practical significance. We can have a statistically significant finding, but the implications of that finding may have no practical application. The researcher must always examine both the statistical and the practical significance of any research finding. For example, we may find that
there is a statistically significant relationship between a citizen's age and satisfaction with city recreation services. It may be that older citizens are 5% less satisfied than younger citizens with city recreation services. But is 5% a large enough difference to be concerned about?

Often times, when differences are small but statistically significant, it is due to a very large sample size; in a sample of a smaller size, the differences would not be enough to be statistically significant.

In a scientific study, a hypothesis is proposed, and then data is collected and analyzed. The statistical analysis of the data will produce a number that is statistically significant if it falls below 5%, which is called the confidence level. In other words, if the likelihood of an event is statistically significant, the researcher can be 95% confident that the result did not happen by chance.

Sometimes, when the statistical significance of an experiment is very important, such as the safety of a drug meant for humans, the statistical significance must fall below 3%. In this case, a researcher could be 97% sure that a particular drug is safe for human use. This number can be lowered or raised to accommodate the importance and desired certainty of the result being correct.

Statistical significance is used to reject or accept what is called the null hypothesis. A hypothesis is an explanation that a researcher is trying to prove. The null hypothesis holds that the factors a researcher is looking at have no effect on differences in the data. Statistical significance is usually written, for example, $t\approx .02, p<.05$. Here, "$t$" stands for the statistic test score and "$p<.05$" means that the probability of an event occurring by chance is less than 5%. These numbers would cause the null hypothesis to be rejected, therefore affirming that the alternative hypothesis is true.

Here is an example of a psychological hypothesis using statistical significance: It is hypothesized that baby girls smile more than baby boys. In order to test this hypothesis, a researcher would observe a certain number of baby girls and boys and count how many times they smile. At the end of the observation, the numbers of smiles would be statistically analyzed.

Two questions arise Tests for statistical significance is used to address the
question: what is the probability that what we think is a relationship between two variables is really just a chance occurrence?

There are two types of tools that are used to address these questions: the first is addressed by tests for statistical significance; and the second is addressed by Measures of Association.

If we selected many samples from the same population, would we still find the same relationship between these two variables in every sample? If we could do a census of the population, would we also find that this relationship exists in the population from which the sample was drawn? Or is our finding due only to random chance?

Tests for statistical significance tell us what the probability is that the relationship we think we have found is due only to random chance. They tell us what the probability is that we would be making an error if we assume that we have found that a relationship exists.

A statistical test is a procedure for deciding whether an assertion (e.g. an Hypothesis) about a quantitative feature of a population is true or false. We test a hypothesis of this sort by drawing a random sample from the population in question and calculating an appropriate statistic on its items. If, in doing so, we obtain a value of the statistic that would occur rarely when the hypothesis is true; we would have reason to reject the hypothesis.

With this procedure it is customary to reject the hypothesis tested when our statistic has a value that is among those that, theoretically, would be expected to occur no more than 5 out of every 100 times that a random sample (of the same size) is drawn from the population in question when the hypothesis is, in fact, true. Much of the text of this tutorial is devoted to explanations of exactly how this kind of theoretical expectation is developed.

Finally, it is noteworthy that the appropriate conduct of any statistical test invariably requires many thoughtful decisions. It is, for example, always necessary to decide what statistic to use, what sample size to employ and what criteria to establish for rejection of the hypothesis tested.

Reading: For more clarification read the following:

Exploring Research: pp, 188-200
1.5.1 What is "Statistical Significance" (p-value)?

The statistical significance of a result is the probability that the observed relationship (e.g., between variables) or a difference (e.g., between means) in a sample occurred by pure chance ("luck of the draw"), and that in the population from which the sample was drawn, no such relationship or differences exist. Using less technical terms, we could say that the statistical significance of a result tells us something about the degree to which the result is "true" (in the sense of being "representative of the population").

More technically, the value of the p-value represents a decreasing index of the reliability of a result (see Brownlee, 1960). The higher the p-value, the less we can believe that the observed relation between variables in the sample is a reliable indicator of the relation between the respective variables in the population. Specifically, the p-value represents the probability of error that is involved in accepting our observed result as valid, that is, as "representative of the population." For example, a p-value of .05 (i.e., 1/20) indicates that there is a 5% probability that the relation between the variables found in our sample is a "fluke." In other words, assuming that in the population there was no relation between those variables whatsoever, and we were repeating experiments such as ours one after another, we could expect that approximately in every 20 replications of the experiment there would be one in which the relation between the variables in question would be equal or stronger than in ours. (Note that this is not the same as saying that, given that there IS a relationship between the variables, we can expect to replicate the results 5% of the time or 95% of the time; when there is a relationship between the variables in the population, the probability of replicating the study and finding that relationship is related to the statistical power of the design. See also, Power Analysis). In many areas of research, the p-value of .05 is customarily treated as a "border-line acceptable" error level.

1.5.2 How to determine that a Result is "Really" Significant

There is no way to avoid arbitrariness in the final decision as to what level of significance will be treated as really "significant." That is, the selection of some level of significance, up to which the results will be rejected as invalid, is
arbitrary. In practice, the final decision usually depends on whether the outcome was predicted a priori or only found post hoc in the course of many analyses and comparisons performed on the data set, on the total amount of consistent supportive evidence in the entire data set, and on "traditions" existing in the particular area of research. Typically, in many sciences, results that yield $p \leq 0.05$ are considered borderline statistically significant, but remember that this level of significance still involves a pretty high probability of error (5%). Results that are significant at the $p \leq 0.01$ level are commonly considered statistically significant, and $p \leq 0.005$ or $p \leq 0.001$ levels are often called "highly" significant. But remember that these classifications represent nothing else but arbitrary conventions that are only informally based on general research experience.

1.5.3 Type I error and Type II error

Even in the best research project, there is always a possibility (hopefully a small one) that the researcher will make a mistake regarding the relationship between the two variables. There are two possible mistakes or errors.

The first is called a Type I error. This occurs when the researcher assumes that a relationship exists when in fact the evidence is that it does not. In a Type I error, the researcher should accept the null hypothesis and reject the research hypothesis, but the opposite occurs. The probability of committing a Type I error is called alpha.

The second is called a Type II error. This occurs when the researcher assumes that a relationship does not exist when in fact the evidence is that it does. In a Type II error, the researcher should reject the null hypothesis and accept the research hypothesis, but the opposite occurs. The probability of committing a Type II error is called beta.

Generally, reducing the possibility of committing a Type I error increases the possibility of committing a Type II error and vice versa, reducing the possibility of committing a Type II error increases the possibility of committing a Type I error. Researchers generally try to minimize Type I errors, because when a researcher assumes a relationship exists when one really does not, things may be
worse off than before. In Type II errors, the researcher misses an opportunity to confirm that a relationship exists, but is no worse off than before.

In this example, which type of error would you prefer to commit?

Research Hypothesis: El Nino has reduced crop yields in County X, making it eligible for government disaster relief.

Null Hypothesis: El Nino has not reduced crop yields in County X, making it ineligible for government disaster relief

If a Type I error is committed, then the County is assumed to be eligible for disaster relief, when it really is not (the null hypothesis should be accepted, but it is rejected). The government may be spending disaster relief funds when it should not, and taxes may be raised.

If a Type II error is committed, then the County is assumed to be ineligible for disaster relief, when it really is eligible (the null hypothesis should be accepted, but it is rejected). The government may not spend disaster relief funds when it should, and farmers may go into bankruptcy.

In this example, which type of error would you prefer to commit?

Research Hypothesis: The new drug is better at treating heart attacks than the old drug

Null Hypothesis: The new drug is no better at treating heart attacks than the old drug

If a Type I error is committed, then the new drug is assumed to be better when it really is not (the null hypothesis should be accepted, but it is rejected). People may be treated with the new drug, when they would have been better off with the old one.

If a Type II error is committed, then the new drug is assumed to be no better when it really is better (the null hypothesis should be rejected, but it is accepted). People may not be treated with the new drug, although they would be better off than with the old one. For more clarification please read this table:
Statistics - Hypothesis Test

Null Hypothesis True Null Hypothesis False

Reject Null Hypothesis

Fail to Reject Null Hypothesis

Reading: For more clarification read the following:

Exploring Research: pp, 192-200
1.6 SELF ASSESSMENT QUESTIONS

Q.1 Define inferential statistics. What does the term “statistically significant” mean?

Q.2 How does the central limit theorem works? And why is it so important to use of Inferential statistics?

Q.3 Explain why a researcher does not set out to prove a hypothesis.

Q.4 What is the difference between type I and type II error?

Q.5 What is a Meta-analysis and why is it important as a research tool?
Statistical Hypothesis Testing

Atifa Durrani
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>29</td>
</tr>
<tr>
<td>Objectives</td>
<td>29</td>
</tr>
<tr>
<td>Types of hypotheses</td>
<td>29</td>
</tr>
<tr>
<td>Hypothesis testing process</td>
<td>32</td>
</tr>
<tr>
<td>Test of Significance</td>
<td>36</td>
</tr>
<tr>
<td>Self Assessment Questions</td>
<td>46</td>
</tr>
</tbody>
</table>
2.1 INTRODUCTION

In this unit the basic concepts of hypothesis testing is presented. Besides this, process of hypotheses testing is also explained with simple examples. You are expected to know that meaning of the new terms introduced in the section. Test of significance along t-test, Z-test and Chi-Square Test are also explained in detail.

2.2 OBJECTIVES

After reading the unit you will be able to:

1. Distinguish between different types of hypothesis
2. Discuss the process of hypothesis testing
3. Compute T, Z and Chi-Square tests

2.3 TYPES OF HYPOTHESIS

There are two types of statistical inferences: estimation of population parameters and hypothesis testing. Hypothesis testing is one of the most important tools of application of statistics to real life problems. Most often, decisions are required to be made concerning populations on the basis of sample information. Statistical tests are used in arriving at these decisions.

There are five ingredients to any statistical test:

(a) Null Hypothesis
(b) Alternate Hypothesis
(c) Test Statistic
(d) Rejection/Critical Region
(e) Conclusion

In attempting to reach a decision, it is useful to make an educated guess or assumption about the population involved, such as the type of distribution.
2.3.1 Statistical Hypotheses

They are defined as assertion or conjecture about the parameter or parameters of a population, for example the mean or the variance of a normal population. They may also concern the type, nature or probability distribution of the population.

Statistical hypotheses are based on the concept of proof by contradiction. For example, say, we test the mean of a population to see if an experiment has caused an increase or decrease. We do this by proof of contradiction by formulating a null hypothesis.

2.3.2 Null Hypothesis

It is a hypothesis which states that there is no difference between the procedures and is denoted by $H_0$. For the above example the corresponding $H_0$ would be that there has been no increase or decrease in the mean. Always the null hypothesis is tested, i.e., we want to either accept or reject the null hypothesis because we have information only for the null hypothesis.

2.3.3. Alternative Hypothesis

It is a hypothesis which states that there is a difference between the procedures and is denoted by $H_A$.

2.4 HYPOTHESIS TESTING

Social science research, and by extension business research, uses a number of different approaches to study a variety of issues. This research may be a very informal, simple process or it may be a formal, somewhat sophisticated process. Regardless of the type of process, all research begins with a generalized idea in the form of a research question or a hypothesis. A research question usually is posed in the beginning of a research effort or in a specific area of study that has had little formal research. A research question may take the form of a basic question about some issue or phenomena or a question about the relationship between two or more variables. For example, a research question might be: "Do flexible work hours improve employee productivity?" Another question might be: "How do flexible hours influence employees' work?"
A hypothesis differs from a research question; it is more specific and makes a prediction. It is a tentative statement about the relationship between two or more variables. The major difference between a research question and a hypothesis is that a hypothesis predicts an experimental outcome. For example, a hypothesis might state: "There is a positive relationship between the availability of flexible work hours and employee productivity."

Hypotheses provide the following benefits:

1. They determine the focus and direction for a research effort.

2. Their development forces the researcher to clearly state the purpose of the research activity.

3. They determine what variables will not be considered in a study, as well as those that will be considered.

4. They require the researcher to have an operational definition of the variables of interest.

The worth of a hypothesis often depends on the researcher's skills. Since the hypothesis is the basis of a research study, it is necessary for the hypothesis be developed with a great deal of thought and contemplation. There are basic criteria to consider when developing a hypothesis, in order to ensure that it meets the needs of the study and the researcher. A good hypothesis should:

1. Have logical consistency. Based on the current research literature and knowledge base, does this hypothesis make sense?

2. Be in step with the current literature and/or provide a good basis for any differences. Though it does not have to support the current body of literature, it is necessary to provide a good rationale for stepping away from the mainstream.

3. Be testable. If one cannot design the means to conduct the research, the hypothesis means nothing.

4. Be stated in clear and simple terms in order to reduce confusion.
2.4.1 HYPOTHESIS TESTING PROCESS

Hypothesis testing is a systematic method used to evaluate data and aid the decision-making process. Following is a typical series of steps involved in hypothesis testing:

1. State the hypotheses of interest
2. Determine the appropriate test statistic
3. Specify the level of statistical significance
4. Determine the decision rule for rejecting or not rejecting the null hypothesis
5. Collect the data and perform the needed calculations
6. Decide to reject or not reject the null hypothesis

Each step in the process will be discussed in detail, and an example will follow the discussion of the steps.

2.4.1.1 Stating the hypotheses.

A research study includes at least two hypotheses—the null hypothesis and the alternative hypothesis. The hypothesis being tested is referred to as the null hypothesis and it is designated as $H_0$. It also is referred to as the hypothesis of no difference and should include a statement of equality ($=, \geq, \text{or} \leq$). The alternative hypothesis presents the alternative to the null and includes a statement of inequality ($\neq$). The null hypothesis and the alternative hypothesis are complementary.

The null hypothesis is the statement that is believed to be correct throughout the analysis, and it is the null hypothesis upon which the analysis is based. For example, the null hypothesis might state that the average age of entering college freshmen is 21 years:

$H_0$: The average age of entering college freshman $= 21$ years
If the data one collects and analyzes indicates that the average age of entering college freshmen is greater than or less than 21 years, the null hypothesis is rejected. In this case the alternative hypothesis could be stated in the following three ways: (1) the average age of entering college freshman is not 21 years (the average age of entering college freshmen ≠ 21); (2) the average age of entering college freshman is less than 21 years (the average age of entering college freshmen < 21); or (3) the average age of entering college freshman is greater than 21 years (the average age of entering college freshmen > 21 years).

The choice of which alternative hypothesis to use is generally determined by the study's objective. The preceding second and third examples of alternative hypotheses involve the use of a "one-tailed" statistical test. This is referred to as "one-tailed" because a direction (greater than [>] or less than Product B performance), or vice versa (Product A performance late shift employee production), rather than simply knowing that these employees have different levels of productivity (early shift employee production ≠ late shift employee production).

Both the alternative and the null hypotheses must be determined and stated prior to the collection of data. Before the alternative and null hypotheses can be formulated it is necessary to decide on the desired or expected conclusion of the research. Generally, the desired conclusion of the study is stated in the alternative hypothesis. This is true as long as the null hypothesis can include a statement of equality. For example, suppose that a researcher is interested in exploring the effects of amount of study time on tests scores. The researcher believes that students who study longer perform better on tests. Specifically, the research suggests that students who spend four hours studying for an exam will get a better score than those who study two hours. In this case the hypotheses might be:

**H₀** the average test scores of students who study 4 hours for the test = the average test scores of those who study 2 hours.

**H₁** the average test score of students who study 4 hours for the test < the average test scores of those who study 2 hours.

As a result of the statistical analysis, the null hypothesis can be rejected or not rejected. As a principle of rigorous scientific method, this subtle but important
rejecting a true hypothesis or failing to reject a false hypothesis.

Rejecting a null hypothesis that is true is called a Type I error and failing to reject a false null hypothesis is called a Type II error. The probability of committing a Type I error is termed $\alpha$ and the probability of committing a Type II error is termed $\beta$. As the value of $\alpha$ increases, the probability of committing a Type I error increases. As the value of $\beta$ increases, the probability of committing a Type II error increases. While one would like to decrease the probability of committing both types of errors, the reduction of $\alpha$ results in the increase of $\beta$ and vice versa. The best way to reduce the probability of decreasing both types of error is to increase sample size.

The probability of committing a Type I error, $\alpha$, is called the level of significance. Before data is collected one must specify a level of significance or the probability of committing a Type I error (rejecting a true null hypothesis). There is an inverse relationship between a researcher's desire to avoid making a Type I error and the selected value of $\alpha$; if not making the error is particularly important, a low probability of making the error is sought. The greater the desire is to not reject a true null hypothesis, the lower the selected value of $\alpha$. In theory, the value of $\alpha$ can be any value between 0 and 1. However, the most common values used in social science research are .05, .01, and .001, which respectively correspond to the levels of 95 percent, 99 percent, and 99.9 percent likelihood that a Type I error is not being made. The tradeoff for choosing a higher level of certainty (significance) is that it will take much stronger statistical evidence to ever reject the null hypothesis.

2.4.1.4 Determining the decision rule.

Before data are collected and analyzed it is necessary to determine under what circumstances the null hypothesis will be rejected or fail to be rejected. The decision rule can be stated in terms of the computed test statistic, or in probabilistic terms. The same decision will be reached regardless of which method is chosen.

2.4.1.5 Collecting the data and performing the calculations.

The method of data collection is determined early in the research process.
Once a research question is determined, one must make decisions regarding what type of data is needed and how the data will be collected. This decision establishes the bases for how the data will be analyzed. One should use only approved research methods for collecting and analyzing data.

2.4.1.6 Deciding whether to reject the null hypothesis.

This step involves the application of the decision rule. The decision rule allows one to reject or fail to reject the null hypothesis. If one rejects the null hypothesis, the alternative hypothesis can be accepted. However, as discussed earlier, if one fails to reject the null he or she can only suggest that the null may be true.

Reading: for more information read the following:

| Exploring Research  | p191-198 |

2.4.1.7 P-Value of the Test

The p-value is the probability of observing a value at least as extreme as the value of the test statistic given that the null hypothesis is true. We can compute the p-value of a test manually only when the sampling distribution is normal. For all other tests we need the computer to produce the p-value.

2.5 TESTS OF SIGNIFICANCE

As you have studies this concept in unit 1 it is being further clarified here.

Two questions arise about any hypothesized relationship between two variables:

1) What is the probability that the relationship exists

2) If it does, how strong is the relationship?

There are two types of tools that are used to address these questions: the first is addressed by tests for statistical significance; and the second is addressed by Measures of Association.

Tests for statistical significance are used to address the question: what is the probability that what we think is a relationship between two variables is really just a chance occurrence?
If we selected many samples from the same population, would we still find the same relationship between these two variables in every sample? If we could do a census of the population, would we also find that this relationship exists in the population from which the sample was drawn? Or is our finding due only to random chance?

Tests for statistical significance tell us what the probability is that the relationship we think we have found is due only to random chance. They tell us what the probability is that we would be making an error if we assume that we have found that a relationship exists.

We can never be completely 100% certain that a relationship exists between two variables. There are too many sources of error to be controlled, for example, sampling error, researcher bias, problems with reliability and validity, simple mistakes, etc.

But using probability theory and the normal curve, we can estimate the probability of being wrong, if we assume that our finding a relationship is true. If the probability of being wrong is small, then we say that our observation of the relationship is a statistically significant finding.

Statistical significance means that there is a good chance that we are right in finding that a relationship exists between two variables. But statistical significance is not the same as practical significance. We can have a statistically significant finding, but the implications of that finding may have no practical application. The researcher must always examine both the statistical and the practical significance of any research finding.

For example, we may find that there is a statistically significant relationship between a citizen's age and satisfaction with city recreation services. It may be that older citizens are 5% less satisfied than younger citizens with city recreation services. But is 5% a large enough difference to be concerned about?

Often times, when differences are small but statistically significant, it is due to a very large sample size; in a sample of a smaller size, the differences would not be enough to be statistically significant.
2.5.1 Steps in Testing for Statistical Significance

1) State the Research Hypothesis
2) State the Null Hypothesis
3) Select a probability of error level (alpha level)
4) Select and compute the test for statistical significance
5) Interpret the results

Reading: For more information read the following:

Exploring Research p 191-198

2.5.1 T-Test

The t-test is probably the most commonly used Statistical Data Analysis procedure for hypothesis testing. The t-test was described by 1908 by William Sealy Gosset for monitoring the brewing at Guinness in Dublin. Guinness considered the use of statistics a trade secret, so he published his test under the pen-name 'Student' -- hence the test is now often called the 'Student's t-test'.

A t-test is any statistical hypothesis test in which the test statistic follows a Student's $t$ distribution if the null hypothesis is true. It is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known. When the scaling term is unknown and is replaced by an estimate based on the data, the test statistic (under certain conditions) follows a Student's $t$ distribution. Actually, there are several kinds of t-tests, but the most common is the "two-sample t-test" also known as the "Student's t-test" or the "independent samples t-test". The two sample t-test simply tests whether or not two independent populations have different mean values on some measure.

The t-test (or student's t-test) gives an indication of the separateness of two sets of measurements, and is thus used to check whether two sets of measures are essentially different (and usually that an experimental effect has
been demonstrated). The typical way of doing this is with the null hypothesis that means of the two sets of measures are equal.

The t-test assumes:

- A normal distribution (parametric data)
- Underlying variances are equal (if not, use Welch's test)

It is used when there is random assignment and only two sets of measurement to compare.

There are two main types of t-test:

- Independent-measures t-test: when samples are not matched.

A single-sample t-test compares a sample against a known figure, for example where measures of a manufactured item are compared against the required standard.

Calculation

The value of t may be calculated using packages such as SPSS. The actual calculation for two groups is:

\[ t = \frac{\text{experimental effect}}{\text{variability}} \]
\[ = \frac{\text{difference between group means}}{\text{standard error of difference between group means}} \]

The t-test is a basic test that is limited to two groups. For multiple groups, you would have to compare each pair of groups, for example with three groups there would be three tests (AB, AC, BC), whilst with seven groups there would need to be 21 tests.

The basic principle is to test the null hypothesis that the means of the two groups are equal.

A significant problem with this is that we typically accept significance with each t-test of 95% (p=0.05). For multiple tests these accumulate and hence reduce the validity of the results.
2.5.2 Z-Test

The Z-test compares sample and population means to determine if there is a significant difference. It requires a simple random sample from a population with a Normal distribution and where the mean is known - test is a statistical test where normal distribution is applied and is basically used for dealing with problems relating to large samples when \( n \geq 30 \). \( n \) = sample size

For example suppose a person wants to test if both tea & coffee are equally popular in a particular town. Then he can take a sample of size say 500 from the town out of which suppose 280 are tea drinkers. To test the hypothesis, he can use Z-test. Thus \( z \) is a measure of how far away a measurement is from the mean, measured in standard deviations.

Calculation:
\[
 z = \frac{X - \bar{X}}{S}
\]

Where \( X \) is a measured value, \( \bar{X} \)-bar is the mean of all measured values and \( S \) is the standard deviation of all measured values.

Example

Saima gets a mark of 64 in a physics test, where the mean is 50 and the standard deviation is 8. Raazia gets a mark of 74 in a chemistry test, where the mean is 58 and the standard deviation is 10.

Saima’s \( z = \frac{(64 - 50)}{8} = 1.75 \)

Raazia’s \( z = \frac{(74 - 58)}{10} = 1.6 \)

Although Raazia’s score is higher, Saima’s score is further above the mean, and it might be concluded that Saima has achieved greater success.

The \( z \)-score provides a simple measure by which different measures can be compared in terms of their deviation from the mean. This is often called standardization. The \( z \)-score in use generally assumes parametric data.

2.5.3 Chi Square Test

The chi-square \( (\chi^2) \) test measures the alignment between two sets of frequency measures. These must be categorical counts and not percentages or ratios measures (for these, use another correlation test).
Note that the frequency numbers should be significant and be at least above 5 (although an occasional lower figure may be possible, as long as they are not a part of a pattern of low figures).

Goodness of fit

A common use is to assess whether a measured/observed set of measures follows an expected pattern. The expected frequency may be determined from prior knowledge (such as a previous year's exam results) or by calculation of an average from the given data. The null hypothesis, \( H_0 \) is that the two sets of measures are not significantly different.

Independence

The chi-square test can be used in the reverse manner to goodness of fit. If the two sets of measures are compared, then just as you can show they align, you can also determine if they do not align. The null hypothesis here is that the two sets of measures are similar.

The main difference in goodness-of-fit vs. independence assessments is in the use of the Chi Square table. For goodness of fit, attention is on 0.05, 0.01 or 0.001 figures. For independence, it is on 0.95 or 0.99 figures (this is why the table has two ends to it).

Calculation

Chi-squared, \( c^2 = \text{SUM}( (\text{observed} - \text{expected})^2 / \text{expected}) \)

\[ c^2 = \text{SUM}( (f_o - f_e)^2 / f_e ) \]

...where \( f_o \) is the observed frequency and \( f_e \) is the expected frequency.

Note that the expected values may need to be scaled to be comparable to the observed values. A simple test is that the total frequency/count should be the same for observed and expected values.

In a table, the expected frequency, if not known, may be estimated as:

\[ f_e = (\text{row total}) \times (\text{column total}) / n \]

...where \( n \) is the total of all rows (or columns).

The result is used with a Chi Square table to determine whether the comparison shows significance.
In a table, the degrees of freedom are:

\[ df = (R - 1) \times (C - 1) \]

...where \( R \) is the number of rows and \( C \) is the number of columns.

**Example**

Goodness of fit

English test grade distributions have changed from last year, with grade B's somewhat lower. Is this significant?

The table below shows the calculation. First, the expected values are created by scaling last year's results to be equivalent to this year. Then the test statistic is calculated as \( \text{SUM}((O - E)^2/E) \).

<table>
<thead>
<tr>
<th></th>
<th>Grade A</th>
<th>Grade B</th>
<th>Grade C</th>
<th>Grade D</th>
<th>Grade E</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>This year, ( O )</td>
<td>23</td>
<td>32</td>
<td>20</td>
<td>15</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Last year</td>
<td>25</td>
<td>20</td>
<td>15</td>
<td>25</td>
<td>10</td>
<td>95</td>
</tr>
<tr>
<td>Scaled last year, ( E )</td>
<td>26</td>
<td>21</td>
<td>16</td>
<td>26</td>
<td>11</td>
<td>100</td>
</tr>
<tr>
<td>( O - E )</td>
<td>-3.3</td>
<td>10.9</td>
<td>4.2</td>
<td>-11.3</td>
<td>-0.5</td>
<td></td>
</tr>
<tr>
<td>( (O - E)^2 )</td>
<td>11.0</td>
<td>119.8</td>
<td>17.7</td>
<td>128.0</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>( (O - E)^2/E )</td>
<td>0.4</td>
<td>5.7</td>
<td>1.1</td>
<td>4.9</td>
<td>0.0</td>
<td>12.1</td>
</tr>
</tbody>
</table>

Chi-square is found to be \( 12.1 \) and the degrees of freedom are \( (5-1) = 4 \) (there are five possible grades). Looking this up in the Chi Square table shows the probability is between 5% \( (9.49) \) and 1% \( (13.28) \), so \( H_0 \) is adequately falsified and a significant change can be claimed.
Independence

A year group in school chooses between Pak studies and Urdu as below. Is there any difference between boys' and girls' choices?

**Observed**

<table>
<thead>
<tr>
<th></th>
<th>Chose Pak. Studies</th>
<th>Chose Urdu</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys</td>
<td>43</td>
<td>55</td>
<td>98</td>
</tr>
<tr>
<td>Girls</td>
<td>52</td>
<td>54</td>
<td>106</td>
</tr>
<tr>
<td>Total</td>
<td>95</td>
<td>109</td>
<td>204</td>
</tr>
</tbody>
</table>

**Expected = (row tot * col tot)/overall tot**

<table>
<thead>
<tr>
<th></th>
<th>Chose Pak. Studies</th>
<th>Chose Urdu</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys</td>
<td>45.6</td>
<td>52.4</td>
<td>98</td>
</tr>
<tr>
<td>Girls</td>
<td>49.4</td>
<td>56.6</td>
<td>106</td>
</tr>
<tr>
<td>Total</td>
<td>95</td>
<td>109</td>
<td>204</td>
</tr>
</tbody>
</table>

\[(\text{observed} - \text{expected})^2/\text{expected}\]
Chi-square is 0.55. There are (2-1)*(2-1) = 1 degree of freedom. Checking the Chi Square table shows 0.55 is between 0.004 and 3.84, so no conclusion can be drawn about independence or similarity between boys' and girls' choices.

**Reporting**

Chi-square is reported in the following form:

\[ c^2 (3, N = 125) = 10.2, \ p = 0.012 \]

Where:

3 - the degrees of freedom

125 - subjects in the sample

10.2 - the \( c^2 \) test statistic

.012 - the probability of the null hypothesis being true
Discussion

This test compares observed data with what we would expect to get (if the null hypothesis of no difference was true). It is based on the principle that if the two variables are not related (for example gender is not related to deafness) then measures applied to each variable will give similar results (for example about the same proportion of men and women being found to use a hearing aid), with any variation between the results being purely caused by chance. If the experimental measures are significantly different, then some relationship can be claimed.

A reason that percentages do not work is because they are fractions and low numbers will not work. In practice, you can often get away with percentages by converting them into larger numbers.

The measurement is unusual in that it has a square on numerator and a non-square on the denominator. Squaring removes negatives and exaggerates outliers. This increases the effect that chi-square has in analyzing the difference between two data sets.

Note that the test only reports whether two sets of figures are similar. It says nothing about the nature of the similarity.

A chi-gram is a bar-chart plot of a set of chi-square calculations and can visually show how chi-square varies across a set of related measurements.

The Chi-square test is non-parametric.

Reading: for further information read the following:

| Exploring Research p 190-198 |

2.5 SELF ASSESSMENT QUESTIONS

Q.1 research study was conducted to examine the differences between older and younger adults on perceived life satisfaction. A pilot study was conducted to examine this hypothesis. Ten older adults (over the age of 70) and ten younger adults (between 20 and 30) were give a life satisfaction test (known to have high
Q.1 What is your computed answer?

Q.2 What would be the null hypothesis in this study?

Q.3 What would be the alternate hypothesis?

Q.4 What probability level did you choose and why?

Q.5 What is your T-test?

Q.6 Is there a significant difference between the two groups?

Q.7 Interpret your answer.

Q.8 If you have made an error, would it be a Type I or a Type II error? Explain your answer.

Q.2 Researchers want to examine the effect of perceived control on health complaints of geriatric patients in a long-term care facility. Thirty patients are randomly selected to participate in the study. Half are given a plant to care for and half are given a plant but the care is conducted by the staff. Number of health complaints is recorded for each patient over the following seven days. Compute the appropriate t-test for the data provided below.

<table>
<thead>
<tr>
<th>Control over Plant</th>
<th>No Control over Plant</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>35</td>
</tr>
<tr>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td>6</td>
<td>26</td>
</tr>
<tr>
<td>15</td>
<td>24</td>
</tr>
<tr>
<td>18</td>
<td>17</td>
</tr>
</tbody>
</table>

47
1. What is your computed answer?

2. What would be the null hypothesis in this study?

3. What would be the alternate hypothesis?
4. What probability level did you choose and why?

5. What are your degrees of freedom?

6. Is there a significant difference between the two groups?

7. Interpret your answer.

8. If you have made an error, would it be a Type I or a Type II error? Explain your answer.

Q.3 In hypothesis testing, _ is the probability of committing an error of Type II. The power of the test, 1 _ is then:

(a) the probability of rejecting H0 when HA is true
(b) the probability of failing to reject H0 when HA is true
(c) the probability of failing to reject H0 when H0 is true
(d) the probability of rejecting H0 when H0 is true
(e) the probability of failing to reject H0.

Q.4 In a statistical test of hypothesis, what happens to the rejection region when _ the level of significance, is reduced?

Q.5 Which of the following is not correct?

(a) The probability of a Type I error is controlled by the selection of the level.
(b) The probability of a Type II error is controlled by the sample size.
(c) The power of a test depends upon the sample size and the distance between the null and alternate hypothesis.
(d) The p-value measures the probability that the null hypothesis is true.
(e) The rejection region is controlled by the _ level and the alternate hypothesis.

Q.6 In testing statistical hypotheses, which of the following statements is false?

(a) The critical region is the values of the test statistic for which we reject the null
hypothesis.

(b) The level of significance is the probability of type I error.

(c) For testing $H_0: \mu = \mu_0$, $HA: \mu > \mu_0$, we reject $H_0$ for high values of the sample mean $X$.

(d) In testing $H_0: \mu = \mu_0$, $HA: \mu \neq \mu_0$, the critical region is two sided.

(e) The p-value measures the probability that the null hypothesis is true.

Q.7 In a hypothesis testing problem:

(a) the null hypothesis will not be rejected unless the data are not unusual (given that the hypothesis is true).

(b) the null hypothesis will not be rejected unless the p-value indicates the data are very unusual (given that the hypothesis is true)

(c) the null hypothesis will not be rejected only if the probability of observing the data provide convincing evidence that it is true

(d) the null hypothesis is also called the research hypothesis; the alternative hypothesis often represents the status quo.

(e) the null hypothesis is the hypothesis that we would like to prove; the alternative hypothesis is also called the research hypothesis.

Q.8 Discuss the following:

I. Chi-square Test,

II. T-Test

III. Z Test

Q.8 What is difference between Type I and type II error?
Descriptive Methods of Data Analysis
<table>
<thead>
<tr>
<th>Contents</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>53</td>
</tr>
<tr>
<td>Objectives</td>
<td>53</td>
</tr>
<tr>
<td>Descriptive statistics</td>
<td>53</td>
</tr>
<tr>
<td>Measures of variability</td>
<td>58</td>
</tr>
<tr>
<td>Parametric and Non Parametric Tests</td>
<td>67</td>
</tr>
<tr>
<td>Analysis of variance</td>
<td>69</td>
</tr>
<tr>
<td>Regression</td>
<td>70</td>
</tr>
<tr>
<td>Self Assessment Questions</td>
<td>71</td>
</tr>
</tbody>
</table>
3.1 INTRODUCTION

Whatever type of research we undertake (for example, naturalistic observation, case study, surveys or experiments) our efforts can usually generate a considerable amount of data: numbers that represent our research findings and provide the basis for our conclusions. Statistical analyses are the methods most often used to summaries and interpret data. In this unit effort has made to introduce a good number of statistical data analysis methods which are commonly used in descriptive data analyses.

3.2 OBJECTIVES

After reading this unit you will be able to:

1. Define descriptive statistics
2. Perform single variable analysis by using important characteristic of descriptive statistics
3. Discuss mean, median and mode and solve the problems
4. Define index of variability problems related to central tendency

3.3 DESCRIPTIVE STATISTICS

Descriptive statistics are numbers that are used to summarize and describe data. The word data refers to the information that has been collected from an experiment, a survey, an historical record, etc. (By the way, "data" is plural. One piece of information is called a "datum.") If we are analyzing birth certificates, for example, a descriptive statistic might be the percentage of certificates issued in Karachi or the average age of the mother. Any other number we choose to compute also counts as a descriptive statistic for the data from which the statistic is computed. Several descriptive statistics are often used at one time, to give a full picture of the data.
Descriptive statistics are just descriptive. They do not involve generalizing beyond the data at hand. Generalizing from our data to another set of cases is the business of inferential statistics which you'll be studying soon. Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data.

Descriptive Statistics are used to present quantitative descriptions in a manageable form. In a research study we may have lots of measures. Or we may measure a large number of people on any measure. Descriptive statistics help us to simply large amounts of data in a sensible way. Each descriptive statistic reduces lots of data into a simpler summary. For instance, consider a simple number used to summarize how well a batter is performing in baseball, the batting average. This single number is simply the number of hits divided by the number of times at bat (reported to three significant digits). A batter who is hitting 333 is getting a hit one time in every three at bats. One batting .250 is hitting one time in four. The single number describes a large number of discrete events. Or, consider the scourge of many students, the Grade Point Average (GPA). This single number describes the general performance of a student across a potentially wide range of course experiences.

Every time you try to describe a large set of observations with a single indicator you run the risk of distorting the original data or losing important detail. The batting average doesn't tell you whether the batter is hitting home runs or singles. It doesn't tell whether she's been in a slump or on a streak. The GPA doesn't tell you whether the student was in difficult courses or easy ones, or whether they were courses in their major field or in other disciplines. Even given these limitations, descriptive statistics provide a powerful summary that may enable comparisons across people or other units.

Univariate analysis involves the examination across cases of one variable at a time. There are three major characteristics of a single variable that we tend to look at:

- Distribution
- Central tendency
Dispersion
In most situations, we would describe all three of these characteristics for each of the variables in our study.

3.3.1 The Distribution

The distribution is a summary of the frequency of individual values or ranges of values for a variable. The simplest distribution would list every value of a variable and the number of persons who had each value. For instance, a typical way to describe the distribution of college students is by year in college, listing the number or percent of students at each of the four years. Or, we describe gender by listing the number or percent of males and females. In these cases, the variable has few enough values that we can list each one and summarize how many sample cases had the value. But what do we do for a variable like income or GPA? With these variables there can be a large number of possible values, with relatively few people having each one. In this case, we group the raw scores into categories according to ranges of values. For instance, we might look at GPA according to the letter grade ranges. Or, we might group income into four or five ranges of income values.

<table>
<thead>
<tr>
<th>Category</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under 35</td>
<td>9%</td>
</tr>
<tr>
<td>36-45</td>
<td>21</td>
</tr>
<tr>
<td>46-55</td>
<td>45</td>
</tr>
<tr>
<td>56-65</td>
<td>19</td>
</tr>
<tr>
<td>66+</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 1. Frequency distribution table.

One of the most common ways to describe a single variable is with a frequency distribution. Depending on the particular variable, all of the data values may be represented, or you may group the values into categories first (e.g., with age, price, or temperature variables, it would usually not be sensible to determine the frequencies for each value. Rather, the values are grouped into ranges and the
frequencies determined.). Frequency distributions can be depicted in two ways, as a table or as a graph. Table 1 shows an age frequency distribution with five categories of age ranges defined. The same frequency distribution can be depicted in a graph as shown in Figure 2. This type of graph is often referred to as a histogram or bar chart.

![Frequency distribution bar chart](image)

**Table 2. Frequency distribution bar chart.**

Distributions may also be displayed using percentages. For example, you could use percentages to describe the:

- percentage of people in different income levels
- percentage of people in different age ranges
- percentage of people in different ranges of standardized test scores

### 3.3.2 Central Tendency

The central tendency of a distribution is an estimate of the "center" of a distribution of values. There are three major types of estimates of central tendency:

- Mean
- Median
- Mode

The **Mean** or average is probably the most commonly used method of
describing central tendency. To compute the mean all you do is add up all the values and divide by the number of values. For example, the mean or average quiz score is determined by summing all the scores and dividing by the number of students taking the exam. For example, consider the test score values:
15, 20, 21, 20, 36, 15, 25, 15
The sum of these 8 values is 167, so the mean is $\frac{167}{8} = 20.875$.

The **Median** is the score found at the exact middle of the set of values. One way to compute the median is to list all scores in numerical order, and then locate the score in the center of the sample. For example, if there are 500 scores in the list, score #250 would be the median. If we order the 8 scores shown above, we would get:

$$15,15,15,20,20,21,25,36$$

There are 8 scores and score #4 and #5 represent the halfway point. Since both of these scores are 20, the median is 20. If the two middle scores had different values, you would have to interpolate to determine the median.

The **Mode** is the most frequently occurring value in the set of scores. To determine the mode, you might again order the scores as shown above, and then count each one. The most frequently occurring value is the mode. In our example, the value 15 occurs three times and is the model. In some distributions there is more than one modal value. For instance, in a bimodal distribution there are two values that occur most frequently. Notice that for the same set of 8 scores we got three different values -- 20.875, 20, and 15 -- for the mean, median and mode respectively. If the distribution is truly normal (i.e., bell-shaped), the mean, median and mode are all equal to each other.

**3.3.3 Dispersion**

Dispersion refers to the spread of the values around the central tendency. There are two common measures of dispersion, the range and the standard deviation. The **range** is simply the highest value minus the lowest value. In our example distribution, the high value is 36 and the low is 15, so the range is $36 - 15 = 21$.

Reading: Read the following material for information:

Exploring Research pp, 166-170
3.4 MEASURES OF VARIABILITY

Variability is the tendency of the measurement process to produce slightly different measurements on the same test item, where conditions of measurement are either stable or vary over time, temperature, operators, etc.

In other words, Variability is about how measures of individuals within a set differ from one another. Measures of sets representing these differences are as important for statistics and understanding sets as are the measures of central tendency. The use of a measure of variability in conjunction with one of central tendency gives the researcher and the reader important information for understanding the properties of a population and for comparing different populations. We will identify four different measures of variability. The most frequently used measures of variability are the Standard Deviation and Range.

3.4.1 Index of variability

There are three major indexes of variability

1. Standard Deviation
2. Range
3. Variance

3.4.1 Standard Deviation

The measure of precision is a standard deviation. Good precision implies a small standard deviation. This standard deviation is called the short-term standard deviation of the process or the repeatability standard deviation. The Standard Deviation shows the relation that set of scores has to the mean of the sample. Again let’s take the set of scores:

\[15, 20, 21, 20, 36, 15, 25, 15\]

To compute the standard deviation, we first find the distance between each value and the mean. We know from above that the mean is 20.875. So, the
differences from the mean are:

<table>
<thead>
<tr>
<th></th>
<th>-</th>
<th>20.875</th>
<th>=</th>
<th>-5.875</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td></td>
<td>20.875</td>
<td>=</td>
<td>-0.875</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>20.875</td>
<td>=</td>
<td>+0.125</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>20.875</td>
<td>=</td>
<td>-0.875</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>20.875</td>
<td>=</td>
<td>-0.875</td>
</tr>
<tr>
<td>36</td>
<td></td>
<td>20.875</td>
<td>=</td>
<td>15.125</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>20.875</td>
<td>=</td>
<td>-5.875</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>20.875</td>
<td>=</td>
<td>+4.125</td>
</tr>
</tbody>
</table>

15 - 20.875 = -5.875

Notice that values that are below the mean have negative discrepancies and values above it have positive ones. Next, we square each discrepancy:

\[-5.875 \times -5.875 = 34.515625\]
\[-0.875 \times -0.875 = 0.765625\]
\[+0.125 \times +0.125 = 0.015625\]
\[-0.875 \times -0.875 = 0.765625\]
\[15.125 \times 15.125 = 228.765625\]
\[-5.875 \times -5.875 = 34.515625\]
\[+4.125 \times +4.125 = 17.015625\]

\[-5.875 \times -5.875 = 34.515625\]

Now, we take these "squares" and sum them to get the Sum of Squares (SS) value. Here, the sum is 350.875. Next, we divide this sum by the number of scores minus 1. Here, the result is 350.875 / 7 = 50.125. This value is known as the variance. To get the standard deviation, we take the square root of the variance (remember that we squared the deviations earlier). This would be SQRT (50.125) = 7.079901129253.

Although this computation may seem convoluted, it's actually quite simple. To see this, consider the formula for the standard deviation:
\[ \sqrt{\frac{\sum(X - \bar{X})^2}{(n - 1)}} \]

where:
- \(X\) = each score
- \(\bar{X}\) = the mean or average
- \(n\) = the number of values
- \(\sum\) means we sum across the values

In the top part of the ratio, the numerator, we see that each score has the mean subtracted from it, the difference is squared, and the squares are summed. In the bottom part, we take the number of scores minus 1. The ratio is the variance and the square root is the standard deviation. In English, we can describe the standard deviation as:

*The square root of the sum of the squared deviations from the mean divided by the number of scores minus one*

Although we can calculate these univariate statistics by hand, it gets quite tedious when you have more than a few values and variables. Every statistics program is capable of calculating them easily for you. For instance, I put the eight scores into SPSS and got the following table as a result:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>8</td>
</tr>
<tr>
<td>Mean</td>
<td>20.8750</td>
</tr>
<tr>
<td>Median</td>
<td>20.0000</td>
</tr>
<tr>
<td>Mode</td>
<td>15.00</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>7.0799</td>
</tr>
<tr>
<td>Variance</td>
<td>50.1250</td>
</tr>
<tr>
<td>Range</td>
<td>21.00</td>
</tr>
</tbody>
</table>

Which confirms the calculations above.
The standard deviation allows us to reach some conclusions about specific scores in our distribution. Assuming that the distribution of scores is normal or bell-shaped (or close to it!), the following conclusions can be reached:

- approximately 68% of the scores in the sample fall within one standard deviation of the mean
- approximately 95% of the scores in the sample fall within two standard deviations of the mean
- approximately 99% of the scores in the sample fall within three standard deviations of the mean

For instance, since the mean in our example is 20.875 and the standard deviation is 7.0799, we can from the above statement estimate that approximately 95% of the scores will fall in the range of 20.875-(2*7.0799) to 20.875+(2*7.0799) or between 6.7152 and 35.0348. This kind of information is a critical stepping stone to enabling us to compare the performance of an individual on one variable with their performance on another, even when the variables are measured on entirely different scales.

Here is an example which will help you to understand. Example on Standard Deviation

A survey conducted by an automobile company showed the number of cars per household and the corresponding probabilities. Find the standard deviation.

<table>
<thead>
<tr>
<th>Number of cars X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability P(X)</td>
<td>0.32</td>
<td>0.51</td>
<td>0.12</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Choices:
A. 4.24
B. 0.63
C. 0.79
Correct Answer: C

Solution:

Step 1: Representing the data in the table and compute $X \cdot P(X)$ and $X^2 \cdot P(X)$

<table>
<thead>
<tr>
<th>X</th>
<th>P(X)</th>
<th>$X \cdot P(X)$</th>
<th>$X^2 \cdot P(X)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>2</td>
<td>0.51</td>
<td>1.02</td>
<td>2.04</td>
</tr>
<tr>
<td>3</td>
<td>0.12</td>
<td>0.36</td>
<td>1.08</td>
</tr>
<tr>
<td>4</td>
<td>0.05</td>
<td>0.20</td>
<td>0.80</td>
</tr>
</tbody>
</table>

\[
\Sigma X \cdot P(X) = 1.9 \quad \Sigma X^2 \cdot P(X) = 4.24
\]

Step 2: From the table, we get
\[
\mu = \sum X \cdot P(X) = 1.9
\]

Step 3: \[
\sum X^2 \cdot P(X) = 4.24
\]

Step 4: Variance \[
(\sigma^2) = \sum [X^2 \cdot P(X)] - \mu^2
\]

Step 5: Variance \[
(\sigma^2) = 4.24 - 3.61 = 0.63
\]

Step 6: Standard deviation \[
(\sigma) = \sqrt{0.63} = 0.79
\]

Step 7: So, the standard deviation is 0.79.

3.4.2 Variance

Variance also indicates a relationship between the mean of a distribution and the data points; it is determined by averaging the sum of the squared deviations. Squaring the differences instead of taking the absolute values allows for greater flexibility in calculating further algebraic manipulations of the data. In
The formula for variance is

$$\sigma^2 = \left( \frac{\sum (x_i - \overline{x})^2}{n} \right)$$

$$= \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + (x_3 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n}$$,

where \( x_1, x_2, x_3, \ldots, x_n \) are the values in the data set.

Example on Variance

Find the variance of the data set \( \{1, 2, 3, 4, 10\} \).

Choices:

A. 10
B. 9
C. 8
D. 7

Correct answers:

Solution:

Step 1: The mean of the data set \( \{1, 2, 3, 4, 10\} \) is

$$\overline{x} = \frac{1+2+3+4+10}{5} = 4.$$  

[Use the formula for mean.]

Step 2: The standard deviation of the data set is

$$\sigma = \sqrt{\frac{1}{5} (1-4)^2 + (2-4)^2 + (3-4)^2 + (4-4)^2 + (10-4)^2}$$

[Use the formula for mean.]

$$= \sqrt{\frac{50}{5}} = \sqrt{10}$$

Step 3:

Step 4: The variance of the data set is \( \sigma^2 = (\sqrt{10})^2 = 10 \). [Substitute \( \sigma = \sqrt{10} \).]

3.4.3 Range

The range is the distance between the lowest data point and the highest data point. Deviation scores are the distances between each data point and the mean. In other words range (R) is simply the highest score minus the lowest score.
+ 1. The "+1" is added because we want to count the lowest score in the range. To clarify this point, calculate the range for the scores of 10, 11, 12, 13, 14, and 15. The range is 15-10+1=5+1=6. If we simply subtracted 15-10, we would get 5, but there are six scores: Count them. The range gives a quick idea of how far the scores are spread out. For class A, as listed above, the range equals 100-40+1 or 61 while R for class B is 74-66+1 or 9. Class A has almost seven times the variability of Class B.

The range is the difference between the extreme values of a distribution.

In other words, it is the difference between the greatest and the least numbers in a set of data.

Range = Greatest – Least

The range is the simplest measure of variation and is the easiest to compute. Though range is not the optimal measure of variation, it is still widely used in simple descriptions of data.

Example 1

The following are Brian's Math test scores. Find the range of his test scores.

79, 63, 74, 81, 77, 67, 84

Solution:

Step 1: First arrange the test scores in order of their size.

63, 67, 74, 77, 79, 81, 84

Step 2: The lowest score is 63 and the highest score is 84.

So, the range of these test scores is 84 – 63 = 21 points.

Example 2

The heights in cm of ten students are:

157, 152, ___, 151, 160, 156, 155, 162, 158, 163

The maximum height is missing in the data set. Find the maximum height if the range of the data set is 14 cm.
Solution:

Step 1: First arrange the heights in order from least to greatest.
151, 152, 155, 156, 157, 158, 160, 162, 163,

Step 2: The maximum height is missing in the data set. We'll call it 'h'.

Step 3: Minimum height = 151 cm

Step 4: Range = Maximum height – Minimum height

Step 5: 14 = h – 151 [Substitute the values.]

Step 6: h = 14 + 151 [Solve the equation for 'h'.]

Step 7: h = 165

Step 8: So, the maximum height is 165 cm.


Reading: for more information read the following:

Exploring Research pp 170-178
3.4.4 Cross Tabulation

(Most of the statistical material related to cross-tabulation is covered under Chi-square.)

Cross-tabulation is about taking two variables and tabulating the results of one variable against the other variable. An example would be the cross-tabulation of course performance against mode of study:

<table>
<thead>
<tr>
<th></th>
<th>HD</th>
<th>D</th>
<th>C</th>
<th>P</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>FT - Internal</td>
<td>10</td>
<td>15</td>
<td>18</td>
<td>33</td>
<td>8</td>
</tr>
<tr>
<td>PT Internal</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>External</td>
<td>4</td>
<td>3</td>
<td>12</td>
<td>15</td>
<td>6</td>
</tr>
</tbody>
</table>

Each individual would have had a recorded mode of study (the rows of the table) and performance on the course (the columns of the table). For each individual, those pairs of values have been entered into the appropriate cell of the table.

3.4.4.1 What does cross-tabulation tell?

A cross-tabulation gives you a basic picture of how two variables inter-relate.

It helps you search for patterns of interaction. Obviously, if certain cells contain disproportionately large (or small) numbers of cases, then this suggests that there might be a pattern of interaction.

In the table above, the basic pattern is what you would expect as a teacher but, at a general level, it says that the bulk of students get a P rating independent of mode of study. What we normally do is to calculate the Chi-square statistic to see if this pattern has any substantial relevance.
3.5 PARAMETRIC AND NON PARAMETRIC TESTS

There are two types of test data and consequently different types of analysis.

3.5.1 Parametric tests

Parametric data has an underlying normal distribution which allows for more conclusions to be drawn as the shape can be mathematically described it is concerned about decision making method where the distribution of the sampling statistic is known

3.5.2 Non-Parametric

Non-Parametric: decision making method which does not require knowledge of the distribution of the sampling statistic. It is concerned with decision making method which does not require knowledge of the distribution of the sampling statistic

<table>
<thead>
<tr>
<th>Assumed distribution</th>
<th>Parametric</th>
<th>Non-parametric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td></td>
<td>Any</td>
</tr>
<tr>
<td>Assumed variance</td>
<td>Homogeneous</td>
<td>Any</td>
</tr>
<tr>
<td>Typical data</td>
<td>Ratio or Interval</td>
<td>Ordinal or Nominal</td>
</tr>
<tr>
<td>Data set relationships</td>
<td>Independent</td>
<td>Any</td>
</tr>
<tr>
<td>Usual central measure</td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>Benefits</td>
<td>Can draw more conclusions</td>
<td>Simplicity; Less affected by outliers</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choosing</td>
<td>Choosing parametric test</td>
<td>Choosing a non-parametric test</td>
</tr>
<tr>
<td>Correlation test</td>
<td>Pearson</td>
<td>Spearman</td>
</tr>
<tr>
<td>Independent measures, 2 groups</td>
<td>Independent-measures t-test</td>
<td>Mann-Whitney test</td>
</tr>
<tr>
<td>Independent measures, &gt;2 groups</td>
<td>One-way, independent-measures ANOVA</td>
<td>Kruskal-Wallis test</td>
</tr>
<tr>
<td>Repeated measures, 2 conditions</td>
<td>Matched-pair t-test</td>
<td>Wilcoxon test</td>
</tr>
<tr>
<td>Repeated measures, &gt;2 conditions</td>
<td>One-way, repeated measures ANOVA</td>
<td>Friedman's test</td>
</tr>
</tbody>
</table>

As the table shows, there are different tests for parametric and non-parametric data.
3.6 ANALYSIS OF VARIANCE (ANOVA)

ANOVA is a set of statistical methods used mainly to compare the means of two or more samples. Estimates of variance are the key intermediate statistics calculated, hence the reference to variance in the title ANOVA. The different types of ANOVA reflect the different experimental designs and situations for which they have been developed.

The reason for doing an ANOVA is to see if there is any difference between groups on some variable. For example, you might have data on student performance in non-assessed tutorial exercises as well as their final grading. You are interested in seeing if tutorial performance is related to final grade. ANOVA allows you to break up the group according to the grade and then see if performance is different across these grades. ANOVA is available for both parametric (score data) and non-parametric (ranking/ordering) data.

3.6.1 Types of ANOVA

There are generally two types of ANOVA

3.6.1.1 One-way between groups

This is the simplest version of ANOVA. One-way ANOVA tests differences in a single interval dependent variable among two, three, or more groups formed by the categories of a single categorical independent variable.

3.6.1.2 Two-way between groups

Two-way ANOVA analyzes one interval dependent in terms of the categories (groups) formed by two independents, one of which may be conceived as a control variable

3.6.1.3 Multivariate or on-way ANOVA

Multivariate or n-way ANOVA generalize n-way ANOVA deals with n independents. It should be noted that as the number of independents increases, the number of potential interactions proliferates. For example, the grades by tutorial analysis could be extended to see if overseas students performed differently to local students. What you would have from this form of ANOVA is: The effect of
final grade. The effect of overseas versus local. The interaction between final grade and overseas/local

Each of the main effects is one-way tests. The interaction effect is simply asking "is there any significant difference in performance when you take final grade and overseas/local acting together".

3.7 REGRESSION

Regression analysis is a statistical tool for the investigation of relationships between variables. Usually, the investigator seeks to ascertain the causal effect of one variable upon another—the effect of price increase upon demand, for example, or the effect of changes in the money supply upon the inflation rate. To explore such issues, the investigator assembles data on the underlying variables of interest and employs regression to estimate the quantitative effect of the causal variables upon the variable that they influence. The investigator also typically assesses the "statistical significance" of the estimated relationships, that is, the degree of confident A statistical measure that attempts to determine the strength of the relationship between one dependent variable (usually denoted by Y) and a series of other changing variables (known as independent variables).

Regression is often used to determine how many specific factors such as the price of a commodity, interest rates, particular industries or sectors influence the price movement of an asset.

The two basic types of regression are linear regression and multiple regression.

3.7.1 Linear Regression is a common technique linear regression uses one independent variable to explain and/or predict the outcome of Y, while multiple regression uses two or more independent variables to predict the outcome. It is used to determine the extent to which there is a linear relationship between a dependent variable and one or more independent variables. There are two types of linear regression, simple linear regression and multiple linear regression. In simple linear regression a single independent variable is used to predict the value of a dependent variable.

Linear Regression: \( Y = a + bX + u \)
Where:
Y = the variable that we are trying to predict
X = the variable that we are using to predict Y
a = the intercept
b = the slope
u = the regression residual.

3.7.2 Multiple Regression. In multiple regression the separate variables are differentiated by using subscripted numbers. Regression takes a group of random variables, thought to be predicting Y, and tries to find a mathematical relationship between them. This relationship is typically in the form of a straight line (linear regression) that best approximates all the individual data points.

Multiple Regression: \[ Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_iX_i + u \]

Where:
Y = the variable that we are trying to predict
X = the variable that we are using to predict Y
a = the intercept
b = the slope
u = the regression residual.

3.8 SELF ASSESSMENT QUESTIONS

Q.1 Ahmad has been playing golf on the weekends for the past three years. Recently, she started keeping track of her recorded scores. Her scores for June and July at her favorite 9-hole (par 36) golf course are provided below.

45
49
42
56
41
36
Q.2 Compute the mean and the median for the following group of score.

1 2 3 4 10

a) In this case why is the median a more useful average than the mean?

b) Why would not you use the mode as an average?

Q.3 What is the median age of the 7 children whose ages are listed below?

11, 10, 13, 11, 12, 16, 14

Q.4 Find the mean, median and mode for the following set of numbers:

a. 1, 2, 3, 4, 5, 6, 7, 8, 9

b. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

c. -500, 350, 475, -300, -500, 450, 425, -400

Q.5 A pilot logged the following flight hours, as a student-pilot, in order to obtain their

Private pilot’s license
Determine the mean, median and mode for the number of flight hours flown as a student pilot.

Q.6 Compute the Standard Deviation of the following data

0.56 s
0.52 s
0.59 s
0.57 s

Q.6 On an interview for a job, the interviewer tells you that the average annual income of the company’s 25 employees is $60,849. The actual annual incomes of the 25 employees are shown below. What are the mean, median, and mode of the incomes? Was the person telling you the truth?

$17,305, $478,320, $45,678, $18,980, $17,408,
$25,676, $28,906, $12,500, $24,540, $33,450,
$12,500, $33,855, $37,450, $20,432, $28,956,
$34,983, $36,540, $250,921, $36,853, $16,430,
$32,654, $98,213, $48,980, $94,024, $35,671
Q.7 Find the mean variance and standard deviation of the set of the given below numbers

4, 10, 8, 2
3, 15, 6, 9, 2
0, 1, 1, 2, 2, 2, 3, 3, 4
2, 2, 2, 2, 2
1, 2, 3, 4, 5, 6, 7
1, 1, 1, 5, 5, 5
49, 62, 40, 29, 32, 70
1.5, 0.4, 2.1, 0.7, 0.8
Data Collection and Analysis

Atifa Durrani
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>77</td>
</tr>
<tr>
<td>Objectives</td>
<td>77</td>
</tr>
<tr>
<td>Data collection</td>
<td>77</td>
</tr>
<tr>
<td>Data processing; coding typing and editing</td>
<td>89</td>
</tr>
<tr>
<td>Qualitative analysis; what is it?</td>
<td>90</td>
</tr>
<tr>
<td>Quantitative analysis what is it</td>
<td>91</td>
</tr>
<tr>
<td>Data presentation</td>
<td>92</td>
</tr>
<tr>
<td>Self Assessment Questions</td>
<td>100</td>
</tr>
</tbody>
</table>
4.1 INTRODUCTION

Data Collection is an important aspect of any type of research study. In-accurate data collection can impact the results of a study and ultimately lead to invalid results.

Data collection methods for impact evaluation vary along a continuum. At the one end of this continuum are quantitative methods and at the other end of the continuum are Qualitative methods for data collection. Quantitative data – numerical data based on pre-set categories decided upon by the evaluator that are applied in identical (or standardized) ways. These data are reported as numerical indicators. Some methods for collecting quantitative data: surveys/questionnaires, checklists.

4.2 OBJECTIVES

After reading the unit you will be able to:

1. Define data collection and analysis
2. Discuss different types of data
3. Explain essential steps in data collection
4. Explain data processing
5. Describe various forms of data presentation

4.3 DATA-COLLECTION

Data collection is a way of gathering information for use in various studies or decision making situations. Depending on the required outcome or information needed methods of data collection can vary and even be combined to achieve needed results. Raw data can take a variety of forms, including measurements, survey responses, and observations. In its raw form, this information can be incredibly useful, but also overwhelming. Over the course of the data analysis process, the raw data is ordered in a way which will be useful. For example, survey results may be tallied, so that people can see at a glance how many people answered the survey, and how people responded to specific questions.
4.3.1 Why Collect Data?

While some see data collection as a task that diverts time and resources from prevention activities, there are good reasons to collect, and use, data. Consider the following:

- Data can help you focus your prevention efforts. Data help you design, target, and implement prevention activities based on the specific substance use/abuse problem in your community. Prevention programs are more likely to be effective (and cost effective) when they do the following:

- Focus on the people most at risk of using drugs (defined by, for example, age, sex, race, income, and location)

- Respond to specific patterns of substance use/abuse (including the substances used and the age at which people begin using drugs and alcohol)

- Are culturally and socially appropriate to the target audience (that is, take into consideration their language, education, age, and attitudes toward substance use/abuse and related issues)

- Respond to both the risk factors associated with substance use/abuse and the protective factors shown to help people resist becoming involved with drugs or alcohol.

- Data can help mobilize and channel community support for prevention. The need for, and success of, substance use/abuse programs is not always self-evident. Drug and alcohol use/abuse or domestic violence is often
concealed from the public eye. A community and its leaders may not realize that a problem exists, overreact to a dramatic incident that is not representative of community problems, or question the efficiency of, and money spent on, prevention. Data that clarify and illuminate the nature and scope of substance use/abuse in a community help generate support and resources and, as importantly, ensure that these resources are used effectively.

- Data support requests for funding. Grant applications that include data are attractive to both public and private funders. Data show that a community has a real need for the proposed program and that the applicant offers a sophisticated approach to prevention. Programs that use evaluation data to prove they are making a difference are more likely to receive continuation funding than those that offer only anecdotal evidence about their success.

- Funders may require data. States, the federal government, and private funders may require prevention programs and other grantees or funded agencies to collect data. These data may be used to judge program effectiveness, advocate for continued legislative support of these programs, and/or contribute to the field of substance use/abuse prevention by providing consistent data across programs or jurisdictions.

4.3.2 Types of Useful Data

Data that can be used to help select, target, and evaluate prevention efforts include the following:

- Information about the types and extent of drug and alcohol use/abuse in your community. These data are useful for clarifying decisions about the types of prevention activities that are needed.
• Data about the characteristics of people in the community and their use patterns. These data are useful in identifying the most commonly used drugs across different age groups, gender, or race/ethnicities.

• Demographic data that describes your community and the different groups of people living in it. These data can aid in the selection of appropriate prevention programs and help you anticipate needs such as translation services. Demographic data also allow you to compare your community to similar communities, which provides an important perspective, since similar communities can be expected to have similar problems.

• Information about risk and protective factors. Research has revealed a number of risk factors that increase the chances of youth using/abusing substances. Evidence also shows that certain protective factors can help shield youth from problems. Programs need to know which risk and protective factors are present in their community. Risk factors could include parents who abuse drugs or alcohol, or an exaggerated perception among middle school students about the extent of drug and alcohol abuse among their peers. Protective factors might include parents who are closely involved in their children's activities or high attachment to school.

• Information about community context that may contribute to substance use/abuse. For example, certain neighborhoods may contain a high density of alcohol outlets or a large number of billboards that promote alcohol or cigarette use.

• Readiness data. Information about community knowledge and attitudes of the public, community opinion leaders, policymakers, and those with responsibility over the institutions in which prevention activities are
implemented (such as schools, public health clinics, and law enforcement agencies) can help move communities and systems toward prevention efforts. It can reveal which players are willing and able to participate in your prevention efforts, and where education and advocacy must be applied to bring critical participants on board.

4.3.4 Essential Steps in Data Collection

When we need to provide data to anyone, we often think of the data report without looking at the source of the data. Data flows like a stream. When we seek to provide or obtain data technical assistance, it helps to know what part of the stream needs attention. The following “steps” make up the flow of data and include collection, management, reports, and usage. These steps, and their relationships to each other, provide a way for you to assess the quality of your own data collection process, and to control data collection and storage, data quality, data reporting and usage.

One important area that you will not see in these steps is data security, privacy and confidentiality. This topic permeates data at almost every step. Certainly you must be mindful of it during Step 2, when seeking data from others. It is also important in Step 3, when you collect data. This is something you must address as a foundation for any data system you consider.

Step 1 identifying and defining data elements: what data do you need to collect?

Identify what you need to collect for reporting, what your program needs to collect for internal monitoring and planning, and why.

This includes:

- Identifying data elements necessary for reporting
- Identifying data elements necessary for program purposes including improving quality of care, and monitoring and evaluating to improve overall capacity to meet the needs of target populations
- Once identified, use provided definitions for all data elements that are
required for reporting, and communicate agreed upon definitions used for all other data elements, so that everyone understands what they need to collect.

**Step 2** Identify Data Sources: Where can you find the data you need?

- Where does the data you need reside, and in how many different places? Are there places that are easier to get the data from?
- What does it take to get the data when, where and how you need it? This includes: requesting, capturing, recording and storing data from these identified sources. It may also include development of inter-organizational agreements to ensure secure and appropriate access to data.

**Step 3** Data collection: how will you collect your data?

- Developing or modifying appropriate data collection tools and protocols to ensure you are collecting all targeted data as defined.
- Defining, communicating how data should be collected and submitted, and as necessary providing the tools for direct entry into a database or submission of previously collected data sources.
- Ensuring your data are appropriately stored to prevent inappropriate access, loss or theft through system breakdown.
- Implementing and communicating standards of confidentiality, privacy and security in order to protect your clients’ data.

**Step 4** Data Validation and Data Quality Procedures: How do you know the data you get is good and accurately reflects what you are trying to measure or report?
• Design and implement procedures to examine your data to ensure validity, reliability, completeness, timeliness, integrity and confidentiality. These procedures can include communication and training as well as system checks and routine data quality improvement activities.

**Step 5** Data Reporting: Who do you report it to, and how do you report the data you have?

• Follow procedures to correctly and efficiently prepare and submit your data to meet the data reporting requirements of thesis / research report or other funding sources.

**Step 6** Communicating about Data: How do you use the data you have to inform your program about how you are doing?

• How do you use the data you have to inform your program about how you are doing and where you need to go?

• Interpret and present data to inform an audience

• Use the data to inform planning, evaluation, allocations, or quality improvement.

**4.3.5 Methods of Data Collection**

There are four main methods of data collection.

**4.3.5.1 Census.** A census is a study that obtains data from every member of a population. In most studies, a census is not practical, because of the cost and/or time required.

**4.3.5.2 Sample survey.** A sample survey is a study that obtains data from a subset of a population, in order to estimate population attributes.

**4.3.5.3 Experiment.** An experiment is a controlled study in which the researcher attempts to understand cause-and-effect relationships. The study is "controlled" in the sense that the researcher controls (1) how subjects are assigned to groups and (2) which treatments each group receives. In the
analysis phase, the researcher compares group scores on some dependent variable. Based on the analysis, the researcher draws a conclusion about whether the treatment (independent variable) had a causal effect on the dependent variable.

4.3.5.4. Observational study. Like experiments, observational studies attempt to understand cause-and-effect relationships. However, unlike experiments, the researcher is not able to control (1) how subjects are assigned to groups and/or (2) which treatments each group receives.

4.3.6 Some preliminary considerations

The general principles of good research practice apply to data collection methods as to any other area of research. Before you embark on any survey or other data collection method, following important points may be kept in mind:

- Whether the data you seek are already available from an existing source
- What is the most appropriate method (e.g. questionnaire, interview, observation) or combination of methods?
- The practicalities of carrying out the data collection (how, when, where, by whom...?)
- How the data will be prepared for analysis (e.g. data entry procedures, coding).

4.3.7 Type of Data and Levels of Measurement

This is a key consideration. Information can be collected in statistics using qualitative or quantitative data. Qualitative data, such as eye color of a group of individuals, is not computable by arithmetic relations. They are labels that advise in which category or class an individual, object, or process fall. They are called categorical variables.

Quantitative data sets consist of measures that take numerical values for which descriptions such as means and standard deviations are meaningful. They can be put into an order and further divided into two groups: discrete data or continuous data. Discrete data are countable data, for example, the number of
defective items produced during a day's production. Continuous data, when the parameters (variables) are measurable, are expressed on a continuous scale. For example, measuring the height of a person.

The first activity in statistics is to measure or count. Measurement/counting theory is concerned with the connection between data and reality. A set of data is a representation (i.e., a model) of the reality based on numerical and measurable scales. Data are called "primary type" data if the analyst has been involved in collecting the data relevant to his/her investigation. Otherwise, it is called "secondary type" data.

Quantitative data are not just measures with numerical values (e.g., age, income) but any data which relate to the quantity of the measure concerned. Quantitative data can be analyzed in a variety of ways, using spreadsheet functions or specialist statistical packages.

Qualitative data are not amenable to automatic numerical analysis. Specialist packages are available to assist in analyzing such data; these are not considered here. There can be considerable value in qualitative data, and most surveys will include at least one opportunity for respondents to make open ended comments.

4.5 DATA ANALYSIS

Data analysis is a practice in which raw data is ordered and organized so that useful information can be extracted from it. The process of organizing and thinking about data is a key to understanding what the data does and does not contain. There are a variety of ways in which people can approach data analysis, and it is notoriously easy to manipulate data during the analysis phase to push certain conclusions or agendas. For this reason, it is important to pay attention when data analysis is presented, and to think critically about the data and the conclusions which were drawn.

4.5.1 Statistical Data Analysis

Data are not information! Vast amounts of statistical information are available in today's global and economic environment because of continual improvements in computer technology. To compete successfully globally,
decision makers/ researchers must be able to understand the information and use it effectively. Statistical data analysis provides hands on experience to promote the use of statistical thinking and techniques to apply in order to make educated decisions in the business world.

Computers play a very important role in statistical data analysis. The statistical software package, SPSS, which is used in this course, offers extensive data-handling capabilities and numerous statistical analysis routines that can analyze small to very large data statistics. The computer will assist in the summarization of data, but statistical data analysis focuses on the interpretation of the output to make inferences and predictions.

Studying a problem through the use of statistical data analysis usually involves four basic steps.

1. Defining the problem
2. Collecting the data
3. Analyzing the data
4. Reporting the results

4.5.1 Defining the Problem

An exact definition of the problem is imperative in order to obtain accurate data about it. It is extremely difficult to gather data without a clear definition of the problem.

4.5.2 Collecting the Data

We live and work at a time when data collection and statistical computations have become easy almost to the point of triviality. Paradoxically, the design of data collection, never sufficiently emphasized in the statistical data analysis textbook, have been weakened by an apparent belief that extensive computation can make up for any deficiencies in the design of data collection. One must start with an emphasis on the importance of defining the population about which we are seeking to make inferences; all the requirements of sampling and experimental design must be met. Designing ways to collect data is an
important job in statistical data analysis. Two important aspects of a statistical study are:

Population - a set of all the elements of interest in a study

Sample - a subset of the population

Statistical inference is referring to extending your knowledge obtains from a random sample from a population to the whole population. This is known in mathematics as an Inductive Reasoning. That is, knowledge of whole from a particular. Its main application is in hypotheses testing about a given population.

The purpose of statistical inference is to obtain information about population form information contained in a sample. It is just not feasible to test the entire population, so a sample is the only realistic way to obtain data because of the time and cost constraints. Data can be either quantitative or qualitative. Qualitative data are labels or names used to identify an attribute of each element. Quantitative data are always numeric and indicate either how much or how many. For the purpose of statistical data analysis, distinguishing between cross-sectional and time series data is important. Cross-sectional data re data collected at the same or approximately the same point in time. Time series data are data collected over several time periods.

4.5.3 Data

Data can be collected from existing sources or obtained through observation and experimental studies designed to obtain new data. In an experimental study, the variable of interest is identified. Then one or more factors in the study are controlled so that data can be obtained about how the factors influence the variables. In observational studies, no attempt is made to control or influence the variables of interest. A survey is perhaps the most common type of observational study.

4.5.4 Analyzing the Data

Statistical data analysis divides the methods for analyzing data into two categories: exploratory methods and confirmatory methods. Exploratory methods are used to discover what the data seems to be saying by using simple arithmetic
and easy-to-draw pictures to summarize data. Confirmatory methods use ideas from probability theory in the attempt to answer specific questions. Probability is important in decision making because it provides a mechanism for measuring, expressing, and analyzing the uncertainties associated with future events.

4.5.5 Reporting the Results

In the course of organizing the data, trends often emerge, and these trends can be highlighted in the write-up of the data to ensure that readers take note. In a casual survey of ice cream preferences, for example, more women than men might express a fondness for chocolate, and this could be a point of interest for the researcher. Modeling the data with the use of mathematics and other tools can sometimes exaggerate such points of interest in the data, making them easier for the researcher to see.

Charts, graphs, and textual write-ups of data are all forms of data analysis. These methods are designed to refine and distill the data so that readers can glean interesting information without needing to sort through all of the data on their own. Summarizing data is often critical to supporting arguments made with that data, as is presenting the data in a clear and understandable way. The raw data may also be included in the form of an appendix so that people can look up specifics for themselves.

When people encounter summarized data and conclusions, they should view them critically. Asking where the data is from is important, as is asking about the sampling method used to collect the data, and the size of the sample. If the source of the data appears to have a conflict of interest with the type of data being gathered, this can call the results into question. Likewise, data gathered from a small sample or a sample which is not truly random may be of questionable utility. Reputable researchers will always provide information about the data gathering techniques used, the source of funding, and the point of the data collection in the beginning of the analysis so that readers can think about this information while they review the analysis.

Through inferences, an estimate or test claims about the characteristics of a population can be obtained from a sample. The results may be reported in the form of a table, a graph or a set of percentages. Because only a small collection
(sample) has been examined and not an entire population, the reported results must reflect the uncertainty through the use of probability statements and intervals of values.

4.6 DATA PROCESSING: CODING, TYPING, AND EDITING

Data are often recorded manually on data sheets. Unless the numbers of observations and variables are small the data must be analyzed on a computer. The data will then go through three stages:

4.6.1 Coding: the data are transferred, if necessary to coded sheets. Coding is an interpretational and mechanical activity at the same time. No one but the researcher and her/his team members can interpret the data, i.e., decipher the meaning of particular data segments, identify what irrelevant about them, and decide what aspect of the research project's goal each is relevant to. Once this conceptual work has been done, it has to be recorded in the data: the boundaries of the relevant data segment have to be marked and the corresponding aspect or "category" has to be signified by means of a "code". The carving out of relevant passages of text, scenes of a video, parts of a drawing, etc. is commonly called "segmenting" in qualitative analysis, and the resulting data portions are called "segments". The entire activity is called "coding".

4.6.2 Typing The data are typed and stored by at least two independent data entry persons. For example, when the Current Population Survey and other monthly surveys were taken using paper questionnaires, the U.S. Census Bureau used double key data entry.

4.6.3 Editing The data are checked by comparing the two independent typed data. The standard practice for key-entering data from paper questionnaires is to key in all the data twice. Ideally, the second time should be done by a different key entry operator whose job specifically includes verifying mismatches between the original and second entries. It is believed that this "double-key/verification" method produces a 99.8% accuracy rate for total keystrokes. Types of error: Recording error, typing error, transcription error (incorrect copying), Inversion (e.g., 123.45 is typed as 123.54), Repetition (when a number is repeated), Deliberate error.
4.7 Qualitative Analysis - What It Is

This is the process of interpreting data collected during the course of qualitative research.

The blue words link to the relevant pages:

The analysis of the data depends on its type. Within this unit there is further guidance on the analysis of visual data and the analysis of narrative data.

4.7.1 Issues of Analysis

Establishing Trustworthiness

In qualitative research data must be auditable through checking that the interpretations are credible, transferable, dependable and confirmable.

Credibility improved through long engagement with the respondents or triangulation in data collection (internal validity)

Transferability achieved through a thick description of the research process to allow a reader to see if the results can be transferred to a different setting (external validity)

Dependability examined through the audit trail (reliability) e.g. member checking.

Confirmability audit trail categories used e.g. raw data included, data analysis and reduction processes described, data reconstruction and synthesis
including structuring of categories and themes, process notes included, instrument development information included. (Adapted from Lincoln and Guba, 1985)

Researchers usually generate data at a point in time and also write analytical notes to them about that data. These notes are then processed into memos or guiding notes to inform the next bout of data collection. And so leads the merry dance.

4.7.2 Quantitative Analysis - What It Is

This is the process of presenting and interpreting numerical data. Descriptive statistics include measures of central tendency (averages - mean, median and mode) and measures of variability about the average (range and standard deviation). Inferential statistics are the outcomes of statistical tests, helping deductions to be made from the data collected to test hypotheses set and relating findings to the sample or population.

4.7.2.1 Quantitative Analysis - Issues of Analysis

Null and alternative hypotheses. The formal statistical procedure for performing a hypothesis test is to state two hypotheses and to use an appropriate statistical test to reject one of the hypotheses and therefore accept (or fail to reject) the other.

The p-value (level of significance) All statistical tests produce a p-value and this is equal to the probability of obtaining the observed difference, or one more extreme, if the null hypothesis is true. To put it another way - if the null hypothesis is true, the p-value is the probability of obtaining a difference at least as large as that observed due to sampling variation.

Statistical power The use of a significance level of 5% controls the probability of erroneously rejecting the null hypothesis when it is, in fact, true. Rejecting the null hypothesis when it is true is called a Type I error. However, there is another error that can be made - failing to reject the null hypothesis when it is, in fact, not true. This is called a Type II error.

Validity and Reliability Validity and Reliability are the key characteristics in quantitative research that reflect quality and rigor in design. A well written research paper will indicate how validity and reliability have been assessed. (There are similar issues within qualitative research, as per Lincoln & Guba 1985)
Validity refers to the accuracy and truth of the data and findings that are produced. It refers to the concepts that are being investigated; the people or objects that are being studied; the methods by which data are collected; and the findings that are produced. There are several different types of validity which all contribute to the overall credibility of the research.

Reliability is concerned with the consistency and dependability of a measuring instrument, i.e. it is an indication of the degree to which it gives the same answers over time, across similar groups and irrespective of who administers it. A reliable measuring instrument will always give the same result on different occasions assuming that what is being measured has not changed during the intervening period.

4.8 DATA PRESENTATION

Most people show lack of interest or have no time to go through facts and figures given in a daily newspaper or a magazine. But if these figures are graphically presented, they become easier to grasp and catch the eye and have a more lasting effect on the reader’s mind. The graphical representation of data makes the reading more interesting, less time-consuming and easily understandable. The disadvantage of graphical presentation is that it lacks details and is less accurate.

The graphical representation of data makes the reading more interesting, less time-consuming and easily understandable. The disadvantage of graphical presentation is that it lacks details and is less accurate. Facts and figures as such do not catch our attention unless they are presented in an interesting way. Graphical representation of data is one of the most commonly used modes of presentation. The different types of graphs that we are going to study are

- Bar graphs
- Pie charts
- Frequency polygon
- Histogram
- Tables
6.8.1 Bar Graphs

This is the simplest type of graphical presentation of data. The following types of bar graphs are possible: (a) Simple bar graph (b) Double bar graph

Simple Bar Graph

<table>
<thead>
<tr>
<th>Class</th>
<th>IX-A</th>
<th>IX-B</th>
<th>IX-C</th>
<th>IX-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of students</td>
<td>22</td>
<td>26</td>
<td>30</td>
<td>40</td>
</tr>
</tbody>
</table>

Double Bar Graph

<table>
<thead>
<tr>
<th>Class</th>
<th>IX - A</th>
<th>IX - B</th>
<th>IX' - C</th>
<th>IX - D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girls</td>
<td>10</td>
<td>26</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>Boys</td>
<td>12</td>
<td>20</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>46</td>
<td>30</td>
<td>40</td>
</tr>
</tbody>
</table>
4.8.2 Pie Graph or Pie Chart

Sometimes a circle is used to represent a given data. The various parts of it are proportionally represented by sectors of the circle. Then the graph is called a Pie Graph or Pie Chart.

4.8.3 Frequency Polygon

In a frequency distribution, the mid-value of each class is obtained. Then on the graph paper, the frequency is plotted against the corresponding mid-value. These points are joined by straight lines. These straight lines may be extended in both directions to meet the X-axis to form a polygon. The marks scored by 120 students in an examination are as given in the table form a frequency polygon.
<table>
<thead>
<tr>
<th>Marks</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 10</td>
<td>2</td>
</tr>
<tr>
<td>10 - 20</td>
<td>8</td>
</tr>
<tr>
<td>20 - 30</td>
<td>10</td>
</tr>
<tr>
<td>30 - 40</td>
<td>15</td>
</tr>
<tr>
<td>40 - 50</td>
<td>24</td>
</tr>
<tr>
<td>50 - 60</td>
<td>36</td>
</tr>
<tr>
<td>60 - 70</td>
<td>14</td>
</tr>
<tr>
<td>70 - 80</td>
<td>6</td>
</tr>
<tr>
<td>80 - 90</td>
<td>5</td>
</tr>
</tbody>
</table>

![Graph showing the distribution of marks with frequency on the y-axis and marks on the x-axis.]
4.8.4 Histogram

A two dimensional frequency density diagram is called a histogram. A histogram is a diagram which represents the class interval and frequency in the form of a rectangle.

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 5</td>
<td>4</td>
</tr>
<tr>
<td>5 - 10</td>
<td>10</td>
</tr>
<tr>
<td>10 - 15</td>
<td>18</td>
</tr>
<tr>
<td>15 - 20</td>
<td>8</td>
</tr>
<tr>
<td>20 - 25</td>
<td>6</td>
</tr>
</tbody>
</table>

![Histogram Diagram]

4.8.5 Tables

4.8.5.1 One-Way Tables

When a table presents data for one, and only one, categorical variable, it is called a one-way table. A one-way table is the tabular equivalent of a bar chart. Like a bar chart, a one-way table displays categorical data in the form of frequency counts and/or relative frequencies.

- Frequency Tables
When a one-way table shows frequency counts for a particular category of a categorical variable, it is called a frequency table. Below, the bar chart and the frequency table display the same data. Both show frequency counts, representing travel choices of 10 travel agency clients.

<table>
<thead>
<tr>
<th>Choice</th>
<th>USA</th>
<th>Europe</th>
<th>Asia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

- **Relative Frequency Tables**

  When a one-way table shows relative frequencies for particular categories of a categorical variable, it is called a relative frequency table.

  Each of the tables below summarizes data from the bar chart above. Both tables are relative frequency tables. The table on the left shows relative frequencies as a proportion, and the table on the right shows relative frequencies as a percentage.

<table>
<thead>
<tr>
<th>Choice</th>
<th>USA</th>
<th>Europe</th>
<th>Asia</th>
<th>Proportion</th>
<th>0.5</th>
<th>0.3</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choice</td>
<td>USA</td>
<td>Europe</td>
<td>Asia</td>
<td>Percentage</td>
<td>50</td>
<td>30</td>
<td>20</td>
</tr>
</tbody>
</table>

### 4.8.5.2 Two-Way Tables

A common task in statistics is to look for a relationship between two variables. A two-way table (also called a contingency table) is a useful tool for examining relationships between categorical variables. The entries in the cells of a two-way table can be frequency counts or relative frequencies (just like a one-way table).
### Two-Way Frequency Tables

<table>
<thead>
<tr>
<th></th>
<th>Dance</th>
<th>Sports</th>
<th>TV</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>2</td>
<td>10</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td>Women</td>
<td>16</td>
<td>6</td>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>16</td>
<td>16</td>
<td>50</td>
</tr>
</tbody>
</table>

To the right, the two-way table shows the favorite leisure activities for 50 adults - 20 men and 30 women. Because entries in the table are frequency counts, the table is a frequency table.

Entries in the "Total" row and "Total" column are called marginal frequencies or the marginal distribution. Entries in the body of the table are called joint frequencies.

If we looked only at the marginal frequencies in the Total row, we might conclude that the three activities had roughly equal appeal. Yet, the joint frequencies show a strong preference for dance among women; and little interest in dance among men.

### Two-Way Relative Frequency Tables

<table>
<thead>
<tr>
<th></th>
<th>Dance</th>
<th>Sports</th>
<th>TV</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>0.04</td>
<td>0.20</td>
<td>0.16</td>
<td>0.40</td>
</tr>
<tr>
<td>Women</td>
<td>0.32</td>
<td>0.12</td>
<td>0.16</td>
<td>0.60</td>
</tr>
<tr>
<td>Total</td>
<td>0.36</td>
<td>0.32</td>
<td>0.32</td>
<td>1.00</td>
</tr>
</tbody>
</table>

- Relative Frequency of Table
We can also display relative frequencies in two-way tables. The table to the right shows preferences for leisure activities in the form of relative frequencies. The relative frequencies in the body of the table are called conditional frequencies or the conditional distribution.

Two-way tables can show relative frequencies for the whole table, for rows, or for columns. The table to the right shows relative frequencies for the whole table. Below, the table on the left shows relative frequencies for rows; and the table on the right shows relative frequencies for columns.

<table>
<thead>
<tr>
<th></th>
<th>Dance</th>
<th>Sports</th>
<th>TV</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>0.10</td>
<td>0.50</td>
<td>0.40</td>
<td>1.00</td>
</tr>
<tr>
<td>Women</td>
<td>0.53</td>
<td>0.20</td>
<td>0.27</td>
<td>1.00</td>
</tr>
<tr>
<td>Total</td>
<td>0.36</td>
<td>0.32</td>
<td>0.32</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Dance</th>
<th>Sports</th>
<th>TV</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>0.11</td>
<td>0.62</td>
<td>0.50</td>
<td>0.40</td>
</tr>
<tr>
<td>Women</td>
<td>0.89</td>
<td>0.38</td>
<td>0.50</td>
<td>0.60</td>
</tr>
<tr>
<td>Total</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Relative Frequency of Row

Relative Frequency of Column

Each type of relative frequency table makes a different contribution to understanding the relationship between gender and preferences for leisure activities. For example, "Relative Frequency for Rows" table most clearly shows the probability that each gender will prefer a particular leisure activity. For instance, it is easy to see that the probability that a man will prefer dance is 10%; the probability that a woman will prefer dance is 53%; the probability that a man will prefer sports is 50%; and so on.
Such relationships are often easier to detect when they are displayed graphically in a segmented bar chart. A segmented bar chart has one bar for each level of a categorical variable. Each bar is divided into "segments", such that the length of each segment indicates proportion or percentage of observations in a second variable.

The segmented bar chart on the right uses data from the "Relative Frequency for Rows" table above. It shows that women have a strong preference for dance; while men seldom make dance their first choice. Men are most likely to prefer sports, but the degree of preference for sports over TV is not great.

4.9 SELF ASSESSMENT QUESTIONS

Q.1 What is data collation and why data is collected? Discuss in detail.

Q.2 Explain the various types of data and essential steps in data collection.

Q.3 Describe data analysis. Discuss the steps which must be followed while statistical analysis of data.

Q.4 Defines quantitative and qualitative data analysis. What is the difference? Discuss the issues in both data analysis.

Q.5 Discuss the graphic data presentation and its different types.

4.10 BIBLIOGRAPHY


UNIT 5

SAMPLING

Memoona Fayaz
Atifa Durrani
<table>
<thead>
<tr>
<th>Contents</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>103</td>
</tr>
<tr>
<td>Objectives</td>
<td>103</td>
</tr>
<tr>
<td>Population</td>
<td>104</td>
</tr>
<tr>
<td>Sample</td>
<td>105</td>
</tr>
<tr>
<td>Sampling</td>
<td>107</td>
</tr>
<tr>
<td>Types of Sampling</td>
<td>114</td>
</tr>
<tr>
<td>Self Assessment Questions</td>
<td>126</td>
</tr>
<tr>
<td>Bibliography</td>
<td>128</td>
</tr>
</tbody>
</table>
5.1 INTRODUCTION

The quantum of the overall queries in the modern day has so increased that researcher natural scientists as well as social scientists cannot afford to examine every element of the population for each query. (Shah and Shah 1978).

The primary purpose of research is to discover principles that have universal applications. Most of the researches use a small proportion of a population for application of treatment, collection of data for analysis and drawing of conclusion. This is done for a number of practical considerations. Some populations may be so long that measurement of their characteristics would be too expensive and time consuming e.g. girls of primary schools in Pakistan. Collection of data might take so long that by the time the investigator has completed his study, the population itself might undergo a change. In that case the conclusions drawn from the study would not be true in respect of the changed population. The composition of the population might have been affected by increased rate of new entrants, and lowered rates of repeaters and drop (Bukhari, 1990).

In some cases it may seem feasible to obtain the information by taking a complete population, but in other cases, the process of sampling would make it possible to draw generalizations. As such sampling is used more frequently, effort has been made I this unit to highlight importance concept in sampling along with its different types and sampling techniques.

5.2 OBJECTIVES

After studying this unit, you should be able to:

1. Differentiate between population and sample.
2. Describe sampling and its importance I research process.
3. Identify different types of sampling.
4. Distinguish between different types of sampling technique.
5. Analyses the importance of sampling error and how it can be
5.3 POPULATION

According to Martin (1989, p, 125-126) the population or universe is the entire group of items, people or events that is being suited. It is the group about which the researcher wants to generalize from his study of the sample units.

Universe or population differs for each research problem depending upon the nature and type of information sought. In other words, population consists of the people who are related to specify problem under investigation. As defined by Selltiz and others (1965, p, 62) A population is the agate of all the cases that confirm to some designated set specification”.

The population might be all of the female residents of an area, or the entire particular group of females or residents. We often speak in terms of population characteristics. Examples of which are age, sex, income, place of residence, caste, occupation, density, migration, economic activity, and so on.

The definition of population depends on the research objectives established when the research problem is formulated. Thus if the research problems calls for information pertaining to the reading habits of a group or class of people then these people constitute the population (Smith, 1975).

5.3.1 Parameter

A parameter is a value, usually unknown (and which therefore has to be estimated), used to represent a certain population characteristic. For example, the population mean is a parameter that is often used to indicate the average value of a quantity.

Within a population, a parameter is a fixed value which does not vary. Each sample drawn from the population has its own value of any statistic that is used to estimate this parameter. For example, the mean of the data in a sample is used to give information about the overall mean in the population from which that sample was drawn.

Parameters are often assigned Greek letters (e.g. \( \theta \)), whereas statistics are assigned Roman letters (e.g. \( s \)).
5.3.2 Target Population

The target population is the entire group a researcher is interested in; the group about which the researcher wishes to draw conclusions.

Suppose we take a group of men aged 35-40 who have suffered an initial heart attack. The purpose of this study could be to compare the effectiveness of two drug regimes for delaying or preventing further attacks. The target population here would be all men meeting the same general conditions as those actually included in the study.

5.3.3 Types of Population

According to the number of characteristics in which we are interested at a time or which we interest to measure at a particular time. Thakur (1993) divides the population; into three types:

1. Univariate population
2. Bivariate population
3. Multivariate population

1. Univariate population is the one in which one considers only on characteristics at time.

2. When we are measuring two characteristics simultaneously for each member it will be Bivariate population.

3. Multivariate population is the one in which we consider observations on three or more characteristic simultaneously.

5.4 SAMPLE

A simple definition of sample is “A portion of the total number of cases having a given characteristics”. (Shah & Shah 1978, p20). According to Bukharis (1990-p: 52) “A sample is a sub-self objects/things/units taken from their population (complete set) for observation and study”.Thakur (1993 P.71-72) defines the samples as: “By sample we mean the aggregate of objects, persons or elements selected from the universe”.

105
According to Matin (1989P:121) "A sample is a representation of the population and if drawn on random basis can provide accurate data about universe".

The extent to which the generalization is applicable depends on the method used to select the sample and the size of the sample.

Majority of research studies today collect data through samples. To keep the cost, time, and quality of research under control, reliable sampling designs are prepared. After knowing different definitions of a sample make your own general definition of a sample.

5.4.1 Sample Variance

Sample variance is a measure of the spread of or dispersion within a set of sample data.

The sample variance is the sum of the squared deviations from their average divided by one less than the number of observations in the data set. For example, for n observations x1, x2, x3, ..., xn with sample mean

$$\bar{x} = \frac{\sum x_i}{n}$$

the sample variance is given by

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

5.4.2 Sampling Variability

Sampling variability refers to the different values which a given function of the data takes when it is computed for two or more samples drawn from the same population.
5.4.3 Sampling

Sampling consists of a limited number of accesses selected from a particular population for the study of some characteristics(s). The assumption behind the sampling theory is that the generalizations based on as sample are applicable to the population from which the sample is drawn.

![Diagram: The inter-relation of sampling and population]

Fig: The inter-relation of sampling and population

5.4.4 Sampling Distribution

The sampling distribution describes probabilities associated with a statistic when a random sample is drawn from a population. The sampling distribution is the probability distribution or probability density function of the statistic. Derivation of the sampling distribution is the first step in calculating a confidence interval or carrying out a hypothesis test for a parameter.
Suppose that $x_1, \ldots, x_n$ are a simple random sample from a normally distributed population with expected value $\mu$ and known variance $\sigma^2$. Then the sample mean is a statistic used to give information about the population parameter $\mu$; $\bar{x}$ is normally distributed with expected value $\mu$ and variance $\sigma^2/n$.

5.4.5 Principles of Sampling

Burges 1984 points out that there are certain procedures and principles those are common to all approaches. These are described as follows:

1) The researches observations should be theoretically directed.

2) It is important for the researcher to be able to locate and enumerate the main units of study.

3) The sample must be representative of the universe from which it is drawn.

4) Researcher should remains in the field unit a theory is adequately rested or developed.

5) Behavior should be sampled in natural setting are made will be relevant to the theoretical perspective.

6) Sampling should, whenever and wherever possible, be comparative.

7) Researcher should make public their sampling procedures so that an assessment may be made of the selection procedures used.

Certainly there are the principles which may researcher would do well to keep in mind, but it is important to remember that different degrees of emphasis will be given to different principles depending upon the research problem which is at hand together with the substantive and theoretical interests of the researcher. With these principles in mind we turn to some of the main strategies of sampling involved in research.
5.4.6 Advantages of Sampling

According to Shah and Shah (1978) sampling has following advantages;

1. Reduced costs

If data are obtained from only a small properties of the total population, expenditure is smaller than if a complete census is attempted e.g. interviewing all ten thousand students at Punjab University is much more costly than interviewing only one hundred students.

ii. greater Speed

Since sample takes only fewer cases compared to the total population data can be collected and summaries in a relatively shorter period of time. This advantage is of great importance when information is needed urgently.

iii. Greater Scope

In certain situations, highly trained personal and specialized equipment are needed. Both of them are in limited availability in developing countries like Pakistan. In such a situation complete senses in impracticable. Sampling therefore renders greater scope.

iv. Greater Accuracy

Accuracy can best be ensured through the use of qualified personnel, effective supervision of the people and scientific processing of data. These are made possible through sampling.

5.4.7 Disadvantages of Sampling

Thakur (1993) mentioned some disadvantages of sampling as below:

(i) Sampling is not feasible in a situation where knowledge about each element or unit or a statistical universe is needed.
(ii) The sampling procedures must be correctly designed and followed otherwise, what we can call as ‘wild sample’ would crop up with misleading results. The sampling error may be larger then expected if the sampling or incorrectly carried out. This leads to biased data and incorrect generalization.

(iii) There are numerous situations in which the units to be measured are and rare and highly variable. Here a very large sample is required in order to field enough cases for achieving statically reliable information.

(iv) Most of the sampling techniques require the services of sampling experts or statisticians. Knowing statistical techniques to determine size and reliability of a sample and to estimate sampling error is not sufficient.

(v) Complicated sampling plans may ultimately involve almost the same amount of time, money, energy or labor as is required in the census method. (Census method is one in which one counts all of the elements of a given population). Each type of sampling has got its own limitations.

(vi) Social Scientists are faced with additional problems in sampling because of scattered distribution of sample unit non-cooperative nature of respondents in credibility to respondents and so on.

(vii) To know certain population characteristics, like population growth rate population density etc. census of population at regular intervals is more appropriate than studying by sampling.

A sample can be representative of population when it has all the significant characteristics of the population in relative proportion. The events, situations, people, researcher’s theoretical and substantive interests will result in the use of different sampling strategies.
5.4.8 What is the Best Sampling Method?

The best sampling method is the sampling method that most effectively meets the particular goals of the study in question. The effectiveness of a sampling method depends on many factors. Because these factors interact in complex ways, the "best" sampling method is seldom obvious. Good researchers use the following strategy to identify the best sampling method.

- List the research goals (usually some combination of accuracy, precision, and/or cost).
- Identify potential sampling methods that might effectively achieve those goals.
- Test the ability of each method to achieve each goal.

5.4.9 How to Determine a Sample Size / Steps in Selecting a Sample-Size

An appropriate sample size is based on a number of accuracy factors that you must consider. Together they comprise a five step process:

1. Determine Goals
2. Determine desired Precision of results
3. Determine Confidence level
4. Estimate the degree of Variability
5. Estimate the Response Rate

5.4.9.1 Step One: Determine Goals

First, know the size of the population with which you’re dealing. If your population is small (200 people or less), it may be preferable to do a census of everyone in the population, rather than a sample. For a marginally higher cost than a 134-person sample, you can survey the entire population and gain a 0% sampling error. However, if the population from which you want to gather information is larger, it makes sense to do a sample.
Second, decide the methods and design of the sample you’re going to draw and the specific attributes or concepts you’re trying to measure.

Third, know what kind of resources you have available, as they could be a limitation on other steps below such as your level of precision. Once you have this information in-hand, you’re ready to go on to the next step.

5.4.9.2 Step Two: Determine the Desired Precision of Results

The level of precision is the closeness with which the sample predicts where the true values in the population lie. The difference between the sample and the real population is called the sampling error. If the sampling error is ±3%, this means we add or subtract 3 percentage points from the value in the survey to find out the actual value in the population. For example, if the value in a survey says that 65% of farmers use a particular pesticide, and the sampling error is ±3%, we know that in the real-world population, between 62% and 68% are likely to use this pesticide. This range is also commonly referred to as the margin of error. The level of precision you accept depends on balancing accuracy and resources. High levels of precision require larger sample sizes and higher costs to achieve those samples, but high margins of error can leave you with results that aren’t a whole lot more meaningful than human estimation.

5.4.9.3 Step Three: Determine the Confidence Level

The confidence level involves the risk you’re willing to accept that your sample is within the average or “bell curve” of the population. A confidence level of 90% means that, were the population sampled 100 times in the same manner, 90 of these samples would have the true population value within the range of precision specified earlier, and 10 would be unrepresentative samples. Higher confidence levels require larger sample sizes. a 95% of assumed confidence level is standard for most social-science applications, though higher levels can be used. If the confidence level that is chosen is too low, results will be “statistically insignificant”.

112
5.4.9.4 Step Four: Estimate the Degree of Variability

Variability is the degree to which the attributes or concepts being measured in the questions are distributed throughout the population. A heterogeneous population divided more or less 50%-50% on an attribute or a concept, will be harder to measure precisely than a homogeneous population, divided say 80%-20%. Therefore, the higher the degree of variability you expect the distribution of a concept to be in your target audience, the larger the sample size must be to obtain the same level of precision. To come up with an estimate of variability, simply take a reasonable guess of the size of the smaller attribute or concept you’re trying to measure, rounding up if necessary. If you estimate that 25% of the population in your country farms organically and 75% does not, then your variability would be .25. If variability is too difficult to estimate, it is best to use the conservative figure of 50%.

5.4.9.5 Step Five: Estimate the Response Rate

The base sample size is the number of responses you must get back when you conduct your survey. However, since not everyone will respond, you will need to increase your sample size, and perhaps the number of contacts you attempt to account for these non-responses. To estimate response rate that you’re likely to get, you should take into consideration the method of your survey and the population involved. Direct contact and multiple contacts increase response, as does a population which is interested in the issues, involved, or connected to the institution doing the surveying, or, limited or specialized in character. You can also look at the rates of response that may have occurred in similar, previous surveys.

After having an estimate of the percentage you expect to respond, and then divide the base sample size by the percentage of response. For example, if you estimated a response rate of 70% and had a base sample size of 220, then your final sample size would be 315 (220/0.7). Once you have this, you’re ready to begin your sampling!

5.5 TYPES OF SAMPLING

According to Burgers (1984) there are three types of sampling:

1. Probability Sampling
2. Non-probability Sampling
3. Theoretical Sampling

Before we discuss the different types of sampling it is essential to define the term probability which is the base of sampling theory. The general meaning of probability is less than certain for which these exists some evidence. In sampling theory the term probability is used as equipment to the relative frequency. Thus when we say that the probability of a tail on a single toss of a coin is \( \frac{1}{2} \), it means that when one makes several tosses the relative frequency of tail will be about \( \frac{1}{2} \) or 0.5.

5.5.1 Probability Sampling

In probability sampling each element in the population has an equal probability (chance) of being reelected in the sample. E.g. in throwing a dice the probability of getting any one of the six faces or say a specific face with 6 spots is 1 to 6. Major forms of probability sampling methods are:

(i) Simple random sampling
(ii) Systematic sampling
(iii) Stratified sampling
(iv) Cluster sampling
(v) Multistage sampling
(vi) Double cluster sampling
5.5.1.1 Simple Random Sampling

According to Forces and Richer (1973) in a simple random sample, each individual in the population must have an equal chance of being included in the sample, a chance less than one (certainty) and more than zero impossibility.

Thakur (1993) defines that by random sampling is correctly meant the arranging of conditions in such a manner that every item of the whole universe from which we are to select the sample share have the same chance of being selected as any other item. It also makes the selection of every possible combinations of the desired number of cases equally likely.

The sample can be drawn by picking up one of the four slips, each one containing the name of the one family member. However, if a simple random samples of two (n=2) were to be drawn, all possible distinct samples should have an equal chance of being selected. In this case, one possible set of samples could be

\[(1) \quad M, F \quad (2) \quad M, S \quad (3) \quad M, GF \quad (4) \quad F, S \quad (5) \quad F, GF \quad (6) \quad S, GF\]

M=Mother, F=Father, S=Son, D=Daughter
GM= Grand Mother, GF = Grand father

In actual practice, however, and particularly where population and sample size are large, all possible samples are not listed prior to the drawing of samples. Instead of the table of random numbers is used to draw simple random samples of desire derived size. These samples provide equal chance to all possible samples to be selected where populations are large; a table of random member is like used as give in your text.
Advantages of simple random sampling

1. It is the most basic, simple and easy method. It requires less money, time and labour.

2. It provides a representative sample. As the sample size enlarges, it becomes increasingly representative of the universe.

Disadvantages of Simple Random Sampling

1) In most of cases it is difficult to find an up to date list of all the units in the population to be sampled.

2) The task of numbering every unit before the sample is clear is time-consuming and expensive, except in regular interval samples.

3) The units need not only to be numbered but also to be arranged in a specific order.

4) As the size of sample required is often large enough in order to achieve statistical reliability representative ness the cost and time-consumed becomes equal to the stratified sampling, which is much more reliable.

5) From the point of view of field survey it has been claimed that masses selected by random sampling method trend to be too widely dispersed geographically and that the time and cost of going from one address to another are too large.

6) The possibility of obtaining a poor or misleading sample is always present when random selection is used.

5.5.1.2 Systematic Sampling Method

It is also called patterned, senior or chain sampling “according to Smith (1975) systematic sampling is one of the simplest, most direct and least expensive sampling method”.

According to Cohen and Holliday (1984) this method is a modified form
of simple random sampling. It involves selecting subjects from a population list rather than a random fashion. For example, if from a population of say 2000, a sample of selection is chosen at random.

Thakur (1993) describes the advantages and disadvantages of systematic sampling as follows:

**Advantages**

This sampling method is frequently used, because it is simple direct and inexpensive.

When a list of names or items is available, systematic sampling is often an efficient approach. In Particular, when records are numbered consecutively or when persons can be summoned, by the numbers, as in the military, systematic sampling is often very economical methods.

**Disadvantage**

At present time exact mathematical procedures for estimating the precision of the systematic sample do not seem to be available. On the basis of approximate procedures available for estimating the precision of systematic sample, it can be said that under certain condition and with those result for proportionate stratified sampling.

To conclude, the different methods of selecting the sample aim to maintain equal chance to all elements and avoid any sort of bias or preference in selection.

**5.5.1.3 Stratified Sampling**

We have seen before that in case of simple random sampling the accuracy of the sample is increased by increasing the size of the sample through application of stratified random sampling method. The stratified random sampling, method can be applied only when the population characteristics are known.

According to Thakur (1993) “eh eth population is divided into different strata or groups and those samples are selected from each strum by simple random sampling procedure or by regular interval method we call it as stratified random sampling method”.

117
Cohen and Holiday (1982) "state that stratified sampling involves dividing the population into homogeneous groups, each group containing subjects with similar characteristics."

The ability of stratified sampling method may further be versified by illustrating the tips or procedures followed, which are five in your text.

Primary purpose of the stratified sampling method is to increase the representative ness of the sample increasing the size of the sample on the basis of having greater knowledge of the population characteristics. (Thakur, 1993)

Thakur (1993) explains some necessary conditions which must be fulfilled for successful use of the stratified sampling method:

(i) The surveyor must achieve up-to-date accounts, complete information about the characteristics of the population on the basis of which stratification is to be made.

(ii) The surveyor should select only those factors for stratification which are significantly related to the topic under consideration. It is customary to use sex, income, religion, age casts etc. for stratification.

(iii) Under the stratum indoor consideration is large enough so that the sampler will have no great difficulty in under candidates for it, it should not be used.

(iv) Basis of stratification should be such that elements from one stratum are highly homogenous within he stratum so that every small sample size will reveal the characteristics of the stratum.

(v) Each stratum must be different from the other.
Advantages

Thakur (1993) Describes advantages and disadvantages as follows:

(i) The population is first stratified into different group and the elements of the sample are selected from each group.

(ii) With more homogenous population greater precision can be achieved with fewer cases. This saves time in collecting and processing of the data.

(iii) When an interviewer is to select a representative sample, he will choose more correctly if the quota is allocated by strata than if he is left upon his own choice or judgment.

(iv) Samples that are self-selected one to be less biased if stratification is added.

(v) As compared with random samples, stratified samples are geographically more concentrated fanned thus save time, money and in moving groom one address to another.

(vi) When detailed study about population characteristics wanted characteristics is more effective.

(vii) In certain studies it becomes not only essential but also unavoidable. E.g. if and when we want to make a comparative study of different ethnic groups, occupational group of social group stratified sampling would provide a more adequate sample.

(viii) A sample of this kind is likely to approach the average of the universe.

(ix) The advantage of stratified sampling over simple random sampling is best expressed by the or 'homogeneity'.

Disadvantages

(i) Unless there are extreme differences between the strata, the expected proportional representation would be small.

(ii) Even after stratification the sample is selected from each stratum either
by simple random sampling method or by systematic ambling method.

(iii) For the application of the stratified method, one must know the characteristics of the specified population on which the study is to be made.

(iv) In social sciences we are mostly interested in measuring several variables.

(v) The process of stratification becomes more and more complicated and difficult as the number of characteristics to be used for stratification is increase.

Stratified random sampling method can be further sub divided into disproportionate stratified sampling; proportionate stratified sampling and optimum allocation stratified sampling methods.

5.5.1.4 Cluster Sampling

According to Shah and Shah (1978), in cluster sampling the sampling units are to be found in groups of individuals or the natural units such as a household or a family.

Thakur (1993) states in cluster sampling, the stratification is one in a manner that the groups are heterogeneous in nature rather than homogeneous. This means that out of several clusters or groups one, two or more number of clusters is selected by simple stratified random method and their element are studied. If all the elements in these clusters are not to be included in the sample, the ultimate selection from with in the cluster is also carried out on a simple or stratified random sampling basis.

Smith (1975) and Thakur (1993) explain the advantages of cluster sampling as follows:

1. Incase of cluster sampling the cost per element is greatly reduced.

2. It becomes possible to take a larger sample and regain the amount of precision that has been lost due to the method.
3. Its success depends upon the nature of the clusters.

4. Like simple random sampling and stratifies sampling, cluster sampling is also a probability sampling.

5. Here we do not need to list each and every element of the process it can be used in situations where it is impossible to obtain simply by other methods.

6. Most situation in which a cluster sample is appropriate do not require precise specification in advance of the total number of elements to be included.

7. The element of cluster are geographically limited rather than widely spread.

Disadvantages

(i) It is a complicated sample design.

(ii) Its various formulas are complex and require electronics data processing.

(iii) Simple random sampling.

5.5.1.4 Multi-stage Sampling

Smith (1975) states that multistage sampling designs are coming into greater use as sampling experts become more sophisticated.

Thakur (1993) defines that normally a multistage sampling procedure is lone that combines cluster and ratified sampling method.

For example, if one want to study the socio-economic background, attitudes and motivation of slum-dwellers;

- We can first make a list of the cities which would make our cluster.

- From these clusters we can select any number of cities.
- Then each city or cluster would by stratified into different slum area.

- A few slum areas can be selected on the principles of Claus sampling and then further from then we can select a sample of households.

Such a process of continuous cluster sampling is called multistage cluster sampling.

The most impotent criterion for the formation of cluster is that the clusters must consist of elements which are most heterogeneous in nature. In this method of drawing samples the various components of the sample selected are often found to be not essentially in the same proportion it is a drawback of multistage sampling. (Thakur, 1993)

### 5.5.1.5 Double Cluster Sampling

This is a commonly used sampling technique in which a large selection of items is made. The information from this large sample is used as a basis of for selection of a small sample for more intensive study e.g. we may select a sample of 10,000 persons for eliciting information say, about occupational distribution. We may further select a sample of 1,000 of 10,000 selected deadlier for an intensive study about certain aspects of occupation. Double sampling is also referred to as two-phased sampling.

### 5.5.2 NON-PROBABILITY SAMPLING

Shah and Shah (1978) states that “In non-probability sampling the selection of the elements is not based on probability, instead it is based on the judgment of the investigator. Burgee (1984) explains that with non-probability sampling there is no means of estimating the probability to units being included in the sampling.
The use of non-probability sampling

Any discussion of the relative merits of probability and non-probability sampling clearly shows the technical advantages of the former. Through randomization the danger of unknown with non-probability designs there is always a chance that some sampling induced element will distort the results. In addition, with non-probability designs we cannot estimate an interval range in which to expect the population parameter.

There are some very practical reasons for using these “less satisfactory” methods:

One may use non-probability sampling because there is no other feasible alternative. The total population may not be available for study. At the scene of a major event some of the participants cannot be interviewed. In another sense those who are included in the simply may select themselves. In mail surveys those who respond may not represent a true crosses-sect. on of those who received the questionnaires.

Second, whole the probability sampling may be superior in theory there are breakdown in its applications. For example, if interviewers are total to select the sample randomly there is always a question as to how skillfully or diligently they will do it.

Third a random sample that is a true cross-section of the population may not be the objective of the research if there is not desire to generalize to a population parameter then there is less concern about whether or not the sample is fully representative often the researcher has more limited objectives, and may be looking only for a feel of the range of condition. (Emory, 1980)

Finally, the costs and time required for probability sampling may be so large that the investigator abandons the use of this method, Carefully controlled non-probability sampling methods seen to give acceptable results, although there is no way to escape the deficiencies mentioned earlier.
According to Thakur (1993) its major forms are

(a) Accidental Sampling
(b) Quota Sampling
(c) Purposive Sampling

5.5.2.1 Accidental Sampling

Accidental sampling means selecting the units on basis of easy approaches. Suppose one is studying political socialization and political participation among the university and college girl students of Punjab University and that his/her sample size in 100. he or she go to the university campus and would select the first hundred students whom the happens to meet, whether in; the classroom or in the students common room.

Such type of sampling is easy to do, and saves time and money, But the chances of bias are also great.

5.5.2.2 Quota Sampling

In quota sampling the objective is to fill a quota reflection the population (forces and Richer 1973). In quota sampling the interviewers are instructed to interview a specified number of people from each category. In studying people’s status, living condition preferences, opinion, attitudes etc. the relevant basis of stratification usually used are: age, sex, education, place of residence, socio-economic status religion and so on. The required number of elements from each category are determined in the according to the number of elements in each category. Thus and interviewers would need to contact number of persons from different age categories, from different religions or social groups and so on. (Thakur, 1993).

The basic goal of quota sampling is the selection; of a sample that is a true replica of the population about which one wants to generalize. This means that the purpose of quota a sampling is to provide a time representative sample of the population so it is quicker and less complicated than a more precise technique. (Forcese, & Richer, 1973).
5.5.2.3 Purposive Sampling

According to Thakur (1993) in purposive sample one picks up the, cases that are considered to be typical of the population in which one is interested. The cases are judged as typical on the basis of the needs of the researcher. Since the selection of elements is based upon the judgment of the researcher, so purposive sampling. The researcher tries his sample to mature the universe in some of important known characteristics. It is assumed that if the sample is representative with respect to known population characteristics it will also be representative with respect to unknown characteristics. Such a assumption is wrong.

Emory, (1980) explains that purposive samples normally involve a more deliberate effort to secure a sample that conforms to some pre-determine criteria. Defect with this method is that the researcher can easily make error in judging as to which cases are typical any time of judgment sample precludes the calculation of estimates of the likely range of error and this is not dependable for any scientific work.

5.5.3 Theoretical Sampling

Mostly writers have mentioned only two types of sampling that is probability and non-probability sampling; but Glaser and Strauss (1967) have suggested that researcher can also use theoretical sampling. They define theoretical sampling as:” The process of data collection for generating theory whereby the analyst jointly collect, codes and analysis his data and decides what data to collect next and where to find them, in order to develop his theory as it emerges. (Glaser &Strauss, 1967, p.45).

In these circumstances data collection is controlled by the emerging theory and the researcher has to consider what group or sub-group are used in data collection.

Theoretical sampling therefore involves researchers in observing groups with a view to extending, modifying, developing and verifying theory.

it is this approach that Smith and Pohland (1976) consider to be one of the
most significant contribution to field research Glaser and Strauss's term theoretical sampling' formalizes various activities that they consider are important in field studies such as:

- Having enough evidence
- Having enough data in a particular area and
- Deciding when to more on the other related problem.

Finally we can concluded that researchers have a number of sampling strategies from which to choose Researchers are engaged in different activities depending on the sample procedures they adopt. To know about Sampling Error and Sample Size please read exploring research.

5.6 SELF ASSESSMENT QUESTION

Q.1. Describe different types of population.

Q.2. Discriminate between different types of sampling.

Q.3. Following statements are, related to

(i) Random sampling  (ii) Stratified sampling  (iii) Cluster sampling

Indicate against each statement the sampling design related to it.

a. Out of a class of hundred students, having 30 science students, 20 art students, 20 have economics students, 10 commerce students and 20 agriculture students, a sample of twenty students is selected
b. In a study of certain characteristics of secondary school lady teachers, the investigator selected some secondary schools randomly for 105 collections of relevant data.

c. In an electric bulb manufacturing factory any five out of every hundred bulbs are re-checked.

d. Out of class of hundred students, a group of twelve is selected by the teacher in such a way that every student stands equal chance of selection.

e. A woman has four grades of apples with 20%, 40%, and 10% in each grade respectively.

Q.4. What in non-probability sampling. Differentiate between different types of non-probability sampling?

Q.5 your city administration have asked you as a sampling expert to help them find out what form (s) and distributions (s) lead poisoning problem present for a city.

a. Design a purposive sampling technique for finding these lead poisoning cases. Give your rationale for the operationalization of your working universe and the sampling strategy you choose.

b. Design a probability sampling technique for studying the city govt.’s organizational response to lead poisoning. Give your rationale for choosing the particular working universe and sampling strategy.
Q.6 what are the steps you would like to consider before selecting sampling size for your research study? Discuss

Q.7 What are the main steps for the selection of sampling size? Discuss in detail.

Q.8 Define the following:
   a. Population
   b. Sampling Size
   c. Parameter
   d. Sample variance
   e. Target Population

5.7 BIBLIOGRAPHY

1. Burgees R. In the Field, in field: An introduction; to Field Research: George Allen & Unwin: Sydney


UNIT 6

RESEARCH TOOLS FOR MEASURING BEHAVIOUR

Atifa Durrani
<table>
<thead>
<tr>
<th>Contents</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>132</td>
</tr>
<tr>
<td>Objectives</td>
<td>132</td>
</tr>
<tr>
<td>Measurement</td>
<td>133</td>
</tr>
<tr>
<td>Reliability</td>
<td>144</td>
</tr>
<tr>
<td>Validity</td>
<td>150</td>
</tr>
<tr>
<td>Tools of measuring behaviour</td>
<td>155</td>
</tr>
<tr>
<td>Self Assessment Questions</td>
<td>161</td>
</tr>
<tr>
<td>Bibliography</td>
<td>162</td>
</tr>
</tbody>
</table>
6.3 MEASUREMENT

Measurement has been defined in different ways by different social scientists. In simplest
Language it is defined as the process of obtaining a numerical description of the characteristics of a group of persons or objects.

According to Bailey, (1968) “measurement is the process of determining the value or levels either qualitative or quantitative of a particular attribute for a particular unit of analysis”. As per this definition measurement is not essentially confined only to numerical or quantitative description but can be qualitative as well. Many scientists take measurement mainly in the sense of a quantitative study. Helmhstadter (1970) defined measurement as possess some characteristics.

Taylor (1963) defined measurement as “assignment of numerals, according to rules”. A still more elaborate and wider for assigning numbers to objects in such a way as to represent quantities of attributes”. Forces (1973) says in his book “Measurement consists of identifying the values which may be assumed by some variables, and representing these values by some numerical notations. The numerical notation is systematically and consistently assigned, that is, it is assigned according to some set of rules”.

Naming, classifying, and defining what an abstract theoretical concept stand for is a difficult procedure and requires a series of steps, with the help of these steps, abstract meaning of a variable is systematically broken down into more concrete tangible meanings that permit the direct collection of valid data. Under specific conditions and procedure measurement is a reproducible process: that is the measurement of one individual must be capable of replication by others. Then we may speak of the reliability of measurement. In addition, if our measurement routine identifies and measures of the phenomenon which we intend to measures, then we may speak of the validity of measurement. Consistent with this general view of measurement, it is conventional to distinguish four levels of measurement: (1) nominal, (2) ordinal, (3) interval, and (4) ratio.

Thus the numerals assigned to the variables, indicate different conditions of the variables or different values of the variables or different degrees of
intensity of a quality possessed by units.

All above definitions make the following characters of measurement evident.

1. Measurement is a purely descriptive process.

2. Measurement implies that the attributes of persons or objects are possessed in varying degrees and the degree of variation can be measured and represented, and

3. Measurement, in essence, is a numerical process. Common objects of measurement in social sciences are individual’s characteristics. Interactions, inter-relations, participation, motivation attitudes, opinions, etc. public opinion studies are carried on by various public and government agencies. Even the commercial organizations measure people’s opinions and attitudes to know the future market of their products.

4. Attitudinal studies help in predicting about individual’s future

5. Behavior and their possible reaction towards different developmental programmes.

Before we discuss different levels of measurement, we must know about the postulates of measurement.

According to Thakur (1993) a postulate can be define as “an assumption that is an essential pre-requisite for carrying out some option of thinking. It is an option about the relation between the objects being measured”. Usually the postulates are assumed to be true in measurement yet they should be tested whenever possible. The three postulates that concern in measurement can be explained as follows:

1. $a$ is either equal to $b$ or not equal to $b$, but not both i.e., either $(a=b)$ or $(ab)$, but not both. For example, if we classify the individuals on the basis of residence we can classify them in rural, urban, or sub-urban category or on the basis of their employment and can be classified into employed, unemployed, under-employed categories. It means that an individual who belongs to one category cannot belong to other and that all the persons of one category are similar to each other in the characteristic.
2. If \( a \) equals \( b \), and \( b \) equals \( c \) then \( a \) equals \( c \)', i.e. if \((a=b)\) and \((b=c)\) then \((a=c)\). This postulate enables a researcher to establish the equality of the members of set on a characteristic by comparing them.

3. If \( a \) is a greater than \( b \), and \( b \) is greater than \( c \), \( a \) is greater than \( c \) i.e. if \((a>b)\) and \((b>c)\) then \((a>c)\). This indicates that a person from one category has more of the property than a person from another category. But this can not always be easily satisfied. For example, take the criterion of dominance: \( a \) may dominate \( b \) and \( b \) may dominate \( c \), but it is possible that \( a \) does not dominate \( c \).

6.4 **LEVELS OF MEASUREMENT**

The four levels of measurement mostly used in sociology and other social sciences are

6.4.1 **Nominal Measurement**

6.4.2 **Ordinal Measurement**

6.4.3 **Interval Measurement**

6.4.4 **Ratio Measurement**

These different levels of measurement differ from each other in the following way:

1) The data to which they are studied.

2) The processor involved.

3) The structure of measures achieved.

The rules used to assign numerals to objects are the criteria that define the kid of scale and the levels of measurement. The four levels of measurement are discussed below:
6.4.1 Nominal Measurement

Nominal measurement (or scale) is the lowest of measurement. In nominal measurement numbers are used to name, identify or classify persons, objects, groups, etc. For example, a sample of persons being studied may be classified as (a) Muslim, (b) Hindu, (c) Christian, or the same sample may be classified on the basis of sex, rural-urban variables, political party affiliation, etc. All these classifications would be examples of nominal measurement. Classifications of the clinical groups into such as schizophrenia, depression, phobia, etc. also constitute nominal measurements. This category fulfills the first postulate of measurement which is an essential requirement. Numerals may be assigned to different categories which are just names or symbols or labels to represent the different sets and do not essentially carry any specific meaning.

In other words, in nominal measurement, numbers of any two groups are always equivalent but all members of any one group are always equivalent. And this equivalence relationship is reflexive, transitive and symmetrical. Nominal scales are more frequently used in exploratory research were greater emphasis is laid on uncovering in a relationship between two characteristics rather than on specifying the relationship in mathematically precise terms. Here one aim’s to know the pattern of relationship among several characteristics of a person. As discussed earlier, it is a system of classification and does not place the entity along a continuum. It involves a simply count of the frequency of the cases assigned to the various categories, and if desired numbers can be nominally assigned to label each category as in the example below:

Figure: 1 example of a nominal scale

| Which of the following food items do you tend to buy at least once per month? (Please tick) |
|---------------------------------|----------------|----------------|
| Okra                            | Palm Oil       | Milled Rice    |
| Peppers                         | Prawns         | Pasteurized milk |
The drawback of nominal measurement is that it is most elementary and simple. Because of these characteristics some exert the view that nominal measurement is not a measurement at all.

6.4.2 Ordinal Measurement

This is second levels of measurement. I ordinal measurement, members denote the rank order of the objects or the individuals. According to Thakur (1993) ordinal scale “is one that defines the relative position of objects or individuals with respect to a given characteristics “. in this way one measures the characteristics given characteristics but is also able to determine for each individual or object, whether one has more of the attribute in questions than another individual, or the same amount or less. On the basis of the amount of quality possessed, the individual are put into different ranks (1, 2, 3,) or they can be put into different categories.

For ordinal levels of measurement the third postulates that if a is transitivity postulate must be satisfied, which states that if a is greater than b and b is greater than c then a is greater than c. Thus, in ordinal measurement, besides the relationship of equivalence, a relationship of “greater than “or “lesser than” exists because all members of any particular subclass are equivalent to each other and at the same time greater or lesser than the members of other subclass. The relationship of greater than is usually reflexive, transitive, and asymmetrical. Ordinal scales involve the ranking of individuals, attitudes or items along the continuum of the characteristic being scaled. For example, if a researcher asked farmers to rank 5 brands of pesticide in order of preference he/she might obtain responses like those in table 2 below.
Figure: 2 an example of an ordinal scale used to determine farmers' preferences among 5 brands of pesticide.

<table>
<thead>
<tr>
<th>Order of preference</th>
<th>Brand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Rambo</td>
</tr>
<tr>
<td>2</td>
<td>R.I.P.</td>
</tr>
<tr>
<td>3</td>
<td>Killalot</td>
</tr>
<tr>
<td>4</td>
<td>D.O.A.</td>
</tr>
<tr>
<td>5</td>
<td>Bugdeath</td>
</tr>
</tbody>
</table>

From such a table the researcher knows the order of preference but nothing about how much more one brand is preferred to another that is there is no information about the interval between any two brands.

All of the information a nominal scale would have given is available from an ordinal scale. In addition, positional statistics such as the median, quartile and percentile can be determined. It is possible to test for order correlation with ranked data. The two main methods are Spearman's Ranked Correlation Coefficient and Kendall's Coefficient of Concordance. Using either procedure one can, for example, ascertain the degree to which two or more survey respondents agree in their ranking of a set of items. Consider again the ranking of pesticides example in figure 3.2. The researcher might wish to measure similarities and differences in the rankings of pesticide brands according to whether the respondents' farm enterprises were classified as "arable" or "mixed" (a combination of crops and livestock). The resultant coefficient takes a value in the range 0 to 1. A zero would mean that there was no agreement between the two groups, and 1 would indicate total agreement. It is more likely that an answer somewhere between these two extremes would be found.
The only other permissible hypothesis testing procedures are the runs test and sign test. The runs test (also known as the Wald-Wolfowitz). Test is used to determine whether a sequence of binomial data - meaning it can take only one of two possible values e.g. African/non-African, yes/no, male/female - is random or contains systematic 'runs' of one or other value. Sign tests are employed when the objective is to determine whether there is a significant difference between matched pairs of data. The sign test tells the analyst if the number of positive differences in ranking is approximately equal to the number of negative rankings, in which case the distribution of rankings is random, i.e. apparent differences are not significant. The test takes into account only the direction of differences and ignores their magnitude and hence it is compatible with ordinal data.

Socioeconomic status is a good example of ordinal measurement because every member of the upper class is higher in social prestige than every member of the middle and lower class. Like wise, every member of the middle class is higher in social prestige than every member of the lower class. One drawback of ordinal measurement is that ordinal measures are not absolute quantities nor do they convey that the distance between the different rank values are equal. This is because ordinal measurements are not equal-interval measurement nor they incorporate absolute zero point. The second demerit of ordinal measurement is that there is no way to ascertain whether a person has none of the characteristics being measured.

### 6.4.3 Interval Measurement

The interval measurement is also known as equal interval scale. It possesses the characteristics of both nominal and ordinal measurement. Blalock and Blalock Jr. (1982) defined interval scales as “whenever it is possible to compare differences in scores because of the existence of a standardize unit, we refer to the scale as interval scale”. The salient feature of interval measurement is that it has numerically equal distances on the scale which indicate equal distances in the properties of the objects being measured. In other words, the unit of measurement is constant and equal. This is the reason why interval measurement is also called as equal-internal measurement.
It is only with an interval scaled data that researchers can justify the use of the arithmetic mean as the measure of average. The interval or cardinal scale has equal units of measurement, thus making it possible to interpret not only the order of scale scores but also the distance between them. However, it must be recognised that the zero point on an interval scale is arbitrary and is not a true zero. This of course has implications for the type of data manipulation and analysis we can carry out on data collected in this form. It is possible to add or subtract a constant to all of the scale values without affecting the form of the scale but one cannot multiply or divide the values. It can be said that two respondents with scale positions 1 and 2 are as far apart as two respondents with scale positions 4 and 5, but not that a person with score 10 feels twice as strongly as one with score 5. Temperature is interval scaled, being measured either in Centigrade or Fahrenheit. We cannot speak of 50°C being twice as hot as 25°F since the corresponding temperatures on the centigrade scale, 10°C and -3.9°C, are not in the ratio 2:1.

Most of the common statistical methods of analysis require only interval scales in order that they might be used. These are not recounted here because they are so common and can be found in virtually all basic texts on statistics.

6.4.4 Ratio Measurement

It is the highest levels of measurement and has all the properties of nominal, ordinal and interval scales plus an absolute or a true zero point. The zero point is the distinguishing characteristic of the ratio scale. Dennis p. forces and Stephen richer (1973) while scale. Highlighting this significant feature commented on the ratio scale as “a ratio scale exists when we are able to designate a zero point in our measurement – a zero point that is isomorphic (consistent) with empirical reality”. Besides that it is a zero point scale which is important form empirical view point. It is a higher scale in the sense that it tells whether an object has none of a property being measured. Thus a ratio scale becomes isomorphic or consistent with empirical reality because of zero point in it. Numbers on the scale indicate the actual amounts of the property being measured. Therefore, in case of ratio scale measurement it is possible to say not only that a is greater that b and b is greater than c but also the ratio of a to c. In social sciences, scientists are more apt to deal with ratio data than interval data.
The measurement of formal education or income is ratio levels. It is possible to point out an empirical state of no education or no income in ratio level.

The Fahrenheit and centigrade thermometers are the examples of interval scales. The units here represent equal amounts of changes in the volume of a column of mercury under a certain pressure. Besides this, time as shown in our calendar scores on intelligence tests and aptitude tests are good examples of interval measurements.

Social scientists deal with social characteristics which are difficult to be measured in the terms in which thermometer units are measured. For example, some researches want to measure student's class achievement on an interval scale. Suppose, the difference between student a and b and between b and c is equal. But one cannot derive from this statement researcher needs to make measurement of a high level, i.e. in terms, of ratio scales. In social sciences, scientists occasionally realize interval data. For example, the scores from IQ tests are interval level, for the distance between a score of 100 and a score of 120 is equal to the distance between a score of 80 and a score of 100. Opinions or attitudes have been measured directly when a respondent is asked to respond to some stimulus-a statement when or a question by assigning the statement to one of several categories which have been identified only numerically. These numerical classes are supposed to be equally spaced, and the respondent makes his/her accession on the basis of his/her hypothetical zero categories.

The common examples of ratio scale are the measures of weight, width, length, loudness, and so on. It is obvious, therefore, that ratio scales are common among physical sciences rather than among social sciences.

Measurement of social characteristics of ratio levels is mostly not possible. Ratio scale is thus of higher level than the interval scale. It helps us not only to measure the individual's characteristics exactly it is enables us to add, subtract, and multiply because of its empirically isomorphic zero point.
6.5 FUNCTIONS OF MEASUREMENT

Measurement has variety of functions in social and natural sciences. It may help a class teacher to distinguish between the high achievers and the low achievers or it may help an industrial psychologist to place a worker in the right job.

Some primary functions of measurement are as follows:

6.5.1 Selection

Selection of personal in industry or other institutions may be carried out with the help of psychologists. The function of measurement tools in selection is to predict the ability of the individual.

6.5.2 Classification

Measurement also helps in various of classification, which sometimes becomes very necessary for a programme to be carried out effectively. For example, a psychiatrist or a clinical psychologist may wish to classify his or her patient into different categories of mental illness here he/she is helped by the tools of measurement.

6.5.3 Comparison

No two individuals are alike. In other words individuals differ in trait, mental processes, habits, tendencies, education achievements, abilities, etc. whenever two persons are to be compared, measurement comes into use. With the help of appropriate measurement, it is possible to conclude that it is better than b in aptitude tests etc.

6.5.4 Guidance and Counseling

Measurement assists psychologists in guidance and counselling.
Counselling is a sort of a specialized guidance programme and refers to the advice given to an individual so that he/she can arrive at a workable solution to various adjustment problems in life. Measurement can help to make accurate predictions regarding problems of adjustment likely to come up in the future and also diagnose mental disabilities, aberrations, deficiencies, etc.

6.5.5 Research

Measurement helps in research activities too. In fact measurement is the fundamental basis of all psychological and educational researches. In psychological and educational researches, usually, the effects of all other variables or set of variables or set of variables is studied while the effects of all other variables are controlled. In doing so a variety of measurement are resorted to confirm that the design, statistical calculation, results, etc. are in accordance wit the principles of measurement or else research becomes meaningless. Thus measurement may be regarded as the heart of any psychological or educational research.

Reading: For further information read the following

| Exploring Research p, 112-117 |

6.6 MEASUREMENT SCALES AND STATISTICAL TESTS

One of the primary purposes of classifying variables according to their level or scale of measurement is to facilitate the choice of a statistical test used to analyze the data. There are certain statistical analyses which are only meaningful for data which are measured at certain measurement scales. For example, it is generally inappropriate to compute the mean for Nominal variables. Suppose you had 20 subjects, 12 of which were male, and 8 of which were female. If you assigned males a value of '1' and females a value of '2', could you compute the mean sex of subjects in your sample? It is possible to compute a mean value, but how meaningful would that be? How would you interpret a mean sex of 1.4?
When you are examining a Nominal variable such as sex, it is more appropriate to compute a statistic such as a percentage (60% of the sample was male).

When a research wishes to examine the relationship or association between two variables, there are also guidelines concerning which statistical tests are appropriate. For example, let's say a University administrator was interested in the relationship between student gender (a Nominal variable) and major field of study (another Nominal variable). In this case, the most appropriate measure of association between gender and major would be a Chi-Square test. Let's say our University administrator was interested in the relationship between undergraduate major and starting salary of students' first job after graduation. In this case, salary is not a Nominal variable; it is a ratio level variable. The appropriate test of association between undergraduate major and salary would be a one-way Analysis of Variance (ANOVA), to see if the mean starting salary is related to undergraduate major.

Finally, suppose we were interested in the relationship between undergraduate grade point average and starting salary. In this case, both grade point average and starting salary are ratio level variables. Now, neither Chi-square nor ANOVA would be appropriate; instead. Here the relationship between these two variables may be calculated by using the Pearson correlation coefficient. (Which we will be studying in unit 7)

6.7 RELIABILITY

Reliability is one of the important characteristics of any test. In its simplest sense, reliability refers to the precision of the measurement of score. It also refers to the score of a test which remain constant for the same unit of measurement overtime. A test is also said to be consistent (when administered once), if the examinees who obtained high score on one set of items also score high on an equivalent set of items and those who obtain low scores on one set of items also score low on an equivalent set of items. Thus reliability may be defined as the consistency of scores obtained for one set of measures another.
6.7.1 Characteristics of reliability

Anastasi (1968) defined reliability as “the consistency of scores obtained by the same individuals when re-examined with the test on different occasions or with different sets of equivalent items, or under other variable examining conditions”. It is obvious from the above interpretation that reliability is never the property of the test itself rather it is the property of test scores. Kerlinger (1964) described reliability as “the relative absence of errors of measurement in a measuring instrument”. So it would be wrong to expect that a measuring instrument will be perfectly free from errors. Achieving absolute reliability is less possible in social or behavioral subjects like sociology, psychology, anthropology, or education. In social sciences, the success of measurement depends upon the degree or extent to which errors can be eliminated. Fundamentally, we can estimate an instrument’s reliability be means of its replication to the same or similar subjects. That is the reuse of the instrument, given identical or similar phenomenon, should realize identical or similar results. Generally speaking a researcher must be aware that a disparity in measurement result from one application to another may be a result, of actual variation in the phenomenon rather than a weakness of measurement instrument itself.

Important instruments of investigation in the areas of Social or Behavior sciences are questionnaires, schedules, interviews, content analyses and case histories. Their reliability depends upon the way they are constructed and used. The extent to which they are reliable, they are dependable also.

Kerlinger (1964) Selltiz and Jahoda (1965) mentioned three different ways of testing reliability of a measuring instrument which are summarized as under;

1) If repeated study of the same thing with the same or comparable measuring instrument under same condition gives the same or similar result then the instrument be called as being reliable. In case the instrument is reliable the difference in result will occur.
2) When the conditions determining or affecting the results have changed, or when the sample on which the study is carried out is completely different one. But such differences should not be due to the measuring instrument, as for example, due to different observers or due to difference in the mental position of the observer from time to time. This definition indicates the quality of stability of the reliable instrument.

3) Second ways of knowing the reliability of a measuring instrument is by knowing whether the measures obtained from a measuring instrument are true measures of the property being measured. This is an accuracy definition of reliability. Here we are really asking whether the measures are accurate or not.

4) The third way of testing reliability is by measuring the, exact amount of error in a measuring instrument. Reliability can thus be defined as the relative absence of errors of measurement in a measuring instrument.

5) The reliability of a test is also defined from another angle. Whenever we measure something either in social physical sciences the measurement involves some kind of errors. If the measurement is perfectly accurate, i.e., free from all kinds of errors, the reliability will be perfect. But this is an ideal which rarely achieved in either social sciences or in physical sciences. Each measurement contains some error and therefore, reliability is never total. Thus it can be said that each individual score or measurement is made up to two types of scores, issues score and the error score. True score is the score, which is free from error occurring due to the chance factor as well as from other kind of errors. Usually, a person’s true score made by the same person on the same test but his obtained score may vary from trail to trail because error score contributes to obtain score in each trail.
Error score may be of two kinds.

1) Random or chance score

2) Systematic or constant score

1) Random errors score sometimes inflates the score and sometimes depresses the score. Thus it works randomly in both the positive and negative direction. Random or chance errors are also known as the errors of measurement. The mean of all these measurements is considered zero.

2) Systematic errors work constantly in one direction and therefore, they would either tend to inflate or depress the score. The mean of such errors of measurement would not be zero. It is important to note that any test is neither perfectly reliable nor perfectly unreliable. Thus, reliability is not an absolute principal rather it is always a matter of degree.

6.7.2 TYPES OF RELIABILITY

There are three common types of the reliability

1. Test-Re test Reliability

2. Parallel Forms Reliability.

3. Inter-Related Consistency or Internal Reliability

6.7.2.1 Test-Retest Reliability

In test retest reliability the single form of the test is administered twice on the same sample with a reasonable time gap. In this way two administered of the
same test yield the two independent sets of scores. The two sets when correlated, give the values of the reliability coefficient. The reliability coefficient thus obtained is known as the temporal stability coefficient and indicates relative position of the examinees in terms of the scores over a given time period. Time period provided for a test plays very effective role in test-retest reliability. However, there have been some disagreements among experts regarding the time lapse between the two administrations of the test.

When the time is too short, it is likely to lower the reliability coefficient. If the time gap on the other hand is too long it is likely to lower the reliability coefficient. The most appropriate and convenient time gap between the two administrations is a fortnight, which is considered neither too short nor too long. There are evidences to support that this time interval yields a comparatively higher reliability coefficient.

Test-retest reliability has its disadvantages also. Some of them are:

a) This method is a time consuming method of estimating the reliability coefficient. Sometimes the same subject are not easily available for the second the second administrations.

b) In this method it is assumed that, the examinee’s physical and psychological set up remains unchanged in both testing situations. But in reality this is not so.

In fact, the examinee’s health, emotional condition, motivational condition, and his mental set up do not remain perfectly uniform. Not only this, the examinee’s physical and mental make up also change. Besides, some uncontrolled environmental changes may take place during the second administration of the test. All these factors are likely to make the total score of the examinee’s different from the first test and thus, the examinee’s relative position is likely to change, thereby lowering the reliability coefficient. Despite these limitations, the test-retest method is the most appropriate method of estimating reliability of both the speed test as well the power test. For a heterogeneous test, too, the test-retest method is the most appropriate method of computing reliability.
6.7.2.2 Parallel Forms Reliability

Parallel form reliability is known by various names such as alternate forms reliability and comparable forms reliability. Parallel forms reliability requires the test to be developed in two forms, which should be comparable or equivalent. Two forms of the test are administered to the same sample with the time interval of usually a fortnight. Parallel form reliability measures the consistency of the examinee’s scores between two administrative of parallel forms of a single test. The biggest problem in parallel forms test is how to make both forms equivalent in the true sense. Freeman (1972) has listed the following criteria for judging whether the two types of reliability are parallel form reliability or not.

- The number of items in both forms should be the same.
- Items in both forms should have uniformity regarding the content, the range of difficulty, and the adequacy of sampling.
- Distribution of the indices of difficulty of items in both should be similar.
- Means and standard deviations of both forms should be equal or nearly so.
- Mode of administration and scoring of both forms should be uniform. One can thus visualize the difficulty involved in making the two forms of a test parallel because it is very difficult, if not impossible, to meet all the above criteria. More-over, such a test involves too much labour and time because all items in a changed language are written twice for the two separate forms.

6.7.2.3 Inter-related or internal consistency reliability

Internal consistency reliability indicates the homogeneity of the test. If all the items of the test measure the same function or trait, the test is said to be a homogenous one and its internal consistency reliability would be quite high.

The most common method of estimating internal consistency reliability is the split-half method which the test is divided into two equal halves. The common
way of splitting the test is the odd-even method. In this method, all add numbered items (like 1, 3, 5, 7 ....etc.) constitute one part of the test and all even numbered items, (like 2, 4, 6, 8 ...etc.) make another part of the test. Thus, each examinee receives two scores. In this way from single administration of the single form of the test two sets of scores are obtained. Thus the variability produced buy the difference in the two administrations of the same test is automatically eliminated. Therefore, a quick estimate of the reliability is made. The disadvantage is that since both sets of scores are obtained on one occasion, the fluctuations due to changes in the temporary conditions and within the examinee as well as due to temporary changes in the external environment will operate.

6.7.3 Index of reliability

According to Singh (1986) “the true score on a test can be defined as the mean of an unlimited number of measurements, made by the same person on the same test or on the equivalent forms of the same test”. Theoretically, the correlation between the obtained score and true scores should be perfect. But in reality this is an exception rather than a rule. Index of reliability has been defined by Singh (1986) as “the correlation coefficient between the obtained scores and then true counterparts”. This statistic indicates the extent to which we can depend upon obtained scores a measure of true scores.

Index of reliability, thus gives the maximum correlation which the test is capable of yielding in its present form. Index of reliability is statistically equal to the square root of the reliability coefficient of the test.

For more information read the following:

| Exploring Research | p 117-125 |

6.8 VALIDITY

Validity is another important characteristic of a scientific instrument. The term “validity” means truth or fidelity. Thus validity refers to the degree to which a test measures what it claims to measure. Experts regard validity as the best measure of the trait or ability being measured by the test.
6.8.1 Characteristics of validity

In its simplest form Forcense and Richer (1973) defined validity as “By validity we means that we are measuring what we intend to measure”. Any measuring instrument is valid when it measures most accurately the objects or individuals and their characteristics.

Different writers have defined validity in different ways. Selltize (1959) defined validity as “validity of a measuring instrument may be defined as the extent to which differences in scores on it reflect true differences among individuals, groups, or situations in the characteristics which it seeks to measure, or true differences in the same individual, group or situation from one occasion to another, rather than constant or random errors”.

There is often direct and close congruence between the nature of the object measured and the measuring instrument. If we are measuring certain physical properties, like colour of the body or hair or eyes, or measuring some attributes that are direct and concrete in nature like age, sex, income then maintaining validity is not a problem. The length or weight of an object can be measured easily. On the other hand, there are many social characteristics which are highly abstract in nature and can be measured only indirectly like achievement, aspiration, attitude, aggressiveness, socialization, etc. it is difficult to measure these characteristics, i.e. in exact term.

Social scientists have developed certain valid and reliable measuring instruments which measure the validity of tests. A valid and reliable measuring instrument can be said to be one that is able to measure the characteristics both accurately and distinctly. That means the differences in scores on the measuring instruments should reveal true difference among the individuals, objects or units of measurement.

In broad sense, validity is concerned with generalisability. When a test is valid, it means its conclusion can be generalized in relation to the general population. Validity has three important properties.
1. Validity is a relative term. A test is not generally valid. It is valid only for a particular purpose. For example, a test of statistical ability will be valid only for measuring statistical ability because it is put only to the use of measuring that ability and not for measuring the knowledge of physics, history, etc.

2. Validity is not a fixed property of the test because validation is not a fixed process rather an unending process. With the discovery of new meanings the old contents of the test become less meaningful. Therefore they need to be modified radically in the light of the new meanings. Hence, the validity of a test computed in the beginning becomes less dependable. So the test constructor should compute a fresh validity of the test in the light of new meaning attached.

3. Validity like reliability is a matter of degree and not an all-or –none property. A test meant for measuring a particular trait or ability cannot be said either perfectly valid or not valid at all. It is usually more or less valid.

There are three main purposes of testing

1) Representation of certain specified area of contest.

The tester may wish to determine how an examinee performs at present in a sample of situation (or contents) that the tests claim to represent. For example, through an English spelling test, the tester may determine the present level of English spelling among school students.

2 Establishment of functional relationship with a variable available at present or in future

The tester may wish to predict an examinee’s future standing on a certain variables or want to determine his present condition on a particular variable. For
example, on a mechanical aptitude test he may wish to measure mechanical aptitude and predict his future performance in a job of mechanics.

3) Measurement of a hypothetical trait or quality

A tester may want to determine the extent to which an examinee possesses some trait as measured by the test performance. For example, a tester wants to know whether or not an examinee, extroversion, etc. which cannot be observed directly.

6.8.2 Type of validity

There are three major of Validity:

1. Content Validity
2. Criterion Validity
3. Construct Validity

6.8.2.1 Content Validity

Contest validity is the representativeness of the content. That is, content validity is concerned with the relevance of the relevance of the contents or the items, individually and as a whole. Anastasi (1968) described that content validity "involves essentially the systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain to be measured". Content validity is needed in the tests, which are constructed to measure how well the examinee has mastered the specific skills or certain course of study.

Content validity is most appropriately applied to the achievement test or the proficiency test. Difficulty in determining content validity lies in the fact that it is easy to select items from the universe of contents which are more obvious that others which are not. Content validity of a collection of items can always be challenged in respect of how large or how good the contents are.
6.8.2.2 Criterion validity

Criterion validity is very common and popular type of test validity. It is obtained by comparing or correlating the test scores with scores available at the present or to be available in the future. The criterion is defined as an external and inter-dependable measure of essentially the same variable that the test claim to.

6.8.2.3 Construct validity

Construct validity is the third important type of validity. It is more complex and difficult process that content validation and criterion validation. Hence an investigator decides to complete construct validity only when he or she is fully satisfied that neither any valid and reliable criterion is available to him/her or the contents entirely satisfactory and adequate to define the quality of the test.

Thakur (1993) has explained the meaning of construct validity in such manner, “Construct validity means determining which factors or properties can explain the variations in the scores or variance of the test.” In other words, which factors are responsible and should be used to explain the differences among individuals or groups in their test scores on a measuring instrument. Thus an investigator wants to locate the variable or factors which are called as constructs. The abstract in nature cannot be measured directly for measurement. The researcher after locating the variables, construct or attempts to trace- out the nature, type and degree of relationship among different variables.

A few example of construct validity are anxiety, intelligence, verbal fluency, dominance, etc. construct validations contains some problems like systematic examination concerning the definition of the construct, unsuitability and inappropriateness of the measure of the construct etc. This is the reason why construct validity is rarely computed by the test construction.

6.8.3 Factors influencing validity

Validity of a test is influenced by several factors. Some of the import factors are enumerated below:

6.8.3.1 Length of the test

Homogenous lengthening of the test not only increases the reliability but
also the validity of the test. The longer the test, the more reliable and valid it becomes. Thus lengthening the test or repeated administration of the same test increases the reliability and since validity in a homogenous test is dependant upon reliability, it also increases the valid of the test.

6.8.3.2 Range of ability

If the directions of the test are ambiguous, it would be differently interpreted by different examinees. Moreover, such items tend to encourage guessing on the part of the examinee. So the validity of the test would be lowered.

6.8.3.3 Socio-cultural differences

Culture differences in different societies affect the validity of a test. A particular test developed in one culture may both be valid for another culture because of the difference in socio-economic status, sex roles, social norms, etc. however, when a test in cross-cultured, this factor does not affect the validity of the test.

Reading: For more information read the following:

Exploring Research p 125-130

6.9 TOOLS OF MEASURING BEHAVIOUR

The second part of this unit deals with tools of measuring behaviour. As you already know that there are three ways to collect data: 1) administer a standardized instrument, 2) administer a self-developed instrument and 3) record naturally available data. Further the most important thing in research is the selection of type of method of instrument to measure the behavior and how far this method or instrument is valid and reliable. The other important thing is the way of asking a question which will determine the use of a particular method of measuring behavior.

For example, if we want to know how much information people have about AIDS we will choose some kind of “knowledge” test for this or if we want to know what people feel about smoking in public places we will use attitude test.
In this unit we will learn the methods of measuring behavior and their application in research.

What do we mean by a test? Before going in details let us understand and define the concept of test. A test is not necessarily a written set of questions to which an individual responds in order to determine whether he or she "passes". A more inclusive definition of a test is "a means of measuring the knowledge, skill, feeling, intelligence or aptitude of an individual or a group". During our education career we take many test of knowledge of the students. Similarly, we go through aptitude test for admission in professional and technical or colleges? Such tests measure the level of aptitude of different individuals for a particular profession or skill. The result of these tests is that they produce numerical scores that can be used to identify, classify, and evaluate those who are tested.

Why a researcher uses tests? Because tests can help to determine the outcomes of an experiment, can give insight in individual's or groups strengths and weakness. Tests also help in placement and selection of individuals according to their knowledge, attitude, feelings, etc.

6.9.1 TYPES OF TESTS

There are many different types of test available and many different ways to classify them (Mitchell, 1985). In this unit we are concerned with the following:

I. Achievements tests
II. Attitude tests
III. Personality tests

6.9.1.1 Achievements tests

Achievement tests are exams that are designed to determine the degree of knowledge and proficiency exhibited by an individual in a specific area or set of areas. An achievement test is sometimes administered as part of the acceptance process into an educational program or to qualify an individual for employment or a promotion with a current employer. In other applications, the achievement test serves as a tool to measure current knowledge levels for the purpose of placing students in an educational environment where they have the chance to advance at a pace that is suitable for their abilities.
Achievements tests attempt to measure what an individual has learned his/her present level of performance. Most tests used in the schools are achievement tests. They are helpful in determining individual or group status in academic learning. Achievement test scores are used in placing, advancing or retaining students at particular grade level. They are used in diagnosing strength and weaknesses and as a basis for awarding prizes scholarships or degrees.

Many research studies are designed to compare the effectiveness of two or more curriculum approaches or methods of instruction. Effectiveness is usually defined in terms of pupil achievement at the end of study. Sometimes studies are deigned to investigate various aspects of remedial instruction. In such studies diagnostic instrument are occasionally utilized. A diagnostic test is a type of achievement test yielding multiple scores fro each area of achievement: these scores facilitate identification of specific area of deficiency. The Stand for Diagnostic Reading Test (Gay, 1990) is an example of a commonly used diagnostic achievement instrument.

There are two types of achievement tests, a) standardized test and b) teacher/researcher made tests. Standardized test are typically developed by experts and are therefore, well constructed. Individual test items are analyzed and revised until they become suitable for broad application across different settings. One resulting characteristic of standardized test objectivity, in essence, objectivity means that an individual, score is the same, or essentially the same, regardless of who is doing the scoring. Another characteristic is existence of validity and reliability of data. Validity is concerned with what a test measures and for whom it is appropriate, reliability refers the consistency with which the test measurers whatever it measurers (Gay, 1990).

Teacher/researcher made test are designed more specifically and comparatively for smaller number of people and for some specific population. For examples, a teacher takes a mid-term test of psychology course of B.A class.

Reading: For more details read the following

Exploring Research 137-143
Many tests or scales have been devised to get a clear-cut opinion of people in different social/research problems, for example, social distance scales by Bogardus, scales of equal appearing internal by L.L. Thurstone, scales by internal consistency by Likert and other social matric scales of Moreo. In this chapter we are only concerned with a) Thurstone scales and b) Likert scales.

6.9.1.3 Thurston Scales

In an attempt to approximate an interval level of measurement, psychologist Robert Thurstone developed the method of equal-appearing intervals. This technique for developing an attitude scale compensates for the limitation of the Likert scale in that the strength of the individual items is taken into account in computing the attitude score. It also can accommodate neutral statements.

A Thurston scale asks an individual to select from a list of statements that represent different point of view those with which he/she is in agreement. Each item has an associated point value between 1 to 11, point values for each item are determined by averaging the values of the item assigned by a number of "judges". An individual's attitude score is the average point values for all the statement checked by the individual (Gay 1990).

6.9.1.4 Likert Scales

The Likert scale is by far the most popular attitude scale type. A statement is followed by several levels of agreement: strongly agree, agree no opinion, disagree, and strongly disagree. This five-point scale is commonly used, but other scales, from four to ten points, can be used as well.

Likert scales asks an individual to respond to a series of statement by indicating whether she to he strongly agrees (SA), agree (A), is undecided (U), disagree (D), or strongly disagree (SD), with each statement, (Likert R. 1932). Each response is associated with a point values, and an individual's score is determined by summing the point values might be assigned to responses to positive statements: SA=5, A=4, U=3, D=2 SD=1. For negative statement might be, "Short people are entitling to same job opportunities as tall people" (Gay L.R.,
1990). A high point value on a positively stated item would indicate a positive attitude and a high total score on the test would be indicative of positive attitude.

### 6.9.1.5 Personality Tests

The term 'personality' refers to the total functions of an individual who interacts with his environment (Singh, 1986). Such definition automatically includes all traits as the main theme of the personality. The purpose of the measurement of the personality is to describe a person in traits. Personality tests are usually self report interments. The individual checks responses to certain questions and statements. These instruments yield scores which are assumed of have been shown to measure certain personality traits or tendencies. Because of the difficulty, inability or unwillingness of individuals to report their own reactions accurately or objectively, these instruments may be of limited value. Part of the limitation may be due to inadequate theories of personality upon which some of inventories have based. At best they provide useful data upon which to suggest the need for further analysis.

The personality measurement aims at studying the following traits: social Traits i.e. those traits which determine how a person interact with other persons in a society, Motives i.e. includes the non biological drives such as need to earn money and prestige, [Personal conceptions i.e. include those methods which determine people’s attitude towards self and others and Adjustment i.e. includes traits like the freedom from emotional worries to instability and other disruptive behaviors.

**Reading:** For more information read the following

| Exploring Research | p, 137-146 |
6.10  SELF ASSESSMENT QUESTIONS

Q.1  How would you define measurement? Explain with few examples.

Q.2  Describe postulates of measurement. Illustrate your answer with some examples.

Q.3  Discuss in detail the characteristics of measurement.

Q.4  Why measurement is divided into different levels. Describe the levels of measurement?

Q.5  How would you differentiate nominal level from ratio level of measurement?

Q.6  Measurement plays very essential functions in different disciplines of social sciences. Discuss in detail.

Q.7  What is reliability, its definition and characteristics? Illustrate your answers with examples.

Q.8  How would you explain error Scores. Also discuss the factors which affect the error Scores.

Q.9  Discuss the common types of reliability.

Q.10 What do you understand by term index of reliability?

Q.11 What are the factors which influence the reliability of a test scores, give examples?

Q.12 How would you explain validity of a test? Describe its characteristics as well.

Q.13 Why is reliability more “important” than validity in the construction of a test?

Q.14 What purposes should be kept in mind while testing validity?

Q.15 Explain different types of validity.

Q.16 Like reliability, validity is also influenced by some factors. What are they Discuss?

Q.17 Defines tests, and discuss their types in detail.
6.11 BIBLIOGRAPHY


HISTORICAL AND
DESCRIPTIVE RESEARCH

Atifa Durrani
<table>
<thead>
<tr>
<th>Contents</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>166</td>
</tr>
<tr>
<td>Objectives</td>
<td>166</td>
</tr>
<tr>
<td>Historical Research</td>
<td>166</td>
</tr>
<tr>
<td>Definition of Purpose</td>
<td>166</td>
</tr>
<tr>
<td>Historical Research Process</td>
<td>168</td>
</tr>
<tr>
<td>Descriptive Research</td>
<td>169</td>
</tr>
<tr>
<td>Types of Descriptive Research</td>
<td>171</td>
</tr>
<tr>
<td>Self Assessment Questions</td>
<td>184</td>
</tr>
</tbody>
</table>
7.1 INTRODUCTION

A full understating of the reasons for conducting research studies in any field is essential for students to appreciate nature of research work (Matin, 1985). Our problem and difficulties in the field of women studies further necessitates a purposeful and sustained research effort.

Such research efforts are distinguished on the basis of their purposes and approaches, research studies differ significantly in their procedure e.g. selection of samples, tools for data collection and analysis of data. Further the basic concept of correlations, relational coefficient is also explained in the later part of unit.

7.2 OBJECTIVES

After studying this unit you should be able:

- To briefly state the definition, purpose and research process of historical and descriptive research.
- To explain different types of descriptive research.
- To compute correlational coefficient for variables

7.3 HISTORICAL RESEARCH

History is meaningful record of man’s achievement. It is not only a list of chronological events but truthful integrated account of the relationships between persons, events, times, and places. Man uses history to understand the past in the light of past events and developments, historical analysis may be directed towards an individual, an idea, a movement, or an institution (Best, 1970).

7.3.1 Definition and Purpose

According Gay (1987, p.179)“Historical Research is the systematic collection and objective evaluation of data related to past occurrences in order to
test hypothesis concerning causes, effects, or trends of those events that may help to explain present events and anticipate future events. Procedures supplementary to observation in which the researcher seeks to test the authenticity of the reports or observations made by others. The historical method is employed by researchers who are interested in reporting events and/or conditions that occurred in the past. An attempt is made to establish facts in order to arrive at conclusions concerning past events or predict future events.

Historical Research justifies itself when used to find out the solutions of the present day problems on the basis of experiences of the past. The nature of historical research produces exercise of many of the control procedures characteristic of other methods. Historical research also involves systematic, objective data collection and analysis and the confirmation or disconfirmation of the hypothesis (Gay, 1987).

History embraces the entire field of human endeavor and is as bread as life itself. Our past includes many areas of social experiences and activities that have proved more significant than political history e.g., culture, values, ideas, law, religion, literature, art, science, etc. there is general agreement among modern historians concerning the richness of the content of history, including social, cultural, economic and intellectual development, and on a board view of past events extending for beyond the study of polices, diplomatic, and war materials (Good, 1959). E.g., a knowledge of history of woman education can yield insight into the circumstances involved in the evolution of the current studies related to women and gender sensitivities as well as practices and approaches that have been found to be ineffective or infeasible (Gay, 1987). However, Best and Kahn (1992) describe the value of historical research as follow:

The history enables the researcher to detect fads and frills and enables the researcher to view present problems in the light of their origin and growth. Past experiences may serve as a basis for tentative generalization in analyzing current
issues and problems. Science predicts the future course of events and history uses present evidence to retrofits the past.

It is liable to be subjective. In history, rational process of analysis and inference is verbalized. However, the scope of historical research may be mentioned in the areas of biography, development of ideas through the ages, history of institution and organizations, etc.

7.3.2 Historical research process

When undertaking a historical study, a researcher engages in some activities that are common to all investigations but the nature of his subject matter present shim with some peculiar problems and requires him to apply some special standards and techniques.

The steps involved in conducting a historical research study are essentially the same as for other types of research. These steps are as follows:

1. Definition of the problem.
2. Formulation of hypothesis.
3. Data collection.
4. Data analysis.
5. Data synthesis.

Now let us briefly discuss these steps. The selection and development of the problem for a study is the first step in historical research. Historical inquiry begins when some event, experience of the past is questioned. A historian can investigate individuals, institutions, important concepts and ideas that have influenced women education during a specific period of time (Van Dalen, 1985).

It is much better to study in depth a well defined problem with one or more specific, well sated hypotheses, then to investigate either a too-broadly sated problem with a fuzzy hypotheses or a problem for which insufficient data are available. (Gay, 1987) collecting data in historical research, the sources are normally classified into two main categories i.e. primary and secondary sources.
Primary sources constitute first hand information such as original documents and reports by actual participants or direct observers. Secondary sources constitute second hands information such as reference books (encyclopedia for example) or reports by relatives of actual participants or observers.

All sources of historical data must be subjected to rigorous scientific analysis to determine both their authenticity (external criticism) and their accuracy (internal criticism).

In determining the accuracy of the documents there are at least four factors that must be considered:

Knowledge and competence of the author
The time delay between the occurrence and recording of events
Biased motives of the author
Consistency of the data (Gay, 1987)

As with the review of related literature historical data should be organized, synthesized, conclusion and generalization formulated. Since summarization of historical research data involves logical analysis rather than statistical analysis the researcher must take care to be as objective as possible.

Reading: For more information read the following
Exploring Research p. 204-210

7.4 DESCRIPTIVE RESEARCH

Descriptive research involves the collection of data in order to test hypotheses or to answer questions concerning the current status of the subjects of the study. This research is the most commonly used and the basic reason for carrying out descriptive research is to identify the cause of something that is happening. For instance, this research could be used in order to find out what age
group is buying a particular brand of cola, whether a company’s market share differs between geographical regions or to discover how many competitors a company has in their marketplace. However, if the research is to return useful results, whoever is conducting the research must comply with strict research requirements in order to obtain the most accurate figures/results possible.

7.4.1 Definition and purpose

Descriptive research, also known as statistical research, describes data and characteristics about the population or phenomenon being studied. Descriptive research answers the questions who, what, where, when and how.

Although the data description is factual, accurate and systematic, the research cannot describe what caused a situation. Thus, descriptive research cannot be used to create a causal relationship, where one variable affects another. In other words, descriptive research can be said to have a low requirement for internal validity.

The description is used for frequencies, averages and other statistical calculations. Often the best approach, prior to writing descriptive research, is to conduct a survey investigation. Qualitative research often has the aim of description and researchers may follow-up with examinations of why the observations exist and what the implications of the findings are.

This type of research helps in locating existing problem, in securing historical perspective through a series of cross-sectional pictures of similar condition at different times, in suggesting the course of future developments, in developing many tools, in contributing to the advancement of knowledge and in providing background ideas and data from which many more studies may be conducted, such studies are factual and hence supply practical information (Best & Kahn, 1992)
7.4.3 Process of Descriptive research

In descriptive studies, the researcher follows certain procedures and steps and does no present private convictions or data based on casual observation. (Akbar, 1990). According to Van Dalen (1995) the following steps may be take by the researcher to carry out a good descriptive research.

1. Examine the problematic situation.
2. Define the problem and sometimes state the hypotheses.
3. List the assumption upon which these hypotheses and procedures are based.
4. Selected appropriate subject and source material.
5. Select or construct techniques for collecting the data.
6. Establish categories for classifying the data are appropriate for the purpose of the capable to bringing about significant differences or relationship.
7. Validate the data gathering techniques.
8. Make discriminating and objective observations to collect data.
9. Describe analyze and interpret their findings in clear and precise terms.

7.4.4 Types of Descriptive Research

Different research scientists have classified descriptive research in different ways.

Salkind (1992) placed descriptive research in three categories:

Developmental Studies
Correlational studies
Case studies

7.4.4.1 Developmental Studies

As you know already that concept of development is essentially biological and has been most commonly associated with living structures and life processes.
However, the concept has come to be applied to physical system as well as to social institutions, cultural and other system (Good, 1959) explains that the purpose of developmental studies is to discover origin, direction, trend, pattern, rate and patterns of growth with a somewhat more recent interest in cause and inter relationship affecting growth.

Developmental studies are further divided into two types:

1. Longitudinal Method
2. Cross-sectional Method (Salkind, 1992)

1. Longitudinal research studies

Longitudinal research provides data from the same participants over a set time period and as such permits causal pathways (for example to health, illness and mortality), to be determined. There is a new emphasis on accountability in the public health sector. Longitudinal research can help those professionals seeking to implement longitudinal style research in order to meet these new demands. In longitudinal studies the growth of a particular group is followed over a period of months or years.

*Conducting Longitudinal Research* will help both novice and experienced researchers, from academia, government departments, private and public sectors to establish and conduct a longitudinal study. Offering direction and advice concerning the efficient conduct of longitudinal research studies, *Conducting Longitudinal Research* fills a gap in the research methodology literature.

Internationally, longitudinal research has become increasingly important to both the academic community and state policy-makers as it is an important way to examine causal relationships - for example, understanding critical issues associated with ageing.
Conducting Longitudinal Research includes topics from the researchers' experiences:

- Strategies to encourage participants to remain in the study (often for decades)
- Establishing succession planning for key personnel, and
- Handling very large volumes of data.

Conducting Longitudinal Research can be read from cover to cover, although it is designed so that each section stands alone, allowing readers a readily accessible guide to overcoming the diverse challenges that occur during longitudinal research. For example:

- When a participant calls the study team and asks for help with a health problem, what procedure should be used?
- How should the study team deal with angry callers?
- What measures can be set in place to avoid data loss?
- How should the research team ensure that participants are not lost to follow-up?

2. Cross sectional studies

A cross-sectional study is the simplest variety of descriptive or observational epidemiology that can be conducted on representative samples of a population. Simply put, it is a study that aims to describe the relationship between diseases (or other health-related states) and other factors of interest as they exist in a specified population at a particular time, without regard for what may have preceded or precipitated the health status found at the time of the study. For instance, a single cross-sectional study may include questions about smoking behavior, occupational exposure to dusts and fumes, respiratory symptoms
(cough, breathlessness), and physical examinations of physical fitness—including simple tests of lung function. Such a study would throw some light on the relationship of both occupational exposures and smoking behavior to respiratory symptoms and respiratory function. However, it is impossible either to establish causal relationships or to get reliable perspectives on the natural history of respiratory disease from such a study.

Cross-sectional studies must be done on representative samples of the population if generalizations from the findings are to have any validity. These studies gather information about the prevalence of health-related states and conditions, but they cannot distinguish between newly occurring and long-established conditions. All they can do is measure the frequency (prevalence) of conditions and demonstrate associations. They cannot identify cause-and-effect relationships, though they do identify the existence of health problems.

Cross-sectional studies, also known as surveys, are a useful way to gather information on important health-related aspects of people's knowledge, attitudes, and practices (such studies are known as "KAP" surveys). In the area of reproductive health, such a survey might include questions such as: How much do girls and women in their reproductive years know about pregnancy and control over their own fertility? What are their beliefs, values, and attitudes towards making decisions about getting pregnant and about desired family size? How do they control their own fertility? KAP surveys are a good example of a tried and tested form of cross-sectional study. Many have been conducted serially to measure the efficacy of family-planning programs, anti-smoking measures, and other public health and health-promotion interventions. The distinction between a cohort study and a repeated cross-sectional study is that a cohort study is conducted with the same individuals who participate over a long period; repeated
or serial cross-sectional studies, on the other hand, do not necessarily (or even usually) study the same individuals repeatedly.

For more information read the following:

| Exploring Research p, 219-222 |

### 7.4.4.2 Correlational research

Correlation Research a statistical measure of a relationship between two or more variables, gives an indication of how one variable may predict another. Correlational research is used to explore co-varying relationships between two or more variables. A simple definition of a co-varying relationship is as one variable changes so does the other variable(s). The purpose of correlational research is to:

1. to identify variables that relate to one each other (i.e. is there a relationship between family income and grade point average; is there a relationship between part time employment and grade point average);
2. to make predictions of one variable from another variable (i.e. can I.Q. test scores be used to predict student achievement; can SAT scores be used to predict college grade point averages);
3. To examine possible cause and effect relationships between one variable and another.

A caution has to be advised when considering correlational research and cause and effect. Major researchers such as B.F. Skinner posit that while we can make many conclusions identifying a relationship between one or more variables, establishing cause and effect is very difficult and maybe impossible due to the myriad interactions of many variables in social science research.
In education-based correlational studies, data is frequently collected using standardized measures such as test scores. Report presentations almost always use hypotheses in the form of "No relationship exists between variable X and variable Y." Data analysis using correlation coefficients is generally quantitative. Rather than rich descriptive narrative as we might see in descriptive or ethnographic studies, correlation presentations tend to be succinct relying on statistical analyses of correlation coefficients and regression. Of the various quantitative methodologies, correlational research is among the easiest to design and apply. For this reason, it is popular and frequently used in conjunction with other research methodologies.

Some other types of Descriptive Research are: Follow up studies, Survey studies, Trend analyses, Community studies, Casual comparative studies, Documentary analysis etc.

7.4.4.3 Case Study

Case study is an ideal methodology when a holistic, in-depth investigation is needed. Case studies have been used in varied investigations, particularly in sociological studies, but increasingly, in instruction. Yin, Stake, and others who have wide experience in this methodology have developed robust procedures. When these procedures are followed, the researcher will be following methods as well developed and tested as any in the scientific field. Whether the study is experimental or quasi-experimental, the data collection and analysis methods are known to hide some details. Case studies, on the other hand, are designed to bring out the details from the viewpoint of the participants by using multiple sources of data (for further detail read from C-4645).
7.5 WHAT IS A CORRELATION?

Scientists are interested in relationships between variables. When two variables vary together (a change in one is accompanied by a change in the other), we say they are correlated.

A correlation is a number between -1 and +1 that measures the degree of association between two variables (call them X and Y). A positive value for the correlation implies a positive association (large values of X tend to be associated with large values of Y and small values of X tend to be associated with small values of Y). A negative value for the correlation implies a negative or inverse association (large values of X tend to be associated with small values of Y and vice versa). A graph of a collection of pairs of scores is called scattered plot.

7.5.1 Correlational Coefficient

The correlation coefficient a concept from statistics is a measure of how well trends in the predicted values follow trends in past actual values. It is a measure of how well the predicted values from a forecast model "fit" with the real-life data.

The correlation coefficient is a number between 0 and 1. If there is no relationship between the predicted values and the actual values the correlation coefficient is 0 or very low (the predicted values are no better than random numbers). As the strength of the relationship between the predicted values and actual values increases so does the correlation coefficient. A perfect fit gives a coefficient of 1.0. Thus the higher the correlation coefficient the better. In other words, the correlation coefficient measures the strength of a linear relationship between two variables.
The correlation coefficient is always between -1 and +1. The closer the correlation is to +/-1, the closer to a perfect linear relationship. Here is how I tend to interpret correlations.

- 1.0 to -0.7 strong negative association.
- -0.7 to -0.3 weak negative association.
- -0.3 to +0.3 little or no association.
- +0.3 to +0.7 weak positive association.
- +0.7 to +1.0 strong positive association.

The strength of the linear association between two variables is quantified by the correlation coefficient.

Given a set of observations \((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\), the formula for computing the correlation coefficient is given by

\[
 r = \frac{1}{n-1} \sum \left( \frac{x_i - \bar{x}}{s_x} \right) \left( \frac{y_i - \bar{y}}{s_y} \right)
\]

The correlation coefficient always takes a value between -1 and 1, with 1 or -1 indicating perfect correlation (all points would lay along a straight line in this case). A positive correlation indicates a positive association between the variables (increasing values in one variable correspond to increasing values in the other variable), while a negative correlation indicates a negative association between the variables (increasing values in one variable correspond to decreasing values in the other variable). A correlation value close to 0 indicates no association between the variables.
Since the formula for calculating the correlation coefficient standardizes the variables, changes in scale or units of measurement will not affect its value. For this reason, the correlation coefficient is often more useful than a graphical depiction in determining the strength of the association between two variables.

A scatter plot is a graph used to determine whether there is a relationship between paired data. In many real-life situations, scatter plots follow patterns that are approximately linear. If $y$ tends to increase as $x$ increases, then the paired data are said to be a positive correlation. If $y$ tends to decrease as $x$ increases, the paired data are said to be a negative correlation. If the points show no linear pattern, the paired data are said to have relatively no correlation.

A scatter plot is often employed to identify potential associations between two variables, where one may be considered to be an explanatory variable (such as years of education) and another may be considered a response variable (such as annual income). A positive association between education and income would be indicated on a scatter plot by a upward trend (positive slope), where higher incomes correspond to higher education levels and lower incomes correspond to fewer years of education. A negative association would be indicated by the opposite effect (negative slope), where the most highly educated individuals would have lower incomes than the least educated individuals. Or, there might not be any notable association, in which case a scatter plot would not indicate any trends whatsoever. The following plots demonstrate the appearance of positively associated, negatively associated, and non-associated variables:
Example of a strong positive association is the correlation between blood viscosity and packed cell volume is 0.88. Notice that small volumes tend to have low viscosity and large volumes tend to have high viscosity.

Example of a weak positive association is the correlation between blood viscosity and fibrogen is 0.46. Notice that there is also a tendency for small fibrogen values to have low viscosity and for large fibrogen values to have high viscosity. This tendency, however, is less pronounced than in the previous example.
Example of little or no association is the correlation between blood viscosity and plasma protein is -0.10. Low levels of protein are associated with both high and low viscosities. High levels of protein are also associated with both high and low viscosities.

A number of correlation methods are used for different situations which are Pearson R Correlation Coefficient, and Spearman R Correlation Coefficient.

7.5.2 Pearson R Correlation Coefficient

Pearson's product moment correlation coefficient, usually denoted by r, is one example of a correlation coefficient. It is a measure of the linear association between two variables that have been measured on interval or ratio scales, such as the relationship between height in inches and weight in pounds. However, it can be misleadingly small when there is a relationship between the variables but it is a non-linear one.

There are procedures, based on r, for making inferences about the population correlation coefficient. However, these make the implicit assumption that the two variables are jointly normally distributed. When this assumption is not justified, a non-parametric measure such as the Spearman Rank Correlation Coefficient might be more appropriate.

7.5.3.1 Computing the Pearson Correlation Coefficient

One formula for the Pearson correlation coefficient r is as follows:

\[
    r = \frac{\sum XY - (\frac{\sum X)(\sum Y)}{n}}{\sqrt{\left( \sum X^2 - (\frac{\sum X)^2}{n} \right) \left( \sum Y^2 - (\frac{\sum Y)^2}{n} \right)}}
\]

(10.1)

The following numerical example shows how the formula is used:
\[ \sum XY = (1)(2) + (3)(5) + (4)(5) + (4)(8) = 69 \]
\[ \sum X = 1 + 3 + 4 + 4 = 12 \]
\[ \sum Y = 2 + 5 + 5 + 8 = 20 \]
\[ \sum X^2 = 1^2 + 3^2 + 4^2 + 4^2 = 42 \]
\[ \sum Y^2 = 2^2 + 5^2 + 5^2 + 8^2 = 118 \]
\[ r = \frac{69 - \frac{(\sum X)(\sum Y)}{4}}{\sqrt{\left(42 - \frac{(\sum X)^2}{4}\right)\left(118 - \frac{(\sum Y)^2}{4}\right)}} = .866 \]

7.5.4 Spearman R Correlation Coefficient

The Spearman rank correlation coefficient is one example of a correlation coefficient. It is usually calculated on occasions when it is not convenient, economic, or even possible to give actual values to variables, but only to assign a rank order to instances of each variable. It may also be a better indicator that a relationship exists between two variables when the relationship is non-linear.

Commonly used procedures, based on the Pearson’s Correlation Coefficient, for making inferences about the population correlation coefficient make the implicit assumption that the two variables are jointly normally
distributed. When this assumption is not justified, a non-parametric measure such as the Spearman Rank Correlation Coefficient might be more appropriate.

The *Spearman Rank Correlation Coefficient* is a form of the Pearson coefficient with the data converted to rankings (i.e. when variables are ordinal). It can be used when there is non-parametric data and hence Pearson cannot be used.

The raw scores are converted to ranks and the differences ($d_i$) between the ranks of each observation on the two variables are calculated. The Spearman coefficient is denoted with the Greek letter rho ($\rho$).

$$\rho = 1 - \frac{6 \times \text{SUM}(d_i^2)}{n \times (n^2 - 1)}$$

For example two groups, $x$ and $y$, are asked to rank ten items. The correlation between their rankings are then compared as below

<table>
<thead>
<tr>
<th>Item</th>
<th>$x$</th>
<th>$y$</th>
<th>$x-y$</th>
<th>$(x-y)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>5</td>
<td>-2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>6</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>7</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

$n$: 10 \hspace{2cm} \text{sum}((x-y)^2): 18

$n \times (n^2 - 1): 990 \hspace{2cm} 6 \times \text{sum}((x-y)^2): 108$

$\rho = 6 \times \text{sum}((x-y)^2) / n \times (n^2 - 1) \hspace{1cm} 0.11$
The correlation of 0.11 is quite low, showing that they agree very little. The Spearman Coefficient can be used to measure ordinal data (i.e. in rank order), not interval (as Pearson). It effectively works by first ranking the data then applying Pearson's calculation to the rank numbers. This coefficient is also called *Spearman's rho* (after the Greek letter used).

Reading: For more details read the following;

| Exploring Research, p.223-228 |

7.6 **SELF ASSESSMENT QUESTIONS**

Q.1 What is historical research. Describe its *definition* and purpose.

Q.2 Discuss the research process of historical research.

Q.3 Differentiate between the primary and secondary of data in historical research. Illustrate your answer with examples.

Q.4 Explain Descriptive research and its types in detail.

Q.5 Consider the following data:

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>-10</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
</tr>
</tbody>
</table>
(a) Compute the correlation between X and Y.

(b) Compute the correlation between Y and X.

(c) Add 5 to Y, so the new values are 5, 8, 15, 6, and 20. Now compute the correlation between X and Y. Is the correlation smaller, larger, or the same as before?

(d) Multiply Y by 5, so the new values are 0, 15, 50, 5, and 75. Now compute the correlation between X and Y. Is the correlation smaller, larger, or the same as before?

(e) Multiply Y by -1, so that the new values are 0, -3, -10, -1, -15. Now compute the correlation between X and Y. Is the correlation smaller, larger, or the same as before?

Q.6 The following data presents the per capita income of 20 European OECD countries for 1960 and as well as the percentages of the labor force employed in agriculture, industry, and services for each country.

\[ PCINC = \text{Per capita income, 1960 (}$) \]

\[ AGR = \text{Percent of labor force in agriculture, 1960} \]

\[ IND = \text{Percent of labor force in industry, 1960} \]

\[ SER = \text{Percent of labor force in service occupations, 1960} \]

<table>
<thead>
<tr>
<th>COUNTRY</th>
<th>PCINC</th>
<th>AGR</th>
<th>IND</th>
<th>SER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pakistan</td>
<td>1536</td>
<td>13</td>
<td>43</td>
<td>45</td>
</tr>
<tr>
<td>Sweden</td>
<td>1644</td>
<td>14</td>
<td>53</td>
<td>33</td>
</tr>
<tr>
<td>Switzerland</td>
<td>1361</td>
<td>11</td>
<td>56</td>
<td>33</td>
</tr>
<tr>
<td>Luxemburg</td>
<td>1242</td>
<td>15</td>
<td>51</td>
<td>34</td>
</tr>
<tr>
<td>Country</td>
<td>PCINC</td>
<td>AGR</td>
<td>IND</td>
<td>INDSER</td>
</tr>
<tr>
<td>--------------</td>
<td>-------</td>
<td>-----</td>
<td>-----</td>
<td>--------</td>
</tr>
<tr>
<td>U. Kingdom</td>
<td>1105</td>
<td>4</td>
<td>56</td>
<td>40</td>
</tr>
<tr>
<td>Denmark</td>
<td>1049</td>
<td>18</td>
<td>45</td>
<td>37</td>
</tr>
<tr>
<td>W. Germany</td>
<td>1035</td>
<td>15</td>
<td>60</td>
<td>25</td>
</tr>
<tr>
<td>France</td>
<td>1013</td>
<td>20</td>
<td>44</td>
<td>36</td>
</tr>
<tr>
<td>Belgium</td>
<td>1005</td>
<td>6</td>
<td>52</td>
<td>42</td>
</tr>
<tr>
<td>Norway</td>
<td>977</td>
<td>20</td>
<td>49</td>
<td>32</td>
</tr>
<tr>
<td>Iceland</td>
<td>839</td>
<td>25</td>
<td>47</td>
<td>29</td>
</tr>
<tr>
<td>Netherlands</td>
<td>810</td>
<td>11</td>
<td>49</td>
<td>40</td>
</tr>
<tr>
<td>Austria</td>
<td>681</td>
<td>23</td>
<td>47</td>
<td>30</td>
</tr>
<tr>
<td>Ireland</td>
<td>529</td>
<td>36</td>
<td>30</td>
<td>34</td>
</tr>
<tr>
<td>Italy</td>
<td>504</td>
<td>27</td>
<td>46</td>
<td>28</td>
</tr>
<tr>
<td>Japan</td>
<td>344</td>
<td>33</td>
<td>35</td>
<td>32</td>
</tr>
<tr>
<td>Greece</td>
<td>324</td>
<td>56</td>
<td>24</td>
<td>20</td>
</tr>
<tr>
<td>Spain</td>
<td>290</td>
<td>42</td>
<td>37</td>
<td>21</td>
</tr>
<tr>
<td>Portugal</td>
<td>238</td>
<td>44</td>
<td>33</td>
<td>23</td>
</tr>
<tr>
<td>Turkey</td>
<td>177</td>
<td>79</td>
<td>12</td>
<td>9</td>
</tr>
</tbody>
</table>

Using your TI-83 calculator, compute the correlation coefficient of the following pairs: PCINC vs. AGR, PCINC vs. IND and PCINC vs. INDSER.

(a) Which among the three labor sectors provides the strongest linear relationship with per capita income?
(b) If majority of the labor force works in agriculture, would you expect a higher per capita income?

(c) Suppose PCINC (per capita income) is coded in thousands of dollars instead, what happens to the correlation coefficients?

Q.7 Consider the first and second exam scores of 35 & at 216 students:

<table>
<thead>
<tr>
<th>Student</th>
<th>First</th>
<th>Second</th>
<th>Student</th>
<th>First</th>
<th>Second</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>22</td>
<td>19</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>23</td>
<td>23</td>
<td>20</td>
<td>13</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>19</td>
<td>21</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>19</td>
<td>22</td>
<td>23</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>23</td>
<td>24</td>
<td>23</td>
<td>11</td>
<td>21</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
<td>21</td>
<td>24</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>18</td>
<td>25</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>16</td>
<td>26</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
<td>20</td>
<td>27</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>10</td>
<td>28</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>11</td>
<td>22</td>
<td>24</td>
<td>29</td>
<td>21</td>
<td>17</td>
</tr>
<tr>
<td>12</td>
<td>22</td>
<td>22</td>
<td>30</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>13</td>
<td>23</td>
<td>22</td>
<td>31</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>14</td>
<td>18</td>
<td>19</td>
<td>32</td>
<td>22</td>
<td>16</td>
</tr>
<tr>
<td>15</td>
<td>22</td>
<td>23</td>
<td>33</td>
<td>18</td>
<td>16</td>
</tr>
<tr>
<td>16</td>
<td>20</td>
<td>20</td>
<td>34</td>
<td>21</td>
<td>13</td>
</tr>
<tr>
<td>17</td>
<td>20</td>
<td>20</td>
<td>35</td>
<td>19</td>
<td>24</td>
</tr>
<tr>
<td>18</td>
<td>20</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(a) Draw a scatter plot of the data. How are the two exam scores related based on the plot? Would you say that this relationship is strong?

(b) Compute the correlation coefficient between the first and second exam scores. Does this value support your judgment in the previous question?

(c) The Stat 216 director decided to curve the first exam scores by giving away 5 points.
   i. Obtain a new scatter plot and compare this with the old plot.
   ii. What happens to the correlation coefficient? Explain this behavior.

(d) Fifth and eleventh students were found cheating during the second exam. As a result, they were both given zeros in that exam. What will happen now to the correlation coefficient? Can you consider this new value reliable? Explain why.
EXPERIMENTAL DESIGNS

Atif: Durrani
<table>
<thead>
<tr>
<th>Contents</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>191</td>
</tr>
<tr>
<td>Objectives</td>
<td>191</td>
</tr>
<tr>
<td>Experimental designs</td>
<td>191</td>
</tr>
<tr>
<td>Types of experimental designs</td>
<td>195</td>
</tr>
<tr>
<td>Internal and external validity in experiment designs</td>
<td>205</td>
</tr>
<tr>
<td>Self assessment questions</td>
<td>209</td>
</tr>
<tr>
<td>Bibliography</td>
<td>209</td>
</tr>
</tbody>
</table>
8.1 INTRODUCTION

By experiment we refer to is that portion of research in which some variables are manipulated and their effect upon each other are observed, thus, experiment is the process for gaining knowledge by collecting new or fresh observation under controlled conditions. When a researcher fails to find out through simple observation the possible factors operating in a given problem, he or she conducts experiment. The difference between simple observation and experiment is that in experiment, observation is done under controlled situation.

Earlier you have read about the different types of research. Experimental research comes under the classification of research by method. While reading experimental research, you may have noticed that in this type of research cause and effect relationships are established and through experimental method, these relationships are studied. The researcher has direct control over the independent variables. The first active characteristic of experimental research is that it provides an opportunity to the researcher to manipulate some of the variables of concern while controlling the effects of others.

The purpose of experimental design is to test a structured or well-defined hypothesis. Research designs can be described as blue print for the collection and analysis of data. There are different types of experimental designs which are used according to the requirements of the research study. It is also known fact that, experimental Designs have their problems, regarding validity of research study. Therefore, the validity has to be checked otherwise the research cannot be considered valid. External and internal validity are types of validity which may affect experimental designs.

In this unit an effort has been made to describe experimental designs, their types and effect of external and internal validity on the research designs.
8.2 OBJECTIVES
After reading the units you will be able to:
1. Describe experimental designs and their types.
2. Explain internal and external validity of experimental design.

8.3 EXPERIMENTAL DESIGNS
Emory (1970) described a research design as, “it is a strategy for a study as well as the plan by which the strategy is to be carried out. It specifies the methods and procedures for the collection, measurement analysis of data”

An experiment involves the creation of a contrived situation in order that the researcher can manipulate one or more variables whilst controlling all of the others and measuring the resultant effects. Boyd and Westfall (1972) have defined experimentation as:

"...that research process in which one or more variables are manipulated under conditions which permit the collection of data which show the effects, if any, in unconfused fashion."

Experiments can be conducted either in the field or in a laboratory setting. When operating within a laboratory environment, the researcher has direct control over most, if not all, of the variables that could impact upon the outcome of the experiment. For example, an agricultural research station may wish to compare the acceptability of a new variety of maize. Since the taste characteristics are likely to have a major influence on the level of acceptance, a blind taste panels might be set up where volunteers are given small portions of maize porridge in unmarked bowls. The participants would perhaps be given two porridge samples and the researcher would observe whether they were able to distinguish between the maize varieties and which they preferred. In addition to taste testing,
laboratory experiments are widely used by marketing researchers in concept testing, package testing, advertising research and test marketing.

Today a good research can not be carried out without preparation of a research design. Once the hypotheses have been formulated and defined, researcher chooses a research design. There are different types of research designs like descriptive research designs that deals with description of any institution or community, exploratory design is applicable to the familiarity of a new phenomenon or community, etc.

According to Brown (1980) there are two main types of designs for conducting research, “experimental designs through which the researcher intervenes by introducing selected facts for the purpose of observing their effects, and non experimental designs through which the researcher intervenes only to the extent necessary to make observations.”

There are many experimental designs that a researcher can use depending upon the study conditions. Experimental designs provide greater certainty in research study. When experimentation is possible, this is the most effective method for testing a hypothesis. Experimental research designing has been used in every social science research. This research design is scientific in its approach. However, it is not easy to develop experimental design for data collection.

In experimental method, three inter-related conditions are required which must be satisfied, otherwise the experiment may not be performed as desired by the researcher. These conditions are:

1. Control
2. Randomization
3. Replication
8.3.1 Control

It is a basic element of the method. External factors which are not the part of hypothesis are controlled through this way. So that cannot influence the result of the experiment and ultimately the research study.

8.3.2 Randomization

It is the second basic principle. It makes the experiment valid. The statistical test which is used in an experimental situation depends upon some assumptions. One of them is very common that is the observations which should be independent: The independence of the observations is maintained when the samples are randomly drawn from the population. Thus randomization ensures the independence of the observations which, in turn makes a statistical test valid. Sometimes it has been found that complete randomization becomes difficult. This is specially true when an experimenter is dealing with organismic variables Ostle and Mensing (1975) presented a suggesting to deal with this type of situations in a better way as “there are situations in which complete randomization is either impossible or uneconomical. The statistician should not, therefore, adopt the unrelenting position of insisting on complete randomization in every case. On the other hand, neither should agree to the use of complete systematic design.

8.3.3 Replication

The term replication is a fusion of two words, namely duplication and repetition. In experimentation, it refers to the deliberate repetition of an experiment, using a nearly identical procedure with a different set of subjects, in a different setting and at a different time. Winter (1971) described replication as, “A replication of an experiment is an independent repetition under as nearly identical conditions as the name of the experimental material will permit”. In experiment design it implies performing various sub-experiments i.e. replication within a framework of one experimental design.
8.4 TYPES OF EXPERIMENTAL DESIGNS

The most widely accepted classification is based on the characteristic of control of a research design. In these terms there are three basic types of experimental designs:

1. Pre-experimental designs
2. True experimental designs
3. Quasi experimental designs

To make things easier, the following will act as representations within particular designs:

- X: Treatment
- O: Observation or measurement
- R: Random assignment

Beside these, in addition, there are some more sophisticated designs which are used to increase the power of these basic designs. These designs enable the investigator more effectively to control variables, some designs help to study the simultaneous effect of more than independent variables.

8.4.1 Pre-Experimental designs

Pre-experimental designs are presented to illustrate the crudest form of experimentation. They are the weakest designs in terms of their scientific value and measurement power. These designs usually fail to provide control over extraneous effects the outcome of the experiment. In other words, thes designs do not provide a control group or the equivalent of a control group. Some of the types of pre-experimental designs are as under:
8.4.1.1 The one-shot case study

This is a single group studied only once. A group is introduced to a treatment or condition and then observed for changes which are attributed to the treatment.

X 0

The Problems with this design are:

- A total lack of control. Also, it is of very little scientific value as securing scientific evidence to make a comparison, and recording differences or contrasts.
- There is also a tendency to have the error of misplaced precision, where the researcher engages in tedious collection of specific detail, careful observation, testing and etc., and misinterprets this as obtaining good research. However you can not misinterpret that a detailed data collection procedure equals a good design.
- History, maturation, selection, mortality and interaction of selection and the experimental variable are all threats to the internal validity of this design.

8.4.1.2 One group pretest-posttest design

This is a presentation of a pretest, followed by a treatment, and then a posttest where the difference between O_1 and O_2 is explained by X:

O_1 X O_2

However, there exist threats to the validity of the above assertion:

- History: between O_1 and O_2 many events may have occurred apart from X to produce the differences in outcomes. The longer the time lapse between O_1 and O_2, the more likely history becomes a threat.
- Maturation: between $O_1$ and $O_2$ students may have grown older or internal states may have changed and therefore the differences obtained would be attributable to these changes as opposed to X.

- Testing: the effect of giving the pretest itself may affect the outcomes of the second test (i.e., IQ tests taken a second time result in 3-5 point increase than those taking it the first time). In the social sciences, it has been known that the process of measuring may change that which is being measured: the reactive effect occurs when the testing process itself leads to the change in behavior rather than it being a passive record of behavior (reactivity: we want to use non-reactive measures when possible).

- Instrumentation: examples are in threats to validity above

- Statistical regression: or regression toward the mean. Time-reversed control analysis and direct examination for changes in population variability are useful precautions against such misinterpretations. What this means is that if you select samples according to their extreme characteristics or scores, the tendency is to regress toward the mean. Therefore those with extreme high scores appear to be decreasing their scores, and those with extreme low scores appear to be increasing their scores. However this interpretation is not accurate, and to control for misinterpretations, researchers may want to do a time-reversed (posttest-pretest) analysis to analyze the true treatment effects. Researchers may exclude outliers from the analysis.

- Others: History, maturation, testing, instrumentation interaction of testing and maturation, interaction of testing and the experimental variable and the interaction of selection and the experimental variable are also threats to validity for this design.
8.4.1.3 Static-group comparison

This is a two group design, where one group is exposed to a treatment and the results are tested while a control group is not exposed to the treatment and similarly tested in order to compare the effects of treatment.

\[
\begin{array}{c}
X & O_1 \\
& O_2 \\
\end{array}
\]

Threats to validity include:

- Selection: groups selected may actually be disparate prior to any treatment.
- Mortality: the differences between \( O_1 \) and \( O_2 \) may be because of the drop-out rate of subjects from a specific experimental group, which would cause the groups to be unequal.
- Others: Interaction of selection and maturation and interaction of selection and the experimental variable.

8.4.2 True experimental designs

The main characteristic of these designs are that they provide a mean by which a researcher can assure equivalence between experiment and control groups though random assignment to the groups. These designs are:

8.4.2.1 Pre-test post test control group design

This design takes on this form:

\[
\begin{array}{c}
R & O_1 & X & O_2 \\
R & O_3 & O_4 \\
\end{array}
\]

This design controls for all of the seven threats to validity described in detail so far. An explanation of how this design controls for these threats is below.
History: this is controlled in that the general history events which may have contributed to the \( O_1 \) and \( O_2 \) effects would also produce the \( O_3 \) and \( O_4 \) effects. This is true only if the experiment is run in a specific manner: meaning that you may not test the treatment and control groups at different times and in vastly different settings as these differences may effect the results. Rather, you must test simultaneously the control and experimental groups. Intrasession history must also be taken into consideration. For example if the groups truly are run simultaneously, then there must be different experimenters involved, and the differences between the experimenters may contribute to effects.

A solution to history in this case is the randomization of experimental occasions: balanced in terms of experimenter, time of day, week and etc.

- Maturation and testing: these are controlled in that they are manifested equally in both treatment and control groups.
- Instrumentation: this is controlled where conditions control for intrasession history, especially where fixed tests are used. However when observers or interviewers are being used, there exists a potential for problems. If there are insufficient observers to be randomly assigned to experimental conditions, the care must be taken to keep the observers ignorant of the purpose of the experiment.
- Regression: this is controlled by the mean differences regardless of the extremely of scores or characteristics, if the treatment and control groups are randomly assigned from the same extreme pool. If this occurs, both groups will regress similarly, regardless of treatment.
- Selection: this is controlled by randomization.
- Mortality: this was said to be controlled in this design, however upon reading the text, it seems it may or may not be controlled for. Unless the
mortality rate is equal in treatment and control groups, it is not possible to indicate with certainty that mortality did not contribute to the experiment results. Even when even mortality actually occurs, there remains a possibility of complex interactions which may make the effects drop-out rates differ between the two groups. Conditions between the two groups must remain similar: for example, if the treatment group must attend treatment session, then the control group must also attend sessions where either not treatment occurs, or a "placebo" treatment occurs. However even in these there remains possibilities of threats to validity. For example, even the presence of a "placebo" may contribute to an effect similar to the treatment; the placebo treatment must be somewhat believable and therefore may end up having similar results!

The factors described so far effect internal validity. These factors could produce changes which may be interpreted as the result of the treatment. These are called main effects which have been controlled in this design giving it internal validity.

However, in this design, there are threats to external validity (also called interaction effects because they involve the treatment and some other variable the interaction of which cause the threat to validity). It is important to note here that external validity or generalizability always turns out to involve extrapolation into a realm not represented in one's sample.

In contrast, internal validity is solvable within the limits of the logic of probability statistics. This means that we can control for internal validity based on probability statistics within the experiment conducted, however, external validity or generalizability can not logically occur because we can't logically extrapolate to different conditions. (Hume's truism that induction or generalization is never fully justified logically).
External threats include:

- Interaction of testing and X: because the interaction between taking a pretest and the treatment itself may affect the results of the experimental group, it is desirable to use a design which does not use a pretest.

- Interaction of selection and X: although selection is controlled for by randomly assigning subjects into experimental and control groups, there remains a possibility that the effects demonstrated hold true only for that population from which the experimental and control groups were selected. An example is a researcher trying to select schools to observe, however has been turned down by 9, and accepted by the 10th. The characteristics of the 10th school may be vastly different than the other 9, and therefore not representative of an average school. Therefore in any report, the researcher should describe the population studied as well as any populations which rejected the invitation.

- Reactive arrangements: this refers to the artificiality of the experimental setting and the subject's knowledge that he is participating in an experiment. This situation is unrepresentative of the school setting or any natural setting, and can seriously impact the experiment results. To remediate this problem, experiments should be incorporated as variants of the regular curricula, tests should be integrated into the normal testing routine, and treatment should be delivered by regular staff with individual students.

Research should be conducted in schools in this manner: ideas for research should originate with teachers or other school personnel. The designs for this research should be worked out with someone expert at research methodology, and the research itself carried out by those who came up with the research idea.
Results should be analyzed by the expert, and then the final interpretation delivered by an intermediary.

Tests of significance for this design: although this design may be developed and conducted appropriately, statistical tests of significance are not always used appropriately.

- Wrong statistic in common use: many use a t-test by computing two ts, one for the pre-post difference in the experimental group and one for the pre-post difference of the control group. If the experimental t-test is statistically significant as opposed to the control group, the treatment is said to have an effect. However this does not take into consideration how "close" the t-test may really have been. A better procedure is to run a 2X2 ANOVA repeated measures, testing the pre-post difference as the within-subject factor, the group difference as the between-subject factor, and the interaction effect of both factors.

- Use of gain scores and covariance: the most used test is to compute pre-posttest gain scores for each group, and then to compute a t-test between the experimental and control groups on the gain scores. Also used are randomized "blocking" or "leveling" on pretest scores and the analysis of covariance are usually preferable to simple gain-score comparisons.

- Statistics for random assignment of intact classrooms to treatments: when intact classrooms have been assigned at random to treatments (as opposed to individuals being assigned to treatments), class means are used as the basic observations, and treatment effects are tested against variations in these means. A covariance analysis would use pretest means as the covariate
8.4.2.2 Posttest only group control design.

This design is as:

\[
\begin{align*}
R & \times O_1 \\
R & \quad O_2
\end{align*}
\]

This design can be thought of as the last two groups in the Solomon 4-group design. And can be seen as controlling for testing as main effect and interaction, but unlike this design, it doesn't measure them. But the measurement of these effects isn't necessary to the central question of whether or not X did have an effect. This design is appropriate for times when pretests are not acceptable.

Statistical tests for this design: the simplest form would be the t-test. However covariance analysis and blocking on subject variables (prior grades, test scores, etc.) can be used which increase the power of the significance test similarly to what is provided by a pretest.

8.4.2.3 Solomon four-group design

The design is as:

\[
\begin{align*}
R & \ O_1 \times O_2 \\
R & \ O_3 \quad O_4 \\
R & \quad O_5 \\
R & \ O_6
\end{align*}
\]
In this design, subjects are randomly assigned to four different groups: experimental with both pre-posttests, experimental with no pretest, control with pre-posttests, and control without pretests. By using experimental and control groups with and without pretests, both the main effects of testing and the interaction of testing and the treatment are controlled. Therefore generalizability increases and the effect of X is replicated in four different ways.

Statistical tests for this design: a good way to test the results is to rule out the pretest as a "treatment" and treat the posttest scores with a 2X2 analysis of variance design-pre-tested against un-pre-tested.

Reading: for more information read the following:

| Exploring Research p 234-238 |

8.4.3 Quasi experimental designs

Quasi-experimental designs were developed to deal with the messy world of field research, where it is not always practical, ethical, or even possible to randomly assign persons to experimental and control groups. Quasi-experimental designs are commonly employed in the evaluation of educational programs when random assignment is not possible or practical. Although quasi-experimental designs need to be used commonly, they are subject to numerous interpretation problems.

1. Non equivalent control group
2. Separate sample pretest posttest design
3. Single group time series
8.4.3.1 Non-equivalent control group

The nonequivalent, posttest only design consists of administering an outcome measure to two groups or to a program/treatment group and a comparison. For example, one group of students might receive reading instruction using a whole language program while the other receives a phonetics-based program. After twelve weeks, a reading comprehension test can be administered to see which program was more effective.

A major problem with this design is that the two groups might not be necessarily the same before any instruction takes place and may differ in important ways that influence what reading progress they are able to make. For instance, if it is found that the students in the phonetics groups perform better, there is no way of determining if they are better prepared or better readers even before the program and/or whether other factors are influential to their growth

8.4.3.2 Separate sample pretest posttest design

The nonequivalent group, pretest-posttest design partially eliminates a major limitation of the nonequivalent group, posttest only design. At the start of the study, the researcher empirically assesses the differences in the two groups. Therefore, if the researcher finds that one group performs better than the other on the posttest, s/he can rule out initial differences (if the groups were in fact similar on the pretest) and normal development (e.g. resulting from typical home literacy practices or other instruction) as explanations for the differences.

Some problems still might result from students in the comparison group being incidentally exposed to the treatment condition, being more motivated than students in the other group, having more motivated or involved parents, etc. Additional problems may result from discovering that the two groups do differ on the pretest measure. If groups differ at the onset of the study, any differences that occur in test scores at the conclusion are difficult to interpret.
as being able to measure what one wants to measure. But a researcher usually
does not know the true position of an individual variable that he or she is typing to
measure. To any scientist the question of validity is of utmost importance. In impel
terms it may be defined as, being able to measure what one wants to measure,
according to shah and shah (1978)"there is no direct way of determining the
validity. The validity is to be judged by the extent to which its results are
comparable with other relevant evidence. There are generally two kinds of
validity that a researcher need to ensure:
1. Internal validity
2. External validity

8.5.1 Internal validity Campbell and Stanley (1972) defined internal validity as,
"the basic minimum without which any experiment is uninterruptible." In this
kind of validity, the researcher look forward to this answer, did not experimental
treatment make a difference in the specific experimental situation that he or she is
studying?

8.5.2 External validity According to Cook and Campbell (1979), "external
validity refers to the extent to which research finding can be generalized beyond
the experimental setting." In simple words if the research data is applicable to the
experimental settings; only, it would be for little value. Gay (1990) defined
external validity as, "the conditions that result are generalized, or applicable to
groups of the experimental setting" external validity in simple words, a researcher
wants to know to what of populations, settings and treatment variables, be
generalized and used to predict relationship outside the experimental setting.
- Selection-maturation interaction: the selection of comparison groups and maturation interacting which may lead to confounding outcomes, and erroneous interpretation that the treatment caused the effect.
- John Henry effect: John Henry was a worker who outperformed a machine under an experimental setting because he was aware that his performance was compared with that of a machine.

Factors which jeopardize external validity

- Reactive or interaction effect of testing: a pretest might increase or decrease a subject's sensitivity or responsiveness to the experimental variable. Indeed, the effect of pretest to subsequent tests has been empirically substantiated (Wilson & Putnam, 1982, Lana, 1959).
- Interaction effects of selection biases and the experimental variable
- Reactive effects of experimental arrangements: it is difficult to generalize to non-experimental settings if the effect was attributable to the experimental arrangement of the research.
- Multiple treatment interference: as multiple treatments are given to the same subjects, it is difficult to control for the effects of prior treatments

Reading: for more information read the following:

| Exploring Research | p 239-245 |
8.6 SELF ASSESSMENT QUESTION

Q.1 How would you explain experimental designs?
Q.2 Why experimental designs are used in research. Explain
Q.3 Describe the three basic conditions of experimental method.
Q.4 Discuss in detail the types of research designs.
Q.5 Write short notes on the following:
   1. Internal validity
   2. External validity
   3. Quasi experimental design
   4. True experimental design
Q.6 What are the main threats to internal and external validity. Write in detail.
Q.7 Why is the balance between external and internal validity necessary for an acceptable research.

8.7 BIBLIOGRAPHY


UNIT 9

USING PERSONAL COMPUTER IN DATA ANALYSIS

Najeeba Batool
<table>
<thead>
<tr>
<th>Contents</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>214</td>
</tr>
<tr>
<td>Objectives</td>
<td>214</td>
</tr>
<tr>
<td>Word Processing</td>
<td>215</td>
</tr>
<tr>
<td>Spread Sheet</td>
<td>217</td>
</tr>
<tr>
<td>The Data Base</td>
<td>220</td>
</tr>
<tr>
<td>Self Assessment Questions</td>
<td>221</td>
</tr>
<tr>
<td>Bibliography</td>
<td>222</td>
</tr>
</tbody>
</table>
9.1 INTRODUCTION

About six decades ago, scientists began to use computers for research, mathematics, and technology. When computers became available commercially, only the largest business organizations acquired them. The uses of computers were quite limited in those days. Today, however, it would be hard to name areas in which computers are not being used. Computers have entered almost every aspect of our daily life, for example, the newspapers are printed through computer, the telephone, electricity billing is done through computers, etc. Think of many household items that contain some type of computer. Television, video recorders, stereos, microwaves, all are computer controlled.

Now a day in cities many people are buying personal computers for home use. Many people are their personal computers for playing games, but home computers are goes beyond entertainment. People are realizing its importance in gathering information. Using the telephone lines, they are using electronic mail and internet facilities. In short computer has become so important in our lives that every one must have basic knowledge of computer.

Knowing the profile of our students of feminist Research Methodology, we are assuming that you know something about personal computer operations no matter what type of computer is available to you. In research, though the computer can do statistical analysis of data apart form word processing and creating a data basic. In this unit we are concerned with word processing, spreadsheet and database only.

9.2 OBJECTIVES

After reading this unit you will be able to:

1. Describe the importance of computer in our life.
2. Explain the use of personal computer in research.
3. Demonstrate and produce a document on the word processor.
4. Define spreadsheet and how it can be used in research.
5. Describe data creating and suing data base file for management of research records.
WORD PROCESSING

Personal computers are widely use in word processing. Words processing is the use of computer equipment in preparing text, it involves writing, editing, and printing.

In research, the compute/word processors are used to produce written materials regarding preparation of research proposals letters, questionnaires, and research reports. The use of word processor in research has become crucial because people attach great importance to the quality of reports. With the use of different world processing software, high quality report with graphics and designs can be prepared. Another advantage of word processor as compared to type writer is that text can be edited and changes can be made before the final print which is not possible with the type writer. In this way the word processors have increased accuracy and efficiency and also have enhanced the quality of the output.

There are number of word processing software available in the market but the most modern and versatile software is of Microsoft Corporation. In your reference book, Exploring Research, the example of Microsoft word has been used which is operated through windows. In Pakistan word for windows is mostly used therefore, we expect that you will not encounter much problem in understanding its use.

Apart from thousands of other functions which a word processor can perform, there are five essential and basic steps to produce a document on a word processor;

1. Creation of a new file by using mouse in the file menu.
2. The typing or entry of the text or material in the newly created file.
3. The editing of the text which involves spelling check, formatting and incorporation of changes, etc.
4. Saving of the editing text by giving name to the file.
5. The printing of the final product/document.
What else can a word processor do?

Besides the above mentioned five essential functions a word processor can perform a large variety of functions which can make our work easier and good looking. Following are a few key functions:

### 9.3.1 File Access

In file access a word processor can:

1. Open files.
2. Find files.
3. Search files.
4. Create new files.
5. Save files.
6. Close files.

### 9.3.2 Document Editing

The word processor can:

1. Format paragraphs.
2. Write in upper/lower case.
3. Use different fonts of writing.
4. Use different size of writing.
5. Border paragraphs/lines.
6. Cut, paste, and copy text.
7. Give page numbers, headers, and footers to a document.
8. Highlight and underline text.
Additional Features

On a word processor, following can be done:-

1. Data presentation in table
2. Data presentation in graphs
3. Data presentation with pictures

Activity

1. Create a new file for your research proposal, name and save it after typing and editing.
2. Open different option present in main menu and identify different function which you can use in making and refining your document.

9.4 SPREAD SHEET

Generally a spreadsheet is used to record business transactions and to perform calculations. A spreadsheet is actually in a ledger sheet having columns in which numbers are written to keep them in line. Without a computer and spread sheet programme, we would use a pencil, a piece of paper and perhaps a calculator to solve mathematical problems. Computer can calculate at the speed of electricity, which is a useful capability with more complicated formulas. Let’s understand the importance of a spreadsheet on computer by an example. Suppose that after finishing your tax returns, you realize you did not include income you received from a temporary job. Every calculation following that part of the form in incorrect. With a spreadsheet programme, you can simply insert the forgotten number and the programme re-calculates all the totals. With the ability to calculate, print, store, merge, and sort numeric information, a spreadsheet is an extremely useful tool.

We can define a computer spreadsheet as “A grid of columns and rows used to store and manipulate numeric information by means of a computer”. The grid appears on the screen and data is stored in the memory. Probably, the most significant advantage of an electronic spreadsheet over a traditional spreadsheet is
the ability to store not only number, but also formulae for calculating numbers. One number in a formula can be changed easily without re-entering the entire formula.

There are three components of a spreadsheet i.e.; worksheet for manipulation of numbers including data analysis, data base for storing and retrieving data, and graphic component for graphic representation of data.

9.4.1 Uses of Spreadsheet

Spreadsheet is widely used in business and also where financial matters are involved. In business, people involved in decision making frequently use spreadsheets. Managers are responsible for making sure their companies run smoothly, and spreadsheet helps them to manage money, goods, and employees. The spreadsheet stores data such as sales figures, expenses, pay roll amount, and prices. A manager can enter formulae to calculate profit and losses.

A number in a spreadsheet can be changed to show how this change can affect the other numbers. The spreadsheet programme saves time by automatically re-calculating formulae when the number is changed in a cell.

Spreadsheets are also being used in research to arrange data in a systematic fashion and to analyze it. Data of a sample population can be sorted by giving numbers to individuals and by recording their responses in different columns. When data entry is complete, a researcher can analyze data by using different formulae/operations in spreadsheet.

To start work on a spreadsheet, a file is opened, then the data are entered into a cell/row and these file are named and saved. While using a spreadsheet a researcher should not worry about straight lines, keeping data in columns and rows---it will be automatically done by the programme. Moreover, data is a row and column can be deleted and changed without disturbing the entire data/spreadsheet.
9.4.2 Using Formulae and Functions:

As spreadsheet is used to manipulate and analyses data, therefore certain formulae and functions are used to produce desired outcome. A formula is a mathematical expression that can contain numbers from other cell as well as constant numbers. If one number in a cell changed, the programme automatically re-calculates any formula that uses that number. If we use a manual spreadsheet instead, we would have to re-copy all the numbers by hand. A spreadsheet programme calculates and display it in seconds what it would take a personal hours to do manually.

Spreadsheet functions are almost similar to formulae. The advantage of suing functions is hat they reduce time involved in calculation. Another distinct feature is hat eh database function allow criteria to be defined. For example, the function SUM adds all the cells above or to left of it. The SUM function may still be used within the database. However, if we desire a sum of only those number that exceed age limit of 25-30 years, the DSUM function will be used by defining the above criteria.

9.4.3 Marking Graphs in Spreadsheet

Graphic presentation of numbers or data has become very popular in business as well as in research now a day. A graph or a picture can convey the message more clearly and effectively as compared to presentation of figures in tables or paragraphs. In a word processor, the facility of making graphs is usually available but processing of data take time. In spreadsheet making a graph is very simple because systematized data in already available we just have to highlight the data and have to select the graph option. We can choose from a number of designs of graphs available, we can add text to graphs, we can able figures on the graphs, etc.

Activity
1. Make a spreadsheet of your monthly income and expenses.
2. Calculate the SUM of the money you spend on yourself.
3. Create a chart/graph of monthly income-expenses statement.
9.5 THE DATABASE

Database is equally important in research as the word processor and spreadsheet are. A database is a central repository of related information. To paraphrase this, a database is a physical grouping of a collection of individual, but related, bits and pieces of information. As an example if we want to maintain information about each and every individual employed in Allama Iqbal Open University, we will need to create a data about all the employees. This base of data could contain, for example, information about each employee’s employment number, name, salary, year of hiring, and data of last promotion, etc. This base will subsequently provide us with immediate access to the type of information we are seeking. Database is being mentioned for every subject from astronomy to zoology. Computers, because of their speed and accuracy, are the information processor, the physical means of creating and subsequently accessing these databases. The internet is an example of computerized database. It contains information in almost all subjects a human mind can think of. A data base is a collection of information which is arranged as character, field, record and file. Before creating a database in you commuter, please clear you concept of the following

9.5.1 Character

If you want to store your name in a computer system, you will have to supply to the system with the letters comprising you name. These letters are individual characters in your name.

9.5.2 Field

Each information created through the use of characters is called a field of information. For example, you may have created your name field, your organization field or your salary field.

9.5.3 Record

Now, if you have stored all the information about yourself in the form of field, it has become a record. In database you can create many records.
9.5.4 File

If you can do the same above mentioned for some of your family members you will have managed to create a file of information comprising several records of information.

Activity:

1. Create a database for personal records of yourself and your family members.

2. Suppose you have conducted a research study on employment opportunities for women in your city. For this study you have developed a questionnaire. Please create database for:

   a) personal data of women
   b) employment status of women interviewed
   c) options available for employment

9.6 SELF ASSESSMENT QUESTIONS

1. List five of the advantages of using personal computers as a research tool.

2. A researcher is studying the effects of maternal employment on adolescents. List five ways in which a researcher would use a database to track and organize the data. Be specific as to the what kinds of data might be collected and how they would be used.

3. Please define the following:
   - Word processing
   - Spreadsheet
9.7 BIBLIOGRAPHY


