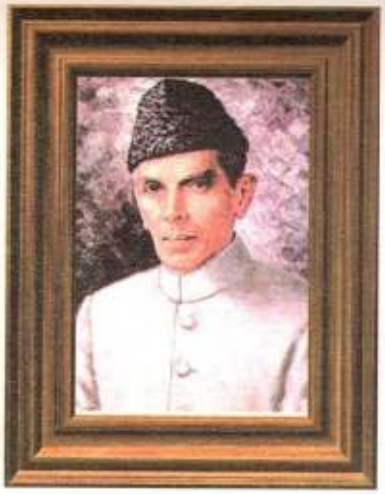
A19983632
ACADEMIC YEAR 2017-18

STATISTICS 11

2017-18

PUNJAB CURRICULUM AND TEXTBOOK BOARD, LAHORE



"Education is a matter of life and death for Pakistan. The world is progressing so rapidly that without requisite advance in education, not only shall we be left behind others but may be wiped out altogether."

(September 26, 1947, Karachi)

Quaid-e-Azam
Muhammad Ali Jinnah
Founder of Pakistan

قومی ترانہ

پاک سَرزمین شاد باد　　کِشورِ حسین شاد باد
تُو نِشانِ عزمِ عالی شان　　ارضِ پاکستان
مرکزِ یقین شاد باد
پاک سَرزمین کا نِظام　　قُوّتِ اُخوتِ عوام
قوم ، مُلک ، سلطنت　　پایندہ تابندہ باد
شاد باد منزلِ مُراد
پَرچمِ ستارہ و ہِلال　　رہبرِ ترقّی و کمال
ترجمانِ ماضی، شانِ حال　　جانِ اِستقبال
سایۂ خدائے ذُوالجلال

All textbooks of the Punjab Curriculum and Textbook Board carry a square shaped security sticker on the title page. The sticker exhibits a unique colour shift due to change of angle (orange to green) in the logo of Punjab Curriculum and Textbook Board, Lahore. If you scratch protective coating on the white strip above QR Code with a coin, it will reveal the registered trade name "PCTB". Moreover, if QR Code is checked with Android Application "News Reader" from your android mobile, serial number on sticker will be seen inside QR Code. This is a hallmark of the authenticity of original textbook of the Punjab Curriculum and Textbook Board. Look for the security sticker while purchasing the textbooks. If there is no sticker on the title page or it has been tampered with, do not buy such a book.

بِسۡمِ اللّٰہِ الرَّحۡمٰنِ الرَّحِیۡمِ

A Textbook of

STATISTICS

For Class

11

PUNJAB CURRICULUM AND
TEXTBOOK BOARD, LAHORE

All rights reserved with the Punjab Curriculum & Textbook Board, Lahore
Prepared by: Punjab Curriculum & Textbook Board, Lahore.
Approved by: Federal Ministry of Education, Curriculum Wing, Islamabad.
Vide letter No F.2-1/99-Maths (Stat)

CONTENTS

Authors:

Dr. Faqir Muhammad
Professor & Controller Examinations,
Allama Iqbal Open University, Islamabad.

Mr. Amjad Mehmood
Lecturer, Punjab College,
Lahore.

Editor:
Mr. Mazhar Hayat
Subject Specialist,

Director (Manuscripts): Dr. Mobeen Akhtar
Deputy Director (Graphics): Raza-ur-Rehman

Supervised by:
Madiha Mehmood
Subject Specialist

ARTIST **Aisha Waheed**

Published by: Idara Farogh Urdu, Lahore
Printed by: Qudrat Ullah Printers, Lahore

Date of Printing	Edition	Impression	No of Copies	Price
May 2017	10th	10th	12,000	96.00

1 Introduction to Statistics

1.1 Introduction

The word statistics was first used by a German scholar Gotifried Achenwall about the middle of the 18th century as the science of Statecraft concerning the collection and use of data by the state. According to another pioneer statistician Yule, the word statistics occurred at the earliest in the book "The elements of universal erudition" by Baron (1770). It was used again with rather a wider definition in 1787 by E.A.W. Zimmermann in "A political survey of the present state of Europe". It appeared in Encyclopedia Britannica in 1797 and was used by Sir John Sinclair in Britain in a series of volumes published between 1791 and 1799 giving a statistical account of Scotland. In the 19th century, the word statistics acquired a wider meaning, covering numerical data of almost any subject whatever and also interpretation of data through appropriate analysis. The word **data** is used for numerical facts and a single numerical fact is **datum**.

Since early 1920's with the growth in the experimental sciences there was a need for reliable scientific methods for analyzing the results of experiments and surveys. The modern subject of statistics evolves with many of the early developers of statistical methodology being experimenters themselves with the incentive of their own practical problems.

Although the gathering and presenting of information is still an important part of statistics, the modern statistics is quite different from the early days. Statistics now-a-days includes probability theory and applied mathematics. These days computers have made to perform statistical analysis routinely which could not have been contemplated in the past. Computers do not form a part of statistical theory but may be useful in applying statistical theory to solve a practical problem.

Statistics has been defined as **the mathematical science of making decisions and drawing conclusions from data in situations of uncertainty**. It includes

designing of experiments, collection, organization, summarization, analysis and interpretation of numerical data.

In the above definition of statistics, we have only considered its scientific meaning. In day to day usage, the word statistics refers to *numbers* or *facts*, such as statistics of births, statistics of deaths and statistics of road accidents etc. In some other situations, it has a symbolic meaning such as "do not become a statistics on the next weekend". The word statistics is also the plural of 'statistic' which is a statistical term and is a quantity calculated from the sample values.

Before proceeding further, some statistical terms and notations need to be defined and discussed. We explain each of the terms through the following example:

Example 1.1: It was observed that out of 500 college students surveyed, 300 were females. Is there evidence that more students in this college are females.

1.1.1. Definitions

Population: The total group under discussion or the group to which the results will be generalized is called population. In the example 1.1, the set of all students in the college is our population.

Sample: Sometimes the measurement of interest can not be made on the whole population, then we choose a subset of population to draw inference about the population. If the inferences from sample to population are going to be meaningful, it is imperative that the sample should be representative of the population. In the example 1.1, the set of 500 students being a subset of all college student, is our sample.

Ratio: The ratio of A to B is the fraction A/B. In the example 1.1, there are 300 females and the remaining 200 are males. So, the male female ratio is 200/300.

Proportion: A proportion is a special ratio of a part to its total. In the example 1.1, the proportion of females and males are 300/500 and 200/500 respectively. Proportion becomes percentage when multiplied by 100.

The female percentage in the example 1.1 is $(300/500) \times 100$ i.e., 60%. 60% means 60 out of 100. The symbol **%** is abbreviation for **percent**.

Parameter: It is a quantity computed from a population when the entire population is available. Parameters are fixed or constant quantities and are not usually known. In the example 1.1, the proportion of female students in the population is our parameter.

Statistic: It is a quantity computed from a sample. In the example 1.1, the proportion of the female students in the sample of students is a statistic. Statistic is variable because it varies from sample to sample.

Sampling Variability: As sample is a representative part of a population; there may be more than one samples in a population. Therefore, all samples from the same population may not be identical e.g. with reference to the example 1.1, another sample of 500 students surveyed may not contain 300 female students.

Experiment: Any study in which the scientist can control the allocation of treatments to the experimental units is called an experiment. Every unit must be capable of receiving every treatment, and the decision as to which unit receives which treatment is determined by an allocation mechanism. This mechanism could be that the scientist observes the unit and then decides which treatment to apply, which would be an unsatisfactory allocation mechanism because of the possibility of the subjective bias, or it could be to assign the treatments according to a rule which would be scientifically acceptable. So, all allocation mechanisms do not lead to acceptable inferences.

Sample Surveys: In a sample survey, there are no treatments. The units in the population under study are listed in a frame and a sample of units is selected from the frame using a selection mechanism. So, the distinguishing feature of a survey is the control over the selection of units.

The features which distinguish experiments and sample surveys are control over allocation and control over selection.

Constant: Quantities which don't vary from individual to individual are called constants e.g.

$$\pi \approx 3.14159,\ e \approx 2.71828,\ 4,\ 25 \text{ etc.}$$

Order Statistic: The order statistic (OS) of data $Y_1, Y_2, Y_3, \ldots, Y_n$ is just the arrangement of data in order of magnitude. It is denoted by $Y_{(1)}, Y_{(2)}, Y_{(3)}, \ldots, Y_{(n)}$

$Y_{(1)}$ is the minimum of $Y_1, Y_2, Y_3, \ldots, Y_n$ and $Y_{(n)}$ is the maximum of $Y_1, Y_2, Y_3, \ldots, Y_n$

$Y_{(1)}$ is minimum order statistic and $Y_{(n)}$ is maximum order statistic. If the data are arranged in increasing order of magnitude, the data are said to be arranged in ***ascending order*** and if the data are arranged in decreasing order of magnitude, the data are said to be arranged in ***descending order***.

Model: It is a mathematical statement used in studying the results of an experiment or predicting the behaviour of future repetitions of the experiment.

The models involve probability distributions which describe the variability in the characteristic of interest in the population and therefore, the variability we might expect to get in different samples.

In the example 1.1, binomial model is appropriate where parameter is the proportion of students who are females in the population.

A common model for describing the makeup of an observation states that it consists of a mean plus an error so an observation can be described by means of simplest model as:

$$Y_i = \mu + \epsilon_i \qquad (1.1)$$

where, Y_i represents any individual observation, μ (a Greek letter read as meu) represents average or mean of the population and ϵ_i (a Greek letter read as epsilon) represents a random error.

Random Error (ϵ_i)**:** Random error is the chance variation in an observational process. It should not be confused with its synonym in common usage mistake, which means human error.

Equation (1.1) can be written as:

$$\epsilon_i = Y_i - \mu \qquad (1.2)$$

ϵ_i's are usually assumed to be from a population having zero mean. As the random error sum to zero, so, there would be approximately equal number of positive and negative deviations.

The term, $(Y_i - \mu)$ is known as *deviation* of an observation Y_i from mean μ.

Equation (1.1) represents the simplest form of the linear additive model. The μ may be a single mean.

1.1.2 Notations

Sigma (Σ): It is a Greek letter and is used as a short-hand notation for sum. For example,

i.e., if we have to add five numbers Y_1, Y_2, Y_3, Y_4, and Y_5, then

$$Y_1 + Y_2 + Y_3 + Y_4 + Y_5 = \sum_{i=1}^{5} Y_i$$

This tells us to sum all Y_i values starting at $i = 1$ (the lower limit) and stopping at $i = 5$ (the upper limit). It is always assumed that consecutive integer values are to be summed unless otherwise specified. Other examples are

i) $\Sigma 2^Y : Y \in D$ where $D = \{2, 3, 4, 7\}$

$$\Sigma 2^Y = 2^2 + 2^3 + 2^4 + 2^7$$

where $\in$ is read as belongs to.

ii) If $P(Y)$ denotes the probability then

$$\Sigma P(Y) = P(2) + P(3) + P(4) + P(7)$$

$$Y \in D$$

Product (π): It is a Greek letter (read as pi) and is used here as a short-hand notation for product. For example,

$$\prod_{i=1}^{n} Y_i = Y_1 \times Y_2 \times Y_3 \times \ldots \times Y_n$$

If we have only three observations

$$Y_1 = 2, Y_2 = 3, Y_3 = 5,$$

then $$\prod_{i=1}^{3} Y_i = Y_1 \times Y_2 \times Y_3 = 2 \times 3 \times 5 = 30$$

***n*!**: Read as n factorial and is defined as:

$$n! = n.(n\text{-}1).(n\text{-}2)\ldots.(3).\ (2).\ (1)$$

$\Rightarrow$ $4! = 4.3.2.1$

$= 24$

and $1! = 1$

Note that $0! = 1$

1.2 Importance Of Statistics In Various Disciplines

Since the information collected in the form of data (observations) from any field will almost always involve some sort of variability (uncertainty), so, this subject has applications in almost all fields of research. The researchers use statistics in the analysis, interpretation and communication of their research findings. Some examples of the questions, which statistics might help to answer with appropriate data are:

i) how much better a yield of wheat do we get if we use a new fertilizer as opposed to a commonly used fertilizer?

ii) are a company's sales figures likely to increase in the next quarter?

iii) what dose of an insecticide be used successfully to monitor an insect population?

iv) what is the likely weather in the coming season?

It is obvious that statistics has its application in almost every field where research is carried out and findings are reported.

When Statistics is applied in Economics, it is called ***Econometrics***, When it is applied in biological sciences, it is called ***Biometry***. Similary, ***Psychometry*** and others. We give a brief account of its application in differnt fields as follows:

1.2.1 Social Sciences

In social sciences, one of the major objectives is to establish relationship that exists between certain variables. This end is achieved through postulating hypothesis and their testing by using different statistical techniques.

Most of the areas of our economy can be studied by econometric models because they help in forecasting and forecasts are important for future planning.

1.2.2 Plant Sciences

The most important aspect of statistics in plant sciences research is its role for efficient planning of experiments and drawing valid conclusions. A technique in statistics known as 'Designs of Experiments' helps introducing new varieties. Optimum plot sizes can be worked out for different crops like wheat, cotton, sugarcane and others under different environmental conditions using statistical techniques.

1.2.3 Physical Sciences

The application of Statistics in physical sciences is widely accepted. The researchers use these methods in the analysis, interpretation and communication of their research findings, linear and nonlinear regression models are used to establish cause and effect relationship between different variables and also these days computers have facilitated the experimentation and it is possible to simulate the process rather than experimentation.

1.2.4 Medical Sciences

The interest may be in the effectiveness of new drugs, effect of environmental factors, heritability, standardization of various records and other related problems. Statistics comes to rescue. It helps to plan the next investigation in order to get trustworthy information from the limited resources. It also helps to analyze the current data and integrate the information with that previously existing.

1.3 Variables

A characteristic that varies from individual to individual in a population is called a variable. The nature abhors constancy, so, the natural phenomena show variability. For example, plant height, number of plants per plot, eye colour (black, blue, green) etc.

Let Y represents the variable and Y_i (read as Y subscript i) represents the ith observation. The variables, $Y_1, Y_2, \ldots, Y_n$ form a set of n observations on the variable Y. For example, we measure the height of 5 wheat plants and observe that the height of first plant is 87 cm, the height of second plant is 90 cm, the height of third plan is 92 cm, the height of fourth plant is 89 cm and the height of fifth plant is 95 cm.
Here, $Y_1 = 87$ cm, and Y_5 is 95 cm.

A variable may be **fixed** or ***mathematical*** when its value can be determined before hand i.e., amount of fertilizer to be applied to a plot, amount of insecticide applied to control insect pests. A variable may be **random** when its value cannot be exactly determined i.e., yield from a plot, the response to an insecticide. Variables are usually of two types:

i) Quantitative variable ii) Qualitative variable

1.3.1 Quantitative variable

A quantitative variable is one which is capable of assuming a numerical value. For example, height of plants, weight of grains or number of students in a class.

Quantitative variables can further be placed into two types depending upon the type of measurement possible.

i) Continuous variables **ii)** Discrete variables

A continuous variable is one that can take all possible values in an interval on the number line. For example, atmospheric pressure, plant height, student height and temperature.

A discrete variable is also known as discontinuous variable. It is one that can take only isolated points on the number line. Usually these values are positive integers as these arise from counting. For example, number of students in a class, number of plants per plot, number of insects in a unit area, number of grains per plant.

1.3.2 Qualitative or Categorical variable

A qualitative variable also known as categorical variable is one which is not capable of taking numerical measurements. An observation is made when an individual is allocated to one of several mutually exclusive categories. Observations falling in each class can only be counted. For example, sex (either male or female), general knowledge (poor, moderate, good), colour (blue, green, red etc.).

Example 1.2: A sample of 5 students from a class was selected and each one of them was asked which brand of soap they use. Their responses were as follows:

Lux, Rexona, Lux, Capry, Rexona

Identify the type of variable?

Solution: As we can only categorize these observations into different brands of soap, so, the observations arise from a categorical variable.

Example 1.3: There are 7 sections of class XI in an intermediate college. The number of students in each section is as follows:

41, 45, 47, 37, 35, 45, 42

Identify the type of variable?

Solution: This data set arises from a quantitative variable because the observations are numerical values. Again the variable is discrete because the observations are isolated points on a number line.

1.4 Descriptive And Inferential Statistics

Generally, rough and crude form of the data is obtained from experiments and surveys that needs to be organized and summarized in order to describe its sense. This is where the ***descriptive statistics*** comes in to help us. It provides procedures for:

i) organizing the data collected from the sample,

ii) summarizing the data. It includes graphic representation and calculation of summary values like measures of central value and measures of variability that we call statistic (mean, proportion, variance).

iii) presenting the summaries in an understandable form for the others.

One may be interested to generalize the results of the data. For example, based upon the descriptive statistics, one might be willing to estimate the value of his measure of a central value, had he gathered data about all the subjects possessing the given character rather than a sample. This procedure of inferring about the characteristics of the population based upon the characteristics of its sample is called ***inferential statistics***.

The discipline strengthening inferential statistics is probability theory. So, our generalizations of results always involve some risks. Thus, we always make probabilistic statements and we really don't prove anything. For example, we say that the probability is high that an experimental variable affected the dependent variable.

The whole issue of descriptive and inferential statistics can be described with the help of what we call ***Statistical Problem***.

The distinguishing feature of a statistical problem is that we are trying to say something about a population based on a sample from the population only, taking into account the variability within the samples. Before we do this, we must make some assumptions about the manner in which the data was produced (these are embodied in a statistical model). Based on the model and data, statistical methods are designed which allow us to work back and make statements about the underlying population. The main aspects of the statistical problem are:

i. a clear definition of the population of interest and objectives,

ii. the design of experiment or the sampling procedure,

iii. the collection, organization and analysis of data,

iv. the selection of suitable model and the process of making statements about the population based on sample information.

Figure 1.1 illustrates clearly the main aspects of statistics. Often the underlying population will be clearly defined, in other situations the population may only be hypothetical corresponding to what might have happened in an infinite series of repetition of the experiment concerned.

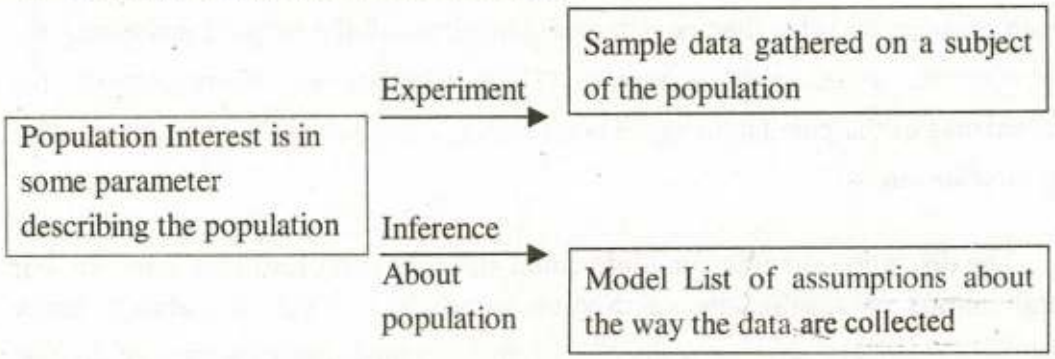


Figure 1.1: Statistical problem – Real world situation

1.5 Sources of Data

Keeping in view the objectives of the study, data are collected by individual research workers or by organizations through sample surveys or experiments. The data collected may be

i) primary data ii) secondary data

By primary data we mean the raw data, which has just been collected from the source and has not gone through any kind of statistical treatment like sorting and tabulation.

By secondary data we mean the data, which has already been collected by someone, that has undergone a statistical treatment like sorting and tabulation etc.

1.5.1 Sources of primary data

The sources of primary data are primary units like basic experimental units, individuals, households and the following methods are usually used to collect data from primary units. The method used depends very much on the nature of the primary unit.

i: Personal Investigation: The researcher conducts the experiment or survey by himself and collects data from it. The data collected is generally accurate and reliable. This is only feasible in case of small scale laboratory, field experiments or pilot surveys and is not practicable for large scale experiments and surveys because it will take too much time.

ii: Through Investigators: The trained investigators are employed to collect the data. In case of surveys, they contact the individuals and fill in the questionnaires after asking the required information. A ***questionnaire*** is an inquiry form having a number of questions designed to obtain information from the respondents. This method is usually employed by most of the organizations. This method gives a reasonably accurate information but it is very costly.

iii: Through questionnaire: The required information is obtained by sending questionnaire to the selected individuals by mail who fill in the questionnaire and return it to the investigator. This method is cheap but non-response rate is very high as most of the respondents don't bother to fill in the questionnaire and send it back.

iv: Through local sources: The local representatives or agents are asked to send requisite information who provide the information based upon their own experience. This method is quick but gives only rough estimates.

v: Through telephone: The information may be obtained by contacting the individuals on telephone. This method is quick and gives accurate information.

vi: Through internet: With the introduction of information technology, the people may be contacted through internet and the individuals may be asked to provide the pertinent information.

It is important to go through the primary data and locate any inconsistent observations before it is given a statistical treatment.

1.5.2 Sources of Secondary Data

The secondary data may be available from the following sources.

i: Government organizations: Federal and Provincial Bureau of Statistics, Crop Reporting Service-Agriculture Department, Census and Registration Organizations etc.

ii: Semi-government organizations i.e., Municipal committees, District Councils, commercial and financial institutions like banks etc.

iii: Teaching and research organizations.

iv: Research journals and newspapers.

v : Internet

Exercise 1

Ans on Page 250

1.1 Define the word statistics and explain its different meanings?

1.2 Define the following terms:

i) Population and sample ii) Parameter and statistic

1.3 Distinguish between qualitative and quantitative variables?

1.4 i) Write the following using a summation sign with appropriate index?

a) $Y_3 + Y_4 + ... + Y_{15}$ b) $Y_1^2 + Y_2^2 + Y_3^2 + Y_4^2$

c) $(Y_1 - \mu)^2 + (Y_2 - \mu)^2 + (Y_3 - \mu)^2$ d) $bY_{20} + bY_{21} + ... + bY_{30}$

ii) Expand the following summation and product signs?

a) $\sum_{i=1}^{5} Y_i$ b) $\sum_{i=5}^{8} (Y_i - \mu)$ c) $\sum_{i=1}^{3} Y_i^2$

d) $\sum_{i=2}^{4} (Y_i - 9)$ e) $\prod_{i=1}^{3} Y_i$ f) $\prod_{i=3}^{6} (aY_i)^2$

g) $\prod_{i=2}^{4} (X_i - Y_i)$

1.5 Classify the following as categorical, discrete or continuous variable:

i) Sex of an insect. ii) Weights of plants.

iii) Major crops of Pakistan. iv) Level of satisfaction.

v) Teaching standards. vi) Temperature measured in Fahrenheit.

1.6 Explain in detail the main aspects of a Statistical problem?

1.7 Define Descriptive and Inferential Statistics and differentiate between them.

1.8 Distinguish between primary and secondary data and give different sources from which these are obtained.

1.9 Fill in the blanks:

i) The purpose of the sample is to draw inference about _____.

ii) Proportion is always _____ or equal to one.

iii) The quantity computed from population is called _____.

iv) The quantity computed from sample is called _____.

v) The quantity which does not vary from individual to individual is called _____.

vi) Sum of the random errors equal to _____.

vii) A variable is called _____ when its value cannot be exactly determined.

viii) A variable that takes numerical values is called _____ variable.

ix) The procedure of inferring about the population characteristics using the sample is called _____.

x) First hand collected data is called _____.

1.10 Against each statement write T for true and F for false statement.

i) The value calculated from the population is called parameter.

ii) Statistics deals with single fact.

iii) Statistics can be placed in relation to each other.

iv) Primary data and ungrouped data are same.

v) A collection of observations is called data.

vi) A variable can assume the same value.

vii) A discrete variable can assume finite values between two given limits.

viii) Height measurements of students is quantitative variable.

ix) The word Statistics is at present used in four ways.

x) A small part of the population is called sample.