

Contents

Chapter	Title	Page
1	Introduction	7
2	Presentation of the data	31
3	Measures of central tendency	73
4	Measures of dispersion	107
5	Probability	143
6	Probability distributions	177
7	Sampling	205
8	Testing of hypothesis (t and Z test)	235
9	Testing of hypothesis (Chi square test)	265
10	Regression and correlation	277
11	ANOVA and Experimental designs	323
12	Index numbers	349
13	Time series	375
	Additional exercise	388
	Statistical Tables	402

CHAPTER 1

INTRODUCTION

Students of other fields consider statistics as uninteresting, non-beneficial and dry subject, due to its non-conceptual teaching.

Learning Goals

- Be familiar with the subject of statistics.
- Be familiar with the techniques used for collection of the data.
- Be familiar with the nature and scale of a variable.

1.1 Statistics

Statistics is a multidisciplinary science, concerned with scientific methods for **collecting, organizing (presenting), summarizing, and analyzing** of data, as well as drawing valid conclusions and making reasonable decisions on the basis of such analysis.

1.2 Meaning of statistics

History tells us the word statistics has been derived from the Latin word "Status" or Italian word "Statista" or "German word "Statistik" or the French word "Statistique" each of which means "Political State".

At present there are three different meanings of the word statistics.

- In singular sense it means that field of study which deals with **collection, presentation, analysis and interpretation** of numerical data.
- In plural sense it refers to numerical facts and figures which are collected by a systematic method.
- In another sense it is plural of the word statistic, a value computed from a sample.

1.3 Descriptive and inferential statistics

Statistics as a subject may be divided into descriptive and inferential statistics.

Introduction

1.3.1 Descriptive statistics

The area of statistics that describes and analyses a given set of data without drawing any conclusions or inferences about it, is called descriptive statistics. It is also known as deductive statistics.

1.3.2 Inferential statistics

The branch of statistics that deals with drawing inferences about the characteristics of the population on the basis of sample information is called as inferential statistics. It is also known as inductive statistics.

Inferential Statistics includes the estimation of population parameters and testing of hypothesis. It is based on probability theory.

1.4 (a) Theoretical Statistics

Theoretical statistics is that branch of statistics in which we formulate statistical methods, formulas and rules to find every day problems' solution in the areas of physical and social sciences. It is also called mathematical statistics.

1.4 (b) Applied statistics

Applied statistics is that branch of statistics in which we find everyday problems' solution in the areas of physical and social sciences by using statistical methods.

Statistics adopts different names when it combines with other disciplines of study.

Other disciplines	New name of Statistics	Other fields
Economics	Econometrics	Managerial economics
Life sciences	Biostatistics/biometry	Bioinformatics, Genetics
Chemistry	Chemometrics	Analytical chemistry
Geography	Geoinformatics Geostatistics	
Psychology	Psychometrics	
Physics	Physical statistics	Statistical mechanics

Introduction

1.4.1 Life sciences and statistics (Biostatistics)

Application of statistical methods to understand the discipline of the life sciences is termed as biostatistics. It is also known as biometry.

Biostatistics/statistics is helpful in understanding the physiology and anatomy. Genetics that is an important area of life sciences is defined as the game of probability. The term regression was originated by Sir Francis Galton (a geneticist) in his paper about inheritance of stature. Bioinformatics based mainly on statistics.

1.4.2 Statistics and economics

Statistics perhaps plays most important role in understanding the economics. Index number, regression analysis, correlation and time series specially are evolved for the sake of economics. Without statistics economics is nothing. Managerial economics and econometrics totally based on application of statistics in economics.

1.4.3 Statistics and chemistry

Chemometrics is application of statistical methods in the field of chemistry. Analytical chemistry is another example for the importance of statistics in chemistry. A researcher or a student of chemistry cannot be independent of factor analysis, principal component analysis and discriminant analysis. Average, measure of dispersion and testing of hypothesis are very important for data analysis of chemistry as in other fields.

1.4.4 Statistics and computer

Nowadays computer technology mainly depends on mathematics and statistics. Especially probability and logistic regression play an important role in artificial intelligence and machine learning process. Well-known features "spell check" and "auto correction" in internet search engines based on probability theory. These features give better results if Bayesian approach to probability is applied. Data scientist is a hybrid of computer engineer and statistician.

1.4.5 Statistics and geography

Statistics applied in geography and geology is known as Geostatistics. Different methods such as correlation technique is applied to estimate the amount of any substance in a

Introduction

mines by drilling a borehole. This technique is called "Kriging" named after a South African mining engineer Danie Krige.

1.5 Population

The totality of observations with which we are concerned is called a population.

Examples: (i) All BS students in a college, (ii) All the books in a library, (iii) All corona patients in a hospital, (iv) Blood of a body and (v) All the trees in a forest.

1.5.1 Parameter

Any numerical value calculated from the population is called parameter or population parameter. Parameter is constant. It is denoted by Greek letter as μ, σ, ρ etc.

1.6 Sample

A sample is a representative part of population which is selected to obtain the information concerning the characteristics of a population.

Examples: (i) Punjab food authority picking a handful of wheat from the bags to check the quality, (ii) A spoon of ice cream selected from a pot to taste the flavour and (iii) A syringe of blood for the test of the patient etc.

1.6.1 Representative Sample

If a sample represents all the qualities of a population it is known as representative sample.

1.6.2 Statistic

Any numerical value calculated from the sample is called statistic or sample statistic. Statistic varies from sample to sample therefore, it is variable. It is denoted by Latin letter as X, S, r etc.

Note: Population and sample sizes are denoted by N and n respectively.

1.7 Importance of statistics

In almost every field of research whether it belongs to physical or social sciences, the collection of data and calculation of uncertainty are very important aspects. Following are some of the uses of statistics in these fields which serve to indicate its importance.

Introduction

- 01) It simplifies complex data to make it easily understandable.
- 02) It presents facts in a numerical form.
- 03) It facilitates comparison of data.
- 04) It studies relationship among different facts.
- 05) It helps in making predictions.
- 06) It helps to draw inferences about the characteristics of the population by using sample information.
- 07) It helps in formulating the policies.
- 08) It helps in testing the laws of other sciences.

1.8 Limitations of statistics

Some limitations of statistics are

- 01) Statistics only deals with aggregate of facts.
- 02) Statistical results are valid (true) only on the average or in the long run.
- 03) Statistics deals with facts that can be numerically measured.
- 04) A person who has an expert knowledge of statistics can handle statistical data efficiently.
- 05) Statistics provides only the tools for analysis.

1.9 Constant

Any value that does not change but remains fixed is called constant.

For example: The number of days in a week, $\pi = 3.1415$, $g = 9.8 \text{ m/sec}^2$ and $e = 2.718281$ etc.

The constants are usually represented by first alphabets of the English language as, a, b, c etc.

1.10 Variable

A characteristic that varies from individual to individual, place to place or time to time is called a variable. Variables are usually represented by last alphabets of the English language as X, Y and Z etc.

Examples:

- (i) Age of a person
- (ii) Weight of a person

Introduction

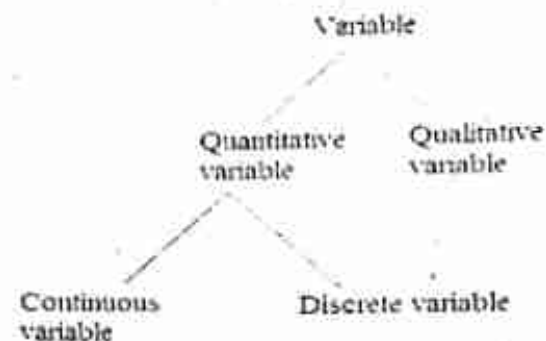
- 03) Weight of a person.
- 04) Eye colour of a person
- 05) RAM in a computer
- 06) Volume of a liquid
- 07) Temperature of a place
- 08) Number of petals in a flower
- 09) Demand of a commodity
- 10) Marital status of an individual
- 11) Hard disk size of a computer
- 12) IQ level of a student
- 13) Behavior of an animal during a specific time interval
- 14) Utility of an item used
- 15) Price of a commodity
- 16) Brand of a computer
- 17) Income of a person
- 18) Amount of an agent in the blood of a person.

1.10.1 Types of variable

There are two types of variable

- (i) Qualitative variable (ii) Quantitative variable

Introduction



1.10.2 Qualitative variable

A qualitative variable is determined when characteristics of interest results in a nonnumeric value. It is also called categorical variable or attribute. For example marital status, gender, pain level or personality type etc.

1.10.3 Quantitative variable

A quantitative variable is determined when characteristics of interest results in a numerical value. In other words physical characteristic that varies is known as quantitative variable. It is simply called a variable, for example age, weight, shoe size or temperature etc.

There are two types of quantitative variable

- (i) Discrete variable (ii) continuous variable

1.10.4 Discrete variable

A quantitative variable that can assume only specified values is called discrete variable. Examples are shoe size, collar size and number of petals in a flower etc.

1.10.5 Continuous variable

A quantitative variable that can assume any value within a given range is called continuous variable. Examples are temperature, height, weight and age etc.

Introduction

1.11 Scale of Measurements

Scale of measurement is defined as a classification that tells about the nature of the data / variable. It was developed by Stanley Smith Steven. There are four types of measurement scales.

1.11.1 Nominal scale: Qualitative data/variables are measured on nominal scale. It labels the variable without any quantitative value. For example data on gender, hair colour and caste follow nominal scale because these are just names to categories the data. It is also called categorical variable scale.

1.11.2 Ordinal scale: Qualitative data/variables with ordering nature are measured on ordinal scale. For example satisfaction level, grade in an examination and education level of a person follow ordinal scale because they can be ordered.

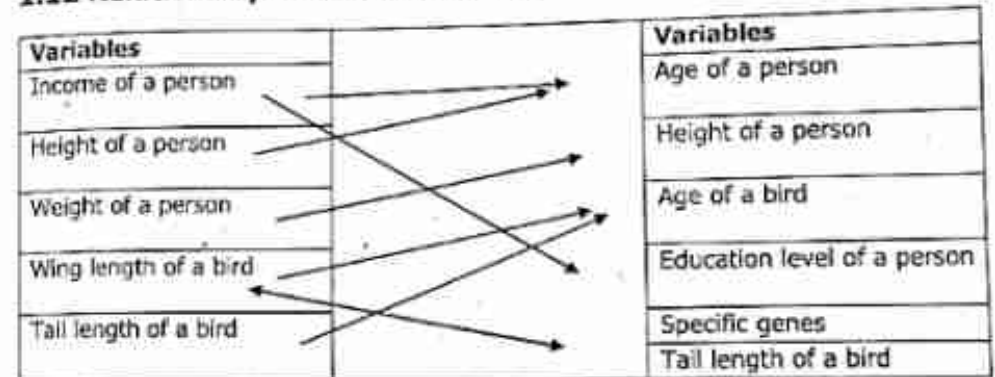
1.11.3 Interval scale: Quantitative data/variables with equal intervals without true zero are measured on interval scale. True zero also known as absolute zero means origin of scale. For example temperature follow interval scale.

1.11.4 Ratio scale: Quantitative data/variables with equal intervals and true zero (no number exists below zero) are measured on ratio scale. For example height, weight and length follow ratio scale.

Scale	Properties
Nominal	It is use for the purpose of only classification.
Ordinal	It is use for the purpose of classification and ranking.
Interval	It is use for the purpose of classification and ranking without true zero.
Ratio	It is use for the purpose of classification and ranking with true zero.

Introduction

1.12 Relationship between variables



Explanation:

Income of a person depends on (is function of) his age and education level.

Height of a person depends on (is function of) his age and genes.

Weight of a person depends on (is function of) his height.

Wing length of a bird depends on (is function of) his age.

Tail length of a bird depends on (is function of) his age.

Wing length and tail length of a bird depends upon each other.

1.12.1 Dependent variable

A variable that depends on other variable(s) is known as a dependent variable or a variable being tested and measured in a scientific experiment. It is also called as explained, regressand, effect and predictand variable.

1.12.2 Independent variable

A variable that represents the potential reason for change in dependent variable or a variable that is changed or controlled in a scientific experiment to test the effects on the dependent variable is known as an independent variable. It is also called as explanatory, regressor, cause and predictor variable.

Introduction

1.13 Data

Data is plural of datum which literally means to give or something given. Series of information is called data.

1.13.1 Types of data

With respect to nature

- 01) Qualitative data.
- 02) Quantitative data.

With respect to collection method:

- 01) Primary data.
- 02) Secondary data.

With respect to formation:

- 01) Cross sectional data
- 02) Time series
- 03) Pooled data

1.13.1.1 Primary data

Data observed or collected directly from first-hand experience is known as primary data. Primary data has not been published yet and is more reliable, authentic and objective. Primary data has not been changed or altered by any one, therefore its validity is greater than secondary data.

1.13.1.2 Sources of primary data

Sources for primary data are limited and at times it becomes difficult to obtain data from primary source due to scarcity of population or lack of cooperation. Following are some of the sources of primary data.

Direct / Indirect Investigation

Experiments: Experiments conducted in laboratories and in any scientific research study are basic sources of primary data.

Survey: Survey is commonly used method in social sciences to collect the primary data. Methods used in survey are:

Introduction

Questionnaire: A list of open ended or closed ended questions is called questionnaire. It is the most commonly used method in survey.

Interview: A face-to-face conversation with the respondent is called interview. This method can be used for collection of the required data.

Registration: An act of recording a name or information on an official/un-official list is known as registration.

Observations : Recording the values by observing a person/individual, with or without letting him know.

Telephone: Telephone can be used for collecting data in survey.

Internet: In this new era of information technology, internet is widely used in the collection of data.

1.13.1.3 Secondary data

Data collected in the past or from other parties is called secondary data.

1.13.1.4 Sources of secondary data

Secondary data is often readily available. Number of secondary sources are more than primary sources in the collection of data.

Published printed sources

There are varieties of published printed sources

Books: Books on any topic that you want to research are most authentic one in secondary sources.

Journals/Periodicals: Journals are most reliable secondary source. They provide up to date data, not available in books, regarding a field of study.

Magazines/Newspapers

Magazines are easily available source, used for collection of secondary data. It is not a reliable source.

Introduction

Published electronic sources

Internet can be used as both primary and secondary source for collection of data. Data can be collected from general websites, weblogs and journals.

General websites: Generally websites are easily reachable secondary sources, but not reliable.

Weblogs: Weblogs are electronic diaries written by an individual. They are reliable to some extent they are like personal diaries..

E-journals: E-journals are more commonly available than printed journals. Some e-journals are free of cost.

Unpublished personal records

Unpublished personal records including diaries and letters are also reliable sources of secondary data. They may also be useful in some cases.

Official / Government records

Government records are very important and reliable secondary sources. They include census data/population statistics, health records and educational institutions records.

Private sector records

They includes banks, chamber of commerce and industries etc.

1.13.1.5 Cross sectional data

Data that comprise observations about a variable for different groups or places in same time interval. For example prices of a commodity for various cities of Pakistan in 2017.

City	Milk Price (2017) in Rs
Lahore	120
Faisalabad	90
Okara	85

Introduction

1.13.1.6 Time series

Data that comprise observations about a variable for same groups or places in different regular time intervals. For example prices of a commodity for the years from 2015 to 2019 in a city.

Year	Milk price (Rs) in Faisalabad
2015	65
2016	70
2017	85
2018	90
2019	100

1.13.1.7 Pooled data

A data that is mixture of cross sectional data and time series is known as pooled data. For example enrolment of students in two different colleges from 2016 – 2018

Year	College 1	College 2
2016	1200	800
2017	1350	900
2018	1400	920

1.14 Error of measurements

Errors in reading, calculating or recording a numerical value. The difference between observed values of a variable recorded under similar conditions and some fixed true value.

[Cambridge Dictionary of Statistics 4th Edition]

There are two types of error (i) biased error and (ii) unbiased error

1.14.1 Biased Error: An error that occurs due to faulty measuring device. It is also known as instrumental, cumulative, systematic error. It does affect the average.

1.14.2 Unbiased error: An error that occurs on the result of repeated measurements and will disappear in the long run. Error caused by a factor that randomly affects the measurement, such as change in environment, noise, tiredness or human mistake. It is also known as residual error, random error, compensating error or accidental error. It does not affect the average.

Introduction

1.15 Significant digits

The digits in a number are called significant digits which give exact and important information. All the digits are significant other than zero. There are some rules for the zero to be significant. To understand the concept of significant digits the table is given below.

Number	27	632	4500	30	7.000	0.36	235.67	0.006	0.039	3.4000
Significant digits	2	3	2	1	4	2	5	1	2	5

1.16 Rounding off numbers

The procedure used for reporting numerical information to fewer decimal places than used during analysis. The rule generally adopted is that excess digits are simply discarded if the first of them is less than five, otherwise the last retained digit is increased by one (replace with zeros in case of whole numbers). The rules for rounding decimal numbers are given below

01) 127.2492341 to three decimal places gives 127.249

02) 25.64682 to two decimal places gives 25.65

If exactly digit 5 is to be dropped then, retained last digit will be increased by one if it is odd and will be remained unchanged if it is even for example

03) 6.465 to two decimal places gives 6.46

04) 6.435 to two decimal places gives 6.44

Introduction

Multiple Choice Questions

1. Statistics is a science of:

- (a) Sources (b) Decision making (c) Collection (d) All

2. Statistics can be divided into branches:

- (a) 1 (b) 2 (c) 3 (d) 4

3. A quantity calculated from population:

- (a) Frequency (b) Statistic (c) Parameter (d) Sample

4. Measurement provides:

- (a) Qualitative data (b) Discrete data (c) Primary data (d) Continuous data

5. Statistics are always:

- (a) Exact (b) Estimated values (c) Constant (d) Population

6. A constant can assume values

- (a) Fixed (b) Not fixed (c) Variable (d) Grouped

7. In which sense statistics mean numerical data:

- (a) Singular (b) Plural (c) Both (a) and (b) (d) None of these

8. Sum of random errors is equal to:

- (a) 3 (b) 2 (c) 1 (d) 0

9. The data in their original form:

- (a) Secondary data (b) Ordered data (c) Ungrouped data (d) unofficial data

10. The data which have already been collected:

- (a) Secondary data (b) Primary data (c) Ungrouped data (d) grouped data

11. A representative part of the population:

- (a) Sample (b) Parameter (c) Statistic (d) average

Introduction

12. Weights of students in a class make:

- (a) Discrete data (b) Continuous data (c) Qualitative data (d) Constant data

13. The branch of statistics that deals with analysis of data.

- (a) Descriptive statistics (b) Inferential statistics
(c) Biostatistics (d) Biometry

14. The branch of statistics which deals with procedures of drawing inference about population on the basis of sample information:

- (a) Descriptive statistics (b) Inferential statistics
(c) Applied statistics (d) Theoretical statistics

15. Statistics deals with:

- (a) Qualitative facts only (b) Single facts (c) Aggregate of facts (d) None

16. Statistics are ----- of the administration:

- (a) Ears (b) Eyes (c) Mouth (d) Hands

17. A collection of all the elements in a group is called:

- (a) Population (b) Sample (c) Data (d) Statistic

18. Technical and trade journals:

- (a) Primary source (b) Secondary source
(c) Unpublished source (d) electronic source

19. Questionnaire is:

- (a) Primary source (b) Secondary source
(c) Published source (d) Official source

20. The word "Statistics" is defined in:

- (a) Singular sense (b) plural sense
(c) Plural of the word statistic (d) All of the above

21. The data collected by NADRA:

- (a) Un-official data (b) Primary data (c) Secondary data (d) Qualitative data

Introduction

22. Statistics is the backbone of :

- (a) Mathematics (b) Computer (c) Economics (d) Research

23. Counting usually provide:

- (a) Continuous data (b) Discrete data (c) Primary data (d) Qualitative data

24. A characteristic that cannot be expressed numerically:

- (a) Continuous variable (b) Quantitative variable
(c) Attribute (d) discrete variable

25. A characteristic that changes from one individual to another:

- (a) Statistic (b) parameter (c) Constant (d) variable

26. Village patwari collecting data about crops:

- (a) Primary data (b) Secondary data (c) Qualitative data (d) Published data

27. Statistical laws are true for:

- (a) Short run (b) Long run (c) Both (a) and (b) (d) None of these

28. A numerical value calculated from sample:

- (a) Parameter (b) Statistic (c) Population (d) sample

29. A constant can assume ----- value(s):

- (a) Different (b) More than one (c) Only one (d) No value at all

30. Hourly temperature recorded by weather bureau:

- (a) Discrete data (b) Qualitative data (c) Secondary data (d) Continuous data

31. Registration is the source of:

- (a) Secondary data (b) Primary data (c) Published data (d) electronic data

32. The number of road accidents on M-1 during a month:

- (a) Discrete variable (b) Continuous variable
(c) Attribute (d) categorical variable

Introduction

33. Statistics has handicap dealing with:

- (a) Qualitative data (b) Quantitative data (c) Discrete data (d) All of the above

34. A variable that makes measurable values:

- (a) Constant (b) Discrete variable (c) Qualitative variable (d) Continuous variable

35. Identify the attribute:

- (a) Ages of patients (b) Temperature of a room
(c) Weights of children (d) Hobbies of students

36. Which of the following is not the example of continuous variable:

- (a) Smoking habits of college students (b) The hobbies of students
(c) The amount of rain fall in Muree (d) Both (a) and (b) but not (c)

37. Proportion becomes percentage when multiplied by:

- (a) 1/10 (b) 1/100 (c) 10 (d) 100

38. Inferential statistics deals with:

- (a) Collection of the data (b) Analysis of the data
(c) Presentation of the data (d) Interpretation of the data

Key

Sr.	Ans	Sr.	Ans	Sr.	Ans	Sr.	Ans	Sr.	Ans	Sr.	Ans
1	b	2	b	3	c	4	d	5	b	6	a
7	b	8	d	9	c	10	a	11	a	12	b
13	a	14	b	15	c	16	b	17	a	18	b
19	a	20	d	21	b	22	d	23	b	24	c
25	d	26	a	27	b	28	b	29	c	30	d
31	b	32	a	33	a	34	d	35	d	36	d
37	d	38	d								

Introduction

Exercise

Q No.1.1: Define statistics and its types descriptive and inferential statistics.

Q No.1.2: Explain the three different meanings of the word statistics.

Q No.1.3: Differentiate the followings terms.

- Qualitative and quantitative variable.
- Discrete and continuous variable.
- Primary and secondary data.
- Cross-sectional data and time series.
- Interval and ratio scale of measurements.

Q No. 1.4: Classify following variables as Attribute, Discrete and continuous, also name scale of measurement.

Variable	Type	Scale
Meal preference		
Political orientation		
Family size		
Income of a person		
Cost of a commodity		
Room number in a hotel		
Germination percentage		
Survival percentage		
Tree height		
Tree growth		
Stem form		
Biomass weight		
Crop yield		
Soil fertility		
Level of essential elements		

Q No.1.5: Define the followings.

Variable, Population, Sample, Parameter and Statistic.

Introduction

Q No.1.6: Write the names of primary and secondary sources related to the collection of the data.

Q No.1.7: What is measurement scale? Identify type and measurement scale for the following variables.

Variable	Type	Scale
Age of person		
Height of a person		
Religion of a person		
Marks obtained by a student		
GPA of a student		
Milk produced by a cow		
Yield of a crop		
Weight of an animal		
No of petals in a flower		
Fertilizer used in a field		
Quality of a seed		
Fertility of a land		
Grade of a student		
No of prayers offered by a person		
Hair colour of a person		

Q No.1.8: From the following table of variables indicate the relationship.

Variable		Variable
Shoes size of a person		Age of a person
Heat		Height of a person
Yield of a crop		Age of a bird
		Quantity of fertilizer used
		Temperature

Introduction

Solution

Q No. 1.4: Classify following variables as Attribute, Discrete and continuous, also name scale of measurement.

Variable	Type	Scale
Meal preference	Attribute	Nominal
Political orientation	Attribute	Nominal
Family size	Discrete	Ratio
Income of a person	Continuous	Ratio
Cost of a commodity	Continuous	Ratio
Room number in a hotel	Attribute	Nominal
Germination percentage	Continuous	Ratio
Survival percentage	Continuous	Ratio
Tree height	Continuous	Ratio
Tree growth	Continuous	Ratio
Stem form	Attribute	Nominal
Biomass weight	Continuous	Ratio
Crop yield	Continuous	Ratio
Soil fertility	Attribute	Ordinal
Level of essential elements	Attribute	Ordinal

Q No.1.7: What is measurement scale? Identify type and measurement scale for the following variables.

Variable	Type	Scale
Age of person	Continuous	Ratio
Height of a person	Continuous	Ratio
Religion of a person	Attribute	Nominal
Marks obtained by a student	Continuous	Ratio
GPA of a student	Continuous	Ratio
Milk produced by a cow	Continuous	Ratio
Yield of a crop	Continuous	Ratio
Weight of an animal	Continuous	Ratio

Introduction

No of petals in a flower	Discrete	Ratio
Fertilizer used in a field	Continuous	Ratio
Quality of a seed	Attribute	Ordinal
Fertility of a land	Attribute	Ordinal
Grade of a student	Attribute	Ordinal
No of prayers offered by a person	Discrete	Ratio
Hair colour of a person	Attribute	Nominal

Q No.1.8: From the following table of variables indicate the relationship.

Variable	Variable
Shoes size of a person	Age of a person
Heat	Height of a person
Yield of a crop	Age of a bird
	Quantity of fertilizer used
	Temperature

Q No.1.9: From the following table of variables indicate the relationship.

Variable	Variable
Price of a commodity	Experience of a person
Quantity sold	Quantity sold
Income of a person	Income of a person
Profit	Profit
Investment	Investment
Quantity demanded	Quantity demanded