

Session 1 - 3

Introduction to Business Statistics

Prof. Jigar M. Shah

Introduction to Business Statistics

- Meaning & Application
 - Origin & Development of Statistics
 - Definition of Statistics
 - Scope & Application of Statistics
 - Limitations & Distrust of Statistics
- Data & Statistics
 - Data Lexicon
 - Scales of Measurement
 - Data Classification
 - Data Sources & Collection
 - Descriptive & Inferential Statistics

Introduction to Business Statistics

Meaning & Application

- Origin & Development of Statistics

Status (Latin)	Statista (Italian)	Statistik (German)	Political State
India	Europe	Germany	England
Chandra Gupta Maurya (324 - 300 B.C.) Birth & death records Akbar (1556 - 1605 AD) Land records & agricultural statistics	17 th century Theory of Probability and Theory of Games & Chance (Modern Statistics) by mathematicians & gamblers of France, Germany & England	18 th century Population records and industrial & agricultural output	19 th century Government revenue & expenditure records 20 th century Application in genetics, biometry, education, agriculture, etc. by Sir Ronald A Fisher (1890 – 1962), Father of Statistics
Sir Ronald A Fisher's contributions won for Statistics a very responsible position among sciences			

Introduction to Business Statistics

Meaning & Application

- Definition of Statistics
 - Different definitions at different times by different authors (based on field of application, either as statistical data or as statistical methods)

Statistics as Statistical Data

Classified facts representing the conditions of the people in a State ... especially those facts which can be stated in numbers or in any other tabular or classified arrangement - [Webster](#)

Numerical statements of facts in any department of enquiry placed in relation to each other - [Bowley](#)

Aggregates of facts affected to a marked extent by multiplicity of causes numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other - [Prof. Horace Secrist](#)

Introduction to Business Statistics

Meaning & Application

- Definition of Statistics

Statistics as Statistical Methods

Science of computing - [Bowley](#) Science of averages - [Bowley](#) Science of estimates & probabilities - [Boddington](#)

Science of measurement of social organism, regarded as a whole in all its manifestations - [Bowley](#)

Science of judging collective, natural or social phenomenon from the results obtained from the analysis or enumeration or collection of estimates - [King](#)

Science which deals with collection, classification & tabulation of numerical facts as the basis for explanation, description & comparison of phenomenon - [Lovitt](#)

Science which deals with the collection, analysis and interpretation of numerical data - [Croton & Cowden](#)

- [Statistics - art and science of collecting, analyzing, presenting, and interpreting data](#)

Introduction to Business Statistics

Meaning & Application

- Scope & Application of Statistics

Scope of Statistics

Planning, Economy, Business, Industry, Mathematics, Biology, Astronomy, Medical Science, Psychology, Education, War, Politics, Environment

Application of Statistics

Planning

"Bengaluru sees over 1,500 new vehicles entering the roads each day in a city that has over 7.5 million vehicles and a population exceeding 10 million"

"The Mumbai Metro Line 3, a 32.32-kilometer long elevated corridor with 32 stations, and expected to reduce travel time between Wadala to Kasarvada by 50 to 75 percent, is being built at a cost of ₹14,549 crore (\$2.26 billion USD) for an expected daily ridership of 8.70 lakhs in 2021 and growing to 12.13 lakhs by 2031"

Introduction to Business Statistics

Meaning & Application

- Scope & Application of Statistics

Application of Statistics

"Consumer pricing index (CPI) data showed that inflation was 7.01 per cent in June 2022, compared to 7.04 per cent in May 2022 and 7.79 per cent in April 2022"

"Economic growth in Organization for Economic Co-operation and Development (OECD) region dipped to 0.1% in Q1 of 2022, compared with 1.2% in Q4 of 2021"

Economy

"According to the provisional estimates of GDP released on May 31, 2022, the Indian economy has fully recovered to the pre-pandemic real GDP level of 2019-20 with real GDP growth in FY 2021-22 at 8.7%, which is 1.5% higher than the real GDP in FY 2019-20"

"China has the world's largest economy with \$27.31 trillion in terms of PPP, followed by the US' with \$22.99 trillion at the second spot, and India's at the third position with \$10.21 trillion, as per World Bank data from 2021"

Introduction to Business Statistics

Meaning & Application

- Scope & Application of Statistics

Application of Statistics

Business

"India has the world's highest data usage per smartphone at an average of 9.8GB per month"

"The cost of EV per kilometre of delivery is less than ₹1/km, while it is ₹3.5/km for petrol-based vehicles"

"Early adopters in India are willing to pay a premium of ₹320 over their average monthly spends for 5G services"

Industry

For quality control to find out parts matching the required specifications, to construct control charts, process capabilities, etc.

Mathematics

Many great mathematicians, Bernouli, Pascal, Laplace, Gauss, R. A. Fisher, etc. have contributed to the field of statistics

Recent advancements in statistical techniques are the outcome of wide applications of advanced mathematics

Introduction to Business Statistics

Meaning & Application

- Scope & Application of Statistics

Application of Statistics

According to Prof. Karl Pearson, the theory of heredity rests on statistical basis

Biology

Biostatistics - statistical techniques used to solve problems related to medicine, health, & biology (e.g., to study the causes, spread & treatments of diseases)

Astronomy

Theory of Gaussian 'Normal Law of Errors' to study movement of stars & planets is developed using the 'Principle of Least Squares'

Medical Science

"India has the world's highest burden of TB, with 27 per cent of all global cases"

"Babies born after bariatric surgery were 200 gm lighter on average than those born to mothers without any surgery"

"Findings of a study indicate that eating just one extra portion of fruits and vegetables a day could have an equivalent effect on mental well-being as around 8 extra days of walking a month (for at least 10 minutes at a time)"

Introduction to Business Statistics

Meaning & Application

- Scope & Application of Statistics

Application of Statistics

Psychology

To measure skills, knowledge, abilities, attitudes, personality traits, etc.

Education

"The average salary across top B-schools was over ₹20 lakh p.a., up 5-16% from last year"

"School enrollment of children in the age group 6-14 is over 96%, and has been 96% or higher since 2010"

War

Decision Theory to assist military & technical personnel to plan maximum destruction with minimum effort

Politics

"From 16 percentage points in 2014, BJP's median victory margin has jumped to 20 percentage points in 2019"

"In the 2019 elections, on average, 14.8 candidates contested per constituency across the country"

Environment

"On average, Greenland's ice sheet is 1.6 miles thick, but there is a lot of variation depending on where you measure"

"Brihanmumbai Municipal Corporation (BMC) said in the last five hours, its weather station has recorded an average rainfall of 43.23 mm in the island city, 64.14 mm in eastern suburbs and 78.21 mm in western suburbs"

Introduction to Business Statistics

Meaning & Application

■ Limitations & Distrust of Statistics

Limitations of Statistics

- Not suited to the study of qualitative phenomenon
- Statistical laws are not exact
- Does not study individuals
- Liable to be misused

Distrust of Statistics

- Figures may be inaccurate or incomplete
- Figures may be molded or manipulated, to conceal the truth

"The number of accidents taking place in the middle of the road is much less than the number of accidents taking place on its side. Hence it is safer to walk in the middle of the road."

"The number of students taking up Mathematics Honors in a University has increased 5 times during the last 3 years. Thus, Mathematics is gaining popularity among university students."

"99% of the people who drink alcohol die before attaining the age of 100 years. Hence drinking is harmful for longevity of life."

Introduction to Business Statistics

Data & Statistics

- Data Lexicon

Term	Meaning
Data	Facts and figures collected, analyzed, and summarized for presentation and interpretation
Data Set	All data collected in a particular study
Elements	Entities on which data are collected
Variable	A characteristic of interest for the elements
Observation	Set of measurements obtained for a particular element

Introduction to Business Statistics

Data & Statistics

- Data Lexicon

Scheme Name	Category	Morningstar Rating	NAV	1-year Return (%)	5-year CAGR (%)
Mirae Asset Emerging Bluechip Fund - Growth	Equity	5-star	54.17	15.17	20.51
IDFC Cash Fund - Growth	Debt	5-star	2295	7.28	7.58
BNP Paribas Liquid Fund - Growth	Debt	5-star	2905.23	7.49	7.59
Axis Liquid Fund - Growth	Debt	5-star	2100.62	7.51	7.67
ICICI Prudential Liquid Fund - Retail Plan - Growth	Debt	4-star	431.27	7.04	7.1

Introduction to Business Statistics

Data & Statistics

- Data Lexicon

Scheme Name	Category	Morningstar Rating	NAV	1-year Return (%)	5-year CAGR (%)
Mirae Asset Emerging Bluechip Fund – Growth	Equity	5-star	54.17	15.17	20.51
IDFC Cash Fund - Growth	Debt	5-star	2295	7.28	7.58
BNP Paribas Liquid Fund - Growth	Debt	5-star	2905.23	7.49	7.59
Axis Liquid Fund - Growth	Debt	5-star	2100.62	7.51	7.67
ICICI Prudential Liquid Fund - Retail Plan - Growth	Debt	4-star	431.27	7.04	7.1
Elements	Variables	Observation	Data Set		

Introduction to Business Statistics

Data & Statistics

- Data Lexicon
 - No. of observations always equals the no. of elements
 - No. of measurements obtained for each element equals the no. of variables
 - Total no. of data items equals the product of no. of observations (or elements) and no. of variables

Introduction to Business Statistics

Data & Statistics

- Scales of Measurement

Scale	Description
Nominal	Data for a variable consists of labels or names used to identify an attribute
Ordinal	If data exhibit the properties of nominal data, & the order / rank of the data is meaningful
Interval	If data exhibit all the properties of ordinal data, & the interval between values is expressed in terms of a fixed unit of measure
Ratio	If data exhibit all the properties of interval data, & the ratio of two values is meaningful

Introduction to Business Statistics

Data & Statistics

■ Data Classification

Qualitative Data (Categorical Data)

- Data that can be grouped by specific categories
- Are measures of '**types**' and may be represented by a name, symbol, or a numeric code
- Signify category to which an item belongs to
- Use either nominal or ordinal scale of measurement
- Qualitative variable - one with qualitative or categorical data
- **E.g.**, responses to questions like 'What color is your car?'
- Special case with only two response options (usually 'yes' & 'no')

Quantitative Data

- Data that use numeric values to indicate **how much or how many**
- Are measures of values or counts
- Expressed as numbers
- Use either interval or ratio scale of measurement
- Quantitative variable - variable with quantitative data
- **E.g.**, responses to questions like 'How many runs will India score in the next match?'

Introduction to Business Statistics

Data & Statistics

- Data Classification
 - Variables with non-numeric values - Qualitative
 - Variables with numeric values - Qualitative or Quantitative
 - Qualitative data can be summarized by counting the number of observations in each category or by computing the proportion of observations in each category
 - Arithmetic operations can be performed on quantitative data

Introduction to Business Statistics

Data & Statistics

- Data Classification

Order No.	Name	State	Amount (₹)	Area Code	Gift?
105-4916-104	Kishore H.	Maharashtra	1840	425409	N
105-0857-311	Shirish P.	Karnataka	1695	563138	Y
105-9684-118	Chirag G.	Gujarat	1520	393135	Y
103-6815-217	Manisha D.	Punjab	1759	148104	N
017-9583-171	Kavita L.	Uttar Pradesh	1557	209728	N

Introduction to Business Statistics

Data & Statistics

■ Data Classification

Cross-Sectional Data

- Data observed at the same or approximately the same point in time

Scheme Name	Category	Morningstar Rating	NAV	1-year Return (%)	5-year CAGR (%)
Mirae Asset Emerging Bluechip Fund - Growth	Equity	5-star	54.17	15.17	20.51
IDFC Cash Fund - Growth	Debt	5-star	2295	7.28	7.58
BNP Paribas Liquid Fund - Growth	Debt	5-star	2905.23	7.49	7.59
Axis Liquid Fund - Growth	Debt	5-star	2100.62	7.51	7.67
ICICI Prudential Liquid Fund - Retail Plan - Growth	Debt	4-star	431.27	7.04	7.1

Time Series Data

- Data observed over several time periods



Introduction to Business Statistics

Data & Statistics

■ Data Classification

Discrete Data

- Data that measure **how many**
- Contains distinct or separate, finite values that have nothing in-between i.e., sub-division is not possible
- Relies on count, includes only those values that can only be counted in whole numbers or integers i.e., data cannot be broken down into fractions or decimals
- **E.g.**, number of students in this class

Continuous Data

- Data that measure **how much**
- Unbroken set of values measured on a scale
- No separation occurs between possible data values
- Can take any value, within a finite or infinite range of possible values
- Relies on measurement, includes values that can be measured and broken down into fractions and decimals according to the measurement precision
- **E.g.**, weight of a person

Introduction to Business Statistics

Data & Statistics

- Data Classification

	Discrete Data	Continuous Data
Definition	A set of data is said to be discrete if the values belonging to the set are distinct & separate (unconnected values)	A set of data is said to be continuous if the values belonging to the set can take any value within a finite or infinite interval
Meaning	Data with clear space between values	Data that falls on a continuous sequence
Nature	Countable	Measurable
Graphical Representation	Bar Graph	Histogram
Function	In graph of discrete function, only separate, distinct points are plotted	In graph of continuous function, points are connected with a continuous line
General Note	Description is usually represented by 'number of ...'	Usually require a measuring device

Introduction to Business Statistics

Data & Statistics

- Data Sources & Collection

Existing Sources

When data needed for an application already exists, the existing data is the used for the application

Secondary Data	data that have already been collected by some researchers or investigators in the past and are available either in published or unpublished form
----------------	--

Impure information, as statistical operations may have already been performed on the data

Internal Company Records	employee records, production records, inventory records, sales records, credit records, customer profile, etc.
--------------------------	--

Private, commercial organizations	Dun & Bradstreet, Bloomberg, Dow Jones & Company, Nielsen, etc.
-----------------------------------	---

Introduction to Business Statistics

Data & Statistics

- Data Sources & Collection

Existing Sources

Industry Associations

SIAM, COAI, ASSOCHAM, NASSCOM, etc.

Government Agencies

Census Bureau, Reserve Bank, Finance Ministry, IMD, MoSPI, etc.

Existing sources of data desirable when time is a constraint

If important data not readily available, additional time & cost involved in obtaining data must be taken into account

Introduction to Business Statistics

Data & Statistics

- Data Sources & Collection

Statistical Studies

When data needed for a particular application are not readily available, data can be obtained by conducting a statistical study

Primary Data data that are collected for the first time by a researcher or investigator for a specific purpose

Pure information, as no statistical operations have been performed on the data, and they are original data

Two types of statistical studies

Experimental

Observational (Non-experimental)

Introduction to Business Statistics

Data & Statistics

- Data Sources & Collection

Statistical Studies

Experimental Statistical Studies

statistical studies where one or more other variables are controlled so that data can be obtained about how they influence the variable of interest
e.g., clinical trials

Observational Statistical Studies

statistical studies where no attempt is made to control any variable and data are noted based on examination, inspection, interview, monitoring, or surveying
e.g., survey

Introduction to Business Statistics

Data & Statistics

- Data Sources & Collection

Data Collection Methods

Experiment	defined as any process that generates data as its outcome, it involves experimental design that identifies one or more factors to be manipulated, changed, controlled or observed, to measure or observe the impact on the variable of interest	
Direct Observation	requires effective, trained observers, and can be costly, time-consuming, and subjective	
Personal Interview	Structured Interview	involves scripted questions
	Unstructured Interview	begins with one or more broadly stated questions, with further questions being based on the responses

Introduction to Business Statistics

Data & Statistics

- Data Sources & Collection

Data Collection Methods

Questionnaire

consists of a list of questions, may be along with the choice of answers, printed or typed in a sequence on a form used for acquiring specific information from the respondents

delivered to the persons concerned (respondents) either by post or e-mail, requesting them to answer the questions and return it

respondents expected to read & understand the questions in the questionnaire and reply in the space provided in the questionnaire itself

translates the required information into a series of questions, that respondents can & will answer

inexpensive, free from interviewer bias, provides enough time to respondents to think & answer, and provides large coverage as it is convenient to reach out to respondents in distant areas

Introduction to Business Statistics

Data & Statistics

- Data Sources & Collection

Data Collection Methods

Schedule

a proforma containing a list of questions filled in by the research worker or enumerators, specially appointed for the purpose of data collection

enumerators go to the informants with the schedule, ask them the questions from the set in the sequence & record the replies in the space provided

enumerators explain the aims & objects of the research to the respondents and interpret the questions to them when required

little expensive as selection, appointment & training of enumerators require a huge amount

used in case of extensive enquiries conducted by the government agencies, big organizations, e.g., - population census

Introduction to Business Statistics

Data & Statistics

- Data Sources & Collection

Data Collection Methods

	Questionnaire	Schedule
Meaning	A technique of data collection that consists of a series of written questions along with alternative answers	A formalized set of questions, statements & spaces for answers, provided to the enumerators who ask questions to the respondents and note their answers
Filled by	Respondents	Enumerators
Response Rate	Low	High
Coverage	Large	Comparatively small

Introduction to Business Statistics

Data & Statistics

- Data Sources & Collection

Data Collection Methods

	Questionnaire	Schedule
Cost	Economical	Expensive
Respondents' Identity	Not known	Known
Success Relies on	Quality of the questionnaire	Honesty & competence of the enumerators
Usage	Only when people are literate and cooperative	Used on both literate and illiterate people

Introduction to Business Statistics

Data & Statistics

- Data Sources & Collection

Data Collection Methods

Open-ended Questions in Questionnaire & Schedule

questions that allow respondents the freedom to respond with any value, words, or statements of their own choosing

provide respondents with greater flexibility in answering a question

responses can be difficult to analyze

Closed-ended Questions in Questionnaire & Schedule

questions that allow respondents the freedom to choose from any of the provided response options

provide respondents with limited flexibility in answering a question

responses can be easily analyzed

Introduction to Business Statistics

Data & Statistics

- Data Sources & Collection

Data Collection Issues

Bias	an effect that alters a statistical result by systematically distorting it (different from a random error that may distort on any one occasion but balances out on the average)
Interviewer Bias	bias interjected by the way interviewer asks questions, by the tone of voice, by the way interviewer looks at the interviewee, etc.
Non-Response Bias	bias interjected due to non-response from a section of the respondents
Selection Bias	bias interjected by way subjects are selected
Observer Bias	bias interjected due to personal perceptions of observer

Introduction to Business Statistics

Data & Statistics

- Data Sources & Collection

Data Collection Issues

Measurement Error	errors made in measuring data
-------------------	-------------------------------

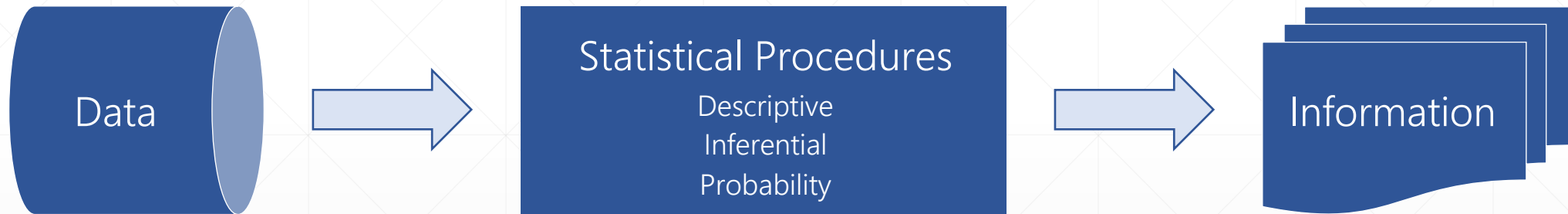
Internal Validity	<p>a characteristic of an experiment in which data are collected in such a way as to eliminate the effects of variables within the experimental environment that are not of interest to the researcher</p> <p>for data to have internal validity, extraneous factors must be controlled</p>
-------------------	---

External Validity	<p>a characteristic of an experiment whose results can be generalized beyond the test environment so that the outcomes can be replicated when the experiment is repeated</p> <p>if results of an experiment can be replicated for groups different from the original population, then there is evidence the results of the experiment have external validity</p>
-------------------	--

Introduction to Business Statistics

Data & Statistics

- Descriptive & Inferential Statistics
 - Role of statistics involves converting data into information using various statistical procedures



Descriptive Statistics

- Consists of methods for organizing and summarizing information
- Includes procedures and techniques specially designed to describe data
- Visual / pictorial description through charts, diagrams, graphs, tables, etc.
- Numerical description through calculation of various measures such as averages, variation, percentiles, etc.

Introduction to Business Statistics

Data & Statistics

- Descriptive & Inferential Statistics

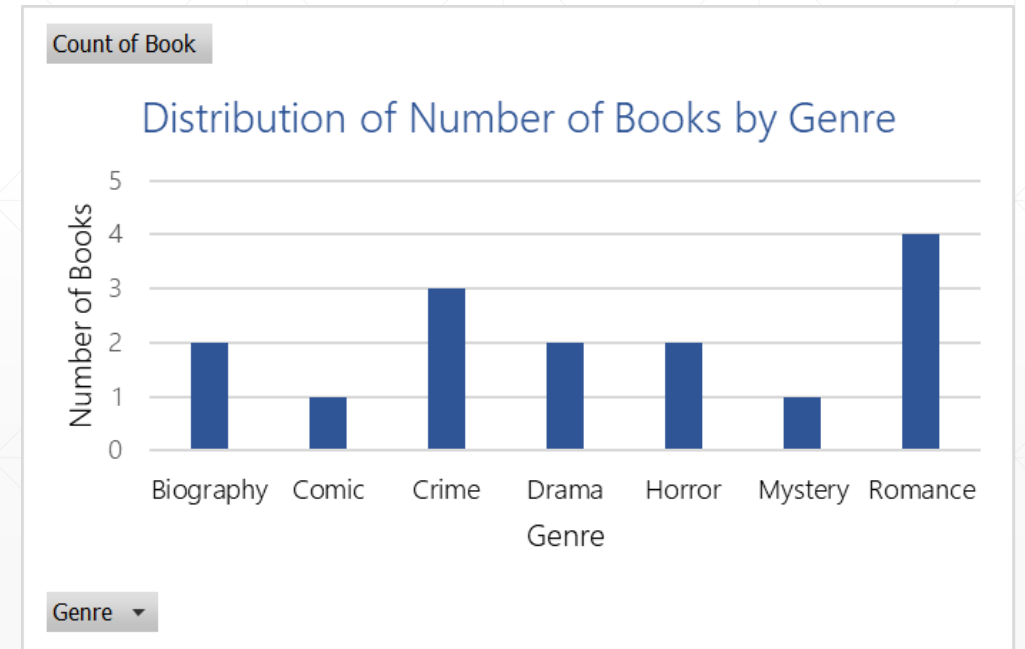
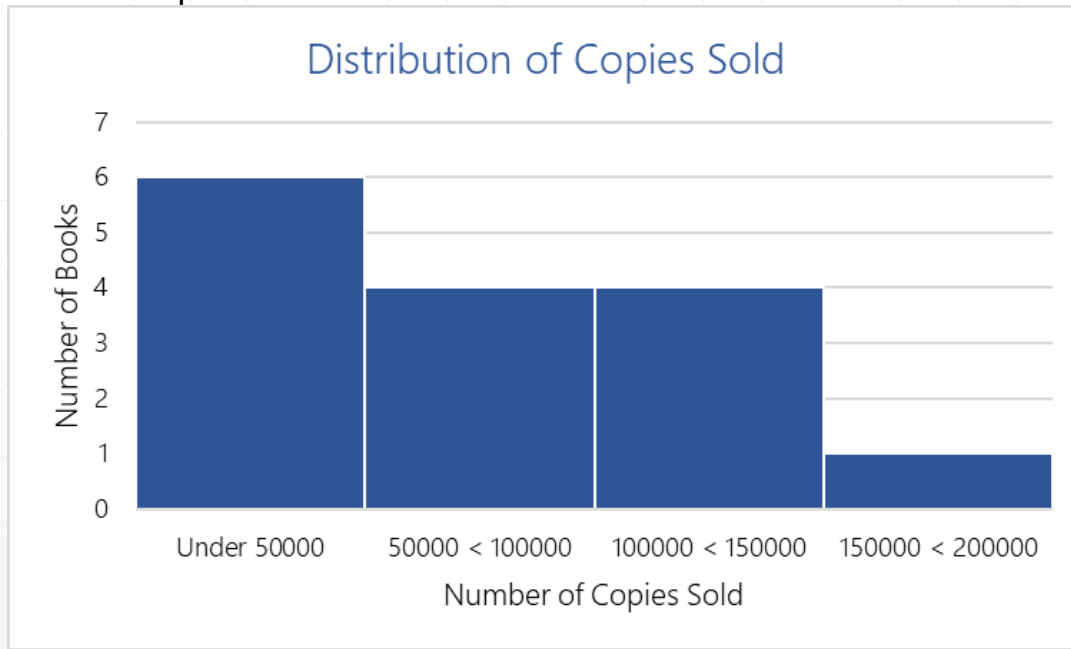
- Descriptive Statistics

Book	Copies Sold	Pages	Year Published	Advertising Budget (₹)	Age of Author	Genre
1	15000	176	2018	25000	49	Comic
2	140000	296	2013	83000	57	Drama
3	75000	483	2017	40000	29	Crime
4	100000	811	2016	29000	37	Mystery
5	26000	302	2018	52000	35	Horror
6	33000	411	2018	33000	43	Romance
7	59000	333	2018	19000	51	Biography
8	103000	602	2015	37000	62	Drama
9	88000	504	2016	51000	33	Romance
10	10000	204	2019	30000	50	Crime
11	9000	376	2019	19000	26	Romance
12	124000	600	2018	41000	40	Horror
13	59000	400	2014	26000	44	Crime
14	183000	597	2016	51000	59	Biography
15	16000	126	2015	28000	38	Romance

Introduction to Business Statistics

Data & Statistics

- Descriptive & Inferential Statistics
 - Descriptive Statistics



Introduction to Business Statistics

Data & Statistics

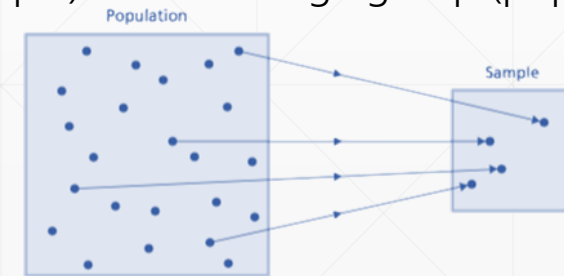
- Descriptive & Inferential Statistics

- Descriptive Statistics

Genre	Frequency	Percent Frequency
Biography	2	13.33
Comic	1	6.67
Crime	3	20.00
Drama	2	13.33
Horror	2	13.33
Mystery	1	6.67
Romance	4	26.67
Total	15	100

- Statistical Inference

- When collecting data for each & every element belonging to a large group of elements (individuals, companies, voters, households, products, consumers, etc.) may be difficult, costly or time-consuming, data is collected from only a small portion (sample) from the large group (population)



Introduction to Business Statistics

Data & Statistics

- Descriptive & Inferential Statistics
 - Statistical Inference

Term	Definition	Term	Definition
Population	Set of all elements of interest in a particular study	Sample	Subset of the population (part of the population from whose elements data is collected)
Census	Process of conducting a survey collect data for the entire population	Sample Survey	Process of conducting a survey to collect data for a sample

Statistical Inference - process through which statistics uses data from a sample to make estimates & test hypotheses about the characteristics of a population

Introduction to Business Statistics

Data & Statistics

- Descriptive & Inferential Statistics
 - Inferential Statistics

Inferential Statistics

- Consists of methods for drawing and measuring the reliability of conclusions about a population based on information obtained from a sample of the population
- Includes inferential procedures that help decision makers draw inferences from a set of data
- Inferential procedures include estimation & hypothesis testing
- E.g., To increase the useful life of its high intensity light bulbs, the product design group at Norris Electronics developed a new light bulb filament. In this case, the population is defined as all light bulbs that could be produced with the new filament. To evaluate the advantages of the new filament, 200 bulbs with the new filament were manufactured & tested. Data collected from this sample shows the number of hours each light bulb operated before filament burnout.

Introduction to Business Statistics

Data & Statistics

- Descriptive & Inferential Statistics
 - Inferential Statistics

Inferential Statistics

- To make an inference about the average hours of useful life for the **population** of all light bulbs that could be produced with the new filament, Norris can use the **sample average** lifetime for the light bulbs
- The sample result can be used to estimate that the average lifetime for the light bulbs in the population is 76 hours

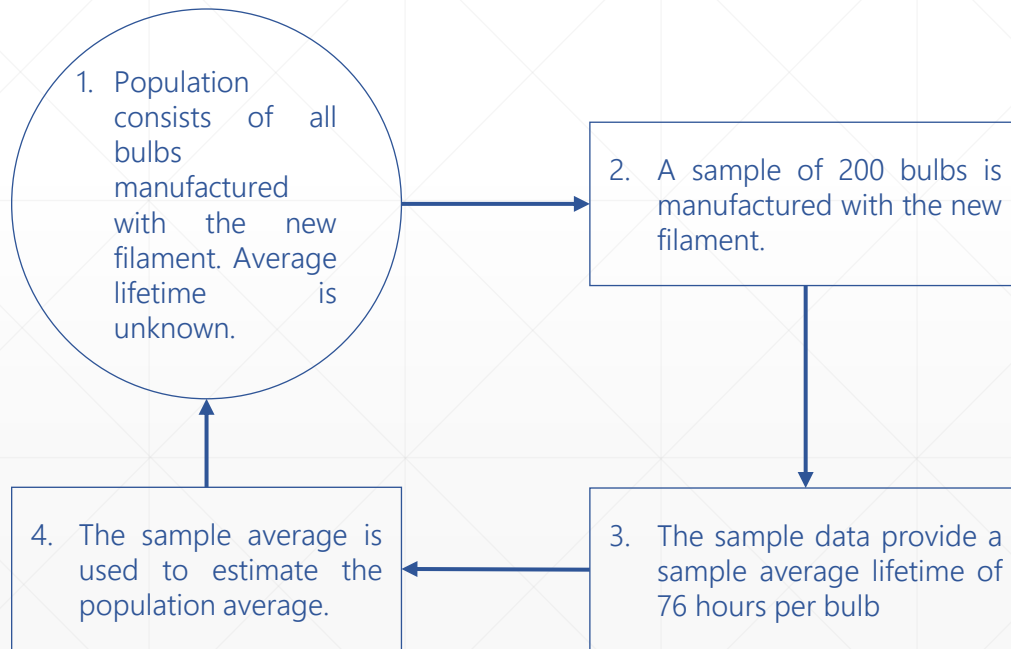
107	66	74	75	81	78	96	68	83	73
73	62	85	90	62	66	75	72	68	77
68	79	73	62	79	66	64	77	72	79
97	86	80	89	83	94	76	60	67	94
76	68	68	71	93	77	72	87	92	63
79	74	78	71	61	63	77	84	89	59
94	61	89	74	65	66	74	75	82	62
59	82	72	70	62	75	65	77	96	71
98	65	58	74	92	68	82	51	77	81
57	98	69	70	65	76	86	45	102	65
54	62	92	65	83	90	66	85	74	73
65	116	78	81	70	78	86	67	91	63
71	65	88	75	70	71	96	87	76	63
70	88	77	62	81	101	89	80	83	89
84	64	103	94	77	78	81	84	66	82
88	79	88	71	72	43	71	93	68	64
62	78	63	85	84	59	85	69	61	85
61	79	68	84	67	67	99	76	73	92
79	77	88	83	59	61	59	89	72	64
98	86	81	63	58	71	92	75	76	73

Hours until burnout for a sample of 200 light bulbs

Introduction to Business Statistics

Data & Statistics

- Descriptive & Inferential Statistics
 - Inferential Statistics



- Estimates about a characteristic of a population based on sample data are accompanied with a statement of the quality, or precision, associated with the estimate
- **Point estimate** of the average life-time for the population - 76 hours with a margin of error of 4 hours
- **Interval estimate** of the average lifetime for all light bulbs - 72 hours to 80 hours
- **Confidence levels** - % chances that interval from 72 hours to 80 hours contains the population average - can be used

Thank You

Prof. Jigar M. Shah