

World of Statistics

P.K. Chaubey
IIPA, New Delhi
pkchaubey@yahoo.com

Lecture Plan

- # Use of the Word Statistics
- # Types of Statistics; Not Mathematical Statistics!
- # Variable and Random Variable
- # Probability
- # Values and Frequencies
- # Presentation of Statistics
- # Statistic(s): one variable/two variables/ not n!
- # Use of right kind of statistics

Statistics

- # Three senses in which the word is used:
 - Discipline
 - Quantitative Data (what is data?)
 - Plural of statistic (sample estimator)
- # Has the State to do anything with statistics? Yes or No. [Achenwall 1749]
- # Note: Yes and No are two values of the variable 'Answer of some question' like Ayes (275) and Noes (265) in Parliament (note for vote)
- # Exercise: Write a Table of above distribution

3

Statistics as Quantitative Data

- # Data
 - Anything that has happened.
 - Data Files and Program Files
- # Quantitative Data
 - Observations (Chemistry titration, agriculture field)
 - Counts (head count, pugmarks)
 - Measurement (Physics screw gauge, temperature)
 - Responses (yes, no; agree, indifferent, disagree, 11-point)
- # Not so quantitative: Blondes 60, Brunets 40
- # Certain things that could have happened (scaffolding)

4

Types of Statistics: Reference to India

- # Descriptive (Hunterian): Director General of Statistical Survey Sir William W. Hunter (1869-81) was a historian who directed the gigantic statistical survey of India and published among other books, the famous Annals of Rural Bengal (three Volumes) and A History of British India (two Volumes). Imperial Gazetteers followed
- # Inferential (Mahalanobisian): A Professor of Physics Prasant Chandra Mahalanobis who per chance saw volumes of Biometrika before leaving England, got interested in use of statistics in a variety of subjects/problems and founded ISI and advocated and organised teaching of Statistics in India, starting from Calcutta University and Presidency College. Devised many statistical tools and survey methods. Created Statistical culture by hosting statisticians, economists, planners in India

5

Types of Statistics

- # Descriptive statistics: Involves the organization, summarization, and display of data; statement of summarized facts; A has x and B has y . At most the difference is $x-y$.
- # Inferential statistics: Involves using a sample to draw conclusions about a population; inferring from those summarized facts, involving some better methods, saying difference $x-y$ is significant, so A is superior/inferior to B. And so on...

6

Descriptive vs. Inferential

- # A says 50% of tobacco users have cancer; a descriptive statistic
- # B says 50% of non-tobacco user have cancer; another descriptive statistic
- # Cancer has nothing to do with tobacco-use, an inference
- # Go further deep: sample of tobacco users was 100 and sample of non-tobacco was 2!
- # Statistics is mass-related, not case-related

7

Variable

- # Attribute whose value changes from observation to observation: Height, weight, final degree, colour you like,
- # No. of children in family, Income of employee, Category of Babus, No. of Villages in Districts in UP, Electricity consumption by Area or Category of consumers , Size of plots in YEIDA, Voltage, Titration readings,
- # Colour of Flower, Complexion of Person, Caste of Person, Official Religion of Country, Shape of flags
- # Pink with Blush, Red with Anger, Pale with Anxiety, White with Fear, Blue with illness.
- # Length of rivers, heights of mountains, rainfall over places
- # (G, g, k, freezing temperature, max of C scale)? Do they vary? Are they variables?

8

Exercise: Tell possible values

- # Boundary:
- # Runs in an over:
- # Tribes in AP:
- # Pay Bands:
- # Monthly Salary...
- # Consumption of liquor...
- # Amount Spent on NREGA month by month..
- # Head of State...

9

Random Variable

Random Variable whose values are associated with probability between 0 and 1, not including them.

Single Variable Case: Write values in one column and associated probabilities/frequencies in the other column

Two Variable Case: How do we write?

10

Probability

- # A number between 0 and 1.
- # Sum of all probabilities=1.
- # Classical Approach; Frequency approach; Axiomatic approach
- # Meaning of zero probability?
- # Impossible event has zero probability but converse is not true.

11

Measurement of Variable

Attribute/characteristic whose value changes

- # Nominal: Blonde and Brunette
- # Ordinal: Above 25, Below 25
- # Numeral
 - Ordinal: 1st, 2nd
 - Cardinal: Interval and Ratio
 - Discrete
 - Continuous

12

Statistical Distribution: Table

- # Variable
- # Frequency
- # Single variable distribution: two columns: (1) values with (2) corresponding frequencies
- # Two-variable distribution: two dimensional table with joint frequencies in cells like caste affiliation and education level or complexion of body and colour of eye/hair
- # Multi-variate distribution with joint frequencies, not easy to write in a tabular form

13

Random Variable: Values and Frequencies

- # 1. No. of Children: Values 0, 1, 2, 3, 4, 5
- # 2. Boundaries: 4, 6
- # 3. Runs: 0, 1, 2, 3, 4, 5, 6
- # 4. Runs: 0, 1, 2,36
- # Frequencies:
 - # 1. Parents/Mothers/Fathers/Families
 - # 2. Balls
 - # 3. Balls
 - # 4. Overs
- # Time Series

14

Visual Presentation of Statistics

- # Frequency Distribution
- # Bar Diagram
- # Pie Chart
- # Histogram
- # Frequency Curve
- # Probability Density Curve

15

Population and Sample

- # Population vs. Sample
- # Parameter vs. Statistic

- # Hypothetical Population: Could-be's
could-have-been's, stochastic selection by
nature or by design
- # Haavelmo (1944): non-experimental data

16

Statistic(s): One variable

- # Maximum
- # Minimum
- # Mean
- # Median
- # Mode
- # Aggregate
- # Min/Max Ratio
- # Share
- # Poverty Line/Mean Income Ratio
- # Birth Rate, Death Rate, Life Expectancy
- # Efficient estimator?
- # Sample mean is an efficient estimator of population mean is but sample maximum is not efficient estimator of population maximum .

17

Statistic(s): Two and More Variables

- # Correlation:
 - Pearson: Linear
 - Spearman: Rank
- # Regression
 - Simple
- # Many Variables
 - Multiple, Partial Correlation
 - Multiple Regression

18

Three Most Popular Statistics

- # Aggregates: GDP, CO₂e, MTOe, Population of Humans, Livestock, Tigers, Blackbucks, Vehicles
- # Proportions: Ratio, Rates: Sex Ratio, Crime Rate, Conviction Rate, Teledensity, Population Density, Level of Urbanization, Literacy Rate,
- # Averages: Per Capita Income, Mean Monthly Consumption Expenditure, Mean Temperature, Sensex Level, Total/Net Fertility Rate, Life Expectancy at Birth
- # Birth Rate, Death Rate, Infant Mortality Rate, Maternal Mortality Rates; What are they?

19

Mean or Expected Mean

- # In Statistics it the expected mean drawn from the sample
- $$E(x) = \text{Summation}_i p_i x_i$$
- # We may note even arithmetic mean in statistics is generally expected mean.
 - # So are other Statistics

20

Save us from Statistics!

- # Benjamin Disraeli (1804-1881): Lies, Damn (Outrageous) Lies and Statistics, phrase popularised by Mark Twain. Also Leonard H. Courtney in 1895 who became President of Royal Statistical Society in 1897!
- # Parallel phrase by a Lawyer about witnesses : Simple Liars, Damn Liars, Experts
- # Should doctor use mean temperature to judge the condition of a patient or its tendency towards normal?
- # As elsewhere, we can use wrong statistics or wrong inference deliberately/innocently

21

Using Right Kind of Statistics

- # Village accountant with knowledge of Statistical computation

22

How to Lie with Statistics

- # Is a serious book on Statistics by Darrel I Huff, an American Journalist, 1954
- # credited with introducing statistics to a generation of college and high-school students on a level that was meaningful, available, and practical, while still managing to teach complex mathematical concepts
- # illustrated volume outlining common errors, both intentional and unintentional, associated with the interpretation of statistics, and how these errors can lead to inaccurate conclusions

23

In the end, not to shock

- # What is a population?
- # Does this class make a population or
- # the collection of your heights is a population and that of your weights another population, that of your services a third one, that of your incomes a fourth one, that of your children/friends a fifth one, that of your gender a sixth one, and of your age a seventh one; yes statistically speaking (but we can have joint distribution too)
- # Average height, average weight, modal service, model degree being pursued, average income, median age, average sex(?)

24

Misuses of Statistics

- # Professional scientists, even mathematicians and professional statisticians, can be fooled by even some simple methods, even if they are careful to check everything.
- # Scientists have been known to fool themselves with statistics due to lack of knowledge of [probability theory](#) and lack of [standardization](#) of their tests.

25

The word "THANKS" is written in a bold, 3D, yellow font with a slight shadow, centered on the slide.

26