

Qualitative & Quantitative Economics

A Statistical Perspective

By Tanujit Chakraborty, PhD (Ongoing)

Indian Statistical Institute, Kolkata

www.ctanujit.com

At Centre For Development Studies, Kerala

Statistics as a Key Technology.....

“Statistics is the universal tool of inductive inference, research in natural and social sciences, and technological applications.

Statistics, therefore, must always have purpose, either in the pursuit of knowledge or in the promotion of human welfare.”

– Professor Prasanta Chandra Mahalanobis

A Roadmap.....

Collection of Data \rightarrow Summarization of Data \rightarrow Analysis of Data \rightarrow Interpretation of Data towards a VALID DECISION.

Problem in Statistics ?

Given a sample (a set of outcomes), we are to say (infer) about the population or the model.

Statistical data

- Uncertainty & Variability.
- Example : Height of Students at CDS Kerala.
- Counter Example : Multiplication table in a tabular form is a quantitative data, but since there is no uncertainty & variability involved in the data so it's not a Statistical Data.
- Statistical data are usually obtained by counting or measuring items.
 - **Primary data** are collected specifically for the analysis desired
 - **Secondary data** have already been compiled and are available for statistical analysis

Qualitative & Quantitative Data

Most data can be put into the following categories:

- **Qualitative** - data are measurements that each fall into one of several categories. (hair color, ethnic groups and other attributes of the population)
- **Quantitative** - data are observations that are measured on a numerical scale (distance traveled to college, number of children in a family, etc.)

Types of Qualitative & Quantitative data

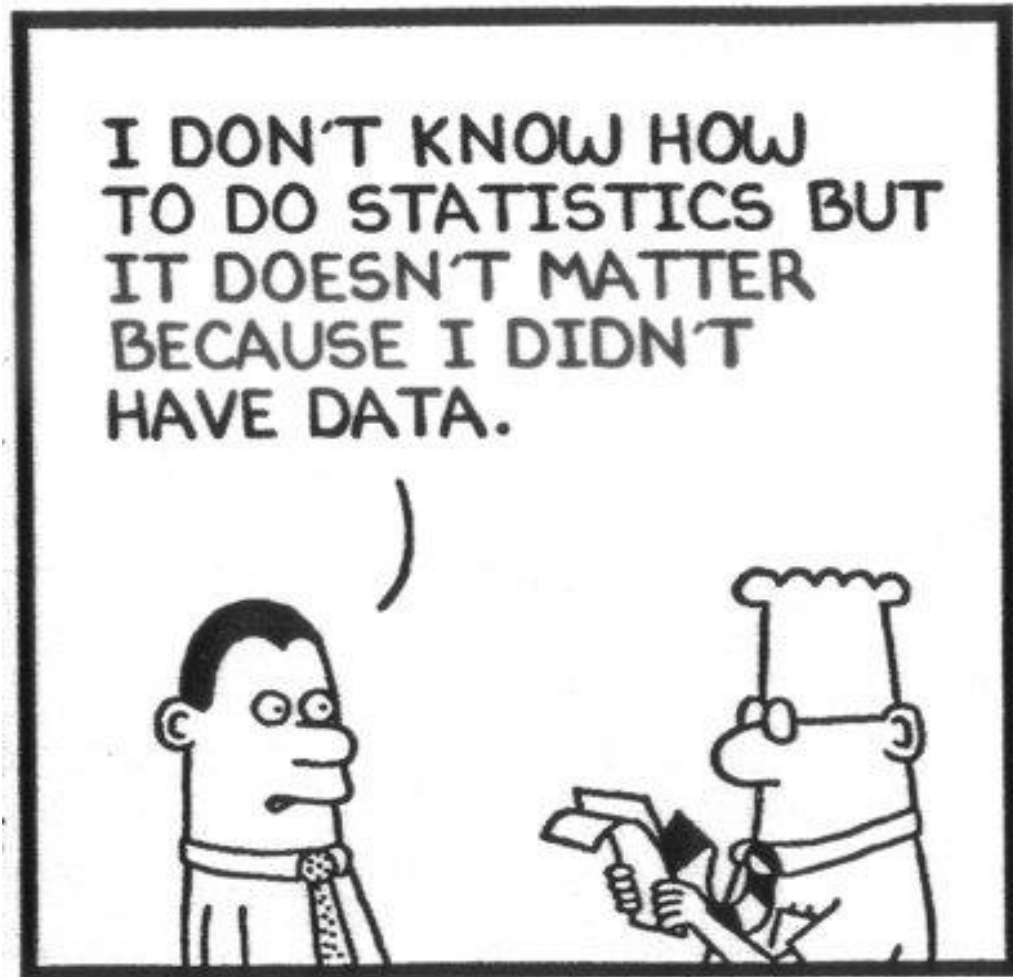
Qualitative data can be separated into two subgroups:

- **Dichotomic** (if it takes the form of a word with two options (gender - male or female))
- **Polynomic** (if it takes the form of a word with more than two options (education - primary school, secondary school and university)).

Quantitative data can be separated into two subgroups:

- **Discrete** (if it is the result of *counting* (the number of students of a given ethnic group in a class, the number of books on a shelf, ...))
- **Continuous** (if it is the result of *measuring* (distance traveled, weight of luggage, ...))

“Figure won't lie, but liar's figure”



Why study statistics?

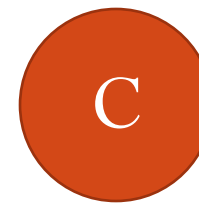
1. Data are everywhere
2. Statistical techniques are used to make many decisions that affect our lives
3. No matter what your career, you will make professional decisions that involve data. An understanding of statistical methods will help you make these decisions effectively

Applications of Statistics in real world problems

- Finance – correlation and regression, index numbers, time series analysis.
- Marketing – hypothesis testing, chi-square tests, nonparametric statistics.
- Development Studies – Gender Inequality, Child Health, Poverty : Econometrics & Applied Statistics.
- Demonetization – Text Mining Study
- Forecasting – Call Centre Call Forecasting using TS.
- Self-Driving Cars, Robotics, etc. as well uses STATISTICS extensively.

Solving these Problems Statistically

- What is Statistical Thinking ?
- How it is useful to solve Social Science Problems ?
- Consider a Problem.
- A 5 Phase Approach to solve.
- 5 Steps :



DEFINE A PROBLEM

- Identify the Problem.
- Define Scope.
- Data Requirement
 - I. **Voice of Customers** (Call Centre Data, Demonetization Data, Customer Service Records)
 - II. Survey Method based Data (Sample Survey (?), Sample Size (?), Sampling Techniques (?))
 - III. Secondary Data (NSSO CSO & other available Data)
- Create a Project Charter at the end.

MEASURING A PROBLEM

- Collect the data (economical, demographic, development related issues, etc.).
- Data Visualization (Tools !)
- Data Scrutiny (Mistakes, Missing data)
- Basic Statistics Check

ANALYSE THE DATA

- Finding Root cause of the problem.
- Prioritize root causes
- Relationship between Ys & Xs.
- Regression (?), Parametric, Non-parametric, Estimation (?), Testing (?), Time Series (?) & Panel Data Analysis (?), Econometric Modelling (?), Health Statistics, Vital Statistics, Data Mining.
- Classification & Clustering Problems.
- Supervised (LR, MLR, Decision Tree, kNN, SVM, ANN) Vs Unsupervised Approaches (Clustering & ARM (Famous Walmart Example) to model the data.

CHECKLIST BEFORE DATA ANALYSIS

- Basic Descriptive Analysis :
 - I. Descriptive Statistics Values (C.T., Dispersion, Shape, Relationship Plots)
 - II. Graphical Display of Data (Line, Bar, Pie, Histogram, Box-Plot, etc.)
- Outlier Tests, Normality Tests & Missing Values Check
- Multicollinearity, Autocorrelation, etc check.
- Variable Selection (PCA, FA, IV, Correlation Matrix, etc)
- Fitting Model & Validation (Adj. R-Square, p-value, MAPE, Confusion Matrix, etc)
- Interpret, Use

List of analyses of categorical data

- Correspondence Analysis (Categorical PCA)
- GLM (Non-Normal Error)
- Linear Discriminant Analysis
- Probit & Logit Models
- Poisson Regression
- Text Mining
- Image Processing

Etc...

List of analyses of numerical data

- Linear Regression
 - GLM
 - Time Series Forecasting
 - I. ARIMA
 - II. ARCH & GARCH Model
 - Panel Data Regression
 - Dummy Variable Regression
 - Polynomial Regression
 - Regression Splines (Non-parametric Regression)
- Etc...

SHORTCUT FOR YOU

Dependent Variable Type (Ys)	Independent Variable Type (Xs)	Modelling Technique
Numerical	Numerical	<ol style="list-style-type: none">1. Linear Regression (Best Subset Regression)2. Non-linear Regression, Regression Splines
Numerical	Categorical + Numerical	<ol style="list-style-type: none">1. Linear Regression with Dummy Variables2. Polynomial Regression with Dummy Variables
Categorical	Numerical	<ol style="list-style-type: none">1. Logistics Regression2. Regression Trees
Categorical	Categorical + Numerical	<ol style="list-style-type: none">1. Logistic Regression with Dummy Variables2. Classification Trees

IMPLEMENT THE FINDINGS

- Design of Experiments
- Feedback Data
- Analysis of Improvement results
- Decision Criteria

CONTROLLING THE PROBLEM

- Monitoring the problem.
- Adjust accordingly
- Write report.
- Publish the paper (😊).
- Sign off.

For other queries mail me at tanujitisi@gmail.com



THANK YOU