my-ebook

Parmeshvar

2025-07-06

Table of contents

1 Introduction

Introduction

DR.Harsh Pradhan, [Phone: +91-9930034241 , Email: harsh.231284@gmail.com], Institute of Management Studies, Banaras Hindu University, Address: 18-GF, Jaipuria Enclave, Kaushambhi, Ghaziabad, India, 2010
Interest: Goal Orientation Job Performance Consumer Behavior Behavioral Finance Bibiliometric Analysis Options as Derivatives Statistics Indian Knowledge System,

Orcid ID
Google Scholar
Youtube ID

Academic Profile

Courses offered:

1. Free online course, four weeks (MOOC), enrollments open: Introduction to Bayesian Data Analysis
2. Short (four-hour) tutorial on Bayesian statistics, taught at EMLAR 2022: here
3. Introduction to (frequentist) statistics
4. Introduction to Bayesian data analysis for cognitive science
5. BDA cover

1.1 Lecture notes

Download from here.

1.2 Moodle website

All communications with students in Potsdam will be done through this website. # Schedule

Week	Lecture	Main Topic	Subtopic	Video	PDF Resource
Week 2	1	Descriptive Statistics	Central Tendency	Video	Week 2.pdf
	2	Descriptive Statistics	Measure of Variability	Video	Same as above
	3	Descriptive Statistics	Describing Data	Video	Same as above
	4	Descriptive Statistics	Probability	Video	Same as above
	5	Descriptive Statistics	Distribution	Video	Same as above
Week 3	1	Descriptive Statistics	Z Table (Normal Distribution)	Video	Week 3.pdf
	2	Descriptive Statistics	Measuring Divergence	Video	Same as above
	3	Inferential Statistics	Sample and Population	Video	Same as above
	4	Inferential Statistics	Model Fit	Video	Same as above
	5	Inferential Statistics	Hypothesis and Error	Video	Same as above
Week 4	1	Terms of Statistics	Terms of Statistics	Video	Week 4.pdf
	2	Terms of Statistics	T-Test	Video	Same as above
	3	Terms of Statistics	T-Test in Detail	Video	Same as above
	4	ANOVA	ANOVA	Video	Same as above
Week 5	1	ANOVA	Example of ANOVA	Video	Week 5.pdf
	2	ANOVA	Types of ANOVA	Video	Same as above

Week	Lecture	Main Topic	Subtopic	Video	PDF Resource
	3	Correlation	Introduction to Correlation	Video	Same as above
	4	Correlation	Regression (Part 1)	Video	Same as above
	5	Correlation	Regression (Part 2)	Video	Same as above
Week 6	1	Correlation	R Script for Regression	Video	Week 6.pdf
	2	Chi Square	Chi Square	Video	Same as above
	3	Chi Square	Chi Square Test	Video	Same as above
	4	Logistic Function	Regression Function	Video	Same as above
	5	Logistic Function	Distribution	Video	Same as above
Week 7	1	Time Series	Intro to Time Series	Video	Week 7.pdf
	2	Time Series	Conditional Probability	Video	Same as above
	3	Time Series	Additional Concepts	Video	Same as above
	4	Time Series	Distribution	Video	Same as above
	5	Time Series	Poisson Distribution	Video	Same as above
	6	Index Numbers	Price & Quantity Index	Video	Same as above
	7	Decision Environments	Risk/Uncertainty, Bayes, Trees	Video	Same as above
	8	Time Series Analysis	Components, Trend, Seasonality	Video	Same as above
	9	Time Series Analysis	Least Squares Method	Video	Same as above
Week 8	1	Effect Size & Documentation	Package/Library	Video	Week 8.pdf

Week	Lecture	Main Topic	Subtopic	Video	PDF Resource
2		Effect Size & Documentation	RStudio vs RKward	Video	Same as above
3		Effect Size & Documentation	Flexplot	Video	Same as above
4		Effect Size & Documentation	Functions	Video	Same as above
5		Effect Size & Documentation	R Shiny & R Markdown	Video	Same as above
6		Effect Size & Documentation	Application with Real Datasets	Video	Same as above
7		Effect Size & Interpretation	Importance in Testing	Video	Same as above
8		Effect Size & Interpretation	Installing dplyr, ggplot2	Video	Same as above
9		Effect Size & Interpretation	Visual Model Interpretation	Video	Same as above
10		Effect Size & Interpretation	Creating/Using Functions	Video	Same as above
11		Effect Size & Interpretation	Report, Dashboard, Interactivity	Video	Same as above

2 Week 1

2.1 Module 1: Introduction to Statistics

2.1.1 Pre-Requisites

- Just an open and eager mind
- Basic understanding of Mathematics or Statistics

2.1.2 Agenda

- Meaning of Statistics
- Nature and Scope
- Uses of Statistics
- Limitations
- Fallacies and Misuse
- Math vs Statistics
- GUI Tools & Transition to Software-based Stats

2.1.3 Meaning of Statistics

Statistics is a science which provides tools for **analysis and interpretation** of raw data collected for decision-making in diverse fields.

It includes four core concepts:

- **Population** – Complete data or total group
- **Sample** – Subset of population
- **Parameter** – Numerical summary from population
- **Statistic** – Numerical summary from sample

2.1.4 Nature of Statistics

- Deals with **numerical facts**
- Focused on **social phenomena** and real-world data
- Organizes, classifies, and analyzes data
- Facilitates **prediction**, **interpretation**, and **decision-making**

2.1.5 Uses of Statistics

- Drawing representative samples
- Summarizing collected data
- Tabulation and systematic arrangement
- Group comparisons
- Determining behavioral relationships
- Estimating chance vs causation
- Application in:
  - Psychology
  - Education
  - Employment surveys
  - Market Research
  - Industrial and Organizational studies

2.1.6 Limitations of Statistics

- Cannot study **qualitative phenomena** without quantification
- Not applicable to individuals
- **Statistical laws are not exact**
- Does not guarantee **causal relationships**
- Vulnerable to misuse

2.2 Misuse of Statistics

- Use of extremely **small or biased** samples
- **Misleading graphs** or visual misrepresentation
- Illogical or **unexpected comparisons**

Fallacies in Statistics

Fallacies may arise from:

- Poor data collection methods
- Vague or manipulated term definitions
- Improper unit selection

- Faulty classification or grouping
- Inappropriate statistical methods

2.3 Module 2: Mathematics vs Statistics

Aspect	Mathematics	Statistics
Nature	Abstract, symbolic reasoning	Applied, data-based reasoning
Focus	Pure logic, proofs	Real-world data, decision-making
Techniques	Algebra, Calculus, Geometry	Probability, Hypothesis testing, Regression
Output	Theorems, functions, formulas	Inferences, predictions, summaries
Tools	Equations, graphs	Charts, tables, models

2.4 Module 3: Software-Based Statistical Revolution

From Paper to Code

Why shift to software?

- **Faster analysis** of massive data
- **Error-free calculations**
- **Anywhere-anytime** access
- **Cloud-based integration**
- Supports **ML/AI**, automation, and deep visualization

2.4.1 Popular Statistical Software

Software	Type	Use Case
R	Script	Core for academic and professional stats
RKWard	GUI	GUI wrapper for R
R Commander	GUI	Menu-based GUI for R
Rattle	GUI	Data mining toolkit in R
Excel	GUI	Basic stats with plugins
Python (pandas)	Script	Modern data science + ML

2.4.2 GUI vs CLI

Feature	GUI (e.g., RKWard)	Command Line (e.g., R Console)
Accessibility	User-friendly	Requires learning syntax
Speed	Slower for heavy tasks	High performance
Learning Curve	Minimal	Moderate to High

Feature	GUI (e.g., RKWard)	Command Line (e.g., R Console)
Customization	Limited	Fully scriptable
Teaching Utility	Good for beginners	Good for understanding logic

2.4.3 Recommended GUI Tools for R

- **RKWard**
- **Rattle**
- **R Commander**
- **R AnalyticFlow**

https://rkward.kde.org

2.4.4 Installing RKWard on Ubuntu

bash sudo apt install kbibtex kate libcurl4-openssl-dev libssl-dev libxml2-dev cmake sudo add-apt-repository ppa:rkward-devel/rkward-stable echo "deb https://ppa.launchpad.net/rkward-devel/rkward-stable/ubuntu jammy main" | sudo tee /etc/apt/sources.list.d/rkward.list sudo apt update sudo apt-get install rkward Awesome. Here's Part 2 of the full markdown, Lines 251–600, continuing the structured content from your Week 1 lecture.

2.5 Module 4: Understanding Variables

2.5.1 What is a Variable?

A **variable** is a characteristic or attribute that can assume different values across individuals or items.

In statistics, variables are categorized for analysis and measurement.

2.5.2 R Definition:

In R, variables are containers for data, created by assignment:

x <- 10 name <- "Harsh" flag <- TRUE

Classification of Variables

A. Qualitative (Categorical)

Type Description Example

Nominal Categories without order Gender (Male, Female) Ordinal Categories with a meaningful order Education Level (UG, PG)

B. Quantitative (Numerical)

Type Description Example

Discrete Countable numbers No. of students Continuous Infinite values in a range Height, Weight

Statistical Data Types (Scale of Measurement)

Data Type Description Examples

Nominal Categories with no order Blood group (A, B, AB, O) Ordinal Ranked categories Satisfaction (Low, Med, High) Interval Numeric scale with no true zero Temperature in Celsius Ratio Numeric scale with true zero Income, Weight, Age

Data Types in R

R Type Description Example Code

Numeric Real numbers x <- 15.3 Integer Whole numbers y <- as.integer(10) Complex Real + imaginary z <- 2+3i Character Text strings c <- “hello” Logical Boolean values b <- TRUE Factor Categorical encoding factor(c(“yes”, “no”, “yes”))

3 Examples in R

x <- 15.6 y <- as.integer(18) z <- 7 + 5i c <- “I am OK” b <- TRUE

Module 5: Data Structures in R

Vectors

A vector is a one-dimensional array of elements.

vec1 <- c(5, 2, 3, 7, 8, 9, 1, 4, 10, 15)

Matrices

Two-dimensional arrays of rows and columns.

mat <- matrix(1:9, nrow=3, ncol=3)

Arrays

Multidimensional generalization of matrices.

arr <- array(1:24, dim=c(3,4,2))

Lists

Collection of different types of elements.

mylist <- list(name=“Alice”, age=30, scores=c(89,90))

Data Frames

Tabular data (like a spreadsheet), each column can have a different type.

df <- data.frame(ID=1:3, Name=c(“A”, “B”, “C”), Score=c(85, 90, 95))

Factors

Used for categorical variables.

gender <- factor(c(“Male”, “Female”, “Male”))

Module 6: Descriptive Statistics

Descriptive statistics summarize and simplify data.

Central Tendency

Measure Formula Meaning

Mean $\bar{x} = \frac{\sum x_i}{n}$ Average Median Middle value in sorted data Central observation Mode Most frequent value Most common observation

Dispersion Measures

Measure Formula Purpose

Range $Range = Max - Min$ Spread of data Variance $s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$ Spread from mean Standard Deviation $s = \sqrt{Variance}$ Average distance from mean

Example in R

x <- c(10, 20, 30, 40, 50) mean(x) median(x) var(x) sd(x)

Module 7: Inferential Statistics

Inferential stats allow us to make conclusions about populations using samples.

Key Concepts

Hypothesis Testing: Assesses assumptions about a population.

Confidence Intervals: Estimate population parameters within a range.

Significance Levels ( ): Commonly 0.05 or 5%

P-Value: Probability of observing the data assuming the null is true.

Hypothesis Types

Type Description

Null Hypothesis No difference / no effect Alternative There is a difference / effect

R Examples

t.test(x) # One-sample t-test t.test(x, y) # Two-sample t-test

Module 8: Visualizing Data

Data visualization helps uncover patterns and insights.

Boxplot

Shows 5-number summary

Identifies outliers

boxplot(x)

Histogram

Frequency distribution of continuous data

hist(x)

Pie Chart

Shows proportion in categories

slices <- c(10, 12, 4, 16, 8) labels <- c("A", "B", "C", "D", "E") pie(slices, labels=labels)

Scatter Plot

Relationship between two variables

plot(x, y)

Ogive (Cumulative Frequency)

4 Create cumulative frequency table manually

Module 9: Spreadsheet Basics

Spreadsheets like Excel or Google Sheets are entry points for data work.

Key Features:

Rows → Observations

Columns → Variables

Supports sorting, filtering

Built-in formulas: =SUM(), =AVERAGE(), etc.

Spreadsheets vs R

Feature Spreadsheet (Excel, GSheets) R / RKWard

Cost Usually licensed Free and open source Flexibility Limited to GUI formulas Full programming capability Graphics Basic Advanced (ggplot2) Reproducibility Low High (script-based)

Module 10: Command Line vs GUI

Command Line (R Console)

5 Windows Command Line

cd .. mkdir new_folder dir

R Console Commands

getwd() setwd("path") install.packages("ggplot2") library(ggplot2)

GUI (RKWard)

Point-and-click interface

No coding needed

View script history and console

Menu for graphs, models, tables

Learning Resources:

Books

Mohanty, B., & Misra, S. (2016). Statistics for Behavioural and Social Sciences

Pandya et al. (2018). Statistical Analysis in Simple Steps using R

Field, A. P. et al. (2012). Discovering Statistics using R

Harris, J. K. (2019) . Statistics with R: Solving Problems using Real-World Data

5.1 Utilizing Statistical Methods for Decision Making

- Use statistical evidence to guide business strategies.
- Make informed policy decisions based on empirical data.
- Report findings clearly for transparency and comprehension.

5.2 Summary

The "Basic Statistics Using GUI-R (RK Ward)" course equips learners with the foundational and practical skills needed for statistical analysis using R. Students will understand theoretical concepts, grasp practical applications, and use RKWard effectively to analyze real-world data.

5.3 Key Takeaways

- Proficiency in defining and using variables and data types.
- Capability to import and manipulate data in RKWard.
- Understanding of basic statistical practices and their applications.
- Skill in visualizing data for effective communication of results.

5.4 Websites

https://rkward.kde.org https://r4stats.com https://cran.r-project.org