

**my-ebook**

Parmeshvar

2025-07-06

# Table of contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Introduction</b>	<b>6</b>
<b>3</b>	<b>Bayesian data analysis for cognitive science</b>	<b>7</b>
3.1	Introduction: What this course is about . . . . .	7
3.2	Teaching . . . . .	7
3.3	Lecture notes . . . . .	8
3.4	Moodle website . . . . .	8
<b>4</b>	<b>Schedule</b>	<b>9</b>
<b>5</b>	<b>Introduction to Basic Statistics Using GUI-R (RKWard)</b>	<b>11</b>
<b>6</b>	<b>Chapter 1: Welcome and Course Overview</b>	<b>12</b>
<b>7</b>	<b>Chapter 2: Agenda and Orientation</b>	<b>13</b>
<b>8</b>	<b>Chapter 3: Meaning and Nature of Statistics</b>	<b>14</b>
<b>9</b>	<b>Chapter 4: Applications and Uses</b>	<b>15</b>
<b>10</b>	<b>Chapter 5: Limitations and Misuse</b>	<b>16</b>
<b>11</b>	<b>Chapter 6: Paper-Based vs. Software-Based Statistics</b>	<b>17</b>
<b>12</b>	<b>Chapter 7: Introduction to Variables and Spreadsheets</b>	<b>18</b>
<b>13</b>	<b>Chapter 8: R and GUI Interfaces</b>	<b>19</b>
<b>14</b>	<b>Chapter 9: Importing Data and Understanding Data Types</b>	<b>20</b>
<b>15</b>	<b>Chapter 10: Statistical Data Types</b>	<b>21</b>
<b>16</b>	<b>Chapter 11: Data Preparation in RKWard</b>	<b>22</b>
<b>17</b>	<b>Chapter 12: Visualizing Data with Plots in RKWard</b>	<b>23</b>
17.1	1. Histogram . . . . .	23

17.2	2. Pie Chart . . . . .	23
17.3	3. Scatter Plot . . . . .	23
17.4	4. Box Plot . . . . .	24
17.5	5. Density Plot . . . . .	24
<b>18</b>	<b>Chapter 13: Summary</b>	<b>25</b>
<b>19</b>	<b>References</b>	<b>26</b>
<b>20</b>	<b>Next Steps</b>	<b>27</b>
<b>21</b>	<b>Basic Statistics Using GUI-R (RKWard)</b>	<b>28</b>
<b>22</b>	<b>Lecture 05: Essential Math for Statistics</b>	<b>29</b>
22.1	Key Concepts . . . . .	29
22.1.1	Number Line & Coordinate Geometry . . . . .	29
22.1.2	Exponential Functions . . . . .	29
22.1.3	Differentiation . . . . .	29
22.1.4	Integration . . . . .	29
<b>23</b>	<b>Lecture 06: Measures of Central Tendency</b>	<b>30</b>
23.1	Definitions . . . . .	30
23.2	Characteristics . . . . .	30
23.3	Suitability by Data Type . . . . .	30
23.4	Skewness . . . . .	30
<b>24</b>	<b>Lecture 07: Measures of Variability</b>	<b>32</b>
24.1	Measures . . . . .	32
24.2	Theoretical Moments . . . . .	32
<b>25</b>	<b>Lecture 08: Mean-Centered &amp; Standardized Data</b>	<b>33</b>
25.1	Mean-Centering . . . . .	33
25.2	Standardization (Z-Score) . . . . .	33
25.3	Applications . . . . .	33
<b>26</b>	<b>Lecture 09: Introduction to Probability</b>	<b>34</b>
26.1	Key Concepts . . . . .	34
26.2	Basic Rule . . . . .	34
26.3	Properties . . . . .	34
26.4	Visual Aids . . . . .	34
<b>27</b>	<b>Lecture 10: Distributions &amp; Binomial Distribution</b>	<b>35</b>
27.1	Bernoulli Trials . . . . .	35
27.2	Binomial Distribution . . . . .	35

27.3	Properties . . . . .	35
27.4	Tools . . . . .	35
<b>28</b>	<b>Lecture 11: Normal Distribution</b>	<b>36</b>
28.1	Characteristics . . . . .	36
28.2	Equation . . . . .	36
28.3	Z-Score . . . . .	36
28.4	Empirical Rule . . . . .	36
28.5	Applications . . . . .	37
<b>29</b>	<b>Summary</b>	<b>38</b>
29.1	Topics Recap . . . . .	38
<b>30</b>	<b>References</b>	<b>39</b>
<b>31</b>	<b>summary</b>	<b>40</b>

# 1 Introduction

## 2 Introduction

**DR.Harsh Pradhan**, Phone: +91-9930034241 , Email: [harsh.231284@gmail.com](mailto:harsh.231284@gmail.com), Institute of Management Studies, Banaras Hindu University, Address: 18-GF, Jaipuria Enclave, Kaushambhi, Ghaziabad, India, 201010

**Interest:** [Goal Orientation](#) [Job Performance](#) [Consumer Behavior](#) [Behavioral Finance](#) [Bibilio-metric Analysis](#) [Options as Derivatives](#) [Statistics](#) [Indian Knowledge System](#),

[Orcid ID](#)

[Google Scholar](#)

[GitHub](#)

[Researcher ID](#)

[Personal Website](#)

[Youtube ID](#)

**Doing a PhD with me:** [README.1st](#)

[Academic Profile](#)

---

## 3 Bayesian data analysis for cognitive science

### 3.1 Introduction: What this course is about

This course provides an introduction to Bayesian data analysis using the probabilistic programming language **Stan**.

We will use a front end software package called **brms**.

This course is for:

- Linguistics (MM5, MM6)
- Cognitive Systems
- Cognitive Science

Please see the [PULS FAQs](#) to find out how the sign-up system works (in German).

We will be using the software [R](#) and [RStudio](#), so make sure you install these on your computer.

Topics to be covered:

1. Basic probability theory, random variable theory (including jointly distributed RVs), probability distributions (including bivariate distributions)
2. Using Bayes' rule for statistical inference
3. An introduction to (generalized) linear models
4. An introduction to hierarchical models
5. Measurement error models
6. Mixture models
7. Model selection and hypothesis testing (Bayes factor and k-fold cross-validation)

### 3.2 Teaching

Science and statistics is/are one unitary thing; you cannot do one without the other. Towards this end, I teach some (in my opinion) critically important classes that provide a solid statistical foundation for doing research in cognitive science. Free online course, four weeks

(MOOC), enrollments open: Introduction to Bayesian Data Analysis. Short (four-hour) tutorial on Bayesian statistics, taught at EMLAR 2022: here Introduction to (frequentist) statistics Introduction to Bayesian data analysis for cognitive science BDA cover

### **3.3 Lecture notes**

Download from [here](#).

### **3.4 Moodle website**

All communications with students in Potsdam will be done through [this website](#).



## 4 Schedule

Week	Lecture	Main Topic	Sub Topic	Video	PDF Resource
Jan 30 + Feb 4	-	Model Selection & Hypothesis Testing	-	-	HW 13
Week 2	1	Descriptive Statistics	Central Tendency	<a href="#">Link</a>	<a href="#">Week 2.pdf</a>
	2	Descriptive Statistics	Measure of Variability	<a href="#">Link</a>	<a href="#">Week 2.pdf</a>
	3	Descriptive Statistics	Describing Data	<a href="#">Link</a>	<a href="#">Week 2.pdf</a>
	4	Probability	-	<a href="#">Link</a>	<a href="#">Week 2.pdf</a>
	5	Distribution	-	<a href="#">Link</a>	<a href="#">Week 2.pdf</a>
Week 3	1	Probability	Z Table (Normal Distribution)	<a href="#">Link</a>	<a href="#">Week 3.pdf</a>
	2	Divergence	Measuring Divergence	<a href="#">Link</a>	<a href="#">Week 3.pdf</a>
	3	Inferential Statistics	Sample and Population	<a href="#">Link</a>	<a href="#">Week 3.pdf</a>
	4	Model Fit	-	<a href="#">Link</a>	<a href="#">Week 3.pdf</a>
	5	Hypothesis Testing	Hypothesis and Error	<a href="#">Link</a>	<a href="#">Week 3.pdf</a>
Week 4	1	Statistical Terms	Terms of Statistics	<a href="#">Link</a>	<a href="#">Week 4.pdf</a>
	2	Hypothesis Testing	T-Test	<a href="#">Link</a>	<a href="#">Week 4.pdf</a>
	3	Hypothesis Testing	T-Test in Detail	<a href="#">Link</a>	<a href="#">Week 4.pdf</a>
	4	ANOVA	ANOVA	<a href="#">Link</a>	<a href="#">Week 4.pdf</a>
Week 5	1	ANOVA	Example of ANOVA	<a href="#">Link</a>	<a href="#">Week 5.pdf</a>
	2	ANOVA	Types of ANOVA	<a href="#">Link</a>	<a href="#">Week 5.pdf</a>

Week	Lecture	Main Topic	Sub Topic	Video	PDF Resource
Week 6	3	Correlation	Introduction to Correlation	Link	<a href="#">Week 5.pdf</a>
	4	Regression	Regression	Link	<a href="#">Week 5.pdf</a>
	5	Regression	Regression	Link	<a href="#">Week 5.pdf</a>
	1	Regression	R Script for Regression	Link	<a href="#">Week 6.pdf</a>
	2	Chi-Square	Chi Square	Link	<a href="#">Week 6.pdf</a>
	3	Chi-Square	Chi Square Test	Link	<a href="#">Week 6.pdf</a>
Week 7	4	Logistic Regression	Logistic Function	Link	<a href="#">Week 6.pdf</a>
	5	Distribution	-	Link	<a href="#">Week 6.pdf</a>
	1	Time Series	Intro to Time Series	Link	<a href="#">Week 7.pdf</a>
	2	Probability	Conditional Probability	Link	<a href="#">Week 7.pdf</a>
	3	Additional Concepts	-	Link	<a href="#">Week 7.pdf</a>
	4	Distribution	-	Link	<a href="#">Week 7.pdf</a>
Week 8	5	Poisson Distribution	-	Link	<a href="#">Week 7.pdf</a>
	1	Libraries & Documentation	Effect Size and Packages	Link	<a href="#">Week 8.pdf</a>
	2	Software Comparison	RStudio vs RKward	Link	<a href="#">Week 8.pdf</a>
	3	Visualization	Flexplot	Link	<a href="#">Week 8.pdf</a>
	4	Programming in R	Functions	Link	<a href="#">Week 8.pdf</a>
	5	R Tools	R Shiny and R Markdown	Link	<a href="#">Week 8.pdf</a>

## **5 Introduction to Basic Statistics Using GUI-R (RKWard)**

## 6 Chapter 1: Welcome and Course Overview

This course offers an introduction to statistics through the RKWard graphical interface of R. Aimed at learners from diverse backgrounds, the course emphasizes practical application over theory. You don't need a strong background in math or computing—just an eagerness to learn.

### Pre-Requisites:

- Curiosity
- Basic awareness of numbers
- No fear of statistics or software

“Aapko darne ki zarurat nahi hai... simple understanding aapko statistics ki data ki aage milegi.”

## 7 Chapter 2: Agenda and Orientation

### Key Themes:

- Difference between Mathematics and Statistics
- Nature, Meaning, and Role of Statistics
- Uses, Limitations, and Common Fallacies

Aspect	Mathematics	Statistics
Nature	Abstract, theoretical	Applied, data-centric
Focus	Concepts, theorems, proofs	Tools, interpretation, decision-making
Tools	Logical reasoning, algebra	Hypothesis testing, regression, probability
Application	General structures	Real-world problems

## 8 Chapter 3: Meaning and Nature of Statistics

### **Definition:**

Statistics is the science of collecting, analyzing, interpreting, and presenting data for decision-making.

### **Core Concepts:**

- Population & Sample
- Parameter & Statistic
- Data classification and tabulation

### **Purpose:**

- Describe and explain phenomena
- Interpret and predict outcomes
- Facilitate scientific and social inquiry

## 9 Chapter 4: Applications and Uses

### Main Uses:

- Summarizing observed data
- Drawing representative samples
- Analyzing relationships and trends
- Supporting decision-making in fields like marketing, psychology, education, and public health

### Important Concepts:

- Data summarization
- Prediction based on patterns
- Comparison across groups
- Scientific objectivity

# 10 Chapter 5: Limitations and Misuse

## **Limitations:**

- Cannot analyze qualitative phenomena
- Not designed for individuals
- Results aren't exact
- Misinterpretation leads to incorrect conclusions

## **Misuse Includes:**

- Small or biased samples
- Misleading graphs
- Invalid comparisons

“Statistics is not a substitute for common sense or understanding the context.”

## **Fallacies Stem From:**

- Poor data collection
- Mislabeling variables
- Improper classification or selection



# 11 Chapter 6: Paper-Based vs. Software-Based Statistics

Traditional exams test pen-paper knowledge, but software-based tools like RKWard make analysis:

- Faster
- Collaborative
- Easier to store and access
- Essential for modern data-centric fields like AI and machine learning

Understanding both paper and digital approaches ensures comprehensive learning.

# 12 Chapter 7: Introduction to Variables and Spreadsheets

## Variables:

- Store information (e.g.,  $x = 5$ )
- Have unique names
- Can be manipulated with commands (e.g.,  $x = x + 2$ )

## Spreadsheets:

- Represent tabular data (rows = observations, columns = variables)
- Familiar formats: Excel, Google Sheets
- Essential in statistical packages

# 13 Chapter 8: R and GUI Interfaces

## Why R?:

- Free and open-source
- Strong community support
- High flexibility
- Powerful graphics and data manipulation capabilities

## GUI Tools in R:

- RKWard (*used in this course*)
- R Commander
- Rattle
- R AnalyticFlow

## Basic Terms:

- **Console:** Type commands & view outputs
- **Working Directory:** File storage location
- **Package:** Predefined or custom functions
- **Script:** Collection of reusable commands
- **Workspace:** All current variables/functions

# 14 Chapter 9: Importing Data and Understanding Data Types

## Using RKWard:

- Import CSV files using GUI
- Data appears in alphabetical order in workspace
- Each header = variable name

## Data Structures:

- Data Frames (most commonly used)
- Matrices
- Vectors
- Lists

## Command Line vs GUI:

- Both achieve the same results
- GUI is user-friendly, command line is customizable

```
mean(my_csv.data$JP_01) # Calculates the mean of variable JP_01
```

# 15 Chapter 10: Statistical Data Types

Statistical Type	Description	R Equivalent
Nominal	Names, labels (e.g., Male/Female)	String
Ordinal	Order/rank (e.g., 1st, 2nd)	Factor
Interval	Ordered + meaningful intervals (e.g., tax slabs)	Numeric
Ratio	Includes absolute zero (e.g., weight)	Numeric

## Others in R:

- Logical (TRUE/FALSE)
- Integer, Complex

Remember: Not all numbers mean quantity. Shirt numbers (like #18) are nominal, not mathematical.

# 16 Chapter 11: Data Preparation in RKWard

- Data must be properly **typed** (e.g., “1” as number vs “1” as label)
- Check alignment: Left = character, Right = number
- **Labels** help collaborators understand variables
- Example: **Gender** = 1 (Male), 0 (Female)
- Must distinguish between numeric calculations and categorical identifiers

## **Best Practices:**

- Define each variable with meaning
- Validate data types
- Store and share workspace for reproducibility

# 17 Chapter 12: Visualizing Data with Plots in RKWard

Data visualization is essential to reveal patterns, trends, and distributions. RKWard offers multiple graphical tools:

## 17.1 1. Histogram

- Depicts the distribution of a single variable
- Can include frequency, relative frequency, and cumulative frequency
- Best for understanding where most data points lie

## 17.2 2. Pie Chart

- Represents categorical data as slices of a circle
- Best when visualizing proportions

## 17.3 3. Scatter Plot

- Plots two variables to examine relationships
- X-axis: Independent variable
- Y-axis: Dependent variable
- Useful in exploring associations or potential causality

## 17.4 4. Box Plot

- Shows data distribution via quartiles
- Median, interquartile range (IQR), and outliers are clearly indicated
- Useful for comparing multiple variables

## 17.5 5. Density Plot

- Smoothed version of a histogram
- Better suited for continuous data with decimal variation

### Key Tips:

- JP\_01 was frequently used as an example variable
- RKWard allows saving and exporting plots easily
- GUI menus guide the user through plot creation

Always choose the plot type that best matches your data and goal: frequency, relationship, or comparison.



# 18 Chapter 13: Summary

This eBook provided a foundation for understanding and applying statistics using the RKWard GUI tool in R. It covered essential concepts from what statistics is, to importing and handling data, understanding types of variables and their measurement levels, and visualizing data using a variety of plots.

Learners were introduced to:

- Basic statistical principles
- Software versus paper-based understanding
- Variable types and spreadsheet usage
- Command line and GUI-based tools
- Data visualization through histogram, pie, scatter, box, and density plots

The course emphasized **conceptual clarity**, **practical tools**, and the **power of visualization**. It prepares learners to interpret, analyze, and present data meaningfully in academic or real-world contexts.

## 19 References

1. Mohanty, B., & Misra, S. (2020). *Statistics for Behavioral and Social Sciences*. PHI Learning.
2. Pandya, D., et al. (2019). *Statistical Analysis in Simple Steps Using R*. Wiley.
3. Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. SAGE Publications.
4. Harris, J. (2021). *Statistics with R: Solving Problems Using Real-World Data*. Pearson.
5. RKWard Project: <https://rkward.kde.org>

## 20 Next Steps

Upcoming lectures will cover:

- Graph creation
- Data visualization tools
- Advanced statistical operations in GUI

# **21 Basic Statistics Using GUI-R (RKWard)**

Lectures 05 to 11: Foundations of Statistical Thinking

## 22 Lecture 05: Essential Math for Statistics

This lecture lays the foundation of mathematical concepts essential to understanding statistics. Topics include coordinate geometry, linear and exponential functions, and basics of calculus.

### 22.1 Key Concepts

#### 22.1.1 Number Line & Coordinate Geometry

- A number line represents a variable's values.
- Cartesian plane: X-axis (independent), Y-axis (dependent).
- Linear equation:  $Y = mX + C$ , where:
  - $m = \frac{Y_2 - Y_1}{X_2 - X_1}$  (slope)
  - $C$  is the Y-intercept (Y when  $X = 0$ )

#### 22.1.2 Exponential Functions

- Form:  $Y = a^X$  or  $Y = e^X$  (where  $e \approx 2.718$ )
- Models rapid, nonlinear growth.

#### 22.1.3 Differentiation

- Measures instantaneous rate of change:

$$\frac{dY}{dX} = \lim_{h \rightarrow 0} \frac{f(X+h) - f(X)}{h}$$

#### 22.1.4 Integration

- Calculates area under the curve:

$$\bar{X} = \int X \cdot f(X) dX$$

---

## 23 Lecture 06: Measures of Central Tendency

Central tendency summarizes a dataset with a single representative value.

### 23.1 Definitions

- **Mean:** Arithmetic average; sensitive to outliers.
- **Median:** Middle value; robust to outliers.
- **Mode:** Most frequent value; best for categorical data.

### 23.2 Characteristics

- Simple, uses all values (mean).
- Median is stable against extreme values.
- Mode is useful for nominal data.

### 23.3 Suitability by Data Type

Measure	Data Type
Mean	Interval/Ratio
Median	Ordinal
Mode	Nominal

### 23.4 Skewness

- Positive skew:  $\text{Mean} > \text{Median}$
- Negative skew:  $\text{Mean} < \text{Median}$

---

## 24 Lecture 07: Measures of Variability

Describes the spread of data.

### 24.1 Measures

- **Range:** Max - Min
- **Interquartile Range (IQR):**  $Q_3 - Q_1$
- **Variance:** Avg. of squared deviations
- **Standard Deviation (SD):**  $\sqrt{\text{Variance}}$
- **Coefficient of Variation (CV):**  $\frac{SD}{Mean}$

### 24.2 Theoretical Moments

Moment	Interpretation
1st	Mean
2nd	Variance
3rd	Skewness
4th	Kurtosis

---



## 25 Lecture 08: Mean-Centered & Standardized Data

Transformation techniques for advanced analysis.

### 25.1 Mean-Centering

- Subtract mean from each data point.
- New mean becomes zero.

### 25.2 Standardization (Z-Score)

- Formula:  $Z = \frac{X - \mu}{\sigma}$
- Mean = 0, SD = 1
- Unitless; allows comparison across distributions.

### 25.3 Applications

- Normalize different scales
  - Preprocessing in statistics and ML
-

# 26 Lecture 09: Introduction to Probability

Foundation of inferential statistics.

## 26.1 Key Concepts

- **Experiment:** Controlled process
- **Event:** Subset of outcomes
- **Sample Space (S):** All possible outcomes

## 26.2 Basic Rule

- $P(A) = \frac{\text{Favorable outcomes}}{\text{Total outcomes}}$

## 26.3 Properties

- $0 \leq P(A) \leq 1$
- $P(S) = 1$
- $P(A^c) = 1 - P(A)$
- If  $A \subseteq B$ , then  $P(A) \leq P(B)$

## 26.4 Visual Aids

- Venn diagrams
  - Frequency Probability distribution linkage
-

# 27 Lecture 10: Distributions & Binomial Distribution

Probability distributions describe variable behavior.

## 27.1 Bernoulli Trials

- Success ( $p$ ) or failure ( $1 - p$ )

## 27.2 Binomial Distribution

- $n$  independent trials
- 

$$P(X = r) = \binom{n}{r} p^r (1 - p)^{n-r}$$

## 27.3 Properties

- Mean:  $np$
- Variance:  $np(1 - p)$

## 27.4 Tools

- Galton board
  - Pascal's triangle
-

## 28 Lecture 11: Normal Distribution

Crucial distribution in statistics.

### 28.1 Characteristics

- Bell-shaped, symmetric about  $\mu$
- Area under curve = 1
- Emerges with large samples

### 28.2 Equation

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

### 28.3 Z-Score

- $Z = \frac{X-\mu}{\sigma}$

### 28.4 Empirical Rule

Range	Coverage
$\pm 1$ SD	68%
$\pm 2$ SD	95%
$\pm 3$ SD	99.7%

## 28.5 Applications

- Heights, test scores, measurement errors
-

## 29 Summary

This section builds statistical fundamentals.

### 29.1 Topics Recap

- **Math:** Derivatives, integrals, exponential functions
  - **Descriptive Stats:** Central tendency, variability
  - **Transformation:** Z-scores, mean-centering
  - **Probability:** Events, rules, visual tools
  - **Distributions:** From discrete (binomial) to continuous (normal)
-

## 30 References

1. Galton, F. (1889). *Natural Inheritance*.
2. Montgomery, D.C., & Runger, G.C. (2014). *Applied Statistics and Probability for Engineers*.
3. Casella, G., & Berger, R.L. (2002). *Statistical Inference*.
4. Lecture transcripts by Dr. Harsh Pradhan, IMS-BHU (Week 2, Lectures 05–11)
5. Course materials and Rkward demonstrations

## 31 summary

This is summary file . 10 june .