

# Mastering AP Statistics with Programming Applications

Mathematics Elevate Series

**Rishabh Kumar**

Elite International Math Mentor

Founder, Mathematics Elevate Academy

Educator, *Math By Rishabh*

Alumnus, IIT Guwahati & the Indian Statistical Institute



# Contents

|           |  |           |
|-----------|--|-----------|
| <b>I</b>  | <b>Unit I. Exploring One-Variable Data</b>                         | <b>1</b>  |
| <b>1</b>  | <b>The Foundations of Data</b>                                     | <b>3</b>  |
| 1.1       | Introducing Statistics: What Can We Learn from Data? . . . . .     | 3         |
| <b>2</b>  | <b>Visualizing Categorical Data</b>                                | <b>7</b>  |
| 2.1       | Organizing with Tables: Frequency and Relative Frequency . . . . . | 7         |
| 2.2       | Graphing with Bar Charts and Pie Charts . . . . .                  | 8         |
| 2.3       | The Language of Variation . . . . .                                | 12        |
| <b>3</b>  | <b>Visualizing Quantitative Data: The Shape of a Story</b>         | <b>17</b> |
| 3.1       | Small Datasets: Dotplots and Stemplots . . . . .                   | 17        |
| 3.2       | Large Datasets: The Power of the Histogram . . . . .               | 18        |
| 3.3       | Describing the Story: Shape, Center, Spread, and Unusual Features  | 19        |
| <b>4</b>  | <b>Numerically Summarizing Quantitative Data</b>                   | <b>23</b> |
| 4.1       | Measures of Center: Mean vs. Median . . . . .                      | 23        |
| 4.2       | Measures of Spread: Standard Deviation and IQR . . . . .           | 24        |
| 4.3       | Boxplots and Identifying Outliers . . . . .                        | 25        |
| 4.4       | Measures of Position: The Z-score . . . . .                        | 26        |
| 4.5       | Putting It All Together: Comparing Distributions . . . . .         | 29        |
| <b>5</b>  | <b>The Normal Distribution: The Cornerstone of Statistics</b>      | <b>33</b> |
| 5.1       | The Normal Model and the Empirical Rule . . . . .                  | 33        |
| 5.2       | Standardizing and Comparing with Z-scores . . . . .                | 35        |
| 5.3       | Finding Proportions for Any Interval . . . . .                     | 36        |
| <b>II</b> | <b>Unit II. Exploring Two-Variable Data</b>                        | <b>41</b> |
| <b>6</b>  | <b>Relationships in Categorical Data: Association</b>              | <b>43</b> |

|            |  |           |
|------------|--|-----------|
| 6.1        | Visualizing Relationships with Two-Way Tables . . . . .                            | 43        |
| 6.2        | Measuring Association with Conditional Distributions . . . . .                     | 44        |
| <b>7</b>   | <b>Describing Relationships in Quantitative Data: Scatterplots and Correlation</b> | <b>49</b> |
| 7.1        | Visualizing Relationships: The Scatterplot . . . . .                               | 49        |
| 7.2        | Measuring Linear Strength: Correlation . . . . .                                   | 51        |
| <b>8</b>   | <b>Modeling with Lines: Linear Regression</b>                                      | <b>55</b> |
| 8.1        | The Least-Squares Regression Line (LSRL) . . . . .                                 | 55        |
| 8.2        | Assessing the Model: Residuals and R-squared . . . . .                             | 57        |
| 8.3        | Cautions and Departures from Linearity . . . . .                                   | 58        |
| <b>III</b> | <b>Unit III. Collecting Data</b>   | <b>61</b> |
| <b>9</b>   | <b>Sampling and Surveys: The Art of Asking Questions</b>                           | <b>63</b> |
| 9.1        | Planning a Study: Populations, Samples, and Bias . . . . .                         | 63        |
| 9.2        | Random Sampling Methods . . . . .  | 65        |
| 9.3        | Potential Problems and Sources of Bias . . . . .                                   | 66        |
| <b>10</b>  | <b>Designing Experiments: Establishing Cause and Effect</b>                        | <b>69</b> |
| 10.1       | The Principles of Experimental Design . . . . .                                    | 69        |
| 10.2       | Common Experimental Designs . . . . .  | 71        |
| 10.3       | Drawing Conclusions from Experiments . . . . .                                     | 72        |
| <b>IV</b>  | <b>Unit IV. Probability, Random Variables, and Probability Distributions</b>       | <b>75</b> |
| <b>11</b>  | <b>Probability: The Study of Chance</b>  | <b>77</b> |
| 11.1       | Introduction to Probability and Simulation . . . . .                               | 77        |
| 11.2       | Basic Probability Rules . . . . .  | 78        |
| 11.3       | Conditional Probability and Independence . . . . .                                 | 79        |
| 11.4       | More Probability Rules . . . . .   | 80        |
| <b>12</b>  | <b>Random Variables: Modeling Numerical Outcomes</b>                               | <b>83</b> |
| 12.1       | Introduction to Random Variables . . . . .   | 83        |
| 12.2       | Mean and Standard Deviation of a Random Variable . . . . .                         | 84        |
| 12.3       | Transforming and Combining Random Variables . . . . .                              | 85        |

---

|   |            |
|---|------------|
| <b>13 Common Discrete Random Variables: Binomial and Geometric</b>      | <b>89</b>  |
| 13.1 The Binomial Distribution . . . . .                                | 89         |
| 13.2 The Geometric Distribution . . . . .                               | 91         |
| <br>  |            |
| <b>V Unit V. Sampling Distributions</b>                                 | <b>95</b>  |
| <br>  |            |
| <b>14 The Big Idea: Sampling Distributions</b>                          | <b>97</b>  |
| 14.1 Why Samples Vary: Sampling Variability . . . . .                   | 97         |
| 14.2 The Central Limit Theorem: The Magic of Inference . . . . .        | 98         |
| 14.3 Good and Bad Estimators . . . . .                                  | 99         |
| <br>  |            |
| <b>15 Modeling the Distribution of Sample Proportions</b>               | <b>103</b> |
| 15.1 The Sampling Distribution of a Sample Proportion . . . . .         | 103        |
| 15.2 The Sampling Distribution of a Difference in Proportions . . . . . | 105        |
| <br>  |            |
| <b>16 Modeling the Distribution of Sample Means</b>                     | <b>109</b> |
| 16.1 The Sampling Distribution of a Sample Mean . . . . .               | 109        |
| 16.2 The Sampling Distribution of a Difference in Means . . . . .       | 112        |
| <br>  |            |
| <b>VI Unit VI. Inference for Categorical Data: Proportions</b>          | <b>115</b> |
| <br>  |            |
| <b>17 Confidence Intervals for a Population Proportion</b>              | <b>117</b> |
| 17.1 The Logic of Confidence Intervals . . . . .                        | 117        |
| 17.2 Constructing and Interpreting a Confidence Interval . . . . .      | 118        |
| <br>  |            |
| <b>18 Significance Tests for a Population Proportion</b>                | <b>121</b> |
| 18.1 The Logic of Significance Tests: Hypotheses and P-Values . . . . . | 121        |
| 18.2 Performing a One-Sample Z-Test for a Proportion . . . . .          | 122        |
| 18.3 Errors in Hypothesis Testing . . . . .                             | 124        |
| <br>  |            |
| <b>19 Comparing Two Proportions</b>                                     | <b>127</b> |
| 19.1 Confidence Intervals for a Difference in Proportions . . . . .     | 127        |
| 19.2 Significance Tests for a Difference in Proportions . . . . .       | 128        |
| <br>  |            |
| <b>VII Unit VII. Inference for Quantitative Data: Means</b>             | <b>131</b> |
| <br>  |            |
| <b>20 Inference for a Single Population Mean</b>                        | <b>133</b> |
| 20.1 The T-Distribution and Confidence Intervals for a Mean . . . . .   | 133        |

---

|   |            |
|---|------------|
| 20.2 Justifying Claims and Performing a T-Test for a Mean . . . . .                   | 135        |
| <b>21 Inference for the Difference Between Two Independent Means</b>                  | <b>139</b> |
| 21.1 Confidence Intervals for a Difference in Means . . . . .                         | 139        |
| 21.2 Significance Tests for a Difference in Means . . . . .                           | 141        |
| <b>22 Inference for Paired Data: Matched Pairs T-Procedures</b>                       | <b>145</b> |
| 22.1 Recognizing and Analyzing Paired Data . . . . .                                  | 145        |
| 22.2 Performing Paired T-Intervals and T-Tests . . . . .                              | 146        |
| <b>VIII Unit VIII. Inference for Categorical Data: Chi-Square</b>                     | <b>151</b> |
| <b>23 The Chi-Square Goodness of Fit Test: Does It Match the Claim?</b>               | <b>153</b> |
| 23.1 The Chi-Square Distribution and the Goodness of Fit Test . . . . .               | 153        |
| 23.2 Performing a Goodness of Fit Test . . . . .                                      | 155        |
| <b>24 Chi-Square Tests for Two-Way Tables: Homogeneity and Independence</b>           | <b>159</b> |
| 24.1 The Test for Homogeneity . . . . .   | 159        |
| 24.2 The Test for Independence . . . . .  | 160        |
| 24.3 Skills Focus: Selecting the Correct Test for Categorical Data . . . . .          | 162        |
| <b>IX Unit IX. Inference for Quantitative Data: Slopes</b>                            | <b>165</b> |
| <b>25 Inference for the Slope of a Regression Line: Is There a Real Relationship?</b> | <b>167</b> |
| 25.1 The Sampling Distribution of the Slope . . . . .                                 | 167        |
| 25.2 Confidence Intervals for the Slope . . . . .                                     | 168        |
| 25.3 Significance Tests for the Slope . . . . .                                       | 169        |
| 25.4 Grand Finale: Selecting the Correct Inference Procedure . . . . .                | 171        |
| <b>X Mock Papers</b>  | <b>175</b> |
| <b>26 Mock Paper 1</b>  | <b>177</b> |
| <b>27 Mock Paper 2</b>  | <b>181</b> |
| <b>28 Mock Paper 3</b>  | <b>185</b> |
| <b>29 Mock Paper 4</b>  | <b>189</b> |

CONTENTS

---

vii

|                                  |            |
|----------------------------------|------------|
| <b>30 Mock Paper 5</b>           | <b>193</b> |
| <b>Common Formulas</b>           | <b>197</b> |
| <b>Data Sets</b>                 | <b>201</b> |
| <b>Standard Normal (Z) Table</b> | <b>203</b> |
| <b>Author's Note</b>             | <b>205</b> |



# Preface

Welcome to the **Mathematics Elevate Series** for AP Statistics, inspired by the AP<sup>®</sup> Statistics syllabus. This book is the culmination of years of teaching, research, and mentoring students across the globe. It is carefully crafted to provide a strong foundation in statistics, equipping students not only to excel in the AP<sup>®</sup> Statistics examination, but also to develop analytical thinking skills that will serve them in higher education and beyond.

The guiding philosophy of this series is simple yet powerful: mastery comes from an *intuitive understanding of core concepts*, reinforced through structured practice, real-world applications, and technological fluency. Each chapter is designed to be clear, engaging, and comprehensive, walking students through the four major themes of AP<sup>®</sup> Statistics: *Exploring Data, Sampling and Experimentation, Anticipating Patterns, and Statistical Inference*.

My goal is not only to prepare you for examination success, but to nurture your curiosity and confidence in statistics—a discipline that empowers you to make sense of the world through data.

I invite you to embark on this journey with dedication and curiosity. With consistent effort and the right guidance, success in AP<sup>®</sup> Statistics - and in future academic endeavors - is within your reach.

**Rishabh Kumar**

Elite International Math Mentor (Online Educator)

Founder, *Mathematics Elevate Academy*

*Math By Rishabh*



# Unit I

## UNIT I. EXPLORING ONE-VARIABLE DATA



# Chapter 1

## The Foundations of Data

Welcome to the start of your journey. Before we can run, we must learn to walk. In statistics, this means starting with the most fundamental question of all: What is data? This chapter will introduce you to the core vocabulary and the foundational mindset of a statistician. We won't just learn definitions; we'll discover *why* these ideas are the building blocks for every single analysis we will ever do, from a simple survey to a complex machine learning model.

### 1.1 Introducing Statistics: What Can We Learn from Data?

Let's start with a question you've probably wondered about. You finish a song on Spotify or a show on Netflix, and instantly, a new recommendation pops up. And it's... surprisingly good. How did it know? Is it magic?

It's not magic. It's **data**.

#### Industry Spotlight

The process Spotify and Netflix use is a classic application of Machine Learning called a **Recommender System**. One popular technique, "collaborative filtering," finds users with similar tastes to you and then recommends content that they liked but you haven't seen yet. The entire system is built on collecting and analyzing the categorical data of user "likes" and "skips."

Every time you "like" a song, skip a track, or re-watch a movie, you are creating a tiny piece of information. On its own, your single "like" doesn't mean much. But when combined with the "likes" and "skips" of millions of other users, a powerful

story emerges. Spotify and Netflix are constantly collecting this information, analyzing it, and using it to build a model of your taste. They are practicing statistics on a massive scale.

In the world around us, this process is happening everywhere, all the time. From the biologist tracking a tagged sea turtle to a political campaign polling voters, from a coffee shop owner counting daily customers to an engineer testing the strength of a new material. They are all using information to discover insights and make better decisions.

This information—these facts, figures, and measurements we gather—is what we call **data**. And **statistics** is the art and science of learning from it. It's the superpower that allows us to find the story hidden within the noise.

At its core, every statistical investigation involves a few key building blocks.

*Vocab*

Let's start with the two most fundamental: **individuals** and **variables**.

*Alert!*

**Individual:**

An object (person, animal, or thing) described by a set of data.

Imagine we're statisticians hired by your school to understand the study habits of its students. Our first step is to define what, or who, we are interested in. In this case, we are interested in the **students**. The students are the *objects* described by our set of data. We call these the **individuals** of the study. Individuals don't have to be people; they could be the apps on your phone, the cars in a parking lot, or the sea turtles in our biologist's study.

Once we know our individuals, we need to decide what characteristics we want to measure about them. What do we want to know about each student? Perhaps we want to know their grade level, their favorite subject, the number of hours they study per week, and their most recent test score. These characteristics are the **variables**. A variable is any characteristic of an individual, and its value can change from one individual to another.

This brings us to a critical distinction, one that will guide almost every choice we make for the rest of this course. Variables come in two main flavors:

**Definition: Types of Variables**

A **categorical variable** (sometimes called a qualitative variable) places an individual into one of several groups or categories.

- *Examples:* Favorite subject (Math, History, English), Grade level (9th, 10th, 11th), Type of car (Sedan, SUV, Truck).

A **quantitative variable** takes numerical values for which it makes sense to perform arithmetic operations like adding or averaging.

- *Examples:* Hours spent studying per week (e.g., 10.5 hours), Test score (e.g., 88), Weight of a car in kilograms.

Think about it: you can't find the "average" favorite subject. But you absolutely can find the average number of hours a student studies. This simple test—"Does it make sense to calculate an average?"—is your best friend for telling these two types of variables apart.

**★ AP® Exam Connection**

As a mentor who has guided students to top institutions like IIT and ISI, I can tell you that this single concept—distinguishing between categorical and quantitative variables—is the most important foundation you will build. It may seem simple, but every graph we create, every summary we calculate, and every conclusion we draw depends on it.

Think of it like being a chef. You must first know if your ingredient is a solid or a liquid before you can decide whether to chop it or pour it. In statistics, you must first know what type of variable you have before you can choose the right tool to analyze it. Master this, and you've taken the first crucial step toward thinking like a true statistician. On the AP® Exam, this is not just a vocabulary question; it's the starting point for nearly every problem.



## Chapter 2

# Visualizing Categorical Data

In the last chapter, we met the two fundamental types of variables. We learned that to think like a statistician, our first job is to identify our ingredients. Now, it's time to become a chef.

Imagine you're the manager of a local coffee shop, and for one hour, you jot down every drink your customers order. At the end of the hour, your notepad looks like a chaotic mess: "Latte, Drip Coffee, Espresso, Latte, Tea, Latte, Drip Coffee..." What can you do with this list? Just looking at it tells you very little.

In this chapter, we'll learn the first and most powerful step in any data analysis: turning a raw list of categorical data into organized tables and insightful graphs. We will learn how to take chaos and create clarity.

### 2.1 Organizing with Tables: Frequency and Relative Frequency

Our first task with the coffee shop data is simple: we need to count. We systematically go through our list, putting a tally mark next to each drink category. This process of counting how often each category occurs creates what we call a **frequency table**.

Let's say after our hour of observation, we have the following data for 50 customers:

### Definition: Frequency and Relative Frequency Tables

A **frequency table** shows the count (or frequency) of individuals in each category.

A **relative frequency table** shows the proportion or percent of individuals in each category. We calculate this by dividing the category's count by the total number of observations.

Table 2.1: Frequency and Relative Frequency of Drinks Sold in One Hour

| Drink Type   | Frequency (Count) | Relative Frequency (%)    |
|--------------|-------------------|---------------------------|
| Drip Coffee  | 15                | $(15 / 50) = 0.30$ or 30% |
| Latte        | 20                | $(20 / 50) = 0.40$ or 40% |
| Espresso     | 5                 | $(5 / 50) = 0.10$ or 10%  |
| Tea          | 10                | $(10 / 50) = 0.20$ or 20% |
| <b>Total</b> | <b>50</b>         | <b>1.00 or 100%</b>       |

Instantly, the story becomes clearer. The Latte is the most popular drink, and Espresso is the least. But why do we need the "relative frequency" column? Imagine another coffee shop across town sold 200 drinks in the same hour. We couldn't compare their raw counts to ours directly. But by comparing percentages (relative frequencies), we can make a fair, apples-to-apples comparison of which drinks are more popular at each location, regardless of the total number of customers.

### ☆ AP® Exam Connection

The concept of relative frequency is absolutely critical on the AP® Exam. You will often be asked to compare a categorical variable across two or more different groups (e.g., comparing social media preferences between 9th graders and 12th graders). Because the group sizes might be different, you **must** use relative frequencies (percents or proportions) to make your comparison. Simply comparing counts will not earn full credit.

## 2.2 Graphing with Bar Charts and Pie Charts

Tables are fantastic for organizing data, but for revealing insights at a single glance, nothing beats a graph. For categorical data, our two main tools are the bar chart

and the pie chart.

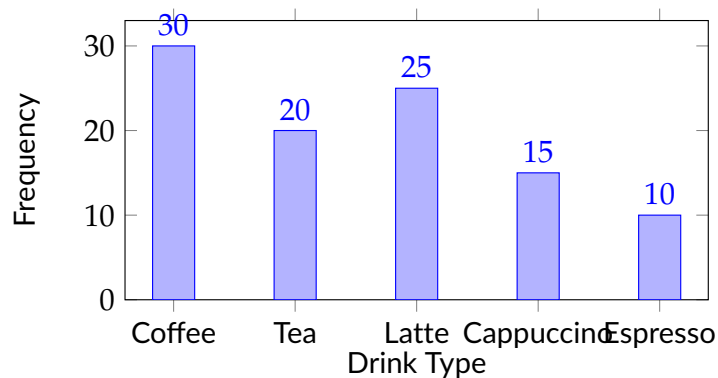


Figure 2.1: A Bar Chart of Coffee Shop Drink Frequencies

### Industry Spotlight

One of the most common activities at tech companies like Google, Amazon, and Meta is called **A/B Testing**. Imagine an e-commerce site wants to see if changing their "Buy Now" button from green to blue will make more people click it.

They show the green button (Version A) to 10,000 users and the blue button (Version B) to another 10,000 users. They record the number of "clicks" for each version. The final result is a simple bar chart comparing the click counts for Version A and Version B. This incredibly simple statistical tool—a basic bar chart—is used to make multi-million dollar business decisions every single day.

A **pie chart** is used when you want to emphasize how each category relates to the whole. The entire circle represents 100

### Triple-Tech Lab

Let's bring our coffee shop data to life using Python and R. These tools make creating professional, publication-quality graphics from raw data incredibly simple.

#### Python with Seaborn:

```
# Import the necessary libraries
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```

# Here's our raw data as a Python list
drinks = ['Latte', 'Drip Coffee', 'Espresso', 'Latte', 'Tea', 'Latte',
↪ 'Drip Coffee',
        'Latte', 'Latte', 'Drip Coffee', 'Tea', 'Latte', 'Drip
↪ Coffee', 'Latte',
        'Espresso', 'Latte', 'Tea', 'Latte', 'Drip Coffee', 'Drip
↪ Coffee', 'Latte',
        'Drip Coffee', 'Latte', 'Tea', 'Tea', 'Latte', 'Drip Coffee',
↪ 'Espresso',
        'Latte', 'Drip Coffee', 'Tea', 'Latte', 'Drip Coffee',
↪ 'Latte', 'Tea',
        'Latte', 'Drip Coffee', 'Tea', 'Latte', 'Drip Coffee',
↪ 'Espresso', 'Latte',
        'Tea', 'Drip Coffee', 'Drip Coffee', 'Tea', 'Espresso',
↪ 'Latte',
        'Drip Coffee', 'Latte']

# Create a pandas DataFrame
df = pd.DataFrame({'drink_type': drinks})

# Create the bar plot. Seaborn's countplot automatically counts the
↪ frequencies!
sns.countplot(x='drink_type', data=df,
↪ order=df['drink_type'].value_counts().index)

# Add a title and display the plot
plt.title('Coffee Shop Sales in One Hour')
plt.show()

```

### R with ggplot2:

```

# Load the ggplot2 library, the gold standard for R graphics
library(ggplot2)

# Create a vector with our raw data
drinks <- c('Latte', 'Drip Coffee', 'Espresso', 'Latte', 'Tea',
↪ 'Latte', 'Drip Coffee',
        'Latte', 'Latte', 'Drip Coffee', 'Tea', 'Latte', 'Drip
↪ Coffee', 'Latte',
        'Espresso', 'Latte', 'Tea', 'Latte', 'Drip Coffee', 'Drip
↪ Coffee', 'Latte',
        'Drip Coffee', 'Latte', 'Tea', 'Tea', 'Latte', 'Drip
↪ Coffee', 'Espresso',

```

## Chapter 30

# Mock Paper 5

### **Advanced Statistics Mock Exam – Free Response Questions (Mock Paper 5)**

#### **SECTION II**

Total Time—2 hours

6 Questions

*Directions:* Show all your work. Indicate clearly the methods you use, because you will be scored on both the correctness of your methods and the accuracy of your results and explanations.

#### Part A

##### Question 1 – Advanced Regression Analysis

A chain of coffee shops recorded the number of cups sold per day for 40 days. Two factors were recorded for each day: temperature (in °C) and daily sales (number of cups sold). Summary statistics are:

- Mean temperature = 22°C, SD = 4°C
- Mean daily sales = 150 cups, SD = 30 cups
- Correlation between temperature and sales = 0.65

- (a) Find the least-squares regression line predicting daily sales from temperature. Interpret the slope and intercept in context.
- (b) Predict daily sales for a day with 28°C.
- (c) Calculate and interpret the coefficient of determination ( $R^2$ ).
- (d) If actual sales for a 28°C day are 210 cups, calculate and interpret the residual.
- (e) Assuming sales are normally distributed, calculate the probability that on a day with 22°C, daily sales exceed 180 cups.
- (f) Discuss limitations of this model for predicting sales on extreme temperatures (e.g., 5°C or 35°C).

### Question 2 – Multi-Sample Comparison (ANOVA-style)

Three fertilizer types (A, B, C) are tested on 45 plants (15 each). The plant heights (in cm) are recorded at the end of the season.

- (a) State the null and alternative hypotheses for testing whether the mean heights differ by fertilizer type.
- (b) Compute or describe the F-statistic formula conceptually for this one-way ANOVA.
- (c) If the calculated  $p$ -value = 0.02, interpret the result in context.
- (d) Describe one assumption necessary for the validity of ANOVA and how to check it.

### Question 3 – Chi-Square Test for Independence

A survey of 200 students records their participation in sports (Yes/No) and preference for online vs. in-person classes. The data are summarized in the table:

|             | Online | In-person |
|-------------|--------|-----------|
| Sports: Yes | 50     | 30        |
| Sports: No  | 40     | 80        |

- 
- (a) State the hypotheses to test if sports participation is independent of class preference.
  - (b) Compute the expected counts for each cell.
  - (c) Calculate the chi-square test statistic (show formula) and determine if the variables are independent at  $\alpha = 0.05$ .
  - (d) Interpret your result in context.

#### Question 4 – Inference for Two Means (Advanced)

A medical researcher compares two treatments for lowering blood pressure. Sample data:

- Treatment 1:  $n_1 = 25$ ,  $\bar{x}_1 = 120$ ,  $s_1 = 8$
- Treatment 2:  $n_2 = 30$ ,  $\bar{x}_2 = 126$ ,  $s_2 = 10$

- (a) State the hypotheses to test if the mean blood pressure differs between treatments.
- (b) Calculate the standard error of the difference in means.
- (c) Construct a 95% confidence interval for the difference in means.
- (d) Based on your interval, what conclusion can you draw about the effectiveness of the treatments?

#### Question 5 – Probability and Sampling Distributions

A factory produces electronic chips with a 2% defect rate. A random sample of 50 chips is inspected.

- (a) Define the random variable  $X$  and its distribution.
- (b) Find  $P(X = 1)$  and  $P(X \leq 2)$ .
- (c) Approximate  $X$  using a normal distribution and calculate  $P(X \geq 3)$ .
- (d) Find the expected number of defective chips and standard deviation.

## Question 6 – Integrated Data Analysis (Challenging)

A hospital records recovery times (days) for patients undergoing two types of surgeries. The summary statistics are:

- Surgery A:  $n = 20$ , mean = 8, SD = 2
  - Surgery B:  $n = 22$ , mean = 10, SD = 3
- (a) Compute a 95% confidence interval for the difference in mean recovery times.
- (b) Test whether Surgery A has a significantly shorter recovery time than Surgery B at  $\alpha = 0.05$ .
- (c) Discuss whether the assumption of equal variances is reasonable and how violation might affect results.
- (d) Suggest a graphical method to visualize and compare the two distributions.

**END OF EXAM**

# Common Formulas

This appendix contains a collection of common formulas used throughout the AP<sup>®</sup> Statistics curriculum. These serve as a quick reference guide for problem-solving and exam preparation.

## Descriptive Statistics

- Mean:  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
- Median: Middle value when data are ordered
- Variance:  $s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$
- Standard Deviation:  $s = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$
- Z-score:  $z = \frac{x - \mu}{\sigma}$

## Probability

- Addition Rule:  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Multiplication Rule:  $P(A \cap B) = P(A) \cdot P(B|A)$
- Conditional Probability:  $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- Complement Rule:  $P(A^c) = 1 - P(A)$

## Random Variables & Distributions

- Expected Value:  $E[X] = \sum x_i p_i$

- Variance:  $\text{Var}(X) = \sum(x_i - \mu)^2 p_i$
- Binomial Distribution:  $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$
- Mean of Binomial:  $\mu = np$
- Standard Deviation of Binomial:  $\sigma = \sqrt{np(1 - p)}$
- Geometric Distribution:  $P(X = k) = (1 - p)^{k-1} p$

### Sampling Distributions

- Mean of Sample Means:  $\mu_{\bar{x}} = \mu$
- Standard Error (mean):  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
- Standard Error (proportion):  $SE = \sqrt{\frac{p(1 - p)}{n}}$

### Confidence Intervals

- One-sample  $z$  interval for mean:  $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$
- One-sample  $t$  interval for mean:  $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$
- One-sample proportion:  $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$
- Two-sample mean difference:  $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
- Two-sample proportion difference:  $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$

### Hypothesis Testing

- Test statistic for mean (z-test):  $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
- Test statistic for mean (t-test):  $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

- 
- Test statistic for proportion:  $z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$
  - Chi-square statistic:  $\chi^2 = \sum \frac{(O - E)^2}{E}$
  - Regression slope  $t$ -test:  $t = \frac{b - \beta_0}{SE_b}$

## Regression

- Least Squares Regression Line:  $\hat{y} = a + bx$
- Slope:  $b = r \cdot \frac{s_y}{s_x}$
- Intercept:  $a = \bar{y} - b\bar{x}$
- Coefficient of Determination:  $R^2 = 1 - \frac{SSE}{SST}$

# Data Sets

This appendix contains the data sets used in the examples and practice problems throughout the book. You can use them to perform your own analysis, replicate examples, and practice with technology tools (graphing calculators, R, Python, or statistical software).

## Dataset 2.1: Student Study Hours

5, 7, 8, 8, 9, 10, 10, 10, 11, 12, 12, 14, 15, 20

## Dataset 3.4: Heights of Students (cm)

150, 152, 155, 158, 160, 162, 163, 165, 168, 170,  
171, 172, 174, 175, 178, 180, 182, 185

## Dataset 5.2: Coin Toss Experiment (Number of Heads in 10 Tosses)

4, 5, 6, 7, 3, 4, 6, 8, 5, 7, 6, 5, 4, 6, 7, 8, 5, 6, 4, 5

## Dataset 7.1: AP<sup>®</sup> Statistics Practice Test Scores

42, 55, 61, 63, 66, 68, 70, 72, 74, 75,  
77, 78, 80, 82, 84, 85, 87, 88, 90, 92



# Standard Normal (Z) Table

The following table provides cumulative probabilities from the far left ( $Z = -\infty$ ) up to the given  $Z$ -score. For negative  $Z$ -scores, use the symmetry rule:

$$P(Z < -z) = 1 - P(Z < z)$$

Table 1: Cumulative probabilities for select positive  $Z$ -scores.

| <b>Z</b>   | <b>.00</b> | <b>.01</b> | <b>.02</b> | <b>.03</b> | <b>.04</b> | <b>...</b> |
|------------|------------|------------|------------|------------|------------|------------|
| <b>0.0</b> | .5000      | .5040      | .5080      | .5120      | .5160      | ...        |
| <b>0.1</b> | .5398      | .5438      | .5478      | .5517      | .5557      | ...        |
| <b>0.2</b> | .5793      | .5832      | .5871      | .5910      | .5948      | ...        |
| <b>0.3</b> | .6179      | .6217      | .6255      | .6293      | .6331      | ...        |
| <b>0.4</b> | .6554      | .6591      | .6628      | .6664      | .6700      | ...        |
| <b>⋮</b>   |            |            |            |            |            |            |

*Note: A complete Z-table is recommended during practice or exams. This excerpt is included here as a reference for common calculations.*



## Author's Note

This book is part of the ongoing **Mathematics Elevate Series**, which is designed to bring clarity, depth, and rigor to high school mathematics learning.

I am continuously working to improve the structure, formatting, and learning experience of this series. Updated editions with refined design, enhanced practice problems, and improved integration with technology are planned.

In particular, I am excited to share that a comprehensive **AP<sup>®</sup> Statistics Book** is currently in development. This upcoming release will provide a complete guide aligned with the official AP<sup>®</sup> Statistics curriculum, offering thorough explanations, exam-style practice, and insights into achieving excellence in the subject.

Your feedback is invaluable in shaping the future editions of this series. If you have suggestions, ideas, or comments, please feel free to reach out through my website:

*[www.mathbyrishabh.com](http://www.mathbyrishabh.com)*

**Rishabh Kumar**  
Elite International Math Mentor  
Math By Rishabh  
Founder, Mathematics Elevate Academy  
Alumnus of IIT Guwahati and the Indian Statistical Institute