



THE PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
UNIVERSITY AMAR TELIDJI OF LAGHOUAT
Faculty of Technology
Department of Science and Technology



COURSE HANDOUT

Presented by: Dr. Mohammed BASSOUDI

Associate Professor (MCA) , University of Laghouat

Email : m.bassoudi@lagh-univ.dz

Course Title

PROBABILITY AND STATISTICS COURSE & EXERCISES WITH SOLUTIONS

Intended for students of
SECOND-YEAR LMD IN SCIENCE AND TECHNOLOGY

Domain: Mathematics – Field: Applied Mathematics
Speciality: Probability and Statistics

Academic year: 2025/2026

Foreword

This course handout is designed to support the course "*Probability and Statistics*" for second-year LMD students in Science and Technology at the University of Laghouat. However, it may also be useful to anyone wishing to learn and, above all, to use the elementary tools of probability and statistics.

The first part of this handout is dedicated to **Statistics**. It covers the basic concepts of descriptive statistics, statistical tables for qualitative and quantitative variables, graphical and numerical representations of data, as well as the characteristics of central tendency and dispersion. It also introduces bivariate statistical series (contingency tables, covariance, correlation, and linear regression).

The second part allows students to understand the fundamental concepts of **Probability**. It consists of three chapters covering combinatorial analysis, probability theory (axioms, conditional probability, independence, and Bayes' theorem), random variables, and the most common probability distributions (Bernoulli, Binomial, Poisson, Uniform, Normal, Exponential).

Each chapter is structured as follows:

- **Theoretical content:** Definitions, properties, and illustrative examples.
- **Exercises:** Core exercises with detailed solutions provided in Appendix A.
- **Additional Exercises:** Supplementary exercises for self-training (without solutions).
- **Additional Engineering Applications:** Practical exercises related to engineering and physical sciences (Civil Engineering, Mechanical Engineering, Electrical Engineering, Hydraulics, Telecommunications, Computer Science, etc.). These exercises aim to bridge the gap between theory and real-world applications.

To produce this handout, I relied on various references, recognized works in the discipline, as well as online resources, which are increasingly present in the dissemination of knowledge.

The statistical graphs presented in this document were primarily generated using the Statistical Package for the Social Sciences (**SPSS**), a widely used software for statistical analysis and data visualization.

Finally, any suggestions or comments that could improve this handout will be most welcome.

*Dr. Mohammed Bassoudi
University of Laghouat, 2026*

Contents

Foreword	ii
List of Figures	vi
Introduction	vii

STATISTICS

1 Basic Definitions	2
I. Introduction	3
II. Statistical Vocabulary	3
II. 1. Population	3
II. 2. Statistical Characteristics	4
II. 3. Modalities	4
II. 4. Frequencies and Statistical Tables	4
III. Types of Statistical Variables	5
III. 1. Qualitative Variables	5
III. 2. Quantitative Variables	6
IV. Summary	7
2 Univariate Statistical Series	12
I. Introduction	13
II. Statistical Tables	13
II. 1. Tables for Qualitative Variables	13
II. 2. Tables for Discrete Quantitative Variables	14
II. 3. Tables for Continuous Quantitative Variables	16
III. Graphical Representations	17
III. 1. Graphs for Qualitative Variables	17
III. 2. Graphs for Discrete Quantitative Variables	19
III. 3. Graphs for Continuous Quantitative Variables	20
IV. Numerical Measures: Central Tendency and Dispersion	22
IV. 1. Introduction	22
IV. 2. Measures of Central Tendency for Discrete Variables	22
IV. 3. Measures of Central Tendency for Continuous Variables	22
IV. 4. Measures of Dispersion	23
IV. 5. Illustrative Examples	24
V. Summary	25

3	Bivariate Statistical Series	33
I.	Introduction	34
II.	Contingency Tables	34
II. 1.	Definition	34
II. 2.	Marginal Distributions	35
II. 3.	Conditional Distributions	35
III.	Scatter Plot	36
III. 1.	Interpreting Scatter Plots	36
IV.	Covariance	37
IV. 1.	Definition	37
IV. 2.	Interpretation of Covariance	37
IV. 3.	Properties of Covariance	37
V.	Linear Correlation Coefficient	38
V. 1.	Properties of Correlation Coefficient	39
V. 2.	Interpreting Correlation Strength	39
V. 3.	Examples of Correlation Patterns	39
VI.	Simple Linear Regression	39
VI. 1.	Method of Ordinary Least Squares (OLS)	40
VI. 2.	Formulas for Regression Coefficients	40
VI. 3.	Prediction Using the Regression Line	41
VII.	Summary	42

PROBABILITY

4	Combinatorial Analysis	49
I.	Introduction	50
II.	Arrangements with Repetition (p -tuple)	50
III.	Arrangements without Repetition	50
IV.	Combinations without Repetition	51
V.	Combinations with Repetition	51
VI.	Binomial Theorem	52
VII.	Summary	52
5	Probability Theory and Conditional Probability	55
I.	Introduction	56
II.	Random Experiments and Sample Spaces	56
III.	Events and Event Operations	56
IV.	Algebra of Events (σ -algebra)	57
V.	Definitions of Probability	57
V. 1.	Classical Approach (Laplace, 1812)	57
V. 2.	Frequentist Approach (Richard von Mises, 1919)	57
V. 3.	Axiomatic Approach (Kolmogorov, 1933)	58
VI.	Probability Spaces	58
VII.	General Probability Theorems	58
VIII.	Conditional Probability	59
IX.	Independence of Events	59
X.	Bayes' Theorem	59
XI.	Summary	60

6	Random Variables and Usual Probability Distributions	64
I.	Introduction	65
II.	Definition of a Random Variable	65
III.	Discrete Random Variables	65
III. 1.	Probability Mass Function (PMF)	65
III. 2.	Cumulative Distribution Function (CDF)	66
III. 3.	Expectation (Mean)	66
III. 4.	Variance and Standard Deviation	66
IV.	Continuous Random Variables	67
IV. 1.	Probability Density Function (PDF)	68
IV. 2.	Cumulative Distribution Function (CDF)	68
IV. 3.	Properties of the CDF	68
IV. 4.	Expectation and Variance for Continuous Variables	69
V.	Discrete Probability Distributions	70
V. 1.	Bernoulli Distribution	70
V. 2.	Binomial Distribution	70
V. 3.	Poisson Distribution	71
VI.	Continuous Probability Distributions	71
VI. 1.	Uniform Distribution	71
VI. 2.	Exponential Distribution	72
VI. 3.	Normal Distribution	73
VI. 4.	Standard Normal Distribution	73
VI. 5.	Standardization (Z-score)	73
VII.	Summary	74
	Appendix A: Solutions to Main Exercises	79
	How to use this appendix	79
	Solutions to Chapter 1	80
	Solutions to Chapter 2	83
	Solutions to Chapter 3	90
	Solutions to Chapter 4	94
	Solutions to Chapter 5	97
	Solutions to Chapter 6	101
	Appendix B: Standard Normal Distribution Table	106
	Bibliography	108

List of Figures

1.1	Statistical study stage	3
1.2	Classification Plan of Statistical Characteristic	7
2.1	Pie chart of Number of employee according to their Job Category.....	18
2.2	Bar chart of employee distribution by job category	18
2.3	bar chart of children per household	19
2.4	Ascending cumulative step curve	19
2.5	Descending cumulative step curve	20
2.6	Histogram for equal class amplitudes	20
2.7	Histogram for unequal class amplitudes	21
2.8	Ascending and descending cumulative curves for continuous data.....	21
3.1	Scatter plot: Study hours vs. Exam score.....	36
3.2	Examples of different correlation patterns.....	40
3.3	Scatter plot with regression line $\hat{y} = 10x + 30$	41
4.1	Combinatorial Analysis Diagram	52
6.1	Uniform distribution on $[0, 10]$ with shaded area $P(X \leq 3)$	72
6.2	Exponential distribution with shaded area $P(X > 15)$	73
6.3	Standard normal distribution with area between -1 and 1	74
6.4	Bar chart for qualitative variable	83
6.5	Pie chart	84
6.6	Bar chart of n_i for quantitative discrete variable.....	85
6.7	Step curve of N^+ and N^- for quantitative discrete variable.....	85
6.8	Histogram $+M_o$ for quantitative continuous variable	86
6.9	Cumulative curve $+M_e$ for quantitative continuous variable	87
6.10	Histogram of unequal class for quantitative continuous variable	88
6.11	Scatter plot	90

Introduction

Statistics is the discipline that studies phenomena through the collection of data and the development of procedures to process, analyze, and interpret these data to make them understandable to everyone. The term "statistics" dates back to the classical Latin word "status" meaning "state". Through a series of evolutions, it gave rise to the French "statistique" (status, stato, statista (1633), statistica (1672), statisticus (1771), statistique). Around the same period, "Statistik" appeared in German, while English used the expression "political arithmetic" until 1798, when the word "statistics" entered the English language's dictionary.

In ancient times, statistics consisted merely of collecting information about states. Nowadays, statistics refers both to a set of observational data and to the activity of collecting, processing, and interpreting that data.

Probability theory allows us to model random phenomena. Initially developed from games of chance, it was later extended to all experimental sciences. This theory also enables the construction of models for these phenomena and the performance of theoretical calculations (for example, one can predict the frequencies of events using a probabilistic model of a game of chance).

The word "chance" comes from the Arabic "az-zahr" (the die). In French, this word first referred to a dice game, and then more generally to any random event. From the 17th century onwards, several principles of probability calculation were advanced, among them the work of Pierre de Fermat, Blaise Pascal, Christian Huygens, Pierre-Simon de Laplace, Abraham de Moivre, Jacob Bernoulli, Denis Simeón Poisson (18th century), Carl Friedrich Gauss, and Henri Poincaré (19th century). It was in 1933 that Andrei Kolmogorov rigorously introduced probability theory, which required the development of measure and integration theories.

This document is structured into two main parts:

Part A: Statistics covers the basic concepts of descriptive statistics, statistical tables for qualitative and quantitative variables, graphical representations, numerical representations, as well as bivariate statistical series.

Part B: Probability covers the fundamental concepts of combinatorial analysis, probability theory, random variables, and the most common probability distributions.

At the end of each chapter, you will find exercises (with solutions in Appendix A), additional exercises for self-training, and engineering applications to strengthen the connection between theory and practice.

Part I

STATISTICS

Contents of this chapter

I.	Introduction	3
II.	Statistical Vocabulary	3
	II. 1. Population	3
	II. 2. Statistical Characteristics	4
	II. 3. Modalities	4
	II. 4. Frequencies and Statistical Tables	4
III.	Types of Statistical Variables	5
	III. 1. Qualitative Variables	5
	III. 2. Quantitative Variables	6
IV.	Summary	7

I. Introduction

Statistics is a scientific discipline concerned with the collection, organization, analysis, interpretation, and presentation of data. It is divided into two main branches:

- **Descriptive Statistics:** Deals with methods for summarizing and presenting data in a clear and understandable way (tables, graphs, numerical measures).
- **Inferential Statistics:** Deals with drawing conclusions and making predictions about a population based on information obtained from a sample.

This chapter introduces the fundamental vocabulary and concepts necessary to understand and construct statistical studies (see figure 1.1).

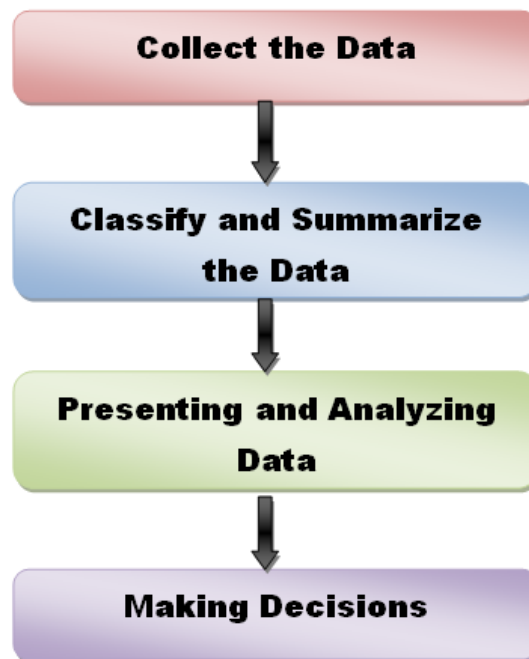


Figure 1.1: Statistical study stage

II. Statistical Vocabulary

To properly describe statistical methods, we must first define some basic notions and the most commonly used terminology.

II. 1. Population

- **Population:** The entire set of individuals or objects of interest in a statistical study.

Example 1 *New high school graduates enrolled in Algerian universities.*

- **Individual (or Statistical Unit):** A single element of the population.

Example 2 *One new high school graduate is an individual of the population.*

- **Sample:** A subset of the population, selected for study when it is impractical to examine the entire population.

Example 3 *New high school graduates enrolled in the Faculty of Technology at Laghouat University is a sample of the population.*

- **Random Sample:** A sample selected in such a way that every member of the population has an equal chance of being included.

II. 2. Statistical Characteristics

- **Characteristic:** What is observed or measured on the individuals of a population.

Example 4 *The age of new high school graduates is a characteristic that can be studied.*

II. 3. Modalities

- Each characteristic can present two or more different situations called **modalities**, denoted by m_i .

Example 5 *The ages of new high school graduates: 18, 19, 20 are modalities.*

- For quantitative variables, modalities are often called **values** and denoted by x_i .

II. 4. Frequencies and Statistical Tables

- **Frequency (Absolute Frequency):** The number of individuals presenting a given modality, denoted by n_i .
- **Total Population Size:** The sum of all frequencies, denoted by N .

$$\sum_{i=1}^k n_i = N \quad (1.1)$$

Example 6

Frequency of new graduates aged 18 = 300

Frequency of new graduates aged 19 = 500

Frequency of new graduates aged 20 = 200

Total population size $N = 1000$

- **Statistical Distribution (or Statistical Series):** The set of modalities and their corresponding frequencies: (m_i, n_i) or (x_i, n_i) .
- **Statistical Table:** A table that organizes modalities and their frequencies.
 - **One-dimensional table:** When only one characteristic is considered.
 - **Two-dimensional (contingency) table:** When two characteristics are considered simultaneously.

- **One-dimensional table format:**

Modalities (x_i)	Frequencies (n_i)
x_1	n_1
x_2	n_2
x_3	n_3
$\vdots$	$\vdots$
x_k	n_k
Total	$N = \sum_{i=1}^k n_i$

Example 7 Consider a population of 1000 new high school graduates enrolled in Algerian universities, studied according to three characteristics: Sex, Age, and Weight (in kg).

1. **One-dimensional tables:**

Sex	Frequency
Male	480
Female	520
Total	1000

Age	Frequency
18	300
19	500
20	200
Total	1000

Weight (kg)	Frequency
[56, 58[	500
[58, 60[	300
[60, 62[	100
[62, 64[	100
Total	1000

2. **Two-dimensional (contingency) table (Weight $\times$ Sex):**

Sex \ Weight	[56, 58[	[58, 60[	[60, 62[	[62, 64[	Total
Male	240	155	45	40	480
Female	260	145	55	60	520
Total	500	300	100	100	1000

III. Types of Statistical Variables

Statistical characteristics can be classified as follows:

III. 1. Qualitative Variables

- A variable is called **qualitative** when its modalities are **not measurable** numerically; they are simply observed and described by words indicating a "state".

Example 8 Sex, profession, eye color are qualitative variables.

- Qualitative variables can be:
 - **Nominal:** Categories have no natural order (e.g., Sex, Blood type).
 - **Ordinal:** Categories have a natural order (e.g., Education level: primary, secondary, university).

III. 2. Quantitative Variables

- A variable is called **quantitative** when its modalities are **measurable** numerically.
- The quantitative variable is then called a **statistical variable** and its modalities are the possible values of the variable.

Example 9 Age, height, weight, number of children are quantitative variables.

- Quantitative variables are divided into two types: **discrete** and **continuous**.

III. 2. a. Discrete Quantitative Variables

Déf. A quantitative variable is **discrete** when it takes only isolated values (usually integers).

Example 10 Number of children per family (0, 1, 2, 3, ...), number of rooms in an apartment (2, 3, 4, ...), number of employees in a company (40, 50, 70, ...).

III. 2. b. Continuous Quantitative Variables

Définition 2

A quantitative variable is **continuous** when it can take any value within an interval of real numbers. In practice, when a variable can take so many different values, we are forced to group them into intervals called **classes**. A class is denoted by:

$$C_i = [e_{i-1}, e_i[, \quad i = 1, \dots, k \quad (1.2)$$

where:

- ▶ e_{i-1} : lower bound of class i
- ▶ e_i : upper bound of class i

Example 11 Weight ([60, 62[, [62, 66[, ...), age ([25, 30[, [30, 35[, ...), surface area ([25, 35[, [35, 45[, ...)) are considered continuous variables.

Définition 3

- ▶ **Class amplitude** (or width): The length of the class interval, denoted by a_i :

$$a_i = e_i - e_{i-1} \quad (1.3)$$

- ▶ **Class midpoint** (or class center): The central value of the class, denoted by x_i :

$$x_i = \frac{e_{i-1} + e_i}{2} \quad (1.4)$$

Example 12 For a weight class $[60, 62[$:

Class amplitude: $a_i = 62 - 60 = 2 \text{ kg}$

Class midpoint: $x_i = \frac{60 + 62}{2} = 61 \text{ kg}$

Remark 1 Class amplitudes can be *equal* or *unequal* depending on the data distribution.

- **Class boundaries:** Care must be taken in defining class limits. In the interval notation $[e_{i-1}, e_i[$, the lower bound is included and the upper bound is excluded. This avoids ambiguity for values that fall exactly on the boundary.

Remark 2 For easier reading of statistical tables, class boundaries are generally chosen as "round" numbers (e.g., $[60, 70[$ rather than $[58.3, 67.9[$).

IV. Summary

In this chapter, we have established the fundamental vocabulary of descriptive statistics:

- **Population, sample, individual** - the basic units of any statistical study.
- **Variables (characteristics)** - what is measured or observed.
- **Modalities (values)** - the possible outcomes of a variable.
- **Frequencies and statistical tables** - how data is organized.
- **Types of variables** - qualitative: nominal/ordinal; and quantitative: discrete/continuous (see figure 1.2).

These concepts form the foundation for all subsequent chapters on data presentation, numerical measures, and probability.

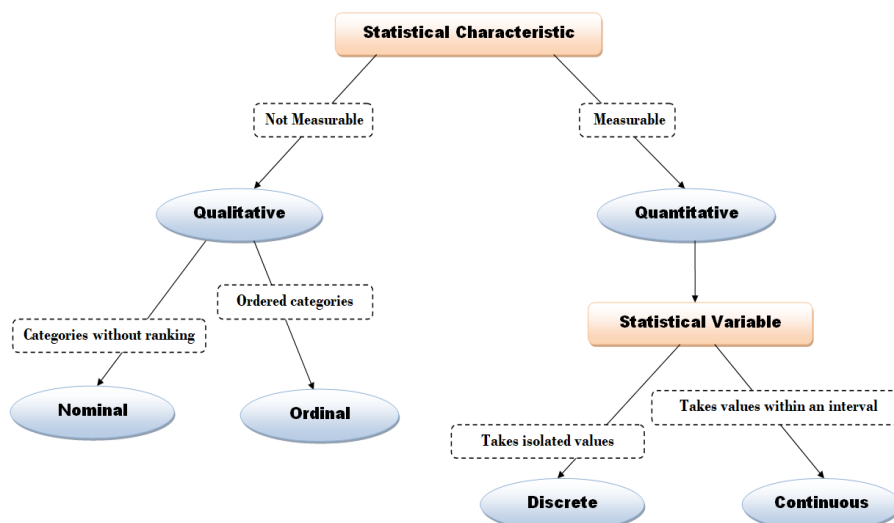


Figure 1.2: Classification Plan of Statistical Characteristic

Exercises

Exercise 1.1.

For each of the following variables, indicate whether it is qualitative (nominal or ordinal) or quantitative (discrete or continuous):

1. Number of students in a class.
2. Blood type (A, B, AB, O).
3. Time to complete an exam (in minutes).
4. Customer satisfaction rating (Very satisfied, Satisfied, Neutral, Unsatisfied, Very unsatisfied).
5. Temperature in degrees Celsius.
6. Marital status.

Solution: see page 80

Exercise 1.2.

A survey of 200 families recorded the number of cars per household:

Number of cars	0	1	2	3 or more
Number of families	30	85	60	25

1. What is the population? The sample?
2. What is the variable studied? Is it qualitative or quantitative? Discrete or continuous?
3. Calculate the total number of families surveyed.

Solution: see page 80

Exercise 1.3.

The following table provides the distribution of 150 students in the Faculty of Technology according to their specialty:

Specialty	Number of Students
Civil Engineering	25
Electrical Engineering	40
Telecommunications	n_3
Mechanical Engineering	30
Process Engineering	20

1. Determine the population, the studied characteristic, and its type.
2. Determine the number of students in "Telecommunications" specialty.

Solution: see page 81

Exercise 1.4.

For each of the following situations, identify the population, the sample, and the variable of interest:

1. A researcher wants to study the average height of students at a university. He measures the height of 150 randomly selected students.
2. A quality control inspector examines 50 light bulbs from a production batch of 10,000 bulbs to check if they are defective.
3. A doctor wants to know the average blood pressure of patients over 60 years old. He records the blood pressure of 200 patients from three different hospitals.

Solution: see page 82

Additional Exercises

Exercise 1.5.

We are studying 20 families based on the number of children they have. The following results are observed:

1, 3, 5, 5, 3, 2, 4, 4, 6, 0, 2, 4, 3, 6, 0, 5, 4, 2, 3, 2.

1. Determine the population, the statistical unit (individual), the statistical variable, and the categories.
2. Complete the following statistical table:

Categories	n_i
0	
1	
2	
3	
4	
5	
6	

Exercise 1.6.

We select the grades of students in the technical sciences (TS) group G1 in the Probability and Statistics module:

5, 6, 7, 1, 0, 3, 4, 6, 6, 5, 3, 1, 7, 4, 5, 5, 3, 4, 4, 3, 5, 4, 6, 5, 4.

1. Determine the population, the statistical variable, and its type.
2. Calculate the mode (M_o), the median (M_e) and the mean ($\bar{X}$); then interpret these results.
Note: The mode, median, and mean will be studied in detail in Chapter 2. This exercise serves as an introduction.

Exercise 1.7.

Complete the following statistical table for the distribution of 50 employees by age group:

Age Group	Frequency n_i	Relative Frequency f_i	Percentage (%)
[20, 30[	10	0.30	10%
[30, 40[	15		
[40, 50[	8		
[50, 60[			
[60, 70[			
Total	50	1.00	100%

Exercise 1.8.

Classify the following variables as qualitative (nominal/ordinal) or quantitative (discrete/continuous):

1. The number of pages in a book.
2. The color of a car.
3. The time it takes to run 100 meters.
4. The rating of a movie (1 star, 2 stars, 3 stars, 4 stars, 5 stars).
5. The number of students absent from class today.
6. The brand of a computer.

Additional Engineering Applications

Exercise 1.9.

A civil engineering laboratory tests the compressive strength of 120 concrete samples. The results are classified into three categories:

Strength Category	Number of samples
Low (10-20 MPa)	25
Medium (20-35 MPa)	60
High (35-50 MPa)	35

1. Identify the population and the variable.
2. What type of variable is this?
3. Calculate the relative frequencies and percentages for each category.

Exercise 1.10.

An electrical engineering lab measures the resistance (in ohms) of 80 resistors produced by three different machines:

Machine	Number of resistors
Machine A	35
Machine B	28
Machine C	17

1. What is the population? The sample?
2. What is the variable studied? Its type?
3. Calculate the percentages for each machine.

2 Univariate Statistical Series

Contents of this chapter

I.	Introduction	13
II.	Statistical Tables	13
II. 1.	Tables for Qualitative Variables	13
II. 2.	Tables for Discrete Quantitative Variables	14
II. 3.	Tables for Continuous Quantitative Variables	16
III.	Graphical Representations	17
III. 1.	Graphs for Qualitative Variables	17
III. 2.	Graphs for Discrete Quantitative Variables	19
III. 3.	Graphs for Continuous Quantitative Variables	20
IV.	Numerical Measures: Central Tendency and Dispersion	22
IV. 1.	Introduction	22
IV. 2.	Measures of Central Tendency for Discrete Variables	22
IV. 3.	Measures of Central Tendency for Continuous Variables	22
IV. 4.	Measures of Dispersion	23
IV. 5.	Illustrative Examples	24
V.	Summary	25

I. Introduction

"The goal of statistics is to extract meaning from data." - David S. Moore

Raw statistical data, as initially collected, is often unusable and difficult to interpret. To make the information **readable** and **understandable**, we must organize and present it in a structured way. This chapter covers:

- Statistical tables for qualitative and quantitative data.
- Frequencies, relative frequencies, and percentages.
- Cumulative frequencies and cumulative relative frequencies.
- Graphical representations: pie charts, bar charts, dot plots, histograms, frequency polygons, and cumulative curves (ogives).

Remark 3 *The choice of an appropriate table or graph depends on:*

1. *The **type of variable** (qualitative vs. quantitative)*
2. *The **sample size** (small vs. large)*
3. *The **audience** (technical vs. general)*
4. *The **purpose** of the analysis (exploration vs. presentation)*

II. Statistical Tables

A statistical table is an organized arrangement of data in rows and columns. It is the most fundamental tool for presenting statistical information.

II. 1. Tables for Qualitative Variables

When dealing with non-measurable (qualitative) characteristics, we arrange them in a table according to the logic that best allows **interpretation**.

Example 13 *Distribution of 150 employees in a company by job category.*

Job Category	Frequency n_i	Relative Frequency f_i	Percentage (%)
<i>Senior Executives</i>	<i>10</i>	<i>0.067</i>	<i>6.7%</i>
<i>Engineers</i>	<i>20</i>	<i>0.133</i>	<i>13.3%</i>
<i>Senior Technicians</i>	<i>30</i>	<i>0.200</i>	<i>20.0%</i>
<i>Technicians</i>	<i>90</i>	<i>0.600</i>	<i>60.0%</i>
Total	150	1.000	100%

II. 2. Tables for Discrete Quantitative Variables

Let X be a discrete quantitative variable that takes values (modalities) $x_1, x_2, \dots, x_k$ with $x_1 < x_2 < \dots < x_k$, and suppose there are n_i individuals having the value x_i .

Déf.1 The **frequency distribution** (or **statistical series**) of a discrete variable is the set of pairs (x_i, n_i) for $i = 1, \dots, k$.

The statistical table for a discrete variable is presented as follows:

Values x_i	Frequencies n_i
x_1	n_1
x_2	n_2
$\vdots$	$\vdots$
x_k	n_k
Total	$N = \sum_{i=1}^k n_i$

Example 14 A survey on the number of children per household gives the following raw data:

5, 0, 2, 1, 3, 2, 1, 2, 5, 2

The discrete variable table is:

Number of children x_i	Number of households n_i
0	1
1	2
2	4
3	1
4	0
5	2
Total	$N = 10$

Note: The value $x = 4$ appears with zero frequency. We still include it because it is a possible value of the variable.

II. 2. a. Frequencies

Déf.2 **Absolute frequency** (or simply **frequency**) of a modality x_i is the number of individuals having that value, denoted by n_i .

Déf.3 The **relative frequency** is the proportion of individuals presenting the same modality in the total population, denoted by f_i :

$$f_i = \frac{n_i}{N}, \quad i = 1, 2, \dots, k \quad (2.1)$$

Remark 4 Relative frequencies are often expressed as **percentages**:

$$f_i \% = \frac{n_i}{N} \times 100\%, \quad i = 1, 2, \dots, k \quad (2.2)$$

Example 15 For Example 14:

x_i	n_i	f_i	$f_i\%$
0	1	$1/10 = 0.1$	10%
1	2	$2/10 = 0.2$	20%
2	4	$4/10 = 0.4$	40%
3	1	$1/10 = 0.1$	10%
4	0	$0/10 = 0.0$	0%
5	2	$2/10 = 0.2$	20%
Total	$N = 10$	1.0	100%

Interpretation: $f_3 = 0.4$ means that 40% of households have exactly 2 children.

Propriété 1 (Sum of Relative Frequencies)

The sum of all relative frequencies equals 1.

$$\sum_{i=1}^k f_i = 1 \quad (2.3)$$

Proof 1 Since $\sum_{i=1}^k n_i = N$, we have:

$$\sum_{i=1}^k f_i = \sum_{i=1}^k \frac{n_i}{N} = \frac{1}{N} \sum_{i=1}^k n_i = \frac{N}{N} = 1$$

II. 2. b. Cumulative Frequencies

Cumulative frequencies are essential for answering questions like "how many individuals have a value less than or equal to a certain threshold?"

Déf.4

For each value x_i , the **ascending cumulative frequency** N_i^+ and **ascending cumulative relative frequency** F_i^+ are defined as:

$$N_i^+ = \sum_{h=1}^i n_h = n_1 + n_2 + \dots + n_i, \quad F_i^+ = \sum_{h=1}^i f_h = f_1 + f_2 + \dots + f_i$$

for $i = 1, \dots, k$.

Interpretation 1 F_i^+ (or N_i^+) represents the proportion (or number) of individuals with a variable value **less than or equal to** x_i .

Déf.5

For each value x_i , the **descending cumulative frequency** N_i^- and **descending cumulative relative frequency** F_i^- are defined as:

$$N_i^- = \sum_{h=k}^i n_h = n_k + n_{k-1} + \dots + n_i, \quad F_i^- = \sum_{h=k}^i f_h = f_k + f_{k-1} + \dots + f_i$$

for $i = 1, \dots, k$.

Interpretation 2 F_i^- (or N_i^-) represents the proportion (or number) of individuals with a variable value **greater than or equal to** x_i .

Example 16 Using the data from Example 14:

x_i	n_i	f_i	N_i^+	F_i^+	N_i^-	F_i^-
0	1	0.1	1	0.1	10	1.0
1	2	0.2	3	0.3	9	0.9
2	4	0.4	7	0.7	7	0.7
3	1	0.1	8	0.8	3	0.3
4	0	0.0	8	0.8	2	0.2
5	2	0.2	10	1.0	2	0.2
Total	10	1.0	/	/	/	/

Interpretation:

- $F_3^+ = 0.7$: 70% of households have 2 children or fewer.
- $F_4^- = 0.3$: 30% of households have 3 children or more.
- $N_6^+ = 10$: All 10 households have 5 children or fewer.
- $N_2^- = 9$: 9 households have 1 child or more.

II. 3. Tables for Continuous Quantitative Variables

For continuous variables, the number of distinct values is too large to list individually. Therefore, we group the data into **classes** (intervals).

Définition 6 (Classes)

A **class** is an interval that groups together individuals with similar values. It is denoted by:

$$C_i = [e_{i-1}, e_i[, \quad i = 1, \dots, k$$

where:

- e_{i-1} : **lower bound** of class i (included)
- e_i : **upper bound** of class i (excluded)

Déf.7

- **Class amplitude** (width): $a_i = e_i - e_{i-1}$
- **Class midpoint** (center): $x_i = \frac{e_{i-1} + e_i}{2}$

Example 17 The weights (in kg) of 40 individuals were recorded and grouped into 5 classes:

Class C_i	Amplitude a_i	Midpoint x_i	Frequency n_i
[60, 70[	10	65	2
[70, 75[	5	72.5	10
[75, 80[	5	77.5	12
[80, 85[	5	82.5	12
[85, 90[	5	87.5	4
Total			$N = 40$

Note: The first class $[60, 70[$ has an amplitude of 10. The remaining classes have amplitude 5.

For continuous variables, we define:

- n_i : number of individuals whose value belongs to class C_i ($\sum n_i = N$)
- $f_i = n_i/N$: relative frequency of class C_i ($\sum f_i = 1$)

Cumulative frequencies for continuous variables are defined similarly, with the understanding that:

- N_i^+ (or F_i^+) is the number (or proportion) of individuals with value **strictly less than** e_i (the upper bound).
- N_i^- (or F_i^-) is the number (or proportion) of individuals with value **greater than or equal to** e_{i-1} (the lower bound).

Example 18 Continuing Example 17:

C_i	n_i	f_i	N_i^+	N_i^-	F_i^+	F_i^-
$[60, 70[$	2	0.05	2	40	0.05	1.00
$[70, 75[$	10	0.25	12	38	0.30	0.95
$[75, 80[$	12	0.30	24	28	0.60	0.70
$[80, 85[$	12	0.30	36	16	0.90	0.40
$[85, 90[$	4	0.10	40	4	1.00	0.10
Total	40	1.00	/	/	/	/

Interpretation:

- $F_3^+ = 0.60$: 60% of individuals weigh less than 80 kg.
- $F_4^- = 0.40$: 40% of individuals weigh 80 kg or more.
- $N_5^+ = 40$: All 40 individuals weigh less than 90 kg.

Remark 5 (Rules for Constructing Statistical Tables) 1. **Title:** Must clearly indicate the phenomenon studied, location, date, scope, and classification criteria.

2. **Headings:** Must clearly indicate whether entries are numbers, percentages, frequencies, or rates.

3. **Units:** Must be clearly specified (e.g., kg, cm, %, etc.).

4. **Source:** Must cite the organization or person who produced the data.

III. Graphical Representations

"A picture is worth a thousand words." - Ancient proverb

III. 1. Graphs for Qualitative Variables

Two main graphical representations dominate for qualitative variables: **pie charts** and **bar charts**.

III. 1. a. Pie Chart (Circular Diagram)

A pie chart divides a circle into sectors corresponding to the observed modalities. The angle of each sector is proportional to the frequency (or percentage) of that modality.

Déf.8

For a modality with absolute frequency n_i or relative frequency f_i , the angle θ_i of its sector is:

$$\theta_i = \frac{n_i}{N} \times 360^\circ = f_i \times 360^\circ$$

Example 19 Using the job category data from Example 13:

Job Category	n_i	f_i	θ_i
Senior Executives	10	0.067	24°
Engineers	20	0.133	48°
Senior Technicians	30	0.200	72°
Technicians	90	0.600	216°
Total	150	1.000	360°

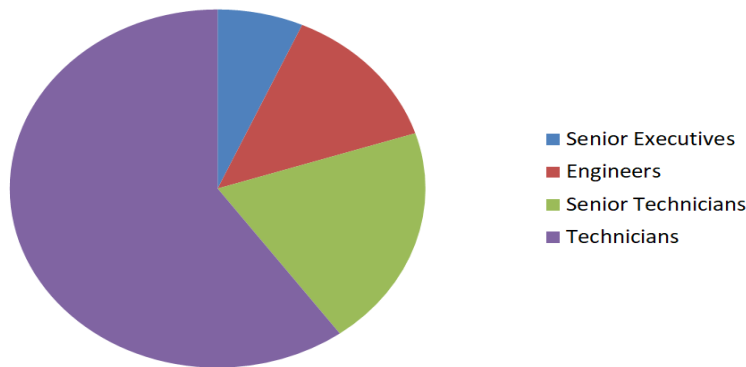


Figure 2.1: Pie chart of Number of employee according to their Job Category

III. 1. b. Bar Chart

Bar charts display modalities on the horizontal axis and vertical bars whose height is proportional to the frequency (or relative frequency) of each modality.

Example 20 For the same job category data:

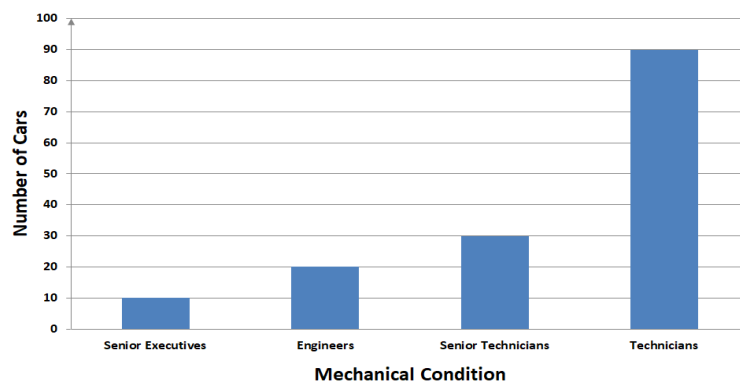


Figure 2.2: Bar chart of employee distribution by job category

III. 2. Graphs for Discrete Quantitative Variables

III. 2. a. Bar Chart

For discrete variables, each value x_i is represented by a vertical bar (or line) whose height is proportional to n_i or f_i .

Example 21 Using the children per household data from Example 14:

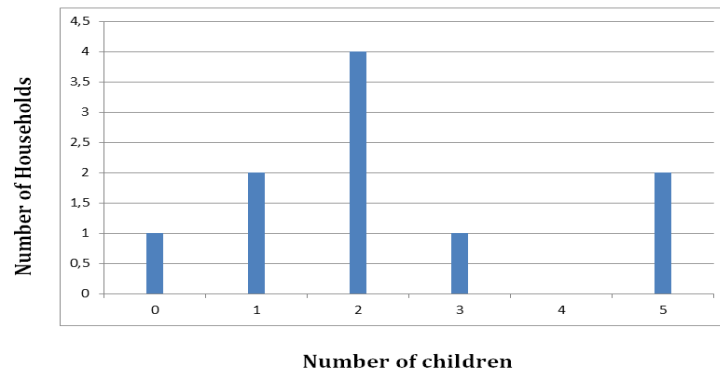


Figure 2.3: bar chart of children per household

III. 2. b. Cumulative Step Curve

To represent ascending cumulative frequencies (N_i^+ or F_i^+), we obtain a **step curve** with horizontal plateaus.

Example 22 Using the cumulative data from Example 16:

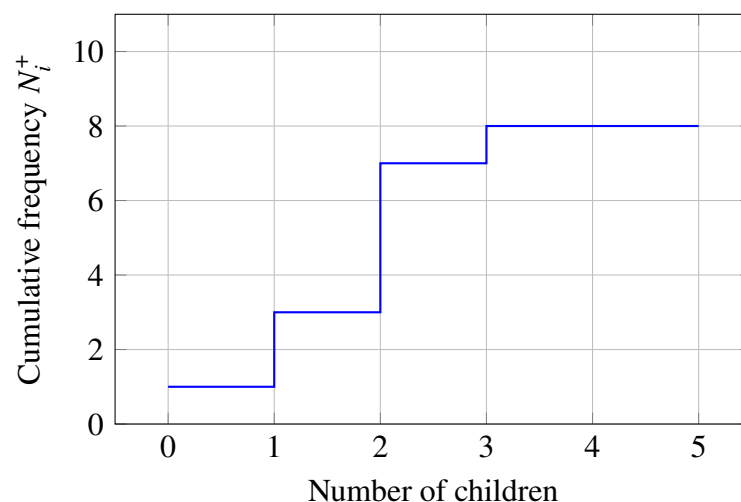


Figure 2.4: Ascending cumulative step curve

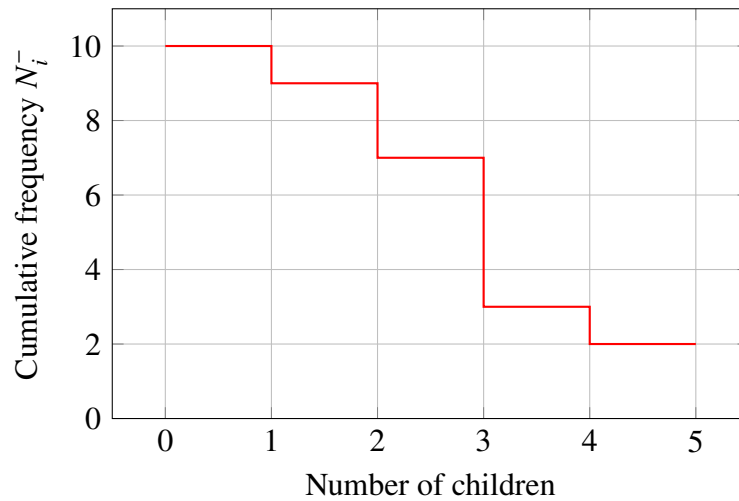


Figure 2.5: Descending cumulative step curve

III. 3. Graphs for Continuous Quantitative Variables

III. 3. a. Histogram

For continuous variables where observations are grouped into classes, the reference diagram is the **histogram**. The **area** of each rectangle represents the frequency of that class.

Case 1: Equal class amplitudes (a_i constant)

When all classes have the same width, the height of each rectangle is proportional to n_i or f_i .

Example 23 *Distance traveled by taxis:*

<i>Distance (km)</i>	<i>Number of taxis</i>	a_i
[10, 20[	9	10
[20, 30[	13	10
[30, 40[	22	10
[40, 50[	10	10
[50, 60[	7	10
[60, 70[	4	10
<i>Total</i>	65	

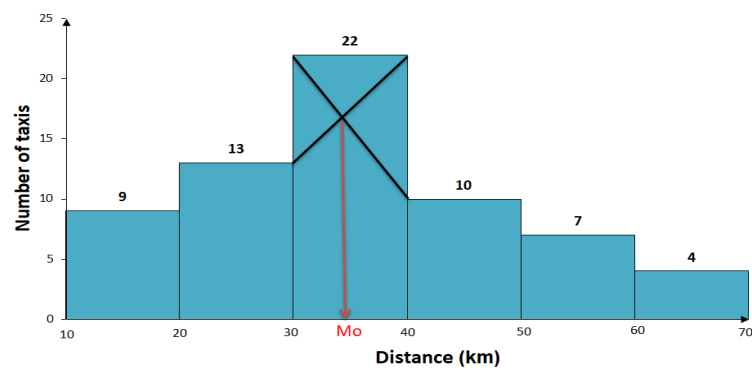


Figure 2.6: Histogram for equal class amplitudes

Case 2: Unequal class amplitudes (a_i not constant)

When amplitudes are unequal, we must use **rectified frequencies** to avoid distorting the visual representation:

$$n'_i = \frac{n_i}{a_i}$$

Example 24 Using the weight data from Example 17:

Class C_i	a_i	n_i	$n'_i = n_i/a_i$
[60, 70[	10	2	0.2
[70, 75[	5	10	2.0
[75, 80[	5	12	2.4
[80, 85[	5	12	2.4
[85, 90[	5	4	0.8

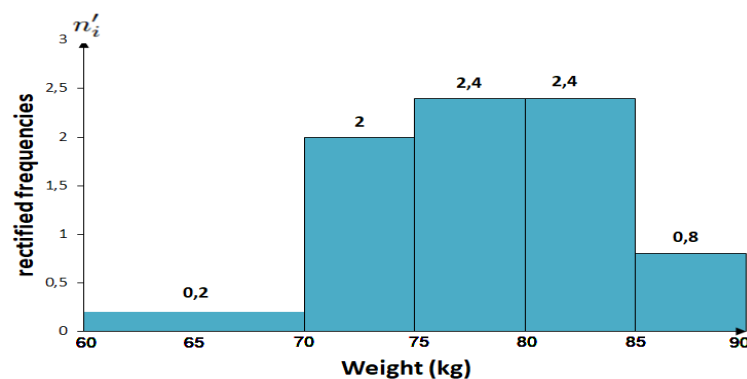


Figure 2.7: Histogram for unequal class amplitudes

III. 3. b. Cumulative Curve for Continuous Variables

For continuous variables, the cumulative curve is obtained by plotting the **upper class boundaries** against the cumulative frequencies (N_i^+ or F_i^+); or the **lower class boundaries** against the cumulative frequencies (N_i^- or F_i^-), then joining the points with a smooth curve.

Example 25 Using the cumulative data from Example 18:

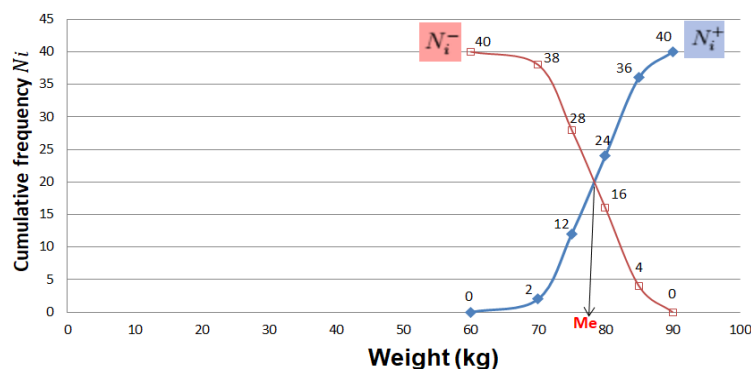


Figure 2.8: Ascending and descending cumulative curves for continuous data

IV. Numerical Measures: Central Tendency and Dispersion

IV. 1. Introduction

In addition to graphical representations, numerical measures provide a concise summary of the main features of a statistical series. These measures are divided into two categories:

- **Measures of central tendency (position):** They indicate the center or typical value around which the data are concentrated (Mean, Mode, Median, Quartiles).
- **Measures of dispersion:** They describe the spread or variability of the data (Range, Interquartile Range, Variance, Standard Deviation).

In this section, we present the formulas for these measures for both **discrete** and **continuous** variables.

IV. 2. Measures of Central Tendency for Discrete Variables

IV. 2. a. Mean

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k n_i x_i$$

IV. 2. b. Mode

The mode is the value x_i that has the highest frequency n_i .

IV. 2. c. Median

The median is the value x_i such that:

$$N_{i-1}^+ \leq \frac{N}{2} \leq N_i^+$$

IV. 2. d. Quartiles

- Q_1 is the value x_i such that $N_{i-1}^+ \leq \frac{N}{4} \leq N_i^+$
- Q_3 is the value x_i such that $N_{i-1}^+ \leq \frac{3N}{4} \leq N_i^+$

IV. 3. Measures of Central Tendency for Continuous Variables

IV. 3. a. Mean

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k n_i x_i$$

where x_i is the class midpoint ($x_i = \frac{e_{i-1}+e_i}{2}$).

IV. 3. b. Mode (Graphical and Analytical)

The mode can be determined graphically from the histogram. As shown in Figure 2.6, the mode corresponds to the highest bar in the histogram. The modal class is the class with the highest frequency (or highest rectified frequency if amplitudes are unequal).

See Figure 2.6 for the graphical determination of the mode.

Analytically, the mode is calculated as:

$$M_o = e_{i-1} + a_i \times \frac{d_1}{d_1 + d_2}$$

where:

- e_{i-1} : lower bound of the modal class
- a_i : width of the modal class
- $d_1 = n_i - n_{i-1}$ (difference with the previous class)
- $d_2 = n_i - n_{i+1}$ (difference with the next class)

IV. 3. c. Median (Graphical and Analytical)

The median can be determined graphically from the ascending cumulative curve. As shown in Figure 2.8, the median is the value corresponding to $F(M_e) = 0.5$ or $N(M_e) = N/2$.

See Figure 2.8 for the graphical determination of the median.

Analytically, the median is calculated as:

$$M_e = e_{i-1} + a_i \times \frac{\frac{N}{2} - N_{i-1}^+}{n_i}$$

IV. 3. d. Quartiles

$$Q_1 = e_{i-1} + a_i \times \frac{\frac{N}{4} - N_{i-1}^+}{n_i}$$

$$Q_3 = e_{i-1} + a_i \times \frac{\frac{3N}{4} - N_{i-1}^+}{n_i}$$

IV. 4. Measures of Dispersion

IV. 4. a. Range

$$R = x_{max} - x_{min}$$

IV. 4. b. Interquartile Range (IQR)

$$IQR = Q_3 - Q_1$$

IV. 4. c. Variance and Standard Deviation

For discrete variables:

$$V(X) = \frac{1}{N} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2$$

$$\sigma(X) = \sqrt{V(X)}$$

For continuous variables:

$$V(X) = \frac{1}{N} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2$$

where x_i is the class center.

IV. 5. Illustrative Examples

IV. 5. a. Example 1: Discrete Variable

Consider the distribution of 400 households by number of vehicles:

x_i	0	1	2	3	4
n_i	30	200	100	60	10

Mean:

$$\bar{x} = \frac{30 \times 0 + 200 \times 1 + 100 \times 2 + 60 \times 3 + 10 \times 4}{400} = \frac{620}{400} = 1.55$$

Mode: $M_o = 1$ (highest frequency: 200)

Median: $M_e = 1$ (since $N_{i-1}^+ = 30 \leq 200 \leq N_i^+ = 230$)

Quartiles: $Q_1 = 1, Q_3 = 2$

Variance:

$$V(X) = \frac{1300}{400} - (1.55)^2 = 3.25 - 2.4025 = 0.8475$$

Standard deviation: $\sigma = \sqrt{0.8475} \approx 0.9206$

IV. 5. b. Example 2: Continuous Variable

Consider the distribution of 65 taxis by mileage (in 10^3 km):

Class C_i	[10, 20[	[20, 30[	[30, 40[	[40, 50[	[50, 60[	[60, 70[
n_i	9	13	22	10	7	4
x_i (midpoint)	15	25	35	45	55	65

Mean:

$$\bar{x} = \frac{9 \times 15 + 13 \times 25 + 22 \times 35 + 10 \times 45 + 7 \times 55 + 4 \times 65}{65} = \frac{2325}{65} \approx 35.77 \text{ km}$$

Mode: Modal class $[30, 40[$, $d_1 = 22 - 13 = 9$, $d_2 = 22 - 10 = 12$

$$M_o = 30 + 10 \times \frac{9}{9 + 12} \approx 34.29 \text{ km}$$

Median: Median class $[30, 40[$

$$M_e = 30 + 10 \times \frac{32.5 - 22}{22} \approx 34.77 \text{ km}$$

Variance:

$$V(X) = \frac{95425}{65} - (35.77)^2 \approx 188.64$$

Standard deviation: $\sigma \approx 13.73 \text{ km}$

V. Summary

This chapter covered the essential tools for organizing and presenting univariate statistical data:

- **Statistical tables** organize data into rows and columns, with absolute frequencies, relative frequencies, and cumulative frequencies.
- **Graphical representations** provide visual summaries:
 - For qualitative data: pie charts and bar charts.
 - For discrete quantitative data:
 - * n_i and f_i are represented by bar charts.
 - * N_i^+ and F_i^+ are represented by ascending cumulative step curve.
 - * N_i^- and F_i^- are represented by descending cumulative step curve.
 - For continuous quantitative data:
 - * n_i and f_i are represented by histograms.
 - * N_i^+ and F_i^+ are represented by ascending cumulative curve.
 - * N_i^- and F_i^- are represented by descending cumulative curve.
- The choice of table or graph depends on the type of variable, sample size, audience, and purpose.
- **Measures of central tendency and dispersion** provide a concise description of the location, spread, and shape of a statistical series.

Measure	Discrete Variable	Continuous Variable
Mean	$\bar{x} = \frac{1}{N} \sum_{i=1}^k n_i x_i$	$\bar{x} = \frac{1}{N} \sum_{i=1}^k n_i x_i$ (using class center)
Mode	Value x_i with the highest frequency n_i	$M_o = e_{i-1} + a_i \times \frac{d_1}{d_1 + d_2}$
Median	Value x_i such that $N_{i-1}^+ \leq \frac{N}{2} \leq N_i^+$	$M_e = e_{i-1} + a_i \times \frac{\frac{N}{2} - N_{i-1}^+}{n_i}$
Variance	$V(X) = \frac{1}{N} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2$	Same formula using class center

Exercises

Exercise 2.1.

The following table provides the distribution of 80 cars in a car park according to their mechanical condition:

Mechanical Condition	Number of Cars
Bad	10
Average	20
Good	n_3
Excellent	30

1. Determine the population, the studied characteristic, and its type.
2. Determine the number of cars in "Good" mechanical condition.
3. Give all possible graphical representations of this distribution.

Solution: see page 83

Exercise 2.2.

The following table gives the distribution of 400 households according to the number of vehicles owned:

Number of Vehicles x_i	0	1	2	3	4
Number of Households n_i	30	200	100	60	10

1. What type of statistical variable is being studied?
2. Complete the statistical table by calculating: relative frequencies (f_i), cumulative frequencies (N_i^+ and N_i^-) and cumulative relative frequencies (F_i^+ and F_i^-); then interpret these values: f_3 , N_3^+ , N_3^- , F_3^+ , F_3^- .
3. Represent the data graphically.
4. Represent the cumulative frequencies (N_i^+ and N_i^-) on the same graph.
5. Calculate the measures of central tendency (Mean, Mode, Median, Quartiles).
6. Calculate the measures of dispersion (Range, Interquartile Range, Variance, Standard Deviation).

Solution: see page 84

Exercise 2.3.

The following data specifies the mileage covered by 65 taxis:

Mileage (in 10^3 km)	[10, 20[	[20, 30[	[30, 40[	[40, 50[	[50, 60[	[60, 70[
Number of taxis	9	13	22	10	7	4

1. Determine the population, the studied characteristic, and its type.
2. Calculate the cumulative relative frequencies (F_i^+ and F_i^-); then interpret F_3^+ and F_3^- .
3. Represent the data graphically.
4. Represent the cumulative relative frequencies graphically.
5. Determine the mode (M_o) and the median (M_e) graphically.
6. Calculate the measures of central tendency and dispersion.

Solution: see page 86

Exercise 2.4.

Using the same statistical series in 2.3, but combining the last two classes:

C_i	[10, 20[	[20, 30[	[30, 40[	[40, 50[	[50, 70[
n_i	9	13	22	10	11

1. Represent the data graphically.
2. Calculate the mode (M_o) and the median (M_e) analytically.
3. Calculate the mean ($\bar{X}$), the variance $V(X)$ and the standard deviation $\sigma(X)$.

Solution: see page 87

Exercise 2.5.

We are studying 20 families based on the number of children they have:

1, 3, 5, 5, 3, 2, 4, 4, 6, 0, 2, 4, 3, 6, 0, 5, 4, 2, 3, 2.

1. Determine the population, the statistical unit, the statistical variable, and the categories.
2. Complete the statistical table.

Solution: see page 88

Additional Exercises

Exercise 2.6.

The following data represents the number of defective items produced per day in a factory over 30 days:

2, 5, 3, 2, 1, 4, 3, 2, 5, 3, 1, 2, 4, 3, 2, 1, 3, 2, 5, 4, 3, 2, 1, 3, 2, 4, 5, 3, 2, 1

1. Construct a frequency table (values, frequencies, relative frequencies, percentages).
2. Calculate cumulative frequencies (ascending and descending).
3. Draw a dot plot.
4. Draw the cumulative step curve (step function).

Solution: see page 83

Exercise 2.7.

The heights (in cm) of 50 students are grouped as follows:

Height (cm)	Number of students
[150, 160[	8
[160, 165[	12
[165, 170[	15
[170, 175[	10
[175, 185[	5

1. Calculate the class midpoints and amplitudes.
2. Compute relative frequencies and cumulative frequencies.
3. Draw the histogram.
4. Draw the cumulative frequency curve (ogive).
5. Draw the frequency polygon.

Solution: see page 84

Exercise 2.8.

A survey of 200 families recorded the number of cars per household:

Number of cars	Number of families
0	30
1	85
2	60
3 or more	25

1. What type of variable is this?
2. Calculate the relative frequencies (percentages).
3. Draw a pie chart.
4. Draw a bar chart.

Exercise 2.9.

The following table gives the distribution of 100 employees by salary range:

Salary (1000 DA)	Number of employees
[20, 30[	15
[30, 40[	25
[40, 50[	35
[50, 60[	20
[60, 80[	5

1. Calculate the class midpoints and amplitudes.
2. Calculate the relative frequencies and percentages.
3. Draw the histogram (note: the last class has a different amplitude).
4. Draw the cumulative frequency curve.

Exercise 2.10.

Thirty (30) special steel specimens are subjected to strength tests:

2	2	3	1	2	1	4	2	3	2
3	2	3	3	4	1	1	4	2	3
2	3	2	2	3	4	3	2	3	2

1. What is the type of statistical variable?
2. Construct the statistical table.
3. Plot the cumulative relative frequencies curves (F_i^+ and F_i^-).
4. Calculate the measures of central tendency.
5. Calculate the measures of dispersion.

Exercise 2.11.

We select the grades of students in the technical sciences group:

5, 6, 7, 1, 0, 3, 4, 6, 6, 5, 3, 1, 7, 4, 5, 5, 3, 4, 4, 3, 5, 4, 6, 5, 4.

1. Determine the population, the statistical variable, and its type.
2. Calculate the mode (M_o), the median (M_e) and the mean ($\bar{X}$); then interpret these results.

Exercise 2.12.

We consider the variable "time lived in housing" (in years):

C_i	[0, 1[	[1, 2[	[2, 3[	[3, 5[	[5, 11[	[11, 16[	[16, 21[	[21, 26[
n_i	35	36	32	25	20	18	16	7

1. What is the type of this variable?
2. Graphically represent this statistical series.
3. Calculate the measures of central tendency.
4. Calculate the measures of dispersion.
5. Due to a data entry error, the upper limit of 26 has been replaced by 66. Does this have an impact on the determination of the median?

Exercise 2.13.

The study of the weight of 50 people (Yule and Sturges rules):

30	43	47	50	52	54	55	56	58	58
61	62	63	63	64	65	66	66	67	68
68	69	69	70	71	72	72	72	73	73
74	74	75	76	76	77	79	79	80	82
82	84	86	87	88	90	91	92	93	98

1. Using Yule's rule ($k = 2.5 \times \sqrt[4]{N}$) and Sturges' rule ($k = 1 + 3.322 \times \log_{10}(N)$), construct the statistical table.
2. Construct the corresponding histogram.
3. Represent the cumulative frequencies (ascending and descending) on the same graph.
4. Determine the value of the weight below which 50% of individuals are located.
5. Determine the percentage of individuals with a weight less than 75 kg.

Solution: see page ??

Additional Engineering Applications

Exercise 2.14.

A mechanical engineering workshop measures the diameter (in mm) of 150 shafts produced by a lathe machine:

Diameter (mm)	Number of shafts
[20, 22[	20
[22, 24[	45
[24, 26[	50
[26, 28[	25
[28, 30[	10

1. What type of variable is this?
2. Calculate the mean, mode, and median.
3. Calculate the variance and standard deviation.
4. Draw the histogram and the cumulative frequency curve.

Exercise 2.15.

A hydraulics engineer measures the flow rate (in L/s) of a water pipe over 50 days:

Flow rate (L/s)	Number of days
[5, 10[	8
[10, 15[	18
[15, 20[	14
[20, 25[	6
[25, 30[	4

1. What type of variable is this?
2. Calculate the mean, median, and mode.
3. Calculate the variance and standard deviation.
4. Draw the histogram.

Exercise 2.16.

An electronics engineer measures the lifetime (in hours) of 60 batteries:

Lifetime (hours)	Number of batteries
[50, 60[	5
[60, 70[	12
[70, 80[	20
[80, 90[	15
[90, 100[	8

1. Calculate the mean, median, and mode.
2. Calculate the variance and standard deviation.
3. Draw the histogram and the frequency polygon.

3 Bivariate Statistical Series

Contents of this chapter

I.	Introduction	34
II.	Contingency Tables	34
II. 1.	Definition	34
II. 2.	Marginal Distributions	35
II. 3.	Conditional Distributions	35
III.	Scatter Plot	36
III. 1.	Interpreting Scatter Plots	36
IV.	Covariance	37
IV. 1.	Definition	37
IV. 2.	Interpretation of Covariance	37
IV. 3.	Properties of Covariance	37
V.	Linear Correlation Coefficient	38
V. 1.	Properties of Correlation Coefficient	39
V. 2.	Interpreting Correlation Strength	39
V. 3.	Examples of Correlation Patterns	39
VI.	Simple Linear Regression	39
VI. 1.	Method of Ordinary Least Squares (OLS)	40
VI. 2.	Formulas for Regression Coefficients	40
VI. 3.	Prediction Using the Regression Line	41
VII.	Summary	42

I. Introduction

In many practical situations, we are interested in studying the **relationship** between two variables rather than a single variable in isolation.

Example 26 *Examples of Bivariate Studies:*

- Is there a relationship between **study time** and **exam score**?
- Does **advertising expenditure** affect **sales revenue**?
- Is there a correlation between **height** and **weight**?
- What is the relationship between **temperature** and **energy consumption**?

This chapter introduces methods for describing and analyzing the relationship between two statistical variables.

Remark 6 *We distinguish two cases:*

1. *Two qualitative variables ? Contingency tables*
2. *Two quantitative variables ? Scatter plots, covariance, correlation, regression*

*This chapter focuses primarily on **quantitative variables**.*

II. Contingency Tables

II. 1. Definition

A **contingency table** is a table that displays the frequency distribution of two qualitative variables simultaneously.

Déf. 1

Consider two qualitative variables X and Y with r and c modalities respectively. The **contingency table** is an $r \times c$ table where each cell (i, j) contains the **absolute frequency** n_{ij} , i.e., the number of individuals that present modality i of X and modality j of Y simultaneously.

$X \setminus Y$	y_1	y_2	$\cdots$	y_c
x_1	n_{11}	n_{12}	$\cdots$	n_{1c}
x_2	n_{21}	n_{22}	$\cdots$	n_{2c}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
x_r	n_{r1}	n_{r2}	$\cdots$	n_{rc}

Example 27 *A survey of 100 students records their **gender** (Male, Female) and their **preferred subject** (Mathematics, Physics, Biology):*

Gender \ Subject	Mathematics	Physics	Biology	Total
<i>Male</i>	25	20	10	55
<i>Female</i>	15	10	20	45
Total	40	30	30	100

Interpretation: $n_{11} = 25$ males prefer Mathematics; $n_{32} = 10$ females prefer Physics, etc.

II. 2. Marginal Distributions

Définition 2

The **marginal frequencies** are obtained by summing the rows and columns of the contingency table:

- **Row marginal:** $n_{i.} = \sum_{j=1}^c n_{ij}$ (frequency of modality i of X)
- **Column marginal:** $n_{.j} = \sum_{i=1}^r n_{ij}$ (frequency of modality j of Y)

The total number of individuals is:

$$N = \sum_{i=1}^r \sum_{j=1}^c n_{ij} = \sum_{i=1}^r n_{i.} = \sum_{j=1}^c n_{.j}$$

Example 28 Continuing the previous example

Gender \ Subject	Math	Physics	Biology	$n_{i.}$
<i>Male</i>	25	20	10	55
<i>Female</i>	15	10	20	45
$n_{.j}$	40	30	30	100

Marginal distribution of Gender: 55% Male, 45% Female.

Marginal distribution of Subject: 40% Math, 30% Physics, 30% Biology.

II. 3. Conditional Distributions

Définition 3

The **conditional distribution** of Y given $X = x_i$ is the distribution of Y restricted to individuals with $X = x_i$. The conditional frequencies are:

$$f_{j|i} = \frac{n_{ij}}{n_{i.}} \quad \text{for } j = 1, \dots, c$$

Similarly, the conditional distribution of X given $Y = y_j$ is:

$$f_{i|j} = \frac{n_{ij}}{n_{.j}} \quad \text{for } i = 1, \dots, r$$

Example 29 Conditional distribution of **Subject** given **Gender = Male**:

$$f_{\text{Math}|M} = \frac{25}{55} \approx 0.455, \quad f_{\text{Physics}|M} = \frac{20}{55} \approx 0.364, \quad f_{\text{Biology}|M} = \frac{10}{55} \approx 0.182$$

Conditional distribution of **Subject** given **Gender = Female**:

$$f_{\text{Math}|F} = \frac{15}{45} \approx 0.333, \quad f_{\text{Physics}|F} = \frac{10}{45} \approx 0.222, \quad f_{\text{Biology}|F} = \frac{20}{45} \approx 0.445$$

Interpretation: Male students prefer Mathematics with 45.5%, while Female students prefer Biology with 44.5%.

III. Scatter Plot

For **quantitative variables**, the first step in studying their relationship is to create a **scatter plot**.

Déf.4

Given n pairs of observations (x_i, y_i) for $i = 1, \dots, n$, a **scatter plot** is a graph where each point $M_i(x_i, y_i)$ is plotted on a Cartesian coordinate system.

Example 30 Consider the following bivariate data:

x (Study hours)	1	2	3	4	5
y (Exam score)	40	50	60	70	80

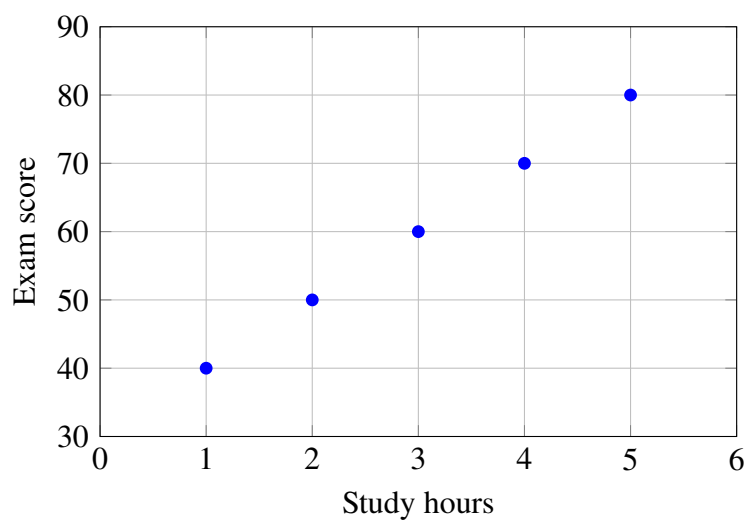


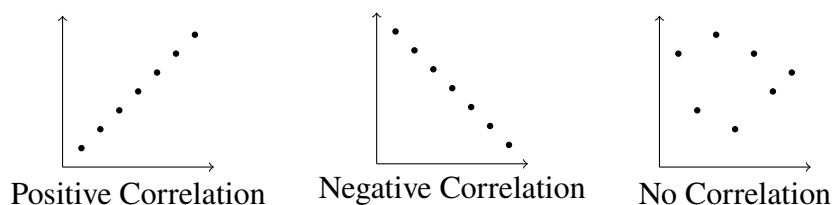
Figure 3.1: Scatter plot: Study hours vs. Exam score

Interpretation: There appears to be a **positive linear relationship** between study hours and exam score.

III. 1. Interpreting Scatter Plots

Scatter plots can reveal:

- **Direction:** positive (upward), negative (downward), or no relationship
- **Form:** linear, curvilinear, or no pattern
- **Strength:** how closely the points follow a pattern
- **Outliers:** points that deviate markedly from the pattern



IV. Covariance

IV. 1. Definition

Covariance measures the **direction** of the linear relationship between two variables.

Définition 5

The **covariance** between two quantitative variables X and Y is defined as:

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (3.1)$$

where $\bar{x}$ and $\bar{y}$ are the means of X and Y respectively.

Remark 7 (Computational Formula) *An equivalent formula that is easier for calculations:*

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} \quad (3.2)$$

IV. 2. Interpretation of Covariance

- $\text{Cov}(X, Y) > 0$: **Positive covariance** ? When X increases, Y tends to increase.
- $\text{Cov}(X, Y) < 0$: **Negative covariance** ? When X increases, Y tends to decrease.
- $\text{Cov}(X, Y) \approx 0$: **No linear relationship** between X and Y .

Remark 8 (Limitation of Covariance) *Covariance depends on the units of measurement. If we change units (e.g., from meters to centimeters), the covariance changes. This makes it difficult to compare covariances across different datasets.*

IV. 3. Properties of Covariance

Propriété 1 (Properties)

For any constants a, b, c, d :

1. $\text{Cov}(X, X) = V(X)$ (variance of X)
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ (symmetry)
3. $\text{Cov}(aX + b, cY + d) = ac \cdot \text{Cov}(X, Y)$
4. $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$

Example 31 (Calculating Covariance) *Using the study hours and exam score data (see 30):*

x_i	y_i	$x_i y_i$	x_i^2
1	40	40	1
2	50	100	4
3	60	180	9
4	70	280	16
5	80	400	25
Σ		1000	55

Step 1: Calculate the means.

$$\bar{x} = \frac{1 + 2 + 3 + 4 + 5}{5} = \frac{15}{5} = 3$$

$$\bar{y} = \frac{40 + 50 + 60 + 70 + 80}{5} = \frac{300}{5} = 60$$

Step 2: Compute $\sum x_i y_i$. (In the table above)

Step 3: Apply the computational formula.

$$\text{Cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}$$

$$\text{Cov}(X, Y) = \frac{1000}{5} - (3 \times 60) = 200 - 180 = 20$$

Interpretation: The positive covariance (20) indicates a positive linear relationship between study hours and exam scores.

V. Linear Correlation Coefficient

The **Pearson correlation coefficient** is a standardized measure of linear relationship that overcomes the unit-dependence problem of covariance.

Définition 6

The **correlation coefficient** r (or r_{XY}) is defined as:

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.3)$$

where σ_X and σ_Y are the standard deviations of X and Y .

V. 1. Properties of Correlation Coefficient

Propriété 2

The correlation coefficient r satisfies:

- $-1 \leq \rho \leq 1$
- $\rho = 1$: Perfect **positive linear** relationship.
- $\rho = -1$: Perfect **negative linear** relationship.
- $\rho = 0$: **No linear** relationship.
- ρ is **unitless** (independent of measurement units).
- ρ is **symmetric**: $\rho_{XY} = \rho_{YX}$.

V. 2. Interpreting Correlation Strength

Correlation coefficient ρ	Interpretation
$\rho = 1$	Perfect positive linear relationship
$0.7 < \rho < 1$	Strong positive relationship
$0 < \rho < 0.7$	Weak positive relationship
$\rho = 0$	No linear relationship
$-0.7 < \rho < 0$	Weak negative relationship
$-1 < \rho < -0.7$	Strong negative relationship
$\rho = -1$	Perfect negative linear relationship

V. 3. Examples of Correlation Patterns

Example 32 (Calculating Correlation) Using the previous example (see 30) with $Cov(X, Y) = 20$:

First, calculate variances:

$$\sigma_X^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{4 + 1 + 0 + 1 + 4}{5} = \frac{10}{5} = 2 \Rightarrow \sigma_X \approx 1.414$$

$$\sigma_Y^2 = \frac{\sum (y_i - \bar{y})^2}{n} = \frac{400 + 100 + 0 + 100 + 400}{5} = \frac{1000}{5} = 200 \Rightarrow \sigma_Y \approx 14.142$$

Then:

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{20}{1.414 \times 14.142} = \frac{20}{20} = 1$$

Interpretation: $\rho = 1$ indicates a perfect positive linear relationship between study hours and exam scores.

VI. Simple Linear Regression

While correlation measures the strength and direction of a linear relationship, **regression** allows us to **predict** one variable from the other.

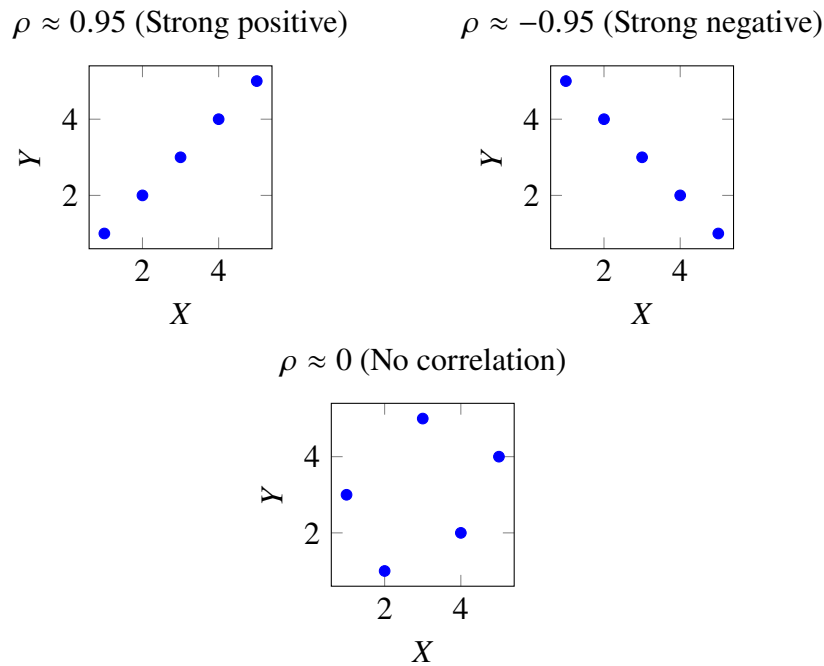


Figure 3.2: Examples of different correlation patterns

VI. 1. Method of Ordinary Least Squares (OLS)

Déf.7

The **least squares regression line** (or **line of best fit**) is the line $\hat{y} = ax + b$ that minimizes the sum of squared vertical distances between observed points and the line:

$$S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (ax_i + b))^2 \rightarrow \text{minimize}$$

VI. 2. Formulas for Regression Coefficients

Propriété 3 (Regression Coefficients)

The coefficients a (slope) and b (intercept) that minimize S are:

$$a = \frac{\text{Cov}(X, Y)}{V(X)} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (3.4)$$

$$b = \bar{y} - a\bar{x} \quad (3.5)$$

Remark 9 (Interpretation)

• **Slope a :** The expected change in Y for a one-unit increase in X .

• **Intercept b :** The expected value of Y when $X = 0$ (if $X = 0$ is meaningful).

Example 33 (Calculating Regression Line) Using the same data (see 30):

$$a = \frac{\text{Cov}(X, Y)}{V(X)} = \frac{20}{2} = 10$$

$$b = \bar{y} - a\bar{x} = 60 - 10 \times 3 = 60 - 30 = 30$$

The regression line is:

$$\hat{y} = 10x + 30$$

Interpretation:

- Slope $a = 10$: Each additional hour of study increases the expected exam score by 10 points.
- Intercept $b = 30$: A student who studies 0 hours is expected to score 30 points.

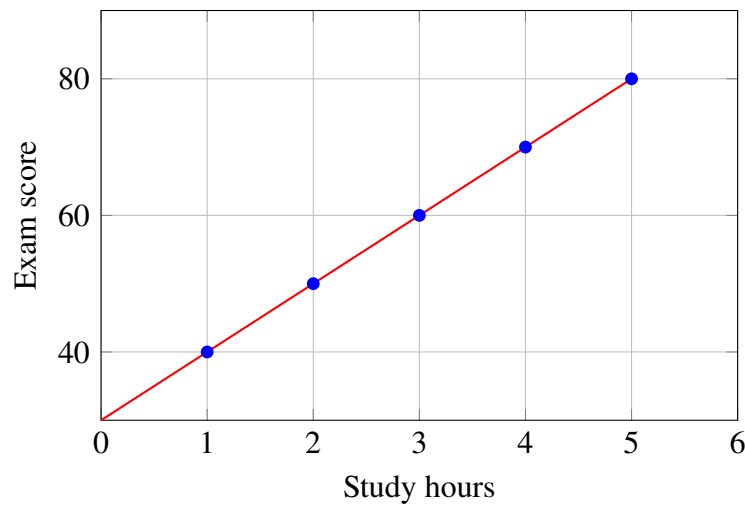


Figure 3.3: Scatter plot with regression line $\hat{y} = 10x + 30$

VI. 3. Prediction Using the Regression Line

Déf.8

For a new value x_0 , the predicted value of Y is:

$$\hat{y}_0 = ax_0 + b$$

Example 34 Using the regression line $\hat{y} = 10x + 30$:

- For a student who studies $x_0 = 3.5$ hours: $\hat{y}_0 = 10(3.5) + 30 = 65$
- For a student who studies $x_0 = 5.5$ hours: $\hat{y}_0 = 10(5.5) + 30 = 85$

Remark 10 (Caution in Prediction)

- **Interpolation** (predicting within the range of X) is generally reliable.
- **Extrapolation** (predicting outside the range of X) can be dangerous and unreliable.
- Correlation does not imply causation.

VII. Summary

This chapter covered methods for analyzing the relationship between two variables:

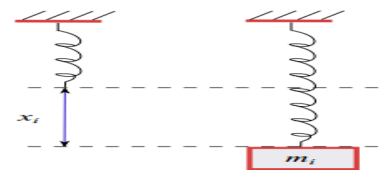
- **Contingency tables** for qualitative variables (marginal and conditional distributions).
- **Scatter plots** for visualizing bivariate quantitative data.
- **Covariance** to measure the direction of linear relationships.
- **Correlation coefficient** to measure both direction and strength (unitless, $-1 \leq r \leq 1$).
- **Simple linear regression** to predict Y from X using the least squares line $\hat{y} = ax + b$.

Exercises

Exercise 3.1.

During a physics practical lab, we have the following table and figure:

x_i (in m)	0	0.5	1.1	1.5	1.9
m_i (in kg)	0	10	20	30	40



Where " m_i " is the different applied masses, and " x_i " is the heights induced from the initial state.

1. Determine the regression equation for m as a function of x .
2. Plot the scatter plot associated with the double statistical series and draw the line m as a function of x .
3. Calculate the linear correlation coefficient, what do you observe?
4. Can we determine x if $m = 50$ kg ?

Solution: see page 90

Exercise 3.2.

A quality control engineer studies the relationship between the diameter (X in mm) and the tensile strength (Y in kN) of a steel cable:

Diameter X (mm)	2	4	6	8	10
Strength Y (kN)	5	12	20	30	42

1. Compute the covariance and the correlation coefficient.
2. Find the regression line of Y on X .
3. Interpret the slope in the context of this problem.
4. Predict the tensile strength for a cable with diameter 7 mm.
5. The manufacturer wants a cable with strength 50 kN. What diameter should be used?

Solution: see page 91

Exercise 3.3.

The following contingency table shows the distribution of 200 employees by **education level** and **satisfaction level**:

Education \ Satisfaction	Low	Medium	High	Total
High School	30	20	10	60
Bachelor	15	35	20	70
Master	5	15	50	70
Total	50	70	80	200

1. Calculate the marginal distributions.
2. Calculate the conditional distribution of satisfaction given education = Master.
3. Calculate the conditional distribution of education given satisfaction = High.
4. Interpret the results.

Solution: see page 92

Exercise 3.4.

An engineer wants to study the relationship between the outside temperature (X in $^{\circ}\text{C}$) and the energy consumption (Y in kWh) of an air conditioning system over 7 days:

Temperature $X (^{\circ}\text{C})$	22	24	26	28	30	32	34
Energy Y (kWh)	18	20	23	27	32	38	45

1. Calculate $\bar{x}$, $\bar{y}$, $V(X)$, $V(Y)$, and $\text{Cov}(X, Y)$.
2. Calculate the correlation coefficient r and interpret it.
3. Find the regression line $\hat{y} = ax + b$.
4. Predict the energy consumption when the temperature is 29°C .
5. Predict the energy consumption when the temperature is 40°C . Is this prediction reliable? Why?

Solution: see page 92

Additional Exercises

Exercise 3.5.

The following data gives the number of hours studied (X) and the final exam score (Y) for 6 students:

X (hours)	2	3	4	5	6	7
Y (score)	50	55	65	70	75	85

1. Draw the scatter plot.
2. Calculate $\bar{x}$, $\bar{y}$, and the covariance $\text{Cov}(X, Y)$.
3. Calculate the variances $V(X)$ and $V(Y)$.
4. Calculate the correlation coefficient r and interpret it.
5. Find the regression line $\hat{y} = ax + b$.
6. Predict the score for a student who studies 4.5 hours.

Solution: see page 90

Exercise 3.6.

Given the following bivariate data:

X	1	2	3	4	5
Y	5	4	3	2	1

1. Draw the scatter plot.
2. Calculate $\text{Cov}(X, Y)$ and r .
3. Find the regression line.
4. What do you notice about X and Y ?

Exercise 3.7.

The following contingency table shows the distribution of 300 engineers in a company by department and years of experience:

Department \ Experience	< 5 years	5 – 10 years	> 10 years	Total
Mechanical	25	35	20	80
Electrical	30	40	30	100
Civil	20	30	25	75
Computer	15	20	10	45
Total	90	125	85	300

1. Calculate the marginal distributions of department and experience.
2. Calculate the conditional distribution of experience given department = Electrical.
3. Calculate the conditional distribution of department given experience > 10 years.
4. Interpret the results: Which department has the most experienced engineers?

Exercise 3.8.

A production manager wants to study the relationship between the quantity produced (X in units) and the production cost (Y in 1000 DA) for 6 different production runs:

Quantity X (units)	10	20	30	40	50	60
Cost Y (1000 DA)	22	30	35	42	48	55

1. Draw the scatter plot.
2. Calculate the covariance and the correlation coefficient.
3. Find the regression line of Y on X .
4. Interpret the slope.
5. Predict the production cost for a quantity of 45 units.
6. Is the correlation strong? Justify your answer.

Exercise 3.9.

A study examines the relationship between advertising expenditure (X , in 1000 DA) and sales (Y , in 1000 DA) for 5 months:

X	10	12	15	18	20
Y	35	40	50	55	60

1. Compute the covariance and correlation coefficient.
2. Find the regression line.
3. Interpret the slope.
4. Predict sales when advertising expenditure is 16,000 DA.

Solution: see page 91

Additional Engineering Applications

Exercise 3.10.

A process engineer studies the relationship between temperature (in $^{\circ}\text{C}$) and reaction yield (in %):

Temperature ($^{\circ}\text{C}$)	30	40	50	60	70	80
Yield (%)	45	52	58	65	70	75

1. Draw the scatter plot.
2. Calculate the covariance and correlation coefficient.
3. Find the regression line of yield on temperature.
4. Predict the yield at 55°C .

Exercise 3.11.

A civil engineer studies the relationship between concrete curing time (in days) and compressive strength (in MPa):

Curing time (days)	3	7	14	21	28
Strength (MPa)	18	25	32	38	42

1. Calculate the covariance and correlation coefficient.
2. Find the regression line of strength on curing time.
3. Predict the strength at 30 days.
4. Interpret the slope in this context.

Exercise 3.12.

An electrical engineer studies the relationship between voltage (in V) and current (in A) in a circuit:

Voltage (V)	5	10	15	20	25
Current (A)	0.5	1.0	1.5	2.0	2.5

1. Draw the scatter plot.
2. Calculate the covariance and correlation coefficient.
3. Find the regression line of current on voltage.
4. What physical law does this illustrate?
5. Predict the current for a voltage of 18 V.

Part II

PROBABILITY

4 Combinatorial Analysis

Contents of this chapter

I.	Introduction	50
II.	Arrangements with Repetition (p -tuple)	50
III.	Arrangements without Repetition	50
IV.	Combinations without Repetition	51
V.	Combinations with Repetition	51
VI.	Binomial Theorem	52
VII.	Summary	52

I. Introduction

Déf.1

Combinatorial analysis is the branch of mathematics that studies the counting of possible arrangements of objects according to specific rules.

Remark 11 *Combinatorics is fundamental to probability theory because to calculate probabilities using the classical approach:*

$$P(A) = \frac{\text{Number of favorable outcomes}}{\text{Total number of possible outcomes}}$$

we need to be able to count outcomes efficiently.

II. Arrangements with Repetition (p -tuple)

Déf.2

An **arrangement with repetition** of p elements chosen from n distinct elements is an ordered p -tuple where each element can be chosen from the n elements, and repetition is allowed.

The number of such arrangements is:

$$n^p$$

Example 35 *How many 4-digit codes can be formed using the digits 0-9?*

$$n^p = 10^4 = 10,000 \text{ codes}$$

Example 36 *How many 3-letter words can be formed using the letters A, B, C (repetition allowed)?*

$$n^p = 3^3 = 27 \text{ words}$$

III. Arrangements without Repetition

Déf.3

An **arrangement without repetition** (or permutation of p elements from n) is an ordered selection of p distinct elements chosen from n distinct elements, where each element can appear at most once.

The number of such arrangements is:

$$A_n^p = \frac{n!}{(n-p)!} = n(n-1)(n-2) \cdots (n-p+1)$$

where $n! = n \times (n-1) \times (n-2) \times \cdots \times 2 \times 1$, and $0! = 1$.

Example 37 *In a race of 10 horses, how many ways to award gold, silver, and bronze medals?*

$$A_{10}^3 = \frac{10!}{(10-3)!} = 10 \times 9 \times 8 = 720 \text{ ways}$$

Déf.4

A **permutation** of all n distinct elements is an arrangement of all n elements in a specific order.

$$P_n = A_n^n = n!$$

Example 38 How many ways can 5 books be arranged on a shelf?

$$P_5 = 5! = 5 \times 4 \times 3 \times 2 \times 1 = 120 \text{ ways}$$

IV. Combinations without Repetition

Déf.5

A **combination without repetition** of p elements chosen from n distinct elements is an **unordered** selection of p distinct elements.

The number of such combinations is:

$$C_n^p = \binom{n}{p} = \frac{n!}{p!(n-p)!}$$

Remark 12 In combinations, order does NOT matter. For example, selecting $\{A, B, C\}$ is the same as $\{C, B, A\}$.

Propriété 1 (Properties of Combinations)

1. $C_n^0 = C_n^n = 1$
2. $C_n^1 = n$
3. $C_n^p = C_n^{n-p}$
4. $C_n^p = C_{n-1}^{p-1} + C_{n-1}^p$ (Pascal's rule)

Example 39 How many ways can a committee of 3 be chosen from 10 people?

$$C_{10}^3 = \binom{10}{3} = \frac{10!}{3! \times 7!} = \frac{10 \times 9 \times 8}{3 \times 2 \times 1} = \frac{720}{6} = 120 \text{ committees}$$

V. Combinations with Repetition

Déf.6

A **combination with repetition** of p elements chosen from n distinct elements is an **unordered** selection where repetition is allowed.

The number of such combinations is:

$$C_{n+p-1}^p = \binom{n+p-1}{p}$$

Example 40 How many ways to choose 4 donuts from 5 types (chocolate, glazed, sprinkle, cream, jelly), where repetition is allowed?

$$C_{n+p-1}^p = C_{5+4-1}^4 = C_8^4 = \frac{8!}{4! \times 4!} = 70 \text{ ways}$$

VI. Binomial Theorem

Théorème 1 (Binomial Theorem)

For any real numbers a and b , and any positive integer n :

$$(a + b)^n = \sum_{p=0}^n C_n^p a^{n-p} b^p$$

Example 41 For $n = 3$:

$$(a + b)^3 = C_3^0 a^3 b^0 + C_3^1 a^2 b^1 + C_3^2 a^1 b^2 + C_3^3 a^0 b^3$$

$$(a + b)^3 = a^3 + 3a^2 b + 3ab^2 + b^3$$

VII. Summary

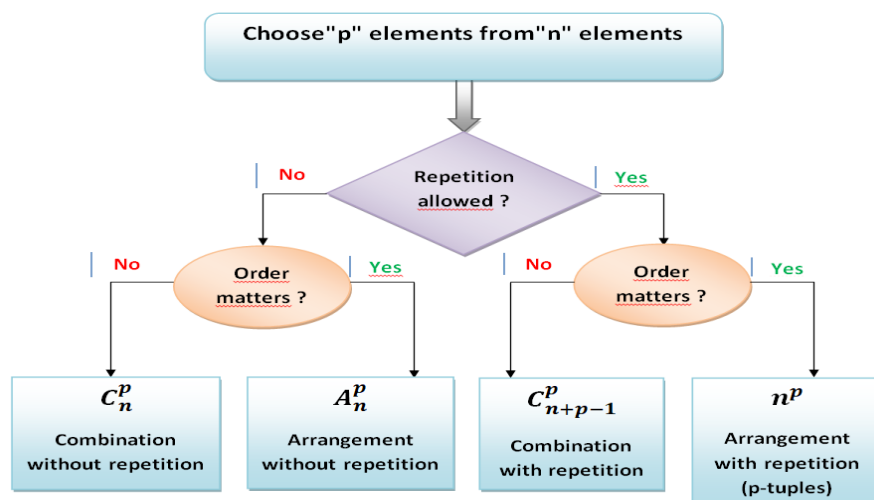


Figure 4.1: Combinatorial Analysis Diagram

Exercises

Exercise 4.1.

How many different 3-digit numbers can be formed using the digits 1, 2, 3, 4, 5 if:

1. Repetition is allowed?
2. Repetition is not allowed?

Solution: see page 94

Exercise 4.2.

In how many ways can 8 people be seated in a row of 8 chairs?

Solution: see page 94

Exercise 4.3.

An urn contains 12 balls: 3 red, 4 blue, and 5 yellow. Three balls are drawn simultaneously at random.

1. How many possible outcomes are there?
2. How many outcomes give exactly 3 red balls?
3. How many outcomes give one ball of each color?

Solution: see page 94

Exercise 4.4.

To form a code, three non-zero digits are randomly selected. How many distinct codes can be formed?

Solution: see page 95

Exercise 4.5.

An urn contains 10 balls numbered from 0 to 9. Successively and without replacement, 4 balls are drawn. Find the number of possible outcomes.

Solution: see page 95

Exercise 4.6.

We want to arrange 4 mathematics books, 3 physics books, and 5 chemistry books in 12 slots. We want to place the books of each category next to each other. How many possible arrangements do we have?

Solution: see page 95

Additional Exercises

Exercise 4.7.

A committee of 4 is to be selected from 12 people. How many different committees are possible?

Exercise 4.8.

How many ways can 3 letters be chosen from the word "MATHEMATICS" with repetition allowed?

Exercise 4.9.

Expand $(x + y)^5$ using the binomial theorem.

Exercise 4.10.

The oral part of a competition consists of a total of 100 topics; candidates randomly draw three subjects and then choose the topic to be addressed from these three. A candidate presents themselves having prepared 60 topics out of the 100. How many different draws are possible?

Additional Engineering Applications

Exercise 4.11.

A telecommunications system transmits a 6-bit digital signal. Each bit can be 0 or 1.

1. How many different signals can be transmitted?
2. How many signals contain exactly three 1s?
3. How many signals contain at least four 1s?

Exercise 4.12.

A mechanical engineering workshop has 8 different machines. In how many ways can 3 machines be selected and arranged in a production line if:

1. Order matters?
2. Order does not matter?

Exercise 4.13.

A hydraulic system has 6 pumps. In how many ways can 4 pumps be selected to operate simultaneously?

Exercise 4.14.

An automation engineer needs to test 5 different control algorithms. If each algorithm can be tested in 3 different modes, how many combinations of algorithm and mode are possible?

5 Probability Theory and Conditional Probability

Contents of this chapter

I.	Introduction	56
II.	Random Experiments and Sample Spaces	56
III.	Events and Event Operations	56
IV.	Algebra of Events (σ -algebra)	57
V.	Definitions of Probability	57
	V. 1. Classical Approach (Laplace, 1812)	57
	V. 2. Frequentist Approach (Richard von Mises, 1919)	57
	V. 3. Axiomatic Approach (Kolmogorov, 1933).....	58
VI.	Probability Spaces	58
VII.	General Probability Theorems	58
VIII.	Conditional Probability	59
IX.	Independence of Events	59
X.	Bayes' Theorem	59
XI.	Summary	60

I. Introduction

Déf.1

Probability is a measure of the likelihood that an event will occur. It is a number between 0 and 1 $0 \leq P \leq 1$, where:

- ▶ 0 indicates impossibility
- ▶ 1 indicates certainty

II. Random Experiments and Sample Spaces

Déf.2

A **random experiment** is a process that:

- 1 Can be repeated under the same conditions
- 2 Has two or more possible outcomes
- 3 The outcome cannot be predicted with certainty in advance

Déf.3

The **sample space** Ω is the set of all possible outcomes of a random experiment.

Example 42 • Tossing a coin: $\Omega = \{H, T\}$ ($H=Heads$; $T=Tails$)

- Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Tossing two coins: $\Omega = \{HH, HT, TH, TT\}$

III. Events and Event Operations

Déf.4

An **event** is a subset of the sample space Ω .

Déf.5

For events A and B :

- ▶ **Union:** $A \cup B$ occurs if A or B occurs
- ▶ **Intersection:** $A \cap B$ occurs if both A and B occur
- ▶ **Complement:** $\bar{A}$ (or A^c) occurs if A does not occur
- ▶ **Difference:** $A \setminus B$ occurs if A occurs but B does not

Example 43 Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$

- $A = \{2, 4, 6\}$ (even numbers)
- $B = \{1, 2, 3\}$ (numbers less than 4)
- $A \cup B = \{1, 2, 3, 4, 6\}$
- $A \cap B = \{2\}$
- $\bar{A} = \{1, 3, 5\}$

IV. Algebra of Events (σ -algebra)

Déf.6 A collection $\mathcal{A}$ of subsets of Ω is called a σ -algebra if:

- 1 $\Omega \in \mathcal{A}$
- 2 If $A \in \mathcal{A}$, then $\bar{A} \in \mathcal{A}$
- 3 If $A_1, A_2, \dots \in \mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

V. Definitions of Probability

V. 1. Classical Approach (Laplace, 1812)

Déf.7 If all outcomes in a finite sample space Ω are equally likely, then the probability of an event A is:

$$P(A) = \frac{|A|}{|\Omega|}$$

Example 44 Rolling a fair die: $P(\text{even}) = \frac{3}{6} = 0.5$

V. 2. Frequentist Approach (Richard von Mises, 1919)

Déf.8 The probability of an event A is the limit of its relative frequency:

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}$$

V. 3. Axiomatic Approach (Kolmogorov, 1933)

Déf.9

A probability P satisfies:

- 1 $0 \leq P(A) \leq 1$
- 2 $P(\Omega) = 1$
- 3 For mutually exclusive events $A_1, A_2, \dots$:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

VI. Probability Spaces

Déf.10

A **probability space** is a triple $(\Omega, \mathcal{A}, P)$ where:

- Ω is the sample space
- $\mathcal{A}$ is a σ -algebra on Ω
- P is a probability measure

VII. General Probability Theorems

Propriété 1 (Probability of the Complement)

$$P(\overline{A}) = 1 - P(A)$$

Propriété 2 (Probability of Union (Inclusion-Exclusion))

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Propriété 3 (Monotonicity)

If $A \subseteq B$, then $P(A) \leq P(B)$.

Example 45 Given $P(A) = 0.4$, $P(B) = 0.5$, $P(A \cap B) = 0.2$:

$$P(A \cup B) = 0.4 + 0.5 - 0.2 = 0.7$$

$$P(\overline{A}) = 1 - 0.4 = 0.6$$

VIII. Conditional Probability

Déf.11

The **conditional probability** of A given B (with $P(B) > 0$) is:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

Example 46 Rolling a die: $A = \{2, 4, 6\}$ (even), $B = \{1, 2, 3\}$ (less than 4)

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2\})}{P(\{1, 2, 3\})} = \frac{1/6}{3/6} = \frac{1}{3}$$

Propriété 4 (Multiplication Rule)

$$P(A \cap B) = P(A | B)P(B) = P(B | A)P(A)$$

IX. Independence of Events

Déf.12

Two events A and B are **independent** if:

$$P(A \cap B) = P(A)P(B)$$

Remark 13 Independence is different from mutual exclusivity. If A and B are mutually exclusive with $P(A) > 0$, $P(B) > 0$, then they cannot be independent.

Example 47 Tossing a coin twice:

- A : First toss is heads, B : Second toss is heads
- $P(A) = 0.5$, $P(B) = 0.5$, $P(A \cap B) = 0.25$
- Since $0.25 = 0.5 \times 0.5$, A and B are independent.

X. Bayes' Theorem

Théorème 1 (Law of Total Probability)

If $B_1, B_2, \dots, B_n$ form a **partition** of Ω (mutually exclusive and $\bigcup_{i=1}^n B_i = \Omega$), then for any event A :

$$P(A) = \sum_{i=1}^n P(A | B_i)P(B_i)$$

Théorème 2 (Bayes' Theorem)

For a partition $B_1, B_2, \dots, B_n$ and an event A with $P(A) > 0$:

$$P(B_i | A) = \frac{P(A | B_i)P(B_i)}{\sum_{j=1}^n P(A | B_j)P(B_j)}$$

Example 48 A factory has two machines:

- Machine 1 produces 60% of items, with 2% defective
- Machine 2 produces 40% of items, with 5% defective

If an item is found defective, what is the probability it came from Machine 1?

Let B_1 = from Machine 1, B_2 = from Machine 2, A = defective.

$$P(B_1 | A) = \frac{0.02 \times 0.6}{0.02 \times 0.6 + 0.05 \times 0.4} = \frac{0.012}{0.012 + 0.02} = \frac{0.012}{0.032} = 0.375$$

XI. Summary

Concept	Formula
Classical Probability	$P(A) = \frac{ A }{ \Omega }$
Complement	$P(\bar{A}) = 1 - P(A)$
Union	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Conditional Probability	$P(A B) = \frac{P(A \cap B)}{P(B)}$
Multiplication Rule	$P(A \cap B) = P(A B)P(B)$
Independence	$P(A \cap B) = P(A)P(B)$
Law of Total Probability	$P(A) = \sum P(A B_i)P(B_i)$
Bayes' Theorem	$P(B_i A) = \frac{P(A B_i)P(B_i)}{\sum P(A B_j)P(B_j)}$

Exercises

Exercise 5.1.

A card is drawn from a standard deck of 52 cards.

1. What is the probability it is a heart?
2. What is the probability it is a king?
3. What is $P(\text{heart} | \text{king})$?

Solution: see page 97

Exercise 5.2.

Two dice are rolled. Find the probability that the sum is 7.

Solution: see page 97

Exercise 5.3.

Let A and B be two events such that:

$$P(A) = 0.4, \quad P(B) = 0.6, \quad P(A \cap B) = 0.3$$

1. Calculate $P(A \cup B)$.
2. Calculate $P(\overline{A} \cap B)$.
3. Calculate $P(A \cap \overline{B})$.
4. Are A and B independent? Justify.

Solution: see page 98

Exercise 5.4.

A chemical analysis laboratory receives a batch of test tubes. These tubes are supplied by three different companies, A , B , and C , in the following proportions: 50%, 30% and 20% respectively. 2% of the tubes manufactured by A , 3% of those manufactured by B , and 4% of those manufactured by C have defects. From the batch that was received, a test tube is selected at random.

1. What is the probability that it is defective?
2. Knowing that the chosen tube is defective, what is the probability that it comes from company A ?

Solution: see page 99

Exercise 5.5.

Let A and B be two events from the same probability space (Ω, F, P) , such that:

$$P(A) = \frac{2}{3}, \quad P(B) = \frac{1}{3}, \quad P(\overline{A} | B) = \frac{1}{4}$$

1. Calculate the value of the conditional probability of A given B .
2. Calculate the value of the conditional probability of B given A .
3. What is the probability that exactly one of the two events occurs?

Solution: see page 100

Additional Exercises

Exercise 5.6.

A bag contains 3 red balls and 5 blue balls. Two balls are drawn without replacement.

1. Find the probability both are red.
2. Find the probability the second is red given the first is blue.

Exercise 5.7.

A test for a disease has 99% sensitivity and 95% specificity. The disease affects 1% of the population. If a person tests positive, what is the probability they actually have the disease? (Use Bayes' theorem)

Exercise 5.8.

The oral part of a competition consists of a total of 100 topics; candidates randomly draw three subjects and then choose the topic to be addressed from these three. A candidate presents themselves having prepared 60 topics out of the 100.

What is the probability that the candidate has revised:

- a) none of the three topics drawn.
- b) one topic out of the three drawn.
- c) at least two topics out of the three drawn.

Exercise 5.9.

An electronics factory has three production lines: Line 1 produces 50% of the total output, Line 2 produces 30%, and Line 3 produces 20%. The defective rates are: 1% for Line 1, 2% for Line 2, and 3% for Line 3. A product is selected at random from the total output and is found to be defective.

1. What is the probability that the defective product came from Line 1?
2. What is the probability that the defective product came from Line 2?
3. What is the probability that the defective product came from Line 3?

Additional Engineering Applications

Exercise 5.10.

A factory has three production lines A, B, and C producing 45%, 30%, and 25% of the total output, respectively. The defective rates are 2%, 3%, and 5% respectively.

1. What is the probability that a randomly selected product is defective?
2. If a product is found defective, what is the probability it came from line A?
3. If a product is found defective, what is the probability it came from line C?

Exercise 5.11.

A quality control test for electronic components shows:

- 95% of components pass the first test
- 90% of components pass both tests
- 60% of components pass the first test given they pass the second test

1. What is the probability that a component passes the second test?
2. Are the two tests independent? Justify.

Exercise 5.12.

An automation system has three sensors. The probabilities of failure are:

- Sensor 1: 0.05
- Sensor 2: 0.04
- Sensor 3: 0.06

1. What is the probability that at least one sensor fails?
2. What is the probability that all sensors work correctly?

6 Random Variables and Usual Probability Distributions

Contents of this chapter

I.	Introduction	65
II.	Definition of a Random Variable	65
III.	Discrete Random Variables	65
III. 1.	Probability Mass Function (PMF)	65
III. 2.	Cumulative Distribution Function (CDF)	66
III. 3.	Expectation (Mean)	66
III. 4.	Variance and Standard Deviation	66
IV.	Continuous Random Variables	67
IV. 1.	Probability Density Function (PDF)	68
IV. 2.	Cumulative Distribution Function (CDF)	68
IV. 3.	Properties of the CDF	68
IV. 4.	Expectation and Variance for Continuous Variables	69
V.	Discrete Probability Distributions	70
V. 1.	Bernoulli Distribution	70
V. 2.	Binomial Distribution	70
V. 3.	Poisson Distribution	71
VI.	Continuous Probability Distributions	71
VI. 1.	Uniform Distribution	71
VI. 2.	Exponential Distribution	72
VI. 3.	Normal Distribution	73
VI. 4.	Standard Normal Distribution	73
VI. 5.	Standardization (Z-score)	73
VII.	Summary	74

I. Introduction

In Chapter 5, we studied probability spaces and events. Now we move from **events** to **numerical values** associated with these events.

A **random variable** is a function that assigns a numerical value to each outcome of a random experiment. For example:

- The number of heads when tossing two coins: $X \in \{0, 1, 2\}$
- The face value when rolling a die: $X \in \{1, 2, 3, 4, 5, 6\}$
- The height of a randomly selected student: $X \in [0, \infty)$

This chapter introduces: **Discrete random variables** finite or countably infinite values, **Continuous random variables** (uncountably infinite values), **Probability distributions** and **cumulative distribution functions**, calculation of **Expectation** and **variance**, and **Usual probability distributions** (Bernoulli, Binomial, Poisson, Uniform, Normal, Exponential).

II. Definition of a Random Variable

Déf.1

A **random variable** is a function $X : \Omega \rightarrow \mathbb{R}$ that assigns a real number to each outcome ω in the sample space Ω .

Example 49 • Tossing two coins: $X = \text{number of heads} \in \{0, 1, 2\}$

- Rolling a die: $X = \text{face value} \in \{1, 2, 3, 4, 5, 6\}$
- Measuring height: $X = \text{height in cm} \in [0, \infty)$

Remark 14 We distinguish two types:

- **Discrete random variable:** Takes a finite or countably infinite number of values.
- **Continuous random variable:** Takes an uncountably infinite number of values (an interval).

III. Discrete Random Variables

III. 1. Probability Mass Function (PMF)

Déf.2

For a discrete random variable X , the **probability mass function** (or probability distribution) is:

$$p(x) = P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

with properties:

1 $p(x) \geq 0$ for all x

2 $\sum_x p(x) = 1$

The probability distribution of a discrete random variable is often presented in a table:

x	x_1	x_2	$\cdots$	x_k
$P(X = x)$	p_1	p_2	$\cdots$	p_k

where $p_i = P(X = x_i)$ and $\sum_{i=1}^k p_i = 1$.

III. 2. Cumulative Distribution Function (CDF)

Déf.3

The **cumulative distribution function** is:

$$F(x) = P(X \leq x) = \sum_{t \leq x} p(t)$$

Properties of the CDF:

1. $F(x)$ is a non-decreasing function.
2. $0 \leq F(x) \leq 1$ for all x .
3. $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow +\infty} F(x) = 1$.
4. $F(x)$ is a step function (constant between values, with jumps at the possible values of X).

General expression of the CDF:

$$F(x) = \begin{cases} 0, & \text{if } x < x_1 \\ p_1, & \text{if } x_1 \leq x < x_2 \\ p_1 + p_2, & \text{if } x_2 \leq x < x_3 \\ \vdots & \vdots \\ 1, & \text{if } x \geq x_n \end{cases}$$

III. 3. Expectation (Mean)

Déf.4

The **expectation** (or mean) of a discrete random variable X is:

$$E(X) = \sum_x x \cdot p(x)$$

III. 4. Variance and Standard Deviation

Déf.5

$$V(X) = E[(X - E(X))^2] = \sum_x (x - \mu)^2 p(x) = E(X^2) - [E(X)]^2$$

$$\sigma_X = \sqrt{V(X)}$$

Propriété 1 (Properties)

1. $E(aX + b) = aE(X) + b$
2. $V(aX + b) = a^2V(X)$

Illustrative Example

Let X be a discrete random variable with the following probability distribution:

x	1	2	3
$P(X = x)$	0.2	0.5	0.3

1. Probability Mass Function (PMF):

$$p(1) = 0.2, \quad p(2) = 0.5, \quad p(3) = 0.3$$

2. Cumulative Distribution Function (CDF):

$$F(x) = \begin{cases} 0, & x < 1 \\ 0.2, & 1 \leq x < 2 \\ 0.7, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

3. Expectation (Mean):

$$E(X) = \sum x \cdot p(x) = 1 \times 0.2 + 2 \times 0.5 + 3 \times 0.3 = 0.2 + 1.0 + 0.9 = 2.1$$

4. Variance:

$$E(X^2) = \sum x^2 \cdot p(x) = 1^2 \times 0.2 + 2^2 \times 0.5 + 3^2 \times 0.3 = 0.2 + 2.0 + 2.7 = 4.9$$

$$V(X) = E(X^2) - [E(X)]^2 = 4.9 - (2.1)^2 = 4.9 - 4.41 = 0.49$$

5. Standard Deviation:

$$\sigma_X = \sqrt{V(X)} = \sqrt{0.49} = 0.7$$

6. Interpretation:

- The expected value is $E(X) = 2.1$, which is the average value of X over many trials.
- The variance $V(X) = 0.49$ measures the spread of the distribution around the mean.
- The standard deviation $\sigma_X = 0.7$ is the square root of the variance.

IV. Continuous Random Variables

IV. 1. Probability Density Function (PDF)

Déf.6

For a continuous random variable X , there exists a function $f(x)$ called the **probability density function** such that:

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

with properties:

1 $f(x) \geq 0$ for all x

2 $P(a \leq X \leq b) = \int_a^b f(x) dx$

IV. 2. Cumulative Distribution Function (CDF)

Déf.7

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

General expression of the CDF:

$$F(x) = \begin{cases} 0, & \text{if } x < a \\ \int_a^x f(t) dt, & \text{if } a \leq x \leq b \\ 1, & \text{if } x > b \end{cases}$$

where $[a, b]$ is the interval where $f(x) > 0$.

Properties of the CDF (same as discrete case):

1. $F(x)$ is a non-decreasing function.
2. $0 \leq F(x) \leq 1$ for all x .
3. $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow +\infty} F(x) = 1$.

IV. 3. Properties of the CDF

Propriété 2 (Important Properties)

1. $P(a < X < b) = F(b) - F(a)$
2. $P(X > a) = 1 - F(a)$
3. $F'(x) = f(x)$ (where derivative exists)
4. F is non-decreasing: $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$
5. $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow +\infty} F(x) = 1$

IV. 4. Expectation and Variance for Continuous Variables

Définition 8

For a continuous random variable with PDF $f(x)$:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$$

$$V(X) = E(X^2) - [E(X)]^2$$

Illustrative Exercise

Let X be a continuous random variable with probability density function:

$$f(x) = \begin{cases} 2x, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

1. Show that $f(x)$ is a valid probability density function.
2. Find the cumulative distribution function $F(x)$.
3. Calculate $P(0.2 < X < 0.6)$.
4. Calculate $E(X)$ and $V(X)$.

Solution:

1. To show that $f(x)$ is a valid PDF, we must verify:

- $f(x) \geq 0$ for all x : Yes, since $2x \geq 0$ on $[0, 1]$.
- $\int_{-\infty}^{\infty} f(x) dx = 1$:

$$\int_0^1 2x dx = [x^2]_0^1 = 1 - 0 = 1$$

Therefore, $f(x)$ is a valid probability density function.

2. The cumulative distribution function $F(x)$:

- For $x < 0$: $F(x) = 0$
- For $0 \leq x \leq 1$:

$$F(x) = \int_0^x 2t dt = [t^2]_0^x = x^2$$

- For $x > 1$: $F(x) = 1$

$$F(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

3. $P(0.2 < X < 0.6) = F(0.6) - F(0.2)$:

$$F(0.6) = (0.6)^2 = 0.36$$

$$F(0.2) = (0.2)^2 = 0.04$$

$$P(0.2 < X < 0.6) = 0.36 - 0.04 = 0.32$$

4. $E(X)$ and $V(X)$:

$$E(X) = \int_0^1 x \cdot 2x \, dx = \int_0^1 2x^2 \, dx = \left[\frac{2x^3}{3} \right]_0^1 = \frac{2}{3} \approx 0.6667$$

$$E(X^2) = \int_0^1 x^2 \cdot 2x \, dx = \int_0^1 2x^3 \, dx = \left[\frac{2x^4}{4} \right]_0^1 = \left[\frac{x^4}{2} \right]_0^1 = \frac{1}{2} = 0.5$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{1}{2} - \left(\frac{2}{3} \right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{9}{18} - \frac{8}{18} = \frac{1}{18} \approx 0.0556$$

$$\boxed{E(X) = \frac{2}{3}}, \quad \boxed{V(X) = \frac{1}{18}}$$

V. Discrete Probability Distributions

V.1. Bernoulli Distribution

Déf.9

A random variable X follows a **Bernoulli distribution** with parameter $p \in [0, 1]$ if:

$$P(X = 1) = p, \quad P(X = 0) = 1 - p = q$$

Notation: $X \sim \text{Bernoulli}(p)$

Propriété 3 (Properties)

$$E(X) = p, \quad V(X) = p(1 - p) = pq$$

Example 50 Tossing a fair coin: $X = 1$ for heads, $X = 0$ for tails.

$$P(X = 1) = 0.5, \quad E(X) = 0.5, \quad V(X) = 0.25$$

V.2. Binomial Distribution

Déf.10

X follows a **binomial distribution** if it counts the number of successes in n independent Bernoulli trials with success probability p :

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n$$

Notation: $X \sim \mathcal{B}(n, p)$ or $X \sim \text{Binomial}(n, p)$

Propriété 4 (Properties)

$$E(X) = np, \quad V(X) = np(1 - p)$$

Example 51 Tossing a fair coin 10 times. Let X = number of heads. Then $X \sim \mathcal{B}(10, 0.5)$:

$$E(X) = 10 \times 0.5 = 5, \quad V(X) = 10 \times 0.5 \times 0.5 = 2.5$$

$$P(X = 5) = C_{10}^5 (0.5)^5 (0.5)^5 = 252 \times 0.03125 \times 0.03125 \approx 0.246$$

V. 3. Poisson Distribution

Déf.11

X follows a **Poisson distribution** with parameter $\lambda > 0$ if:

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Notation: $X \sim \text{Poisson}(\lambda)$

Propriété 5 (Properties)

$$E(X) = \lambda, \quad V(X) = \lambda$$

Remark 15 (Application) The Poisson distribution models:

- Number of emails received per hour
- Number of accidents per day
- Number of phone calls per minute

Example 52 If a call center receives an average of 3 calls per minute ($\lambda = 3$):

$$P(X = 2) = e^{-3} \frac{3^2}{2!} = e^{-3} \times \frac{9}{2} \approx 0.0498 \times 4.5 \approx 0.224$$

VI. Continuous Probability Distributions

VI. 1. Uniform Distribution

Déf.12

X follows a **uniform distribution** on $[a, b]$ if:

$$f(x) = \frac{1}{b - a}, \quad a \leq x \leq b$$

Notation: $X \sim \mathcal{U}(a, b)$

Propriété 6 (Properties)

$$F(x) = \frac{x-a}{b-a}, \quad a \leq x \leq b$$
$$E(X) = \frac{a+b}{2}, \quad V(X) = \frac{(b-a)^2}{12}$$

Example 53 A bus arrives every 10 minutes. The waiting time $X \sim \mathcal{U}(0, 10)$:

$$P(X \leq 3) = F(3) = \frac{3-0}{10-0} = 0.3$$

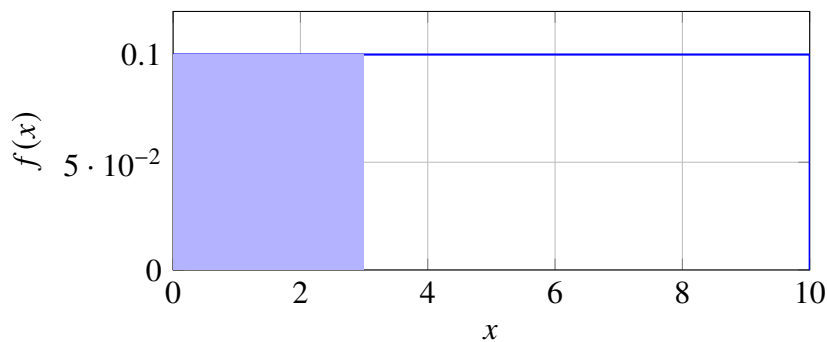


Figure 6.1: Uniform distribution on $[0, 10]$ with shaded area $P(X \leq 3)$

VI. 2. Exponential Distribution

Déf.13

X follows an **exponential distribution** with parameter $\lambda > 0$ if:

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

Notation: $X \sim \text{Exp}(\lambda)$

Propriété 7 (Properties)

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0$$
$$E(X) = \frac{1}{\lambda}, \quad V(X) = \frac{1}{\lambda^2}$$

Example 54 The lifetime of a component follows an exponential distribution with $\lambda = 0.1$ (mean 10 hours):

$$P(X > 15) = e^{-0.1 \times 15} = e^{-1.5} \approx 0.223$$

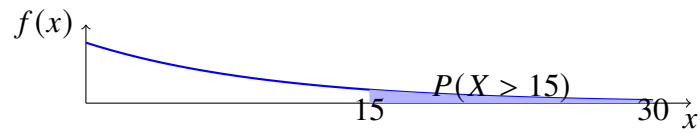


Figure 6.2: Exponential distribution with shaded area $P(X > 15)$

VI. 3. Normal Distribution

Déf.14

X follows a **normal distribution** with mean μ and variance σ^2 if:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R}$$

Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$

Propriété 8 (Properties)

$$E(X) = \mu, \quad V(X) = \sigma^2$$

The normal distribution is symmetric and bell-shaped.

VI. 4. Standard Normal Distribution

Déf.15

The **standard normal distribution** has $\mu = 0$ and $\sigma = 1$, denoted $Z \sim \mathcal{N}(0, 1)$. Its PDF and CDF are:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad \Phi(z) = \int_{-\infty}^z \phi(t) dt$$

VI. 5. Standardization (Z-score)

Propriété 9 (Standardization)

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then:

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Propriété 10 (Computing Probabilities)

$$\begin{aligned}P(X \leq x) &= P\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right) \\P(a < X < b) &= P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right) \\P(X > a) &= 1 - P(X < a) = 1 - \Phi\left(\frac{a - \mu}{\sigma}\right)\end{aligned}$$

Example 55 Suppose $X \sim \mathcal{N}(10, 4)$. Calculate $P(X \leq 12)$:

$$Z = \frac{12 - 10}{2} = 1, \quad P(X \leq 12) = \Phi(1) \approx 0.8413$$

Calculate $P(8 < X < 12)$:

$$\begin{aligned}P(8 < X < 12) &= P\left(\frac{8 - 10}{2} < \frac{X - 10}{2} < \frac{12 - 10}{2}\right) = P(-1 < Z < 1) \\P(-1 < Z < 1) &= \Phi(1) - \Phi(-1) \\&= 0.8413 - (1 - 0.8413) = 0.8413 - 0.1587 = 0.6826\end{aligned}$$

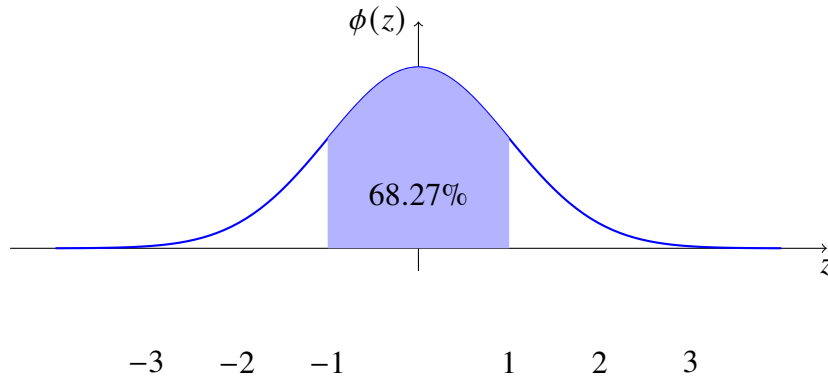


Figure 6.3: Standard normal distribution with area between -1 and 1

Remark 16 (Standard Normal Table) The values of $\Phi(z)$ for different z are tabulated in **Appendix B** (see page 106). To use the table:

1. Compute $Z = \frac{X - \mu}{\sigma}$
2. Look up $\Phi(Z)$ in the standard normal table in Appendix B
3. Use $\Phi(-z) = 1 - \Phi(z)$ for negative values

VII. Summary

We can summarize the main discrete and continuous probability distributions covered in this chapter in the following table:

Distribution	Parameters	$E(X)$	$V(X)$
Bernoulli	$p \in [0, 1]$	p	$p(1 - p)$
Binomial	$n \in \mathbb{N}, p \in [0, 1]$	np	$np(1 - p)$
Poisson	$\lambda > 0$	λ	λ
Uniform	$a < b$	$\frac{a + b}{2}$	$\frac{(b - a)^2}{12}$
Normal	$\mu \in \mathbb{R}, \sigma^2 > 0$	μ	σ^2
Exponential	$\lambda > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

- The Bernoulli distribution is a special case of the Binomial distribution with $n = 1$.
- The Poisson distribution is often used as an approximation of the Binomial distribution when n is large and p is small ($np = \lambda$).
- The Normal distribution is the most important distribution in statistics due to the Central Limit Theorem.
- The Exponential distribution is commonly used to model waiting times and lifetimes.

Exercises

Exercise 6.1.

A multiple-choice test has 20 questions, each with 4 options. If a student guesses randomly:

1. What is the probability of getting exactly 5 correct?
2. What is the expected number of correct answers?
3. What is the variance?

Solution: see page 101

Exercise 6.2.

A bookstore sells an average of 2.5 novels per hour.

1. Find the probability that exactly 3 novels are sold in a given hour.
2. Find the probability that at least 1 novel is sold in a given hour.

Solution: see page 101

Exercise 6.3.

Suppose $X \sim \mathcal{N}(50, 100)$. Compute:

1. $P(X \leq 60)$
2. $P(X > 45)$
3. $P(40 < X < 65)$

Solution: see page 102

Exercise 6.4.

The random variable X represents the number obtained after rolling a six-sided die numbered from 1 to 6. Let the random variable $Y = X(7 - X)$.

1. Find the set of values for the random variable Y .
2. Determine the probability distribution of Y .
3. Determine the cumulative distribution function $F(y)$.
4. Calculate the expected value $E(Y)$ and the variance $V(Y)$.

Solution: see page 103

Exercise 6.5.

Let X be a continuous random variable with f a function defined by:

$$f(x) = \begin{cases} k(1 - x), & \text{if } 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

1. Determine the value of the constant k for f to be a probability density function.
2. Determine the cumulative distribution function $F(x)$.
3. Calculate the expected value $E(X)$ and the variance $V(X)$.
4. Calculate the following probabilities: $P(X \geq \frac{1}{2})$ and $P(\frac{1}{2} \leq X \leq 2)$.

Solution: see page 104

Exercise 6.6.

For a normal distribution with mean $\mu = 50$ and standard deviation $\sigma = 4$ of the random variable X .

1. Calculate $P(X > 54)$.
2. Calculate $P(48 \leq X \leq 52)$.
3. Find the value of a defined by: $P(X \leq a) = 0.70$.

Solution: see page 105

Additional Exercises

Exercise 6.7.

The waiting time at a service counter follows an exponential distribution with mean 5 minutes.

1. Find the probability that a customer waits more than 8 minutes.
2. Find the probability that a customer waits less than 3 minutes.
3. Find the probability that a customer waits between 4 and 7 minutes.

Exercise 6.8.

A random number generator produces numbers uniformly between 0 and 1.

1. Find $P(0.2 \leq X \leq 0.7)$
2. Find $P(X > 0.8)$
3. Find $E(X)$ and $V(X)$

Exercise 6.9.

We roll a die, and suppose that we win \$10 if we get face 1; \$5 if we get face 2 or 3; \$2 if we get face 4; and lose \$5 if we get face 5 or 6. Let X be the random variable representing the gain.

1. Calculate the probability distribution of X and represent it graphically.
2. Find the cumulative distribution function and represent it graphically.
3. Calculate $E(X)$ and $V(X)$.

Exercise 6.10.

Let X be a random variable with a probability density function defined by:

$$f(x) = \begin{cases} \frac{1}{k}, & \text{if } x \in [1, 2[, \\ \frac{1}{4}, & \text{if } x \in [2, 4[, \\ 0, & \text{otherwise.} \end{cases}$$

1. Determine the constant k .
2. Determine the cumulative distribution function $F(x)$.
3. Calculate $E(X)$ and $V(X)$.
4. Calculate the following probabilities: $P(X \geq 3)$, $P(1 < X < 3)$, $P(3 \leq X \leq 5)$.

Additional Engineering Applications

Exercise 6.11.

An electronics manufacturer finds that 4% of components are defective. A sample of 20 components is randomly selected.

1. What is the probability that exactly 2 components are defective?
2. What is the probability that at least 1 component is defective?
3. What is the expected number of defective components?
4. What is the variance?

Exercise 6.12.

The lifetime of a mechanical part follows a normal distribution with mean 5000 hours and standard deviation 200 hours.

1. What is the probability that a part lasts less than 4700 hours?
2. What is the probability that a part lasts between 4800 and 5200 hours?
3. Find the lifetime such that 95% of parts exceed it.

Exercise 6.13.

The number of failures per day in a hydraulic system follows a Poisson distribution with mean $\lambda = 2$.

1. What is the probability of exactly 3 failures in a day?
2. What is the probability of at least 1 failure in a day?
3. What is the expected number of failures in a week?

Exercise 6.14.

The waiting time for a customer at a service counter follows an exponential distribution with mean 5 minutes.

1. What is the probability that a customer waits less than 3 minutes?
2. What is the probability that a customer waits more than 8 minutes?
3. What is the probability that a customer waits between 4 and 7 minutes?

Appendix A: Solutions to Main Exercises

How to use this appendix

STATISTICS

Chapter	Exercise	Page
1	1	80
1	2	80
1	3	81
1	4	82
2	1	83
2	2	84
2	3	86
2	4	87
2	5	88
3	1	90
3	2	91
3	3	92
3	4	92

PROBABILITY

Chapter	Exercise	Page
4	1	94
4	2	94
4	3	94
4	4	95
4	5	95
4	6	95
5	1	97
5	2	97
5	3	98
5	4	99
5	5	100
6	1	101
6	2	101
6	3	102
6	4	103
6	5	104
6	6	105

Solutions to Chapter 1

Solution to Exercise 1.1

Question: For each of the following variables, indicate whether it is qualitative (nominal or ordinal) or quantitative (discrete or continuous).

1. Number of students in a class.

- **Type:** Quantitative, Discrete.
- **Reason:** It takes integer values (0, 1, 2, ...) and cannot be fractional.

2. Blood type (A, B, AB, O).

- **Type:** Qualitative, Nominal.
- **Reason:** Categories have no natural order or ranking.

3. Time to complete an exam (in minutes).

- **Type:** Quantitative, Continuous.
- **Reason:** Time can take any real value within an interval (e.g., 45.5 minutes, 62.75 minutes).

4. Customer satisfaction rating (Very satisfied, Satisfied, Neutral, Unsatisfied, Very unsatisfied).

- **Type:** Qualitative, Ordinal.
- **Reason:** Categories have a natural order (Very satisfied > Satisfied > Neutral > Unsatisfied > Very unsatisfied).

5. Temperature in degrees Celsius.

- **Type:** Quantitative, Continuous.
- **Reason:** Temperature can take any real value within a range (e.g., 22.3°C, 18.7°C).

6. Marital status.

- **Type:** Qualitative, Nominal.
- **Reason:** Categories (Single, Married, Divorced, Widowed) have no natural order.

[Return to exercise](#)

Solution to Exercise 1.2

Question 1: What is the population? The sample?

- **Population:** All families in the surveyed area.
- **Sample:** The 200 families surveyed.

Question 2: What is the variable studied? Is it qualitative or quantitative? Discrete or continuous?

- **Variable:** Number of cars per household.
- **Type:** Quantitative, Discrete.
- **Reason:** Takes integer values (0, 1, 2, 3 or more).

Question 3: Calculate the total number of families surveyed.

$$N = 30 + 85 + 60 + 25 = 200$$

Question 4: Calculate the relative frequencies (percentages) for each modality.

$$f_i = \frac{n_i}{N}, \quad \text{Percentage} = f_i \times 100\%$$

Number of cars	n_i	f_i	Percentage (%)
0	30	$30/200 = 0.15$	15%
1	85	$85/200 = 0.425$	42.5%
2	60	$60/200 = 0.30$	30%
3 or more	25	$25/200 = 0.125$	12.5%
Total	200	1.00	100%

[Return to exercise](#)

Solution to Exercise 1.3

Question 1: Determine the population, the studied characteristic, and its type.

- **Population:** All students in the Faculty of Technology.
- **Studied characteristic:** Specialty.
- **Type:** Qualitative, Nominal (specialties have no natural order).

Question 2: Determine the number of students in "Telecommunications" specialty.

$$N = 25 + 40 + n_3 + 30 + 20 = 150$$

$$115 + n_3 = 150 \Rightarrow n_3 = 35$$

Question 3: Calculate the relative frequencies and percentages for each specialty.

$$f_i = \frac{n_i}{N}, \quad \text{Percentage} = f_i \times 100\%$$

Specialty	n_i	f_i	Percentage (%)
Civil Engineering	25	$25/150 \approx 0.1667$	16.67%
Electrical Engineering	40	$40/150 \approx 0.2667$	26.67%
Telecommunications	35	$35/150 \approx 0.2333$	23.33%
Mechanical Engineering	30	$30/150 = 0.20$	20%
Process Engineering	20	$20/150 \approx 0.1333$	13.33%
Total	150	1.00	100%

[\(Return to exercise\)](#)

Solution to Exercise 1.4

Question: For each situation, identify the population, the sample, and the variable of interest.

1. **A researcher wants to study the average height of students at a university. He measures the height of 150 randomly selected students.**
 - **Population:** All students at the university.
 - **Sample:** The 150 randomly selected students.
 - **Variable:** Height (in cm) - Quantitative, Continuous.
2. **A quality control inspector examines 50 light bulbs from a production batch of 10,000 bulbs to check if they are defective.**
 - **Population:** All 10,000 light bulbs in the production batch.
 - **Sample:** The 50 light bulbs examined.
 - **Variable:** Defective or not (Defective / Non-defective) - Qualitative, Nominal.
3. **A doctor wants to know the average blood pressure of patients over 60 years old. He records the blood pressure of 200 patients from three different hospitals.**
 - **Population:** All patients over 60 years old.
 - **Sample:** The 200 patients from three hospitals.
 - **Variable:** Blood pressure (in mmHg) - Quantitative, Continuous.

(Return to exercise)

Solutions to Chapter 2

Solution to Exercise 2.1

Data: Distribution of 80 cars by mechanical condition.

Mechanical Condition	Number of Cars
Bad	10
Average	20
Good	n_3
Excellent	30
Total	80

Question 1: Population, characteristic, type.

- **Population:** 80 cars in the car park.
- **Characteristic:** Mechanical condition.
- **Type:** Qualitative, Ordinal (Bad < Average < Good < Excellent).

Question 2: Number of cars in "Good" condition.

$$10 + 20 + n_3 + 30 = 80 \Rightarrow 60 + n_3 = 80 \Rightarrow n_3 = 20$$

Question 3: Graphical representations.

Bar chart:

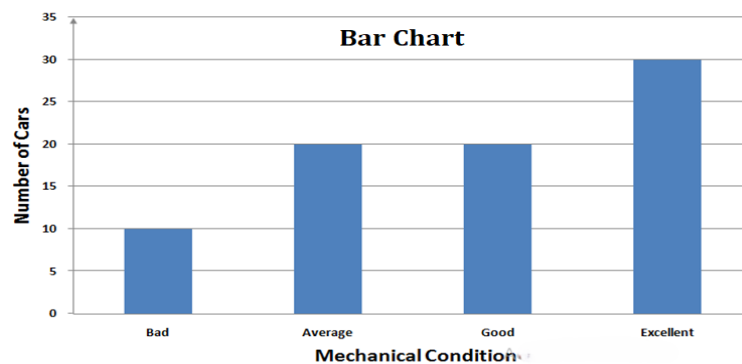


Figure 6.4: Bar chart for qualitative variable

Pie chart:

Angles for each sector: $\theta_i = f_i \times 360^\circ$.

Condition	n_i	f_i	θ_i
Bad	10	0.125	45°
Average	20	0.25	90°
Good	20	0.25	90°
Excellent	30	0.375	135°

Pie chart of cars according to their mechanical condition

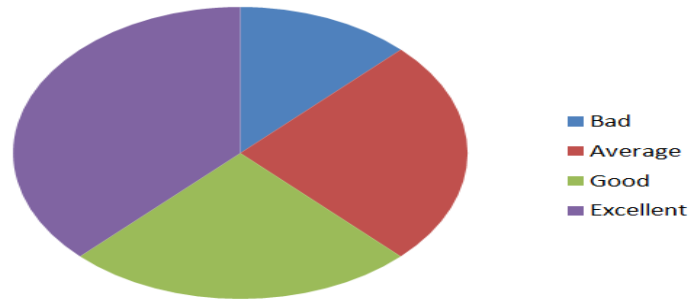


Figure 6.5: Pie chart

(Return to exercise)

Solution to Exercise 2.2

Data: Distribution of 400 households by number of vehicles.

x_i	0	1	2	3	4
n_i	30	200	100	60	10

Question 1: Type of variable: Quantitative Discrete.

Question 2: Complete the statistical table.

x_i	n_i	f_i	N_i^+	N_i^-	F_i^+	F_i^-	$n_i x_i$	$n_i x_i^2$
0	30	0.075	30	400	0.075	1	0	0
1	200	0.5	230	370	0.575	0.925	200	200
2	100	0.25	330	170	0.825	0.425	200	400
3	60	0.15	390	70	0.975	0.175	180	540
4	10	0.025	400	10	1	0.025	40	160
Σ	400	1					620	1300

Interpretation:

- $f_3 = 0.25$: 25% of households own exactly 2 vehicles.
- $N_3^+ = 330$: 330 households own ≤ 2 vehicles.
- $N_3^- = 170$: 170 households own ≥ 2 vehicles.
- $F_3^+ = 0.825$: 82.5% of households own ≤ 2 vehicles.
- $F_3^- = 0.425$: 42.5% of households own ≥ 2 vehicles.

Question 3 & 4: Graphical representations.

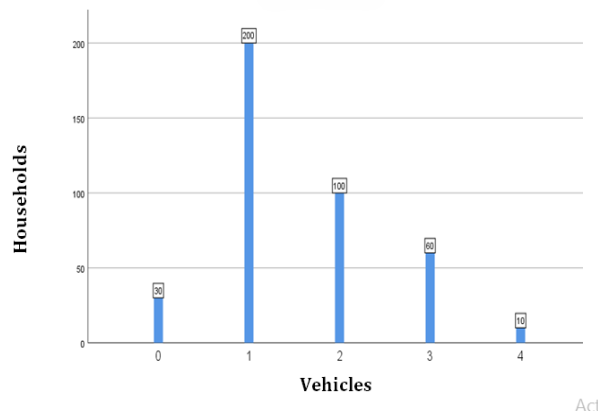


Figure 6.6: Bar chart of n_i for quantitative discrete variable

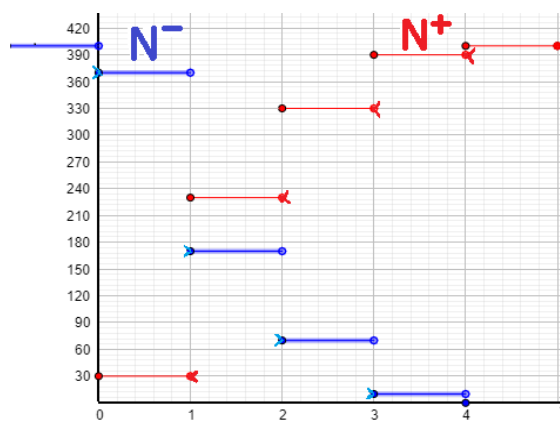


Figure 6.7: Step curve of N^+ and N^- for quantitative discrete variable

Question 5: Measures of central tendency.

$$\bar{X} = \frac{620}{400} = 1.55$$

$$M_o = 1 \quad (\text{highest frequency } n_{\max} = 200)$$

$$M_e = 1 \quad (N_{i-1}^+ = 30 \leq 200 \leq N_i^+ = 230)$$

$$Q_1 = 1, \quad Q_2 = 1, \quad Q_3 = 2$$

Question 6: Measures of dispersion.

$$\text{Range} = 4 - 0 = 4, \quad IQR = 2 - 1 = 1$$

$$V(X) = \frac{1300}{400} - (1.55)^2 = 3.25 - 2.4025 = 0.8475$$

$$\sigma(X) = \sqrt{0.8475} \approx 0.9206$$

(Return to exercise)

Solution to Exercise 2.3

Data: Mileage of 65 taxis.

Mileage (10^3 km)	[10, 20[	[20, 30[	[30, 40[	[40, 50[	[50, 60[	[60, 70[
n_i	9	13	22	10	7	4

Question 1: Population, characteristic, type.

- **Population:** 65 taxis.
- **Characteristic:** Mileage.
- **Type:** Quantitative Continuous.

Question 2: Cumulative relative frequencies.

C_i	n_i	f_i	F_i^+	F_i^-
[10,20[	9	0.1385	0.1385	1
[20,30[	13	0.2	0.3385	0.8615
[30,40[	22	0.3385	0.6770	0.6615
[40,50[	10	0.1538	0.8308	0.3230
[50,60[	7	0.1077	0.9385	0.1692
[60,70[	4	0.0615	1	0.0615

Interpretation:

- $F_3^+ = 0.6770$: 67.70% of taxis have mileage $\leq 40 \times 10^3$ km.
- $F_3^- = 0.6615$: 66.15% of taxis have mileage $\geq 30 \times 10^3$ km.

Question 3 & 4: Graphical representations.

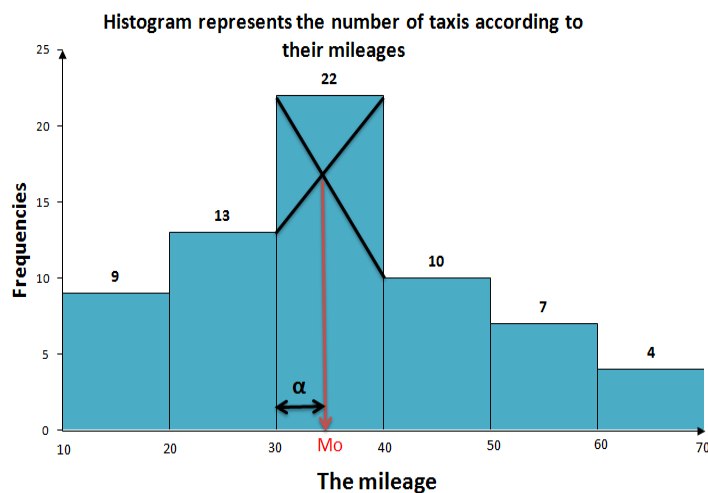


Figure 6.8: Histogram + M_o for quantitative continuous variable

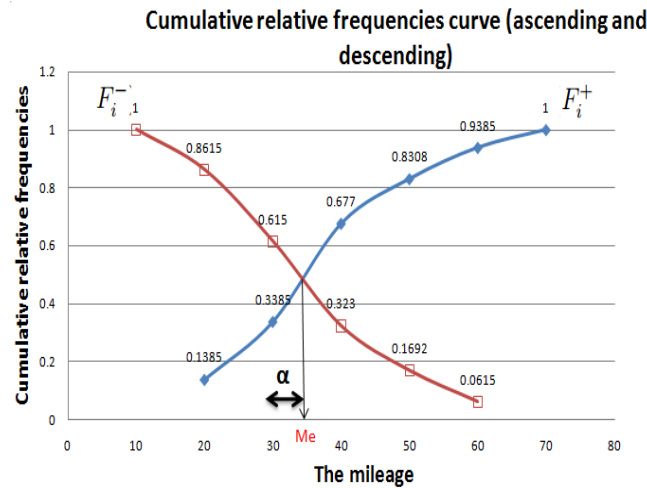


Figure 6.9: Cumulative curve + M_e for quantitative continuous variable

Question 5: Mode and median graphically.

- Mode (from histogram): $M_o \approx 34.3$ km
- Median (from cumulative curve): $M_e \approx 34.8$ km

Question 6: Measures of central tendency and dispersion.

$$\bar{X} = \frac{2325}{65} \approx 35.7692 \text{ km}$$

$$M_o = 30 + 10 \times \frac{9}{9 + 12} \approx 34.286 \text{ km}$$

$$M_e = 30 + 10 \times \frac{0.5 - 0.3385}{0.3385} \approx 34.771 \text{ km}$$

$$Q_1 \approx 25.575 \text{ km}, \quad Q_3 \approx 44.746 \text{ km}$$

$$V(X) \approx 188.6412, \quad \sigma(X) \approx 13.7347 \text{ km}$$

(Return to exercise)

Solution to Exercise 2.4

Data: Same series with combined classes.

C_i	[10, 20[	[20, 30[	[30, 40[	[40, 50[	[50, 70[
n_i	9	13	22	10	11

Question 1: Graphical representation.

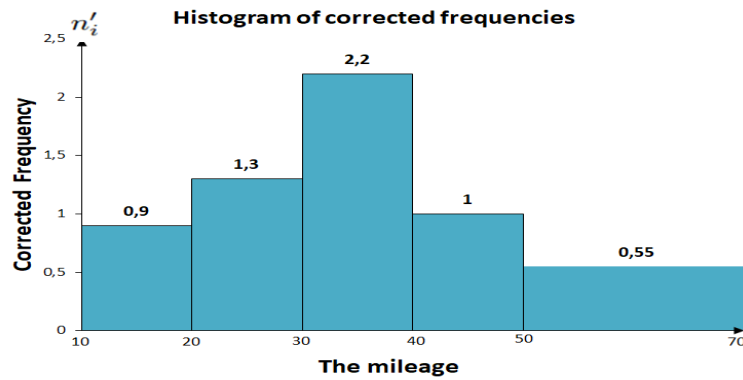


Figure 6.10: Histogram of unequal class for quantitative continuous variable

Question 2: Mode and median analytically.

$$\text{Modal class: } [30, 40[, \quad d_1 = 22 - 13 = 9, \quad d_2 = 22 - 10 = 12$$

$$M_o = 30 + 10 \times \frac{9}{9 + 12} \approx 34.286 \text{ km}$$

$$M_e = 30 + 10 \times \frac{32.5 - 22}{22} \approx 34.773 \text{ km}$$

Question 3: Mean, variance, standard deviation.

$$\bar{X} = \frac{2340}{65} = 36 \text{ km}$$

$$V(X) = \frac{96950}{65} - 36^2 = 1491.5385 - 1296 = 195.5385$$

$$\sigma(X) = \sqrt{195.5385} \approx 13.9835 \text{ km}$$

(Return to exercise)

Solution to Exercise 2.5

Data: 20 families, number of children.

Data: 1, 3, 5, 5, 3, 2, 4, 4, 6, 0, 2, 4, 3, 6, 0, 5, 4, 2, 3, 2

Question 1: Population, statistical unit, variable, categories.

- **Population:** 20 families.
- **Statistical unit:** One family.
- **Variable:** Number of children.
- **Categories:** 0, 1, 2, 3, 4, 5, 6.

Question 2: Complete the statistical table.

x_i	n_i	f_i	N_i^+	F_i^+
0	2	0.10	2	0.10
1	1	0.05	3	0.15
2	5	0.25	8	0.40
3	4	0.20	12	0.60
4	4	0.20	16	0.80
5	3	0.15	19	0.95
6	1	0.05	20	1.00

(Return to exercise)

Note for exercise 2.8, part 1

• **Yule's rule:** $k = 2.5 \times \sqrt[4]{N}$ • **Sturges' rule:** $k = 1 + 3.322 \times \log_{10}(N)$

• **Class width:** $a_i = \frac{\text{Range}}{k}$

Where:

- N Total number of observations.
- k Number of classes.
- a_i Amplitude (class width).

Solutions to Chapter 3

Solution to Exercise 3.1

1. The regression equation for m as a function of x :

x_i	m_i	x_i^2	m_i^2	$x_i \times m_i$
0	0	0	0	0
0.5	10	0.25	100	5
1.1	20	1.21	400	22
1.5	30	2.25	900	45
1.9	40	3.61	1600	76
5	100	7.32	3000	148

- $\bar{x} = \frac{1}{n} \sum_{i=1}^5 x_i = \frac{5}{5} = 1$.
- $V(x) = \frac{1}{n} \sum_{i=1}^5 x_i^2 - \bar{x}^2 = \frac{7.32}{5} - 1^2 = 0.464 \Rightarrow \sigma_x = 0.681$.
- $\bar{m} = \frac{1}{n} \sum_{i=1}^5 m_i = \frac{100}{5} = 20$.
- $V(m) = \frac{1}{n} \sum_{i=1}^5 m_i^2 - \bar{m}^2 = \frac{3000}{5} - 400 = 200 \Rightarrow \sigma_m = 14.142$.
- $\text{Cov}(x, m) = \sigma_{xm} = \frac{1}{n} \sum_{i=1}^5 x_i m_i - \bar{x} \bar{m} = \frac{148}{5} - (1 \times 20) = 9.6$.

$$\begin{cases} \mathbf{a} = \frac{\text{Cov}(x, m)}{V(x)} = \frac{9.6}{0.464} = \mathbf{20.69} \\ \text{and} \\ \mathbf{b} = \bar{m} - a\bar{x} = 20 - 20.69 \times 1 = \mathbf{-0.69}. \end{cases}$$

So, the equation of the linear regression for m on x is: **$m = 20.69 x - 0.69$** .

2. The scatter plot associated with the double statistical series and the line m as a function of x :

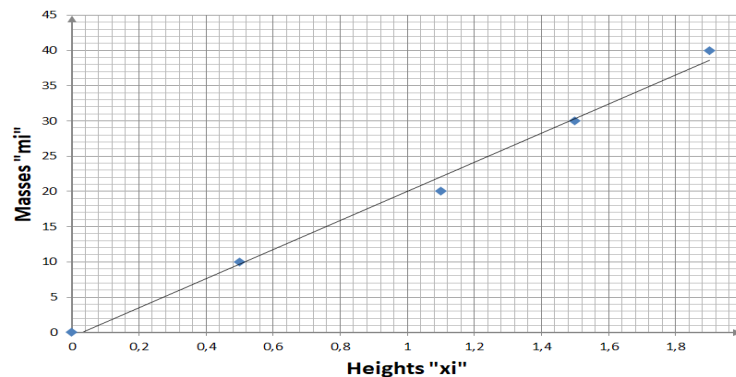


Figure 6.11: Scatter plot

3. The linear correlation coefficient:

$$\rho = \frac{\text{Cov}(x, m)}{\sigma_x \sigma_m} = \frac{9.6}{0.681 \times 14.142} = \mathbf{0.997}.$$

We can conclude that the linear adjustment is **good** ($\rho = 0.997 > 0.7$).

3. Determine x if $m = 50$ kg:

We can estimate x for $m = 50$ from the regression line equation $D(m/x)$:

$$x = 0.048m + 0.033 \Rightarrow x = 0.048(50) + 0.033 \Rightarrow \mathbf{x=2.433 \text{ m}}.$$

(Return to exercise)

Solution to Exercise 3.2

Data: Diameter (X) vs Tensile strength (Y).

X (mm)	2	4	6	8	10
Y (kN)	5	12	20	30	42

Question 1: Covariance and correlation coefficient.

$$\bar{x} = \frac{2 + 4 + 6 + 8 + 10}{5} = \frac{30}{5} = 6$$

$$\bar{y} = \frac{5 + 12 + 20 + 30 + 42}{5} = \frac{109}{5} = 21.8$$

$$\sum x_i y_i = 2 \times 5 + 4 \times 12 + 6 \times 20 + 8 \times 30 + 10 \times 42$$

$$= 10 + 48 + 120 + 240 + 420 = 838$$

$$\text{Cov}(X, Y) = \frac{838}{5} - 6 \times 21.8 = 167.6 - 130.8 = 36.8$$

$$\sum x_i^2 = 4 + 16 + 36 + 64 + 100 = 220$$

$$V(X) = \frac{220}{5} - 6^2 = 44 - 36 = 8$$

$$\sum y_i^2 = 25 + 144 + 400 + 900 + 1764 = 3233$$

$$V(Y) = \frac{3233}{5} - (21.8)^2 = 646.6 - 475.24 = 171.36$$

$$\rho = \frac{36.8}{\sqrt{8} \times \sqrt{171.36}} = \frac{36.8}{2.828 \times 13.09} = \frac{36.8}{37.02} \approx 0.994$$

Question 2: Regression line Y on X .

$$a = \frac{\text{Cov}(X, Y)}{V(X)} = \frac{36.8}{8} = 4.6$$

$$b = \bar{y} - a\bar{x} = 21.8 - 4.6 \times 6 = 21.8 - 27.6 = -5.8$$

$$\hat{y} = 4.6x - 5.8$$

Question 3: Interpret the slope.

For each additional 1 mm increase in diameter, the tensile strength increases by 4.6 kN.

Question 4: Predict strength for diameter 7 mm.

$$\hat{y} = 4.6 \times 7 - 5.8 = 32.2 - 5.8 = 26.4 \text{ kN}$$

Question 5: Diameter for strength 50 kN.

$$50 = 4.6x - 5.8 \Rightarrow 4.6x = 55.8 \Rightarrow x = \frac{55.8}{4.6} \approx 12.13 \text{ mm}$$

(Return to exercise)

Solution to Exercise 3.3

Data: Contingency table - education vs satisfaction.

Education \ Satisfaction	Low	Medium	High	Total
High School	30	20	10	60
Bachelor	15	35	20	70
Master	5	15	50	70
Total	50	70	80	200

Question 1: Marginal distributions.

Marginal of Education: $P(\text{High School}) = 60/200 = 0.3$, $P(\text{Bachelor}) = 70/200 = 0.35$, $P(\text{Master}) = 70/200 = 0.35$

Marginal of Satisfaction: $P(\text{Low}) = 50/200 = 0.25$, $P(\text{Medium}) = 70/200 = 0.35$, $P(\text{High}) = 80/200 = 0.4$

Question 2: Conditional distribution of satisfaction given Education = Master.

$$P(\text{Low} \mid \text{Master}) = 5/70 \approx 0.0714 \quad (7.14\%)$$

$$P(\text{Medium} \mid \text{Master}) = 15/70 \approx 0.2143 \quad (21.43\%)$$

$$P(\text{High} \mid \text{Master}) = 50/70 \approx 0.7143 \quad (71.43\%)$$

Question 3: Conditional distribution of education given Satisfaction = High.

$$P(\text{High School} \mid \text{High}) = 10/80 = 0.125 \quad (12.5\%)$$

$$P(\text{Bachelor} \mid \text{High}) = 20/80 = 0.25 \quad (25\%)$$

$$P(\text{Master} \mid \text{High}) = 50/80 = 0.625 \quad (62.5\%)$$

Question 4: Interpretation.

Among Master's degree holders, 71.43% are highly satisfied. Among highly satisfied employees, 62.5% have a Master's degree. This suggests a strong positive relationship between higher education level and job satisfaction.

(Return to exercise)

Solution to Exercise 3.4

Data: Temperature (X) vs Energy consumption (Y).

$X(^{\circ}C)$	22	24	26	28	30	32	34
$Y(kWh)$	18	20	23	27	32	38	45

Question 1: Calculate $\bar{x}$, $\bar{y}$, $V(X)$, $V(Y)$, $\text{Cov}(X, Y)$.

$$\bar{x} = \frac{22 + 24 + 26 + 28 + 30 + 32 + 34}{7} = \frac{196}{7} = 28$$

$$\bar{y} = \frac{18 + 20 + 23 + 27 + 32 + 38 + 45}{7} = \frac{203}{7} = 29$$

$$\sum x_i^2 = 484 + 576 + 676 + 784 + 900 + 1024 + 1156 = 5600$$

$$V(X) = \frac{5600}{7} - 28^2 = 800 - 784 = 16$$

$$\sum y_i^2 = 324 + 400 + 529 + 729 + 1024 + 1444 + 2025 = 6475$$

$$V(Y) = \frac{6475}{7} - 29^2 = 925 - 841 = 84$$

$$\sum x_i y_i = 22 \times 18 + 24 \times 20 + 26 \times 23 + 28 \times 27 + 30 \times 32 + 32 \times 38 + 34 \times 45$$

$$= 396 + 480 + 598 + 756 + 960 + 1216 + 1530 = 5936$$

$$\text{Cov}(X, Y) = \frac{5936}{7} - 28 \times 29 = 848 - 812 = 36$$

Question 2: Correlation coefficient r .

$$r = \frac{36}{\sqrt{16} \times \sqrt{84}} = \frac{36}{4 \times 9.165} = \frac{36}{36.66} \approx 0.982$$

Interpretation: Very strong positive correlation. As temperature increases, energy consumption increases significantly.

Question 3: Regression line $\hat{y} = ax + b$.

$$a = \frac{\text{Cov}(X, Y)}{V(X)} = \frac{36}{16} = 2.25$$

$$b = \bar{y} - a\bar{x} = 29 - 2.25 \times 28 = 29 - 63 = -34$$

$$\hat{y} = 2.25x - 34$$

Question 5: Predict energy consumption for $X = 29^\circ\text{C}$.

$$\hat{y} = 2.25 \times 29 - 34 = 65.25 - 34 = 31.25 \text{ kWh}$$

Question 6: Predict for $X = 40^\circ\text{C}$.

$$\hat{y} = 2.25 \times 40 - 34 = 90 - 34 = 56 \text{ kWh}$$

Reliability: This prediction is less reliable because $X = 40^\circ\text{C}$ is outside the observed range ($22 - 34^\circ\text{C}$). Extrapolation beyond the data range can be risky.

(Return to exercise)

Solutions to Chapter 4

Solution to Exercise 4.1

Question: How many different 3-digit numbers can be formed using the digits 1, 2, 3, 4, 5?

(a) Repetition allowed:

For a 3-digit number, we have 3 positions: hundreds, tens, units. - Hundreds position: 5 possibilities (1,2,3,4,5) - cannot be 0 anyway - Tens position: 5 possibilities (repetition allowed) - Units position: 5 possibilities (repetition allowed)

$$\text{Total} = 5 \times 5 \times 5 = 5^3 = 125$$

(b) Repetition not allowed:

- Hundreds position: 5 possibilities - Tens position: 4 possibilities (one digit already used) - Units position: 3 possibilities (two digits already used)

$$\text{Total} = 5 \times 4 \times 3 = 60$$

(Return to exercise)

Solution to Exercise 4.2

Question: In how many ways can 8 people be seated in a row of 8 chairs?

This is a permutation of 8 distinct people in 8 distinct chairs.

$$P_8 = 8! = 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 40320$$

(Return to exercise)

Solution to Exercise 4.3

Data: Urn with 12 balls: 3 red (R), 4 blue (B), 5 yellow (Y). Three balls drawn simultaneously.

Question 1: Total number of possible outcomes.

Since order does not matter and repetition is not allowed (balls are distinct within each color):

$$\text{Total outcomes} = C_{12}^3 = \binom{12}{3} = \frac{12!}{3! \times 9!} = \frac{12 \times 11 \times 10}{3 \times 2 \times 1} = 220$$

Question 2: Outcomes giving exactly 3 red balls.

We need to choose 3 red balls from the 3 available red balls, and 0 balls from the other colors:

$$\binom{3}{3} \times \binom{9}{0} = 1 \times 1 = 1$$

Question 3: Outcomes giving one ball of each color.

We choose 1 red from 3, 1 blue from 4, 1 yellow from 5:

$$\binom{3}{1} \times \binom{4}{1} \times \binom{5}{1} = 3 \times 4 \times 5 = 60$$

(Return to exercise)

Solution to Exercise 4.4

Question: To form a code, three non-zero digits are randomly selected. How many distinct codes can be formed?

Non-zero digits: 1, 2, 3, 4, 5, 6, 7, 8, 9 (9 digits total).

Since the code is a 3-digit sequence and digits can be repeated (not stated otherwise):

$$\text{Number of codes} = 9 \times 9 \times 9 = 9^3 = 729$$

Note: If the problem intended *distinct* digits (no repetition), the answer would be:

$$A_9^3 = 9 \times 8 \times 7 = 504$$

(Return to exercise)

Solution to Exercise 4.5

Question: An urn contains 10 balls numbered from 0 to 9. Successively and without replacement, 4 balls are drawn. Find the number of possible outcomes.

Since the balls are drawn **successively** (order matters) and **without replacement** (no repetition):

- First draw: 10 possibilities - Second draw: 9 possibilities (one ball removed) - Third draw: 8 possibilities - Fourth draw: 7 possibilities

$$\text{Number of outcomes} = 10 \times 9 \times 8 \times 7 = 5040$$

Alternatively, using arrangements (permutations):

$$A_{10}^4 = \frac{10!}{(10-4)!} = \frac{10!}{6!} = 10 \times 9 \times 8 \times 7 = 5040$$

(Return to exercise)

Solution to Exercise 4.6

Question: Arrange 4 mathematics books, 3 physics books, and 5 chemistry books in 12 slots. Books of each category must be placed next to each other. Find the number of possible arrangements.

Step 1: Treat each category as a single block. - Block M (Mathematics): 4 books - Block P (Physics): 3 books - Block C (Chemistry): 5 books

We have 3 blocks to arrange: $3! = 6$ ways.

Step 2: Arrange books within each block. - Mathematics books: $4! = 24$ ways - Physics books: $3! = 6$ ways - Chemistry books: $5! = 120$ ways

Step 3: Multiply all possibilities.

$$\text{Total arrangements} = 3! \times 4! \times 3! \times 5!$$

$$= 6 \times 24 \times 6 \times 120$$

$$= 6 \times 24 = 144, \quad 144 \times 6 = 864, \quad 864 \times 120 = 103680$$

103680 arrangements

(Return to exercise)

Solutions to Chapter 5

Solution to Exercise 5.1

Question 1: A card is drawn from a standard deck of 52 cards. What is the probability it is a heart?

A standard deck has 13 hearts.

$$P(\text{heart}) = \frac{13}{52} = \frac{1}{4} = 0.25$$

Question 2: What is the probability it is a king?

A standard deck has 4 kings.

$$P(\text{king}) = \frac{4}{52} = \frac{1}{13} \approx 0.07692$$

Question 3: What is $P(\text{heart} \mid \text{king})$?

Given that the card is a king, there is only 1 king of hearts.

$$P(\text{heart} \mid \text{king}) = \frac{1}{4} = 0.25$$

Note that $P(\text{heart}) = 0.25 = P(\text{heart} \mid \text{king})$, so the events "heart" and "king" are independent.

(Return to exercise)

Solution to Exercise 5.2

Question: Two dice are rolled. Find the probability that the sum is 7.

Total number of possible outcomes: $6 \times 6 = 36$.

Outcomes where the sum is 7:

(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1) (6 outcomes)

$$P(\text{sum} = 7) = \frac{6}{36} = \frac{1}{6} \approx 0.1667$$

(Return to exercise)

Solution to Exercise 5.3

Data: $P(A) = 0.4$, $P(B) = 0.6$, $P(A \cap B) = 0.3$

Question 1: $P(A \cup B)$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.6 - 0.3 = 0.7$$

Question 2: $P(\bar{A} \cap B)$

$$P(\bar{A} \cap B) = P(B) - P(A \cap B) = 0.6 - 0.3 = 0.3$$

Question 3: $P(A \cap \bar{B})$

$$P(A \cap \bar{B}) = P(A) - P(A \cap B) = 0.4 - 0.3 = 0.1$$

Question 4: Are A and B independent?

Two events are independent if $P(A \cap B) = P(A) \times P(B)$.

$$P(A) \times P(B) = 0.4 \times 0.6 = 0.24$$

$$P(A \cap B) = 0.3 \neq 0.24$$

Therefore, A and B are **not independent**.

(Return to exercise)

Solution to Exercise 5.4

Data: Test tubes from companies A, B, C.

$$P(A) = 0.5, \quad P(B) = 0.3, \quad P(C) = 0.2$$

$$P(D | A) = 0.02, \quad P(D | B) = 0.03, \quad P(D | C) = 0.04$$

where D = "defective tube".

Question 1: $P(D)$

Using the Law of Total Probability:

$$P(D) = P(D | A)P(A) + P(D | B)P(B) + P(D | C)P(C)$$

$$P(D) = 0.02 \times 0.5 + 0.03 \times 0.3 + 0.04 \times 0.2$$

$$P(D) = 0.01 + 0.009 + 0.008 = 0.027$$

Question 2: $P(A | D)$

Using Bayes' Theorem:

$$P(A | D) = \frac{P(D | A)P(A)}{P(D)} = \frac{0.02 \times 0.5}{0.027} = \frac{0.01}{0.027} \approx 0.3704$$

(Return to exercise)

Solution to Exercise 5.5

Data: $P(A) = \frac{2}{3}$, $P(B) = \frac{1}{3}$, $P(\bar{A} | B) = \frac{1}{4}$

Question 1: $P(A | B)$

Since $P(\bar{A} | B) = 1 - P(A | B)$, we have:

$$P(A | B) = 1 - P(\bar{A} | B) = 1 - \frac{1}{4} = \frac{3}{4}$$

Question 2: $P(B | A)$

First, calculate $P(A \cap B)$ using the multiplication rule:

$$P(A \cap B) = P(A | B) \times P(B) = \frac{3}{4} \times \frac{1}{3} = \frac{3}{12} = \frac{1}{4}$$

Then:

$$P(B | A) = \frac{P(A \cap B)}{P(A)} = \frac{1/4}{2/3} = \frac{1}{4} \times \frac{3}{2} = \frac{3}{8} = 0.375$$

Question 3: Probability that exactly one of the two events occurs.

$$P(\text{exactly one}) = P(A \cap \bar{B}) + P(\bar{A} \cap B)$$

We already have $P(A \cap B) = \frac{1}{4}$.

$$P(A \cap \bar{B}) = P(A) - P(A \cap B) = \frac{2}{3} - \frac{1}{4} = \frac{8}{12} - \frac{3}{12} = \frac{5}{12}$$

$$P(\bar{A} \cap B) = P(B) - P(A \cap B) = \frac{1}{3} - \frac{1}{4} = \frac{4}{12} - \frac{3}{12} = \frac{1}{12}$$

$$P(\text{exactly one}) = \frac{5}{12} + \frac{1}{12} = \frac{6}{12} = \frac{1}{2} = 0.5$$

(Return to exercise)

Solutions to Chapter 6

Solution to Exercise 6.1

Question: A multiple-choice test has 20 questions, each with 4 options. A student guesses randomly.

Let X = number of correct answers. This follows a binomial distribution:
 $X \sim \mathcal{B}(n = 20, p = 0.25)$.

Question 1: Probability of getting exactly 5 correct.

$$\begin{aligned}P(X = 5) &= \binom{20}{5} (0.25)^5 (0.75)^{15} \\ \binom{20}{5} &= \frac{20!}{5! \times 15!} = \frac{20 \times 19 \times 18 \times 17 \times 16}{5 \times 4 \times 3 \times 2 \times 1} = 15504 \\ (0.25)^5 &= \left(\frac{1}{4}\right)^5 = \frac{1}{1024} \approx 0.00097656 \\ (0.75)^{15} &= \left(\frac{3}{4}\right)^{15} = \frac{3^{15}}{4^{15}} \approx \frac{14348907}{1073741824} \approx 0.01336\end{aligned}$$

$$P(X = 5) \approx 15504 \times 0.00097656 \times 0.01336 \approx 15504 \times 0.00001304 \approx 0.2023$$

Question 2: Expected number of correct answers.

$$E(X) = n \times p = 20 \times 0.25 = 5$$

Question 3: Variance.

$$V(X) = n \times p \times (1 - p) = 20 \times 0.25 \times 0.75 = 20 \times 0.1875 = 3.75$$

(Return to exercise)

Solution to Exercise 6.2

Question: A bookstore sells an average of 2.5 novels per hour.

Let X = number of novels sold per hour. This follows a Poisson distribution:
 $X \sim \text{Poisson}(\lambda = 2.5)$.

Question 1: Probability that exactly 3 novels are sold.

$$P(X = 3) = e^{-\lambda} \frac{\lambda^3}{3!} = e^{-2.5} \times \frac{(2.5)^3}{6}$$

$$e^{-2.5} \approx 0.082085$$

$$(2.5)^3 = 15.625$$

$$\frac{15.625}{6} \approx 2.60417$$

$$P(X = 3) \approx 0.082085 \times 2.60417 \approx 0.2138$$

Question 2: Probability that at least 1 novel is sold.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-\lambda} \frac{\lambda^0}{0!} = 1 - e^{-2.5}$$

$$P(X \geq 1) = 1 - 0.082085 \approx 0.9179$$

(Return to exercise)

Solution to Exercise 6.3

Question: $X \sim \mathcal{N}(50, 100)$. Compute:

Question 1: $P(X \leq 60)$

Standardize: $Z = \frac{X - \mu}{\sigma} = \frac{X - 50}{10}$

$$P(X \leq 60) = P\left(Z \leq \frac{60 - 50}{10}\right) = P(Z \leq 1)$$

From the standard normal table: $\Phi(1) = 0.8413$

$$P(X \leq 60) = 0.8413$$

Question 2: $P(X > 45)$

$$P(X > 45) = 1 - P(X \leq 45) = 1 - P\left(Z \leq \frac{45 - 50}{10}\right)$$

$$= 1 - P(Z \leq -0.5) = 1 - [1 - P(Z \leq 0.5)] = P(Z \leq 0.5)$$

$$\Phi(0.5) = 0.6915$$

$$P(X > 45) = 0.6915$$

Question 3: $P(40 < X < 65)$

$$P(40 < X < 65) = P\left(\frac{40 - 50}{10} < Z < \frac{65 - 50}{10}\right) = P(-1 < Z < 1.5)$$

$$= \Phi(1.5) - \Phi(-1) = \Phi(1.5) - [1 - \Phi(1)]$$

$$\Phi(1.5) = 0.9332, \quad \Phi(1) = 0.8413$$

$$P(40 < X < 65) = 0.9332 - (1 - 0.8413) = 0.9332 - 0.1587 = 0.7745$$

(Return to exercise)

Solution to Exercise 6.4

Question: X = number on a die (1 to 6). $Y = X(7 - X)$.

Question 1: Set of values for Y .

X	1	2	3	4	5	6
$Y = X(7 - X)$	$1 \times 6 = 6$	$2 \times 5 = 10$	$3 \times 4 = 12$	$4 \times 3 = 12$	$5 \times 2 = 10$	$6 \times 1 = 6$

Values of Y : $\{6, 10, 12\}$

Question 2: Probability distribution of Y .

Each face of the die has probability $1/6$.

$$P(Y = 6) = P(X = 1 \text{ or } X = 6) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

$$P(Y = 10) = P(X = 2 \text{ or } X = 5) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

$$P(Y = 12) = P(X = 3 \text{ or } X = 4) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

y	6	10	12
$P(Y = y)$	$1/3$	$1/3$	$1/3$

Question 3: Cumulative distribution function $F(y)$.

$$F(y) = P(Y \leq y) = \begin{cases} 0, & y < 6 \\ \frac{1}{3}, & 6 \leq y < 10 \\ \frac{2}{3}, & 10 \leq y < 12 \\ 1, & y \geq 12 \end{cases}$$

Question 4: $E(Y)$ and $V(Y)$.

$$E(Y) = \sum y \cdot P(Y = y) = 6 \times \frac{1}{3} + 10 \times \frac{1}{3} + 12 \times \frac{1}{3} = \frac{6 + 10 + 12}{3} = \frac{28}{3} \approx 9.3333$$

$$E(Y^2) = 6^2 \times \frac{1}{3} + 10^2 \times \frac{1}{3} + 12^2 \times \frac{1}{3} = \frac{36 + 100 + 144}{3} = \frac{280}{3} \approx 93.3333$$

$$V(Y) = E(Y^2) - [E(Y)]^2 = \frac{280}{3} - \left(\frac{28}{3}\right)^2 = \frac{280}{3} - \frac{784}{9} = \frac{840}{9} - \frac{784}{9} = \frac{56}{9} \approx 6.2222$$

(Return to exercise)

Solution to Exercise 6.5

Question: $f(x) = k(1 - x)$ for $0 \leq x \leq 1$.

Question 1: Determine k so that f is a PDF.

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad \Rightarrow \quad \int_0^1 k(1 - x) dx = 1$$

$$\int_0^1 (1 - x) dx = \left[x - \frac{x^2}{2} \right]_0^1 = 1 - \frac{1}{2} = \frac{1}{2}$$

$$k \times \frac{1}{2} = 1 \quad \Rightarrow \quad k = 2$$

Question 2: Cumulative distribution function $F(x)$.

For $x < 0$: $F(x) = 0$

For $0 \leq x \leq 1$:

$$F(x) = \int_0^x 2(1 - t) dt = 2 \left[t - \frac{t^2}{2} \right]_0^x = 2 \left(x - \frac{x^2}{2} \right) = 2x - x^2$$

For $x > 1$: $F(x) = 1$

$$F(x) = \begin{cases} 0, & x < 0 \\ 2x - x^2, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

Question 3: $E(X)$ and $V(X)$.

$$\begin{aligned} E(X) &= \int_0^1 x \cdot 2(1 - x) dx = 2 \int_0^1 (x - x^2) dx = 2 \left[\frac{x^2}{2} - \frac{x^3}{3} \right]_0^1 \\ &= 2 \left(\frac{1}{2} - \frac{1}{3} \right) = 2 \times \frac{1}{6} = \frac{1}{3} \approx 0.3333 \end{aligned}$$

$$\begin{aligned} E(X^2) &= \int_0^1 x^2 \cdot 2(1 - x) dx = 2 \int_0^1 (x^2 - x^3) dx = 2 \left[\frac{x^3}{3} - \frac{x^4}{4} \right]_0^1 \\ &= 2 \left(\frac{1}{3} - \frac{1}{4} \right) = 2 \times \frac{1}{12} = \frac{1}{6} \approx 0.1667 \end{aligned}$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{1}{6} - \left(\frac{1}{3} \right)^2 = \frac{1}{6} - \frac{1}{9} = \frac{3}{18} - \frac{2}{18} = \frac{1}{18} \approx 0.0556$$

Question 4: $P(X \geq \frac{1}{2})$ and $P(\frac{1}{2} \leq X \leq 2)$.

$$P\left(X \geq \frac{1}{2}\right) = 1 - F\left(\frac{1}{2}\right) = 1 - \left[2 \times \frac{1}{2} - \left(\frac{1}{2}\right)^2 \right] = 1 - \left(1 - \frac{1}{4} \right) = 1 - \frac{3}{4} = \frac{1}{4}$$

$$P\left(\frac{1}{2} \leq X \leq 2\right) = F(2) - F\left(\frac{1}{2}\right) = 1 - \frac{3}{4} = \frac{1}{4}$$

(Return to exercise)

Solution to Exercise 6.6

Question: $X \sim \mathcal{N}(\mu = 50, \sigma = 4)$.

Question 1: $P(X > 54)$

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 50}{4}$$

$$P(X > 54) = P\left(Z > \frac{54 - 50}{4}\right) = P(Z > 1)$$

From standard normal table: $\Phi(1) = 0.8413$

$$P(Z > 1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$$

Question 2: $P(48 \leq X \leq 52)$

$$P(48 \leq X \leq 52) = P\left(\frac{48 - 50}{4} \leq Z \leq \frac{52 - 50}{4}\right) = P(-0.5 \leq Z \leq 0.5)$$

$$= \Phi(0.5) - \Phi(-0.5) = \Phi(0.5) - [1 - \Phi(0.5)] = 2\Phi(0.5) - 1$$

$$\Phi(0.5) = 0.6915$$

$$P(48 \leq X \leq 52) = 2 \times 0.6915 - 1 = 1.383 - 1 = 0.3830$$

Question 3: Find a such that $P(X \leq a) = 0.70$

$$P(X \leq a) = 0.70 \quad \Rightarrow \quad P\left(Z \leq \frac{a - 50}{4}\right) = 0.70$$

From standard normal table, the z -value corresponding to 0.70 is approximately $z = 0.52$.

$$\frac{a - 50}{4} = 0.52 \quad \Rightarrow \quad a - 50 = 2.08 \quad \Rightarrow \quad a = 52.08$$

(Return to exercise)

Appendix B: Standard Normal Distribution Table

Standard Normal Table

Note: An entry in the table is the area under the curve to the left of z , $P(Z \leq z) = \Phi(z)$

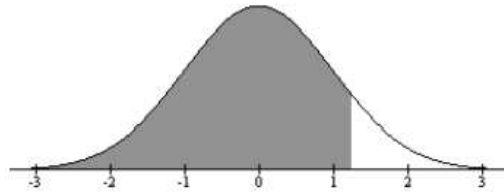


Table I.1 Cumulative Normal Distribution Function $\Phi(z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5674	0.5714	0.5753
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.70	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.80	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.90	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.00	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.10	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.20	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.30	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.40	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.50	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.60	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.70	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.80	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.90	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.00	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.10	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.20	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.30	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.40	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Negative values of z can be obtained from the relationship $\Phi(-z) = 1 - \Phi(z)$.

Bibliography

- [1] Agresti, A., & Franklin, C. (2017), *Statistics: The Art and Science of Learning from Data*, 4th ed., Pearson.
- [2] Bentobache, M. (2018), *Cours de Statistique et Probabilités pour les Sciences et Techniques*, Polycopiée de cours, Université Amar Téliidji de Laghouat, Année Universitaire 2018-2019.
- [3] Chekroun, A. (2017), *Statistiques descriptives et exercices: Rappels de cours et exercices corrigés*, Polycopiée de cours, Université Abou Bekr Belkaid Tlemcen, Année Universitaire 2017-2018.
- [4] Cramér, H. (1946), *Mathematical Methods of Statistics*, Princeton University Press.
- [5] Devore, J. L. (2015), *Probability and Statistics for Engineering and the Sciences*, 9th ed., Cengage.
- [6] Fahsi, A. (2019), *Statistique Descriptive*, Département de Mathématiques, Faculté des Sciences et Techniques Mohammeda, Université de Hassan II de Casablanca.
- [7] Feller, W. (1968), *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd ed., Wiley, New York, London.
- [8] Fisher, R. A. (1925), *Statistical Methods for Research Workers*, Oliver & Boyd.
- [9] Kerboua, M. (2015), *Cours, exercices et examens: Statistique et Probabilité (1^{re} LMD MI, 2^{me} LMD ST)*, Polycopiée de cours, Université de Guelma, Année Universitaire 2015-2016.
- [10] Kolmogorov, A. N. (1933), *Foundations of the Theory of Probability*, Springer.
- [11] Lakhdari, I. E. (2021), *Introduction aux probabilités et statistique descriptive*, Support de cours, Université Mohamed Khider Biskra, Année Universitaire 2021/2022.
- [12] Laplace, P. S. (1812), *Théorie Analytique des Probabilités*, Paris.
- [13] Lecoutre, J.-P. (2008), *Statistique et probabilités: Rappels de cours, QCM, exercices corrigés, sujets d'annales*, 4th ed., Dunod.
- [14] Montfort, A. (1997), *Introduction à la statistique*, Ecole Polytechnique, France.
- [15] Nemiche, M. (2015), *Exercices Corrigés: Statistique et Probabilités*, Faculté des Sciences d'Agadir (STU 3), Université Ibn Zohr Agadir.
- [16] Ross, S. M. (2019), *A First Course in Probability*, 10th ed., Pearson.

- [17] Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2008), *Mathematical Statistics with Applications*, 7th ed., Brooks/Cole.
- [18] Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2017), *Probability and Statistics for Engineers and Scientists*, 9th ed., Pearson.