MOHAMED KHIDER UNIVERSITY OF BISKRA
FACULTY OF NATURAL AND LIFE SCIENCES AND EARTH SCIENCES
DEPARTMENT OF COMMON CORE IN NATURAL AND LIFE SCIENCES

L2

BIOSTATISTICS

CHAPTER 1: DESCRIPTIVE STATISTICAL ANALYSIS

Prepared by:
Dr. Ghedjemis Fatiha

Academic Session: 2025/2026

Contents

Chapter 1

DESCRIPTIVE STATISTICAL ANALYSIS

1.1 Statistical Variables and Data Visualization

1.1.1 Introduction

Statistics is a scientific method used to reduce, analyze, and interpret large biological datasets. In the life sciences, we call this **Biostatistics**. It allows us to manage the inherent "biological variability" found in every species.

1.1.2 Core Statistical Concepts

To begin our study, we must define the fundamental units of any statistical analysis within a biological context.

Definition:

- **Statistical Population:** The entire group of biological entities (e.g., all individuals of a species, all cells in a culture) that we wish to study.
- **Sample:** A representative subset of the population actually observed during an experiment.
- **Individual:** The smallest entity on which measurements are taken (a single plant, a patient, or a leaf).
- **Characteristics and Modalities:** In biostatistics, a characteristic is any feature that allows us to identify and classify individuals. The different possible situations called categories or modalities. However, for a statistical study to be valid, the modalities of these traits must be mutually exclusive. This means that a single individual cannot belong to more than one modality for the same characteristic at once (for example, a cell cannot be both "infected" and "non-infected" in the same measurement).

Solved Example: Concepts

Example on Botany: A study on the growth of date palm trees (Phoenix dactylifera) in the Biskra oasis.

- **Population:** All date palm trees located in the Biskra region.
- **Sample:** 150 date palm trees randomly selected from three different farms.
- **Individual:** One single date palm tree.
- **Characteristic:** The total height of the tree (measured in meters).
- **Modalities:** Different heights:
  1. From 2 to 5 m
  2. From 5 to 8 m
  3. From 8 to 11 m

Example on Genetics: Observing the phenotypic expression of eye color in a laboratory population of fruit flies (Drosophila melanogaster).

- **Population:** All fruit flies produced in the laboratory's current breeding cycle.
- **Sample:** 400 fruit flies collected from a specific incubator.
- **Individual:** One single fruit fly.
- **Characteristic:** Eye color.
- **Modalities:**
  1. Red
  2. White
  3. Sepia

Example on Microbiology: Evaluating the contamination of a water source by counting bacterial colonies on agar plates.

- **Population:** The total volume of water in a specific reservoir (represented by all possible 1ml samples).
- **Sample:** 60 Petri dishes (each containing $1ml$ of the reservoir water).
- **Individual:** One single Petri dish.
- **Characteristic:** The number of Colony Forming Units (CFU).
- **Modalities:**
  1. $0, 1, 2, 3, ...n$ colonies.

Example on Physiology: Assessment of the severity of a viral infection in a group of hospital patients.

- **Population:** All patients diagnosed with the specific virus in Algeria during 2024.
- **Sample:** 120 patients admitted to the University Hospital of Biskra.
- **Individual:** One single patient.
- **Characteristic:** The severity of the clinical symptoms.
- **Modalities:** These are qualitative but follow a logical ranking:
  1. Asymptomatic
  2. Mild
  3. Moderate
  4. Severe

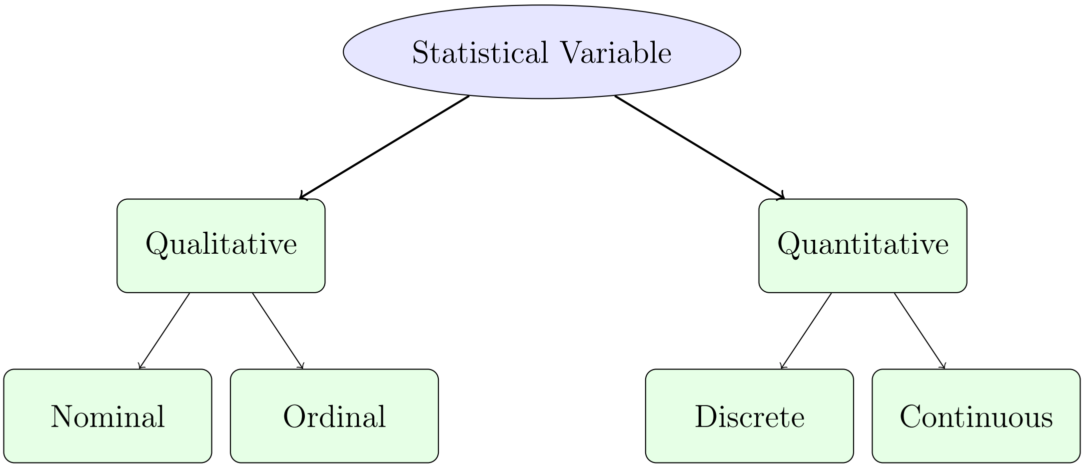
1.1.3 Classification of Variables

A **Characteristic** (or Variable) is the specific property we measure. In biology, these are classified based on their nature:

1. **Qualitative Variables:** Describe a state or attribute that cannot be measured numerically.
   - *Nominal:* Categories with no natural order, e.g. Blood types: A, B, AB, O; Eye color.
   - *Ordinal:* Categories with a logical progression, e.g., Disease severity: Mild, Moderate, Severe.
2. **Quantitative Variables:** Represent counts or measurements.
   - *Discrete:* Resulting from counting, e.g., Number of offspring in a litter; number of bacterial colonies.
   - *Continuous:* Resulting from measurement on a scale, e.g., Glucose concentration in mmol/L; height of a sunflower.

Before calculating parameters, we must identify the nature of the data.

Diagram 1.1: The Hierarchy of Biological Data



Solved Example: Data Classification

Identify the type of the following variables:

1. Species of trees in a forest.
2. pH level of soil samples.
3. Number of heartbeats per minute.
4. Stages of larval development (1st instar, 2nd instar, etc.).

---

1. Qualitative variable (nominal).
2. Quantitative variable (discrete).
3. Quantitative variable (continuous).
4. Qualitative variable (ordinal).

1.2 Analysis of Qualitative Variables

Qualitative variables describe attributes or states (e.g., Blood Type, Phenotype).

1.3 Definitions of Frequencies

Definition: Frequency for Qualitative Data

- **Absolute Frequency (n_i):** The number of individuals belonging to the i^{th} category.
- **Relative Frequency (f_i):** The proportion of the total sample in that category. $f_i = \frac{n_i}{n}$.
- **Cumulative Frequency ($F_i\nearrow$):** For **Ordinal** variables (like Infection Severity: Low $<$ Medium $<$ High), the cumulative frequency represents the proportion of individuals at or below a certain state.
  Note: For **Nominal** qualitative variables (like Eye Color), cumulative frequencies are **meaningless** because there is no order.

Table 1.1: Statistical Distribution of Educational Attainment

Satisfaction Level	n_i	f_i	$F_i\nearrow$
Very Dissatisfied	10	0.05	0.05
Dissatisfied	20	0.10	0.15
Somewhat Dissatisfied	30	0.15	0.30
Neutral	50	0.25	0.55
Somewhat Satisfied	40	0.20	0.75
Satisfied	35	0.175	0.925
Very Satisfied	15	0.075	1.000
Total	**200**	**1.000**	—

Example 1.1.

1.3.1 The Mode (M_o)

Definition: Mode

The Mode is simply the modality (category) with the **highest frequency** n_i.

Example 1.2. ***Bacterial Resistance***
Data: *100 colonies tested.*
Results: *Resistant (75), Sensitive (25).*

- ***Frequencies:*** $n_{Res} = 75$, $n_{Sen} = 25$.
- ***Rel. Frequencies:*** $f_{Res} = 0.75$, $f_{Sen} = 0.25$.

- ***cumulative relative frequencies:*** $F_{Res} \overset{\nearrow}{=} 0.75$, $F_{Sen} \overset{\nearrow}{=} 1$.
- ***Mode:*** *The highest frequency is* 75.
  Thus, $M_o =$ *Resistant.*

Example 1.3. ***Genetics***
Phenotype counts in $n = 400$ *flies: Normal (300), Mutant (100).*
1. Determine Frequencies and Relative Frequencies.
2. Calculate Mode.

1.

Phenotype	***Frequency*** (n)	***Relative Frequency*** (f)	***Percentage (%)***
Normal	*300*	$\frac{300}{400} = 0.75$	*75%*
Mutant	*100*	$\frac{100}{400} = 0.25$	*25%*
Total	$n = 400$	***1.00***	***100%***

2. ***The Mode*** *is the modality* ***Normal.***

1.3.2 Graphical Representations for Qualitative Data

Here are the primary types of graphs for qualitative variables:

1. Bar Chart

- **Description:** The most common representation for categorical data.
- **Definition:** Each modality is represented by a bar. The height of the bar corresponds to the category's frequency (n_i) or relative frequency (f_i).
- **Key Feature:** The bars are separated by spaces.

2. Pie Chart

- **Description:** Used to show the relationship of parts to a whole.
- **Definition:** A circle divided into "slices" or sectors. The size of each sector (the angle) is proportional to the relative frequency of the category.
- **Formula:** The angle of each slice is calculated as:

$$\text{Angle} = f_i \times 360°$$

- **Best Use:** It is most effective when there are few categories (usually fewer than 6).

Exercise 1: Genetics - Phenotypes

A researcher counts fruit flies by eye color: Red (140), White (40), Sepia (20). 1. Identify the variable type.
2. Determine the Mode.
3. Draw the pie chart and the bar chart.

Exercise Solution: Fruit Fly Eye Color

1. Variable Type The variable "Eye Color" is a **Qualitative Nominal** variable.

- It is *Qualitative* because it describes a quality or category (color).
- It is *Nominal* because there is no natural mathematical order between Red, White, and Sepia.

2. The Mode The Mode (M_o) is the modality with the highest frequency.

- Frequency of Red = 140
- Frequency of White = 40
- Frequency of Sepia = 20

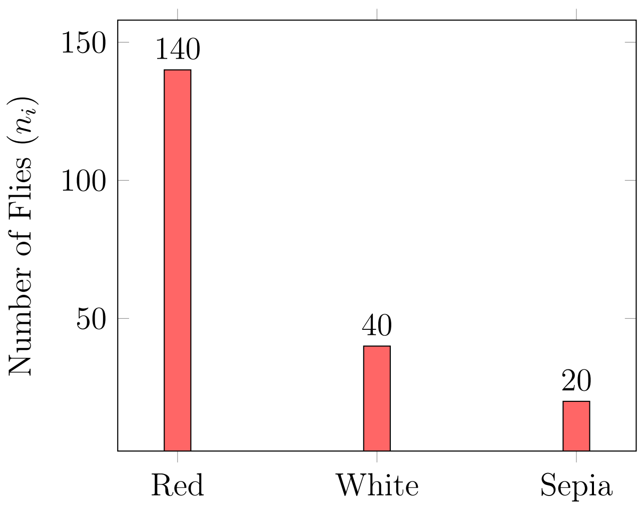
Mode: Red (since 140 is the highest frequency).

3. Calculations for Charts To draw the charts, we first calculate the total ($N = 200$), relative frequencies (f_i), and pie chart angles.

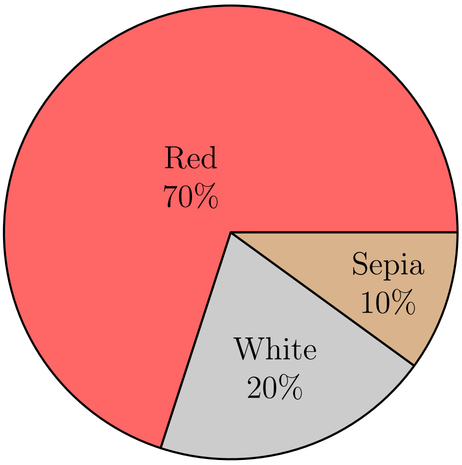
Eye Color	**Freq. (n_i)**	**Rel. Freq. (f_i)**	**Angle ($f_i \times 360°$)**
Red	140	0.70	252°
White	40	0.20	72°
Sepia	20	0.10	36°
Total	**200**	**1.00**	**360°**

4. Graphical Representations

Bar Chart

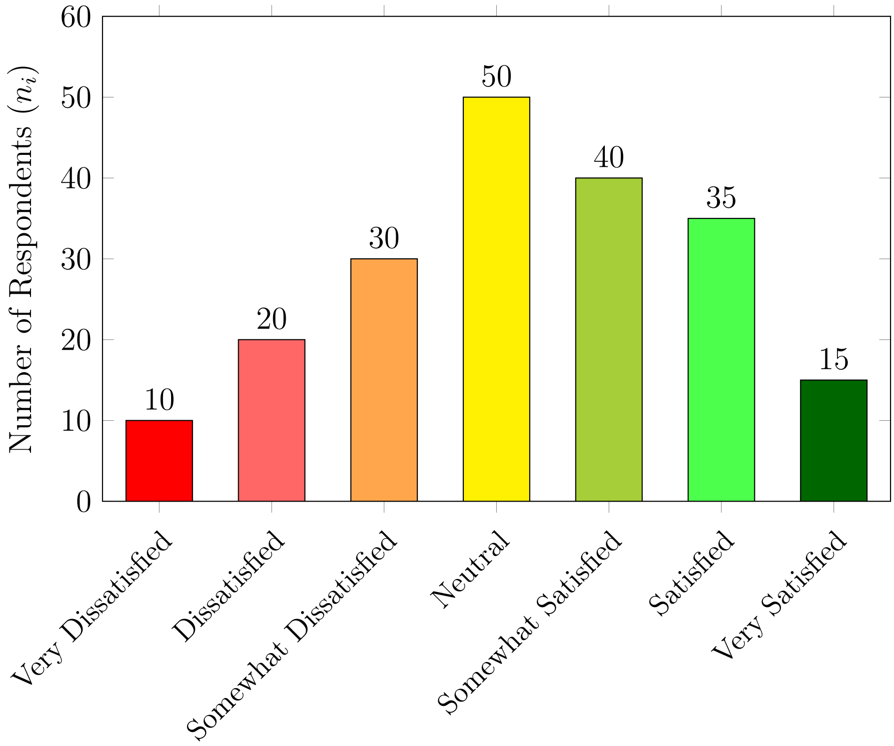


Pie Chart

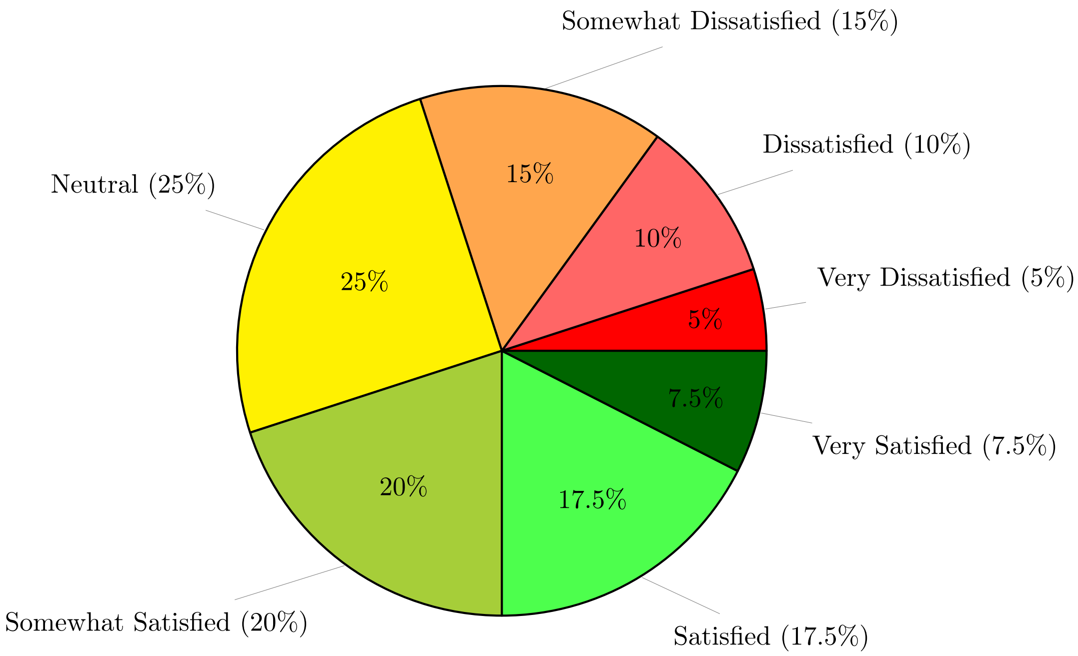

Red
70%
Sepia
10%
White
20%

Graphical Representation: for the Example 1.1 (Satisfaction Survey)

1. Bar Chart



2. Pie Chart The pie chart illustrates the relative frequency (f_i) for each satisfaction level.



1.4 Analysis of Discrete Quantitative Variables

Discrete variables result from counting (e.g., number of cells, number of seeds).

1.4.1 Definitions of Frequencies

Definition: Frequency for Discrete Data

- **Absolute Frequency (n_i):** Count of a specific integer value x_i.
- **Relative Frequency (f_i):** $f_i = n_i/n$.
- **Cumulative Frequency ($F_i\nearrow$):** The sum of relative frequencies up to value x_i. It represents the proportion of individuals with a value **less than or equal to** x_i.

$$F_i = f_1 + f_2 + \cdots + f_i$$

1.4.2 Parameters of Central Tendency

Calculation Method:

- **The Mode (M_o)** is the value x_i that occurs most often (highest n_i).
- **Mean ($\bar{x}$):** $\bar{x} = \sum n_i x_i/n$ or $\bar{x} = \sum f_i x_i$.
- **Median (M_e):** The first value where $F_i \geq 0.50$.
- **Quartiles (Q_1, Q_3):** First values where $F_i \geq 0.25$ and $F_i \geq 0.75$.

Solved Example: Seed Counting

Data: Number of seeds in 5 pods: $\{4, 6, 6, 7, 6\}$.

- **Frequencies:** $x = 4(n = 1)$, $x = 6(n = 3)$, $x = 7(n = 1)$.
- **Mode:** The value $x = 6$ appears 3 times (highest frequency). $M_o = 6$ seeds.
- **Mean** $\bar{x} = 4 \times 1/5 + 6 \times 3/5 + 7 \times 1/5 = 5.8$
- **Median**
  1. Step 1: Arrange data in ascending order:
  $$\{4, 6, 6, 6, 7\}$$
  2. Step 2: Find the position of the median: Since $n = 5$ is odd, the median is at the $(\frac{n+1}{2})$-th position:
  $$\text{Position} = \frac{5+1}{2} = 3\text{rd position}$$
  3. Step 3: Identify the value: The 3rd value in the ordered set is **6**

1.4.3 Dispersion Parameters

Definition:

- **Variance** (V): $\frac{\sum n_i x_i^2}{n} - \bar{x}^2$.
- **Standard Deviation** (σ): $\sqrt{V}$.
- **Coefficient of Variation** (Cv): $\sigma/\bar{x}$.
- **Range** (R) **:** It is the difference between the maximum and minimum values.

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

- **Inter-Quartile Range** (IQR) **:** It is the difference between the third quartile (Q_3) and the first quartile (Q_1).

$$IQR = Q_3 - Q_1$$

Exercise 2: Microbiology

Bacterial counts in 19 dishes:

x_i (Count)	10	11	12	13
n_i (Freq)	2	7	6	4

1. Determine Cumulative Frequencies.
2. Calculate Mean, Median, Mode.
3. Find Q_1, Q_3.
4. Give the dispersion parameters.

1. Frequency Distribution Table Given the total number of dishes $n = 19$.

x_i	n_i	f_i	F_i	$x_i \cdot n_i$	$x_i^2 \cdot n_i$
10	2	0.105	0.105	20	200
11	7	0.368	0.473	77	847
12	6	0.316	0.789	72	864
13	4	0.211	1.000	52	676
Total	**19**	**1.000**	–	**221**	**2587**

2. A. Mean ($\bar{x}$) The mean is calculated using the formula:

$$\bar{x} = \frac{\sum(x_i \cdot n_i)}{n} = \frac{221}{19} \approx 11.63$$

B. Mode The mode is the value with the highest frequency (n_i).

- Highest frequency is 7.

- This corresponds to $x_i = 11$.
- Mode = 11.

C. Median (Q_2) Median = 12.

3. Quartiles (Q_1, Q_3) The quartiles are found by looking at the Cumulative Frequency $(F_i\nearrow)$.

- $Q_1 = 11$.
- $Q_3 = 12$.

4. Dispersion parameters

A. Variance (σ^2)

The variance is calculated using the formula:

$$\sigma^2 = \frac{\sum(x_i^2 \cdot n_i)}{n} - \bar{x}^2$$

$$\sigma^2 = \frac{2587}{19} - (11.6316)^2$$

$$\sigma^2 = 136.1579 - 135.2941 = 0.8638$$

B. Standard Deviation (σ)

The standard deviation is the square root of the variance:

$$\sigma = \sqrt{\sigma^2} = \sqrt{0.8638} \approx 0.9294$$

C. Coefficient of Variation (CV)

$$CV = \left(\frac{\sigma}{\bar{x}}\right) \cdot 100$$

$$CV = \left(\frac{0.9294}{11.6316}\right) \cdot 100 \approx 7.99\%$$

1.4.4 Graphical Representations

Definition:

1. **Frequency Bar Chart:** A graph where the x-axis represents the discrete values (x_i) and the y-axis represents the frequencies (n_i or f_i). Because the data is discrete, the bars are drawn separated (not touching) to indicate that the variable is not continuous.

2. **Cumulative Frequency Step Function (Staircase Graph):** This graph plots the cumulative frequency (F_i). The graph consists of horizontal segments that "jump" at each discrete value of x, resembling a staircase. It is used to determine percentiles, medians, and quartiles.

Example:

A biologist counts the number of specific protein structures (x_i) in a sample of $n = 19$ cells.

Frequency Distribution Table Using the counts provided: $n_1 = 2, n_2 = 7, n_3 = 6, n_4 = 4$.

x_i **(Structures)**	n_i	f_i	F_i	$x_i \cdot n_i$	$x_i^2 \cdot n_i$
10	2	0.105	0.105	20	200
11	7	0.368	0.473	77	847
12	6	0.316	0.789	72	864
13	4	0.211	1.000	52	676
Total	**19**	**1.000**	–	**221**	**2587**

Central Tendency Parameters

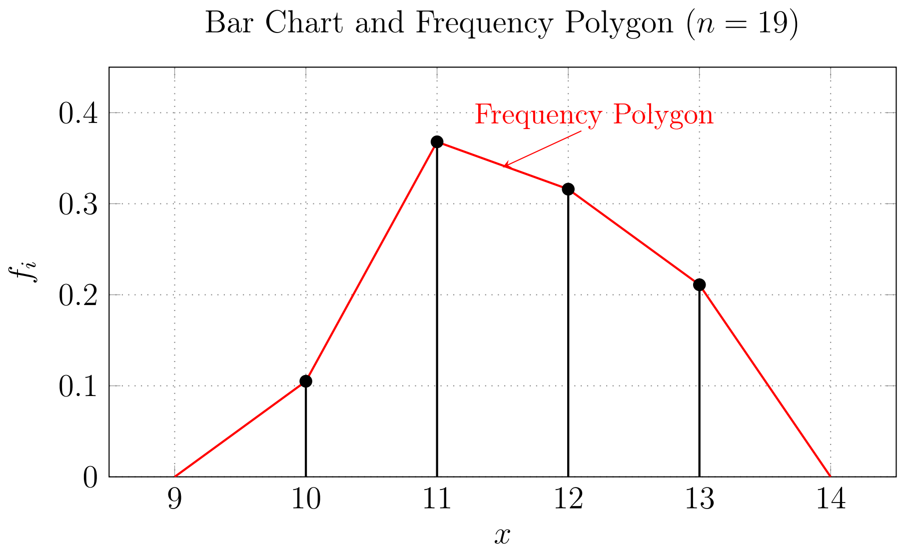
- **Mean** ($\bar{x}$)**:** $\bar{x} = \frac{\sum(x_i \cdot n_i)}{n} = \frac{221}{19} \approx 11.63$
- **Mode** (Mo)**:** The value with the highest frequency ($n_i = 7$) is $x_i = 11$.
- **Median** (Me)**:** For $n = 19$, the position is $\frac{19+1}{2} = 10^{th}$ value. Looking at F_i, the 10^{th} value is 12.
- **Quartiles:** Q_1 (5^{th} value) is 11; Q_3 (15^{th} value) is 12.

Dispersion Parameters

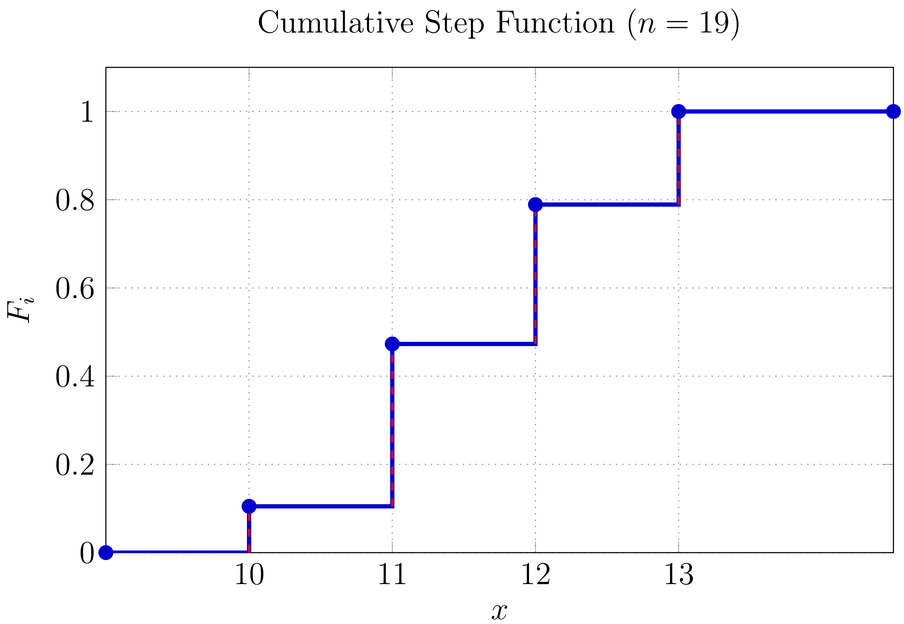
- **Variance** (σ^2)**:** $\sigma^2 = \frac{\sum x_i^2 n_i}{n} - \bar{x}^2 = \frac{2587}{19} - (11.6316)^2 = 0.8638$
- **Standard Deviation** (σ)**:** $\sigma = \sqrt{0.8638} \approx 0.929$
- **Coefficient of Variation** (CV)**:** $CV = (\frac{0.929}{11.63}) \times 100 \approx 7.99\%$

Graphical Representations of the Exercise

Figure 1.1: Bar chart and Frequency Polygon



This "staircase" graph represents the cumulative relative frequencies (F_i) for the discrete variable.



1.5 Analysis of Continuous Quantitative Variables

Continuous variables result from measurements and can take any value within an interval (e.g., height, weight, temperature, concentration).

1.5.1 Definitions and Grouping

Definition: Grouping into Classes

When dealing with continuous data, we group values into **intervals** or **classes** $[L_i, L_{i+1}[$.

- **Midpoint (c_i):** The midpoint of the interval: $c_i = \frac{L_i+L_{i+1}}{2}$.
- **Class Width (w):** The difference between the upper and lower limits.

1.5.2 Grouping Data: Sturge's Rule

When raw continuous data is collected, it must be organized into classes. To determine the optimal number of intervals, we follow a systematic process.

Definition: The Grouping Process

1. **Range (R):** The difference between the maximum and minimum values.

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

2. **Number of Classes (k):** Calculated using **Sturge's Rule**:

$$k = 1 + 3.322 \log_{10}(n)$$

(Note: k is usually rounded to the nearest whole number.)

3. **Class Width (w):** The size of each interval.

$$w = \frac{R}{k}$$

Solved Example: Grouping Leaf Lengths

A researcher measures the length of $n = 30$ leaves. The minimum length is 5 cm and the maximum is 20 cm.

- **Step 1: Range**

$$R = 20 - 5 = 15 \text{ cm}$$

- **Step 2: Number of Classes**

$$k = 1 + 3.322 \log_{10}(30) \approx 1 + 3.322(1.477) \approx 5.9$$

We choose $k = 6$ classes.

- **Step 3: Class Width**

$$w = \frac{15}{6} = 2.5 \text{ cm}$$

Resulting Intervals: $[5, 7.5[, [7.5, 10[, [10, 12.5[, [12.5, 15[, [15, 17.5[, [17.5, 20]$.

Example 1.4. ***Biomass Measurement Analysis***

Data set: $\{11.0, 12.2, 13.0, 13.4, 14.8, 15.5, 15.5, 16.2\}$ *($n = 8$)*

1. Calculations (Sturge's Rule)

- ***Number of classes (k):***

$$k = 1 + 3.322 \log_{10}(8) \approx 1 + 3.322(0.903) \approx 3.99 \rightarrow \mathbf{4}$$

- ***Range (R):***

$$R = Max - Min = 16.2 - 11.0 = \mathbf{5.2}$$

- ***Class Width (w):***

$$w = \frac{R}{k} = \frac{5.2}{4} = \mathbf{1.3}$$

2. Frequency Distribution Table

Table 1.2: Frequency and Relative Frequency of Plant Weights

Weight Class (g)	Frequency (f)	Relative Freq.	Percentage
$[11.0, 12.3[$	2	0.250	25.0%
$[12.3, 13.6[$	2	0.250	25.0%
$[13.6, 14.9[$	1	0.125	12.5%
$[14.9, 16.2]$	3	0.375	37.5%
Total	**8**	**1.000**	**100%**

Exercise 4: Data Organization

You have measured the weight of $n = 50$ lab mice. The lightest mouse is 18g and the heaviest is 34g.
1. Calculate the Range (R).
2. Use Sturge's Rule to find the number of classes (k).
3. Determine the class width (w).

Solution:

1. **Range:** $R = 34 - 18 = 16$g.
2. **Classes (k):** $k = 1 + 3.322 \log_{10}(50) \approx 1 + 3.322(1.699) \approx 6.64$. We round to $k = 7$.
3. **Width (w):** $w = 16/7 \approx 2.28$. We can round this to a convenient value like $w = 2.3$ or 2.5.

1.5.3 Central Parameters Tendency for Continuous Data

Calculation Method:

- **Mean ($\bar{x}$):** Calculated using middle points: $\bar{x} = \frac{\sum n_i c_i}{n}$.
- **Mode:** The **modal class** is the interval with the highest frequency.

$$M_o = L + A \times \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right)$$

  L: lower bound, A: amplitude, $\Delta_1 = n_{modal} - n_{prev}$, $\Delta_2 = n_{modal} - n_{next}$.

- **Median (M_e):** Found by linear interpolation within the first class $[a_i, a_{i+1}[$ where $F_i \geq 0.50$:

$$M_e = a_i + \left(\frac{0.50 - F_{i-1}}{f_i} \right) \times w$$

  Where a_i is the lower limit of the median class.

- **First Quartile** Found by linear interpolation within the first class $[a_i, a_{i+1}[$ where $F_i \geq 0.25$:

$$Q_1 = a_i + \left(\frac{0.25 - F_{i-1}}{f_i} \right) \times w$$

  Where a_i is the lower limit of the first quartile class.

- **Third Quartile** Found by linear interpolation within the first class $[a_i, a_{i+1}[$ where $F_i \geq 0.75$:

$$Q_3 = a_i + \left(\frac{0.75 - F_{i-1}}{f_i} \right) \times w$$

  Where a_i is the lower limit of the third quartile class.

Exercise 3: Ecology

Height of $n = 20$ seedlings (in cm):

Interval (cm)	[0, 5[	[5, 10[	[10, 15[	[15, 20[
Frequency (n_i)	3	8	6	3

1. Calculate the central tendency parameters.

Solution: **Frequency Table** Total $n = 20$.

Interval	c_i	n_i	f_i	F_i	$n_i \cdot c_i$
[0, 5[	2.5	3	0.15	0.15	7.5
[5, 10[	7.5	8	0.40	0.55	60.0
[10, 15[	12.5	6	0.30	0.85	75.0
[15, 20[	17.5	3	0.15	1.00	52.5
Total	–	**20**	**1.00**	–	**195**

1. Mean ($\bar{x}$) Using the midpoints (c_i):

$$\bar{x} = \frac{\sum n_i c_i}{n} = \frac{(3 \times 2.5) + (8 \times 7.5) + (6 \times 12.5) + (3 \times 17.5)}{20}$$

$$\bar{x} = \frac{7.5 + 60 + 75 + 52.5}{20} = \frac{195}{20} = \mathbf{9.75} \text{ cm}$$

2. Mode (M_o) The modal class is $[5, 10[$ because it has the highest frequency ($n_i = 8$).

- $L = 5,\ A = 5$
- $\Delta_1 = n_{modal} - n_{prev} = 8 - 3 = 5$
- $\Delta_2 = n_{modal} - n_{next} = 8 - 6 = 2$

$$M_o = 5 + 5 \times \left(\frac{5}{5+2}\right) = 5 + 5 \times 0.714 = \mathbf{8.57} \text{ cm}$$

3. Median (M_e): Target $F_i \geq 0.50 \rightarrow$ Class $[5, 10[$. ($a_i = 5, F_{i-1} = 0.15, f_i = 0.40, w = 5$)

$$M_e = 5 + \left(\frac{0.50 - 0.15}{0.40}\right) \times 5 = 5 + \left(\frac{0.35}{0.40}\right) \times 5 = \mathbf{9.375}$$

4. First Quartile (Q_1): Target $F_i \geq 0.25 \rightarrow$ Class $[5, 10[$. ($a_i = 5, F_{i-1} = 0.15, f_i = 0.40, w = 5$)

$$Q_1 = 5 + \left(\frac{0.25 - 0.15}{0.40}\right) \times 5 = 5 + \left(\frac{0.10}{0.40}\right) \times 5 = \mathbf{6.25}$$

5. Third Quartile (Q_3): Target $F_i \geq 0.75 \rightarrow$ Class $[10, 15[$. ($a_i = 10, F_{i-1} = 0.55, f_i = 0.30, w = 5$)

$$Q_3 = 10 + \left(\frac{0.75 - 0.55}{0.30}\right) \times 5 = 10 + \left(\frac{0.20}{0.30}\right) \times 5 = \mathbf{13.33}$$

1.5.4 Dispersion Parameters for Continuous Data

Calculation Method:

- **Range (R):** Difference between the maximum and minimum values.

$$R = \text{Upper Bound}_{max} - \text{Lower Bound}_{min}$$

- **Interquartile Range (IQR):** The spread of the middle 50% of the data.

$$IQR = Q_3 - Q_1$$

- **Variance (s^2):** The average of the squared deviations from the mean.

$$s^2 = \sum f_i(c_i - \bar{x})^2 \quad \text{or} \quad s^2 = \left(\sum f_i c_i^2\right) - \bar{x}^2$$

- **Standard Deviation (s):** The square root of the variance.

$$s = \sqrt{s^2}$$

- **Coefficient of Variation (CV):** Measures relative dispersion as a percentage.

$$CV = \left(\frac{s}{\bar{x}}\right) \times 100\%$$

Application: Ecology Example ($n = 20$, $\bar{x} = 9.75$)
We use the midpoints (c_i) and relative frequencies (f_i) calculated previously:

Interval	f_i	c_i	$(c_i - \bar{x})^2$	$f_i(c_i - \bar{x})^2$
[0, 5[	0.15	2.5	52.5625	7.884
[5, 10[	0.40	7.5	5.0625	2.025
[10, 15[	0.30	12.5	7.5625	2.269
[15, 20[	0.15	17.5	60.0625	9.009
Total	**1.00**			**21.187**

1. Range and Interquartile Range

- **Range:** $R = 20 - 0 = \mathbf{20}$ cm
- **IQR:** $IQR = Q_3 - Q_1 = 13.33 - 6.25 = \mathbf{7.08}$ cm

2. Variance and Standard Deviation

- **Variance (s^2):**

$$s^2 = \sum f_i(c_i - 9.75)^2 = \mathbf{21.187}$$

- **Standard Deviation (s):**

$$s = \sqrt{21.187} \approx \mathbf{4.60} \text{ cm}$$

3. Coefficient of Variation (CV)

$$CV = \left(\frac{4.60}{9.75}\right) \times 100 = \mathbf{47.18\%}$$

1.5.5 Graphical Representations (Continuous)

Definition:

1. **Histogram:** Adjacent rectangles where the area is proportional to the frequency. Unlike the bar chart, there are **no gaps** between bars.
2. **Frequency Polygon:** Created by connecting the midpoints (c_i) of the top of the histogram bars.
3. **Cumulative Curve:** A continuous line graph connecting the points (L_{i+1}, F_i). It shows the "flow" of data.

Figure 2.1: Histogram and Frequency Polygon

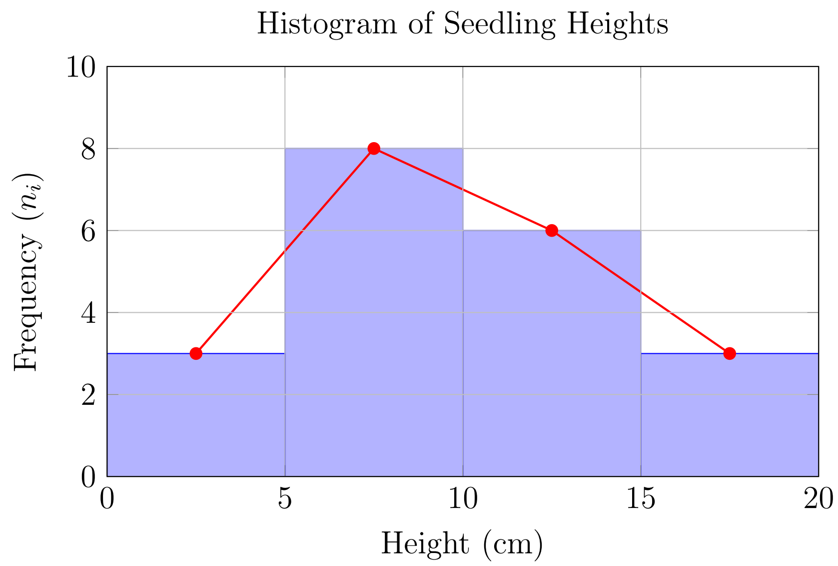


Figure 2.2: Cumulative Frequency Curve