STATISTICAL LABORATORY

Lisbon School of Economics & Management
Universidade de Lisboa

Applied Mathematics for Economics and Management
1st Year/1st Semester
2025/2026

CONTACT

Professor: Elisabete Fernandes
E-mail: efernandes@iseg.ulisboa.pt

https://doity.com.br/estatistica-aplicada-a-nutricao

https://basiccode.com.br/produto/informatica-basica/

PROGRAM

1. Fundamental Concepts of Statistics

2. Exploratory Data Analysis

3. Organizing and Summarizing Data

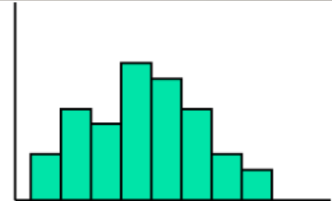4. Association and Relationships Between Variables
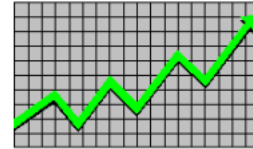
5. Index Numbers

6. Time Series Analysis

LECTURE 2: EXPLORATORY DATA ANAYSIS

DESCRIPTIVE STATISTICS

- Present data
  - e.g., Tables and graphs
- Summarize data
  - e.g., Sample mean $= \dfrac{\sum X_i}{n}$

Newbold et al (2013)

DATA REPRESENTATION
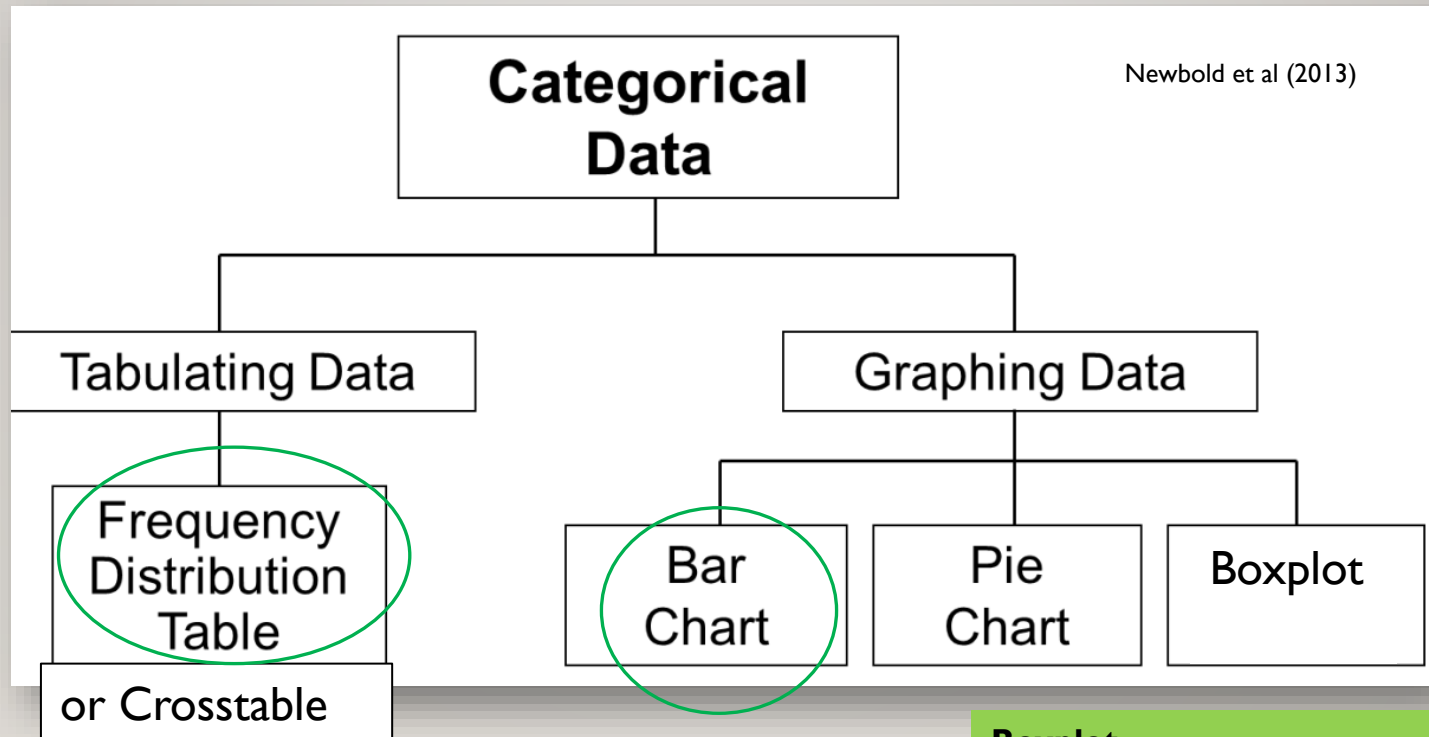

- **Tables**: frequency distributions.

- **Graphs**: bar chart, pie chart, histogram, boxplot, line chart, etc.

- **Choice depends on**: type of variable & analysis purpose.

TABLES AND GRAPHS FOR CATEGORICAL VARIABLES

Newbold et al (2013)

Categorical Data

- Tabulating Data
  - Frequency Distribution Table (or Crosstable)
- Graphing Data
  - Bar Chart
  - Pie Chart
  - Boxplot

- **Frequency Table**: shows the distribution of one categorical variable. **Example:** Gender (Male: 40 and Female: 60).
- **Contingency Table / Crosstabulation**: shows the joint distribution of two categorical variables. **Example:** Gender × Education level.

Boxplot
- Not suitable for **nominal variables**, because they **lack order**.
- Can be used for **numerical variables** or **ordinal variables with five or more categories**.

FREQUENCY DISTRIBUTION

Frequency Distribution

A **frequency distribution** is a table used to organize data. The left column (called classes or groups) includes all possible responses on a variable being studied. The right column is a list of the frequencies, or number of observations, for each class. A **relative frequency distribution** is obtained by dividing each frequency by the number of observations and multiplying the resulting proportion by 100%.

Newbold et al (2013)

FREQUENCY DISTRIBUTION TABLE EXAMPLE

Relative Frequeny(percentage)

Summarize data by category
Example: Hospital Patients by Unit

Absolute Frequeny

Hospital Unit	Number of Patients	Percent (rounded)
Cardiac Care	1,052	11.93
Emergency	2,245	25.46
Intensive Care	340	3.86
Maternity	552	6.26
Surgery	4,630	52.50
Total:	8,819	100.0

(Variables are categorical)

Newbold et al (2013)

BAR CHART EXAMPLE

Hospital Unit	Number of Patients
Cardiac Care	1,052
Emergency	2,245
Intensive Care	340
Maternity	552
Surgery	4,630
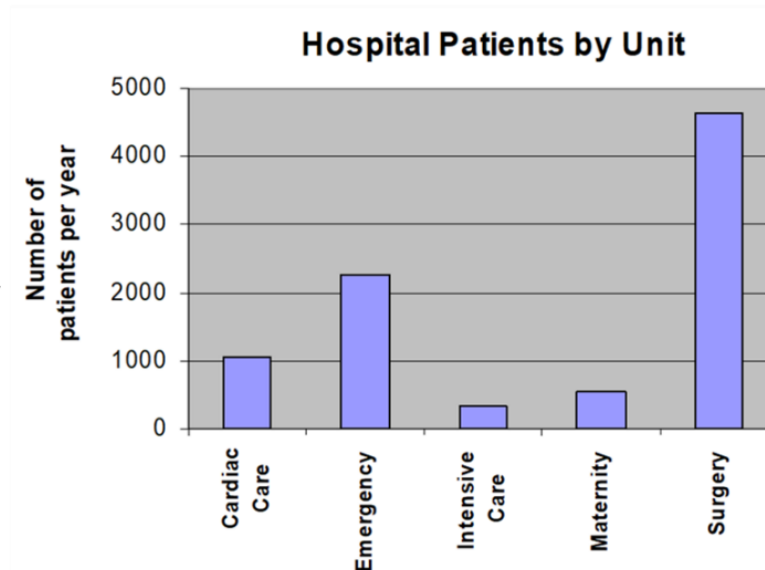
Newbold et al (2013)

PIE CHART EXAMPLE (SUPPLEMENTARY)

Hospital Unit	Number of Patients	% of Total
Cardiac Care	1,052	11.93
Emergency	2,245	25.46
Intensive Care	340	3.86
Maternity	552	6.26
Surgery	4,630	52.50
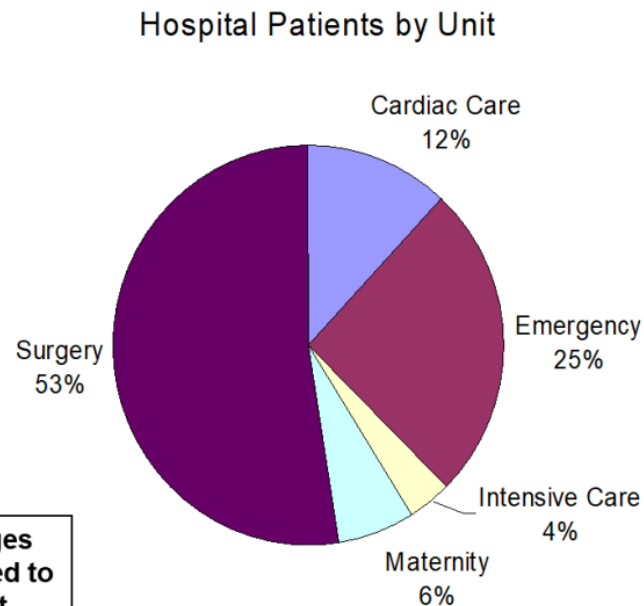
(Percentages are rounded to the nearest percent)

Hospital Patients by Unit

Cardiac Care 12%

Emergency 25%

Intensive Care 4%

Maternity 6%

Surgery 53%

Newbold et al (2013)

- **Bar charts and Pie charts** are often used for qualitative (categorical) data.
- Height of bar or size of pie slice shows the frequency or percentage for each category.

EXERCISE: FREQUENCY TABLE & BAR CHARTS

🖍️ **Exercise**

A survey was conducted among 30 university students about their preferred mode of transportation to campus. The responses are shown below:

🚶 Walking, 🚴 Bicycle, 🚌 Bus, 🚗 Car, 🚇 Metro   or underground

Data (sample):

Bus, Car, Walking, Bus, Metro, Car, Car, Bicycle, Walking, Bus,

Bus, Car, Metro, Car, Walking, Walking, Bicycle, Bus, Car, Metro,

Walking, Car, Bicycle, Car, Bus, Bus, Metro, Walking, Walking, Car

---

Tasks:

1. **Construct a frequency table** for the categorical variable *Mode of Transportation*. Include:
   - Absolute frequency
   - Relative frequency (perc
2. **Draw a bar chart** to represent the distribution of transportation modes.

EXERCISE 1: SOLUTION

✅ Answer: Frequency Table

$$Relative\ Frequency = \frac{Absolute\ Frequency}{n}$$

Mode of Transportation	Absolute Frequency	Relative Frequency
Bus	7	0.233
Car	9	0.300
Walking	7	0.233
Metro/Underground	4	0.133
Bicycle	3	0.100
Total	30	1.000

EXERCISE 1: SOLUTION

✅ Answer: Frequency Table with Absolute and Relative Cumulative Frequencies (supplementary)

Mode of Transportation	Absolute Frequency	Relative Frequency	Cumulative Absolute Frequency	Cumulative Relative Frequency
Bus	7	0.233	7	0.233
Car	9	0.300	16	0.533
Walking	7	0.233	23	0.766
Metro/Underground	4	0.133	27	0.899
Bicycle	3	0.100	30	0.999 ≈ 1.000
Total	30	1.000	—	—

EXERCISE 2: SOLUTION

✅ Answer: Bar Chart (Absolute Frequency)
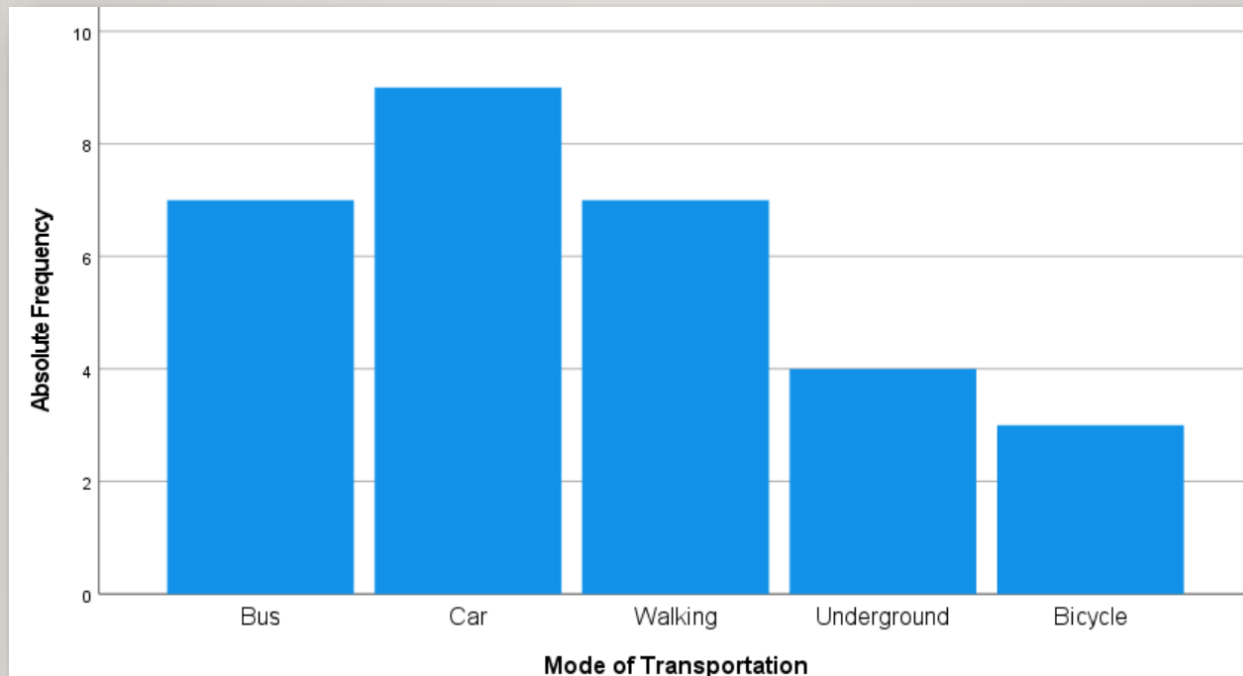
SPSS output

EXERCISE 2: SOLUTION
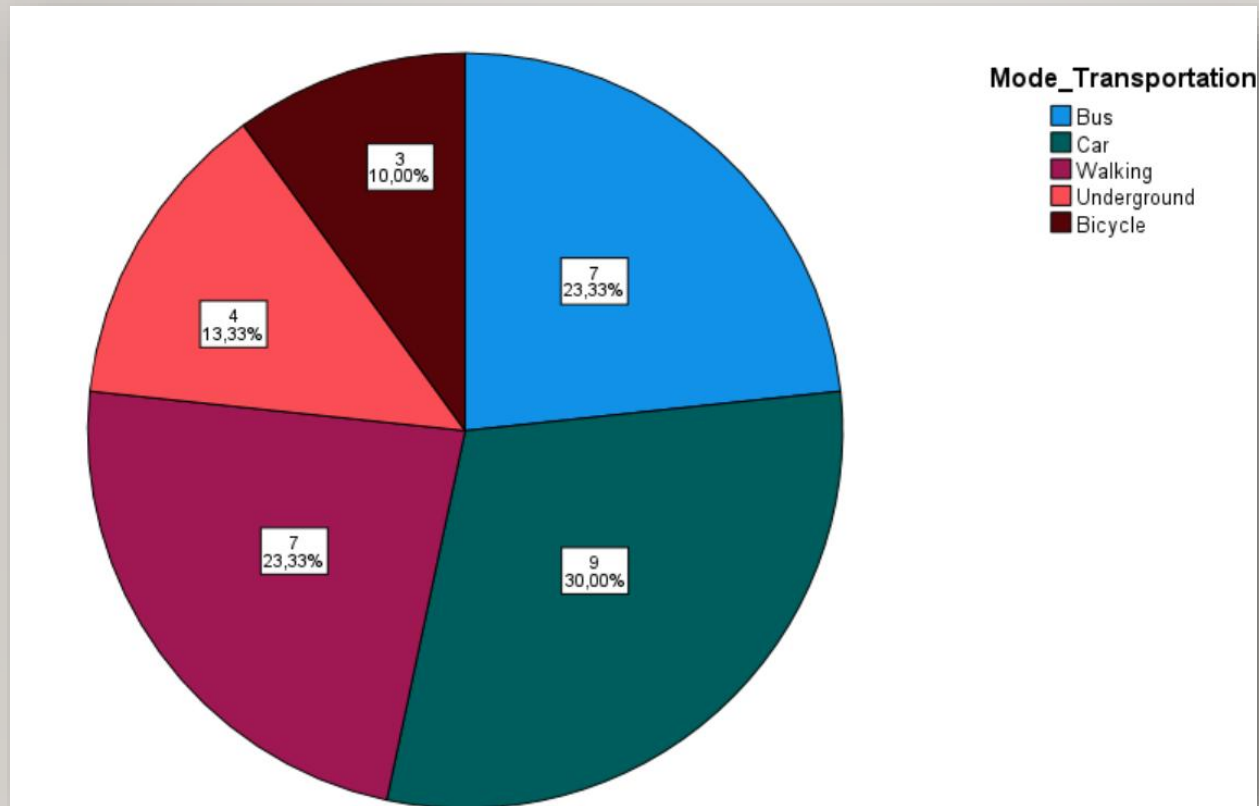
SPSS output

GRAPHS TO DESCRIBE NUMERICAL VARIABLES

Numerical Data

Frequency Distributions and Cumulative Distributions

Stem-and-Leaf Display

Boxplot

Histogram — Continuous Variables

Ogive
Bar Chart — Discrete Variables
Scatterplot

Newbold et al (2013)

- **Univariate (one variable):** Stem-and-Leaf, Boxplot, Histogram, and Ogive
- **Bivariate (two variables):** Scatterplot
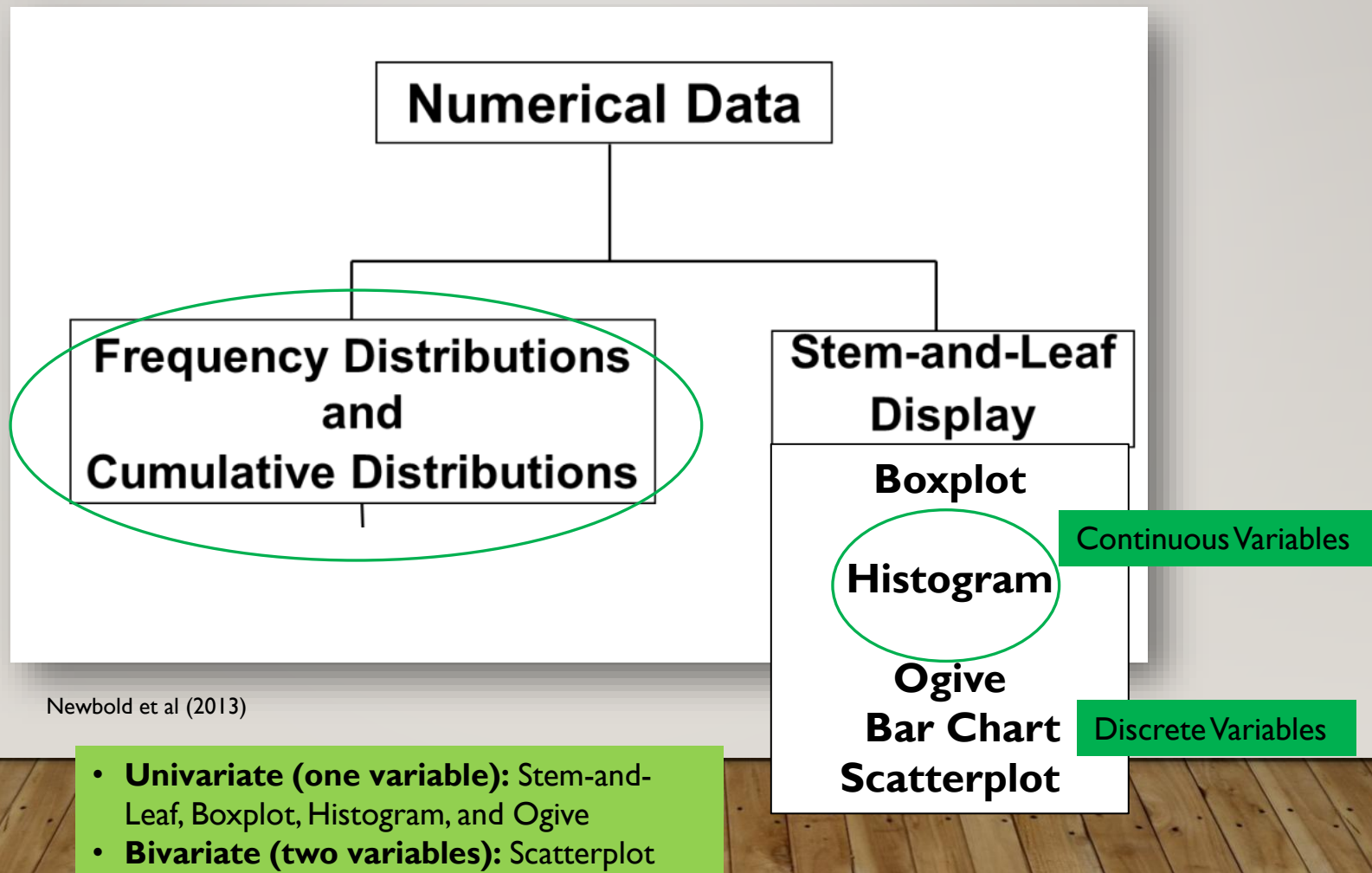
RULES FOR BUILDING CLASSES (FREQUENCY TABLE & HISTOGRAM)

1. Number of classes (k):

Sturges' Rule:

The number of classes k for a histogram is the **smallest integer** such that:

$$2^k \geq n$$

where n is the sample size.

or

"Square root rule for the number of classes: The number of classes (k) can be estimated as the square root of the number of observations (n):

$$k \approx \sqrt{n}$$

where n is the total number of data points."

2. Class width (h):

$$h = \frac{\max - \min}{k}$$

- Always round class width, h, upward.
- Classes must be inclusive and nonoverlapping.

3. Where to Start the First Class? (3 options)

- At the minimum observed value
  - e.g. data from 12 to 87, $k = 8, h \approx 9.4$
  - Classes: [12, 21.4), [21.4, 30.8), ...
- Round down to a "nice" number (e.g. multiple of 5 or 10)
  - Easier to read → start at 10, width = 10
  - Classes: [10, 20), [20, 30), ...
- General rule
  - Lower limit ≤ minimum
  - Upper limit ≥ maximum

CLASS INTERVALS

- Each class grouping has the same width

- Determine the width of each interval by

$$w = \text{interval width} = \frac{\text{largest number} - \text{smallest number}}{\text{number of desired intervals}}$$

- Use at least 5 but no more than 15-20 intervals

- Intervals never overlap

- Round up the interval width to get desirable interval endpoints

Newbold et al (2013)

FREQUENCY DISTRIBUTION EXAMPLE

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Class Interval	Class Midpoint	Absolute Frequency	Relative Frequency
[10, 20]	15	3	0.150
]20, 30]	25	7	0.350
]30, 40]	35	4	0.200
]40, 50]	45	4	0.200
]50, 60]	55	2	0.100
Total	—	20	1.000

Sample size: n = 20
Number of classes: k = 5 (Sturges´ Rule)
Width of each interval = (Max-Min)/k = (58 -12) / 5 = 9.2 ~ 10

Sturges' Rule:

The number of classes k for a histogram is the **smallest integer** such that:

$$2^k \geq n$$

where n is the sample size.

HISTOGRAM EXAMPLE

Class Interval	Class Midpoint	Absolute Frequency	Relative Frequency
[10, 20]	15	3	0.150
]20, 30]	25	7	0.350
]30, 40]	35	4	0.200
]40, 50]	45	4	0.200
]50, 60]	55	2	0.100
Total	—	20	1.000

- If all class widths are equal, histogram heights can simply represent absolute or relative frequencies.
  If class widths differ, using absolute frequency ÷ class width as the bar height ensures that the area of each bar correctly represents the absolute frequency.

(No gaps between bars)

Absolute Frequency

Classes

Temperature in Degrees

EXERCISE 1.32

1.32 Consider the following data:

17	62	15	65
28	51	24	65
39	41	35	15
39	32	36	37
40	21	44	37
59	13	44	56
12	54	64	59

a. Construct a frequency distribution.
b. Construct a histogram.

Newbold et al (2013)

EXERCISE 1.32 A): SOLUTION

✅ Answer:

Class	Midpoint	Absolute Frequency	Relative Frequency
[10,21]	15.5	6	6/28 ≈ 0.214
]21,32]	27	2	2/28 ≈ 0.071
]32,43]	38	9	9/28 ≈ 0.321
]43,54]	49	4	4/28 ≈ 0.143
]54,65]	60	7	7/28 ≈ 0.250

Sample size: n = 28
Number of classes: k = 5 (Sturges´s Rule)
Width of each interval = (Max-Min)/k = (65 -12) / 5 = 10.6 ~ 11

EXERCISE 1.32 B): SOLUTION

✅ Answer:

Class	Midpoint	Absolute Frequency	Relative Frequency
[10,21]	15.5	6	$6/28 \approx 0.214$
]21,32]	27	2	$2/28 \approx 0.071$
]32,43]	38	9	
]43,54]	49	4	
]54,65]	60	7	

Excel output

Absolute Frequency

Classes

EXERCISE 1.36

1.36 The following table shows the ages of competitors in a charity tennis event in Rome:

Age	Percent
18–24	18.26
25–34	16.25
35–44	25.88
45–54	19.26
55+	20.35

a. Construct a relative cumulative frequency distribution.
b. What percent of competitors were under the age of 35?
c. What percent of competitors were 45 or older?

Newbold et al (2013)

EXERCISE 1.36 A): SOLUTION

✅ Answer:

a) Relative cumulative frequency distribution

To get the **cumulative percentage**, we add the percentages successively:

Age	Percent	Cumulative Percent
18–24	18.26	18.26
25–34	16.25	18.26 + 16.25 = 34.51
35–44	25.88	34.51 + 25.88 = 60.39
45–54	19.26	60.39 + 19.26 = 79.65
55+	20.35	79.65 + 20.35 = 100.00

The **cumulative percentage column** is the **relative cumulative frequency**.

EXERCISE 1.36 B): SOLUTION

✅ Answer:

b) Percent of competitors under the age of 35
- Competitors under 35 include **18–24** and **25–34**.
- Add the percentages:

$$18.26 + 16.25 = 34.51\%$$

34.51% of competitors were under 35.

EXERCISE 1.36 C): SOLUTION

✅ Answer:

c) Percent of competitors 45 or older

- Competitors 45 or older include **45–54** and **55+**.
- Add the percentages:

$$19.26 + 20.35 = 39.61\%$$

39.61% of competitors were 45 or older.

THANKS!

Questions?