STATISTICAL METHODS

Master in Industrial Management,

Operations and Sustainability (MIMOS)

2nd 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. Descriptive Data Analysis
3. Introduction to Inferential Analysis
4. Parametric Hypothesis Testing
5. Non-Parametric Hypothesis Testing
6. Linear Regression Analysis

HOMEWORK OF LECTURE 2: SOLUTIONS

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.
c. Construct an ogive.
d. Construct a stem-and-leaf display.

Newbold et al (2013)

EXERCISE 1.32 A): SOLUTION

✅ Answer:

Class Interval	Absolute Frequency	Relative Frequency	Cumulative Absolute Frequency	Cumulative Relative Frequency
[10, 19)	5	0.179	5	0.179
[19, 28)	2	0.071	7	0.250
[28, 37)	4	0.143	11	0.393
[37, 46)	8	0.286	19	0.679
[46, 55)	2	0.071	21	0.750
[55, 66)	7	0.250	28	1.000

Since the upper limit of the last class was 64, the value 65 would not be included.
To solve this, we chose the following approach:

If you are working with integer data, a very simple solution is to set the last boundary as **max + 1** (here 65+1=66) and keep the class intervals in the form [a,b).

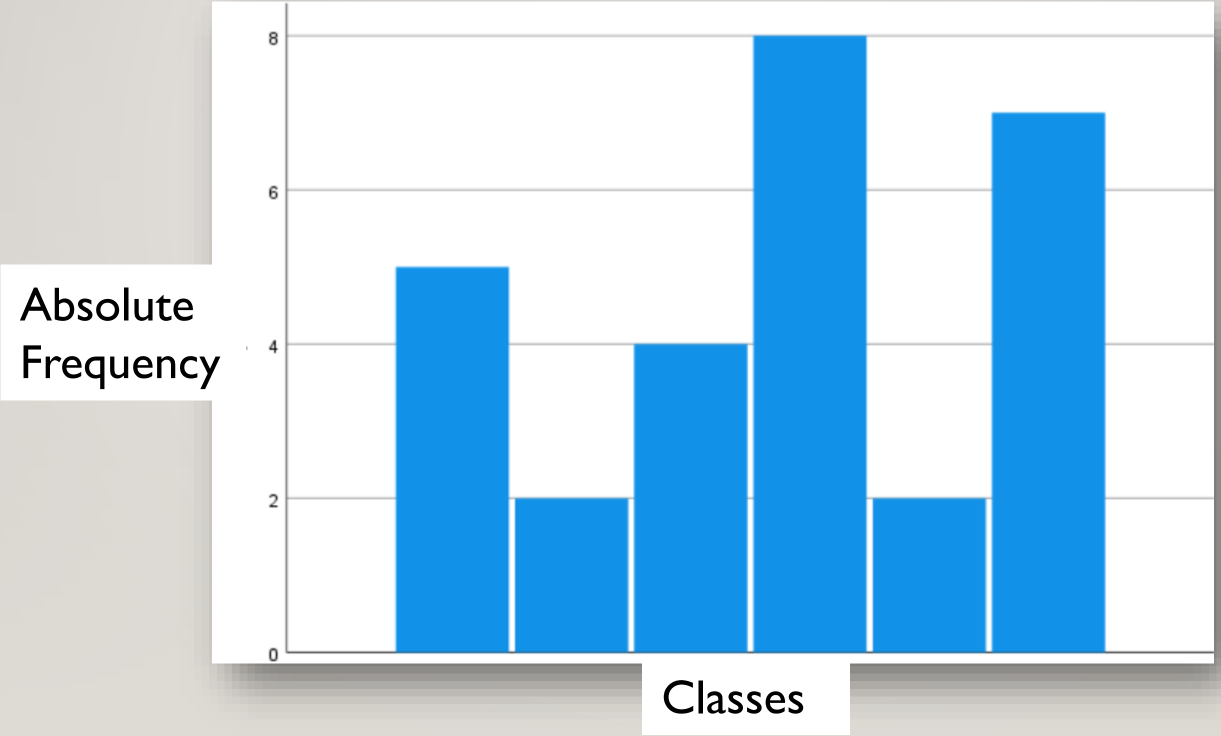
n = 28 (sample size)
k = $\sqrt{28} = 5.29 \sim 6$ (number of classes)

Width of each interval = (65 -12) / 6 = 8.8 ~ 9

EXERCISE 1.32 B): SOLUTION

✅ Answer:

SPSS Output: Histogram



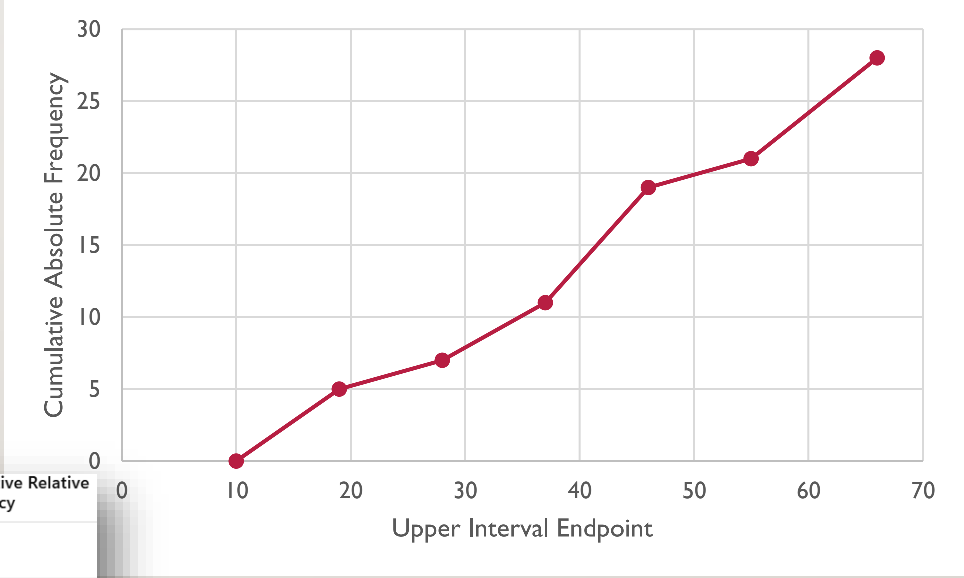
EXERCISE 1.32 C): SOLUTION

Answer:

Upper interval endpoint	Cum Freq
10	0
19	5
28	7
37	11
46	19
55	21
56	28

Class Interval	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
[10, 19)	5	0.179	5	0.179
[19, 28)	2	0.071	7	0.250
[28, 37)	4	0.143	11	0.393
[37, 46)	8	0.286	19	0.679
[46, 55)	2	0.071	21	0.750
[55, 66)	7	0.250	28	1.000

Excel Output: Ogive



Drag the **upper class boundaries** to the **x-axis**.
Drag the **cumulative frequency** to the **y-axis**.

EXERCISE 1.32 D): SOLUTION

✅ Answer:

Step 1 – Data

Copiar código

```
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
```

Step 2 – Determine stems

- Tens digit → **stem**
- Units digit → **leaf**

Step 3 – Assign leaves to stems (unsorted)

- Stem 1 → Leaves: 7, 5, 5, 3, 2 (frequency = 6)
- Stem 2 → Leaves: 8, 4, 1 (frequency = 3)
- Stem 3 → Leaves: 9, 5, 2, 6, 7, 7, 9 (frequency = 7)
- Stem 4 → Leaves: 1, 0, 4, 4 (frequency = 4)
- Stem 5 → Leaves: 1, 4, 6, 9, 9 (frequency = 5)
- Stem 6 → Leaves: 2, 5, 5, 4 (frequency = 4)

EXERCISE 1.32 D): SOLUTION

✅ Answer:

Step 4 – Order leaves within each stem

- Stem 1 → 2, 3, 5, 5, 5, 7
- Stem 2 → 1, 4, 8
- Stem 3 → 2, 5, 6, 7, 7, 9, 9
- Stem 4 → 0, 1, 4, 4
- Stem 5 → 1, 4, 6, 9, 9
- Stem 6 → 2, 4, 5, 5

Step 5 – Final stem-and-leaf plot with frequencies

markdown

```
Stem | Leaf          | Frequency
-------------------------------
1    | 2 3 5 5 5 7   | 6
2    | 1 4 8         | 3
3    | 2 5 6 7 7 9 9 | 7
4    | 0 1 4 4       | 4
5    | 1 4 6 9 9     | 5
6    | 2 4 5 5       | 4
```

Step 6 – Interpretation

- Most data are concentrated in the 30s.
- Frequencies show which stems contain more values.
- Leaves are ordered to make it easier to read the distribution.

EXERCISE 1.32 D): SOLUTION

Answer:

SPSS Output: Steam and Leaf Plot

d) Stem-and-Leaf Display

```
Stem | Leaf          | Frequency
------------------------------
1    | 2 3 5 5 5 7   | 6
2    | 1 4 8         | 3
3    | 2 5 6 7 7 9 9 | 7
4    | 0 1 4 4       | 4
5    | 1 4 6 9 9     | 5
6    | 2 4 5 5       | 4
```

👉 Here, the **stem** is the tens digit, and the **leaf** is the ones digit.

For example, `1 | 2 3 5 5 7` means 12, 13, 15, 15, 17.

Data Gráfico de Ramos e Folhas

Frequency	Stem &	Leaf
5,00	1 .	23557
3,00	2 .	148
7,00	3 .	2567799
4,00	4 .	0144
5,00	5 .	14699
4,00	6 .	2455

Largura do ramo: 10

Cada folha: 1 caso(s)

EXERCISE 2.8

2.8 The ages of a sample of 12 students enrolled in an on-line macroeconomics course are as follows:

21	22	27	36	18	19
22	23	22	28	36	33

a. What is the mean age for this sample?
b. Find the median age.
c. What is the value of the modal age?

Newbold et al (2013)

EXERCISE 2.8 A): SOLUTION

✅ Answer:

Data (12 students):

21, 22, 27, 36, 18, 19, 22, 23, 22, 28, 36, 33

a. Mean age

$$\text{Mean} = \frac{\text{Sum of all ages}}{n}$$

$$\text{Sum} = 21 + 22 + 27 + 36 + 18 + 19 + 22 + 23 + 22 + 28 + 36 + 33$$

$$= 307$$

$$\text{Mean} = \frac{307}{12} \approx 25.58$$

✅ Mean ≈ **25.6 years**

EXERCISE 2.8 B): SOLUTION

✅ Answer:

b. Median age

1. Order data:

   18, 19, 21, 22, 22, 22, 23, 27, 28, 33, 36, 36

2. Since $n = 12$ (even), median = average of 6th and 7th values.

- 6th = 22
- 7th = 23

Position of Median = (n+1)/2 = (12+1)/2 = 6.5

$$\text{Median} = \frac{22 + 23}{2} = 22.5$$

✅ Median = **22.5 years**

EXERCISE 2.8 C): SOLUTION

✅ Answer:

Data (12 students):

21, 22, 27, 36, 18, 19, 22, 23, 22, 28, 36, 33

c. Modal age

The mode is the most frequent value.

- 22 appears **3 times** (most frequent).

✅ Mode = **22 years**

EXERCISE 2.14

2.14 Calculate the coefficient of variation for the following sample data:

10 8 11 7 9

Newbold et al (2013)

EXERCISE 2.14: SOLUTION

Answer:

1. Sample mean:

$$\bar{x} = \frac{10+8+11+7+9}{5} = \frac{45}{5} = 9$$

2. Sample variance (denominator $n-1=4$):

$$s^2 = \frac{(10-9)^2+(8-9)^2+(11-9)^2+(7-9)^2+(9-9)^2}{4}$$

$$s^2 = \frac{1+1+4+4+0}{4} = \frac{10}{4} = 2.5$$

3. Standard deviation:

$$s = \sqrt{2.5} \approx 1.58$$

4. Coefficient of Variation (CV):

$$CV = \frac{s}{\bar{x}} \times 100 = \frac{1.58}{9} \times 100 \approx 17.6\%$$

The coefficient of variation is **17.6%**.

EXERCISE 2.15

2.15 The ages of a random sample of people who attended a recent soccer match are as follows:

23	35	14	37	38	15	45
12	40	27	13	18	19	23
37	20	29	49	40	65	53
18	17	23	27	29	31	42
35	38	22	20	15	17	21

a. Find the mean age.
b. Find the standard deviation.
c. Find the coefficient of variation.

Newbold et al (2013)

EXERCISE 2.15 A): SOLUTION

Answer:

Data (n = 30):

23, 35, 14, 37, 38, 15, 45, 12, 40, 27, 13, 18, 19, 23, 37,
20, 29, 49, 40, 65, 53, 18, 17, 23, 27, 29, 31, 42, 35, 38, 22, 20, 15, 17, 21

a. Mean age

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\sum x_i = 863 \quad \text{and} \quad n = 30$$

$$\bar{x} = \frac{863}{30} = 28.77$$

Mean = 28.77 years

EXERCISE 2.15 B): SOLUTION

Answer:

Data (n = 30):

23, 35, 14, 37, 38, 15, 45, 12, 40, 27, 13, 18, 19, 23, 37,

20, 29, 49, 40, 65, 53, 18, 17, 23, 27, 29, 31, 42, 35, 38, 22, 20, 15, 17, 21

b. Standard deviation

Formula:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Step 1: Compute squared deviations.

$$\sum (x_i - \bar{x})^2 = 4636.90$$

Step 2: Divide by $n - 1 = 29$.

$$\frac{4636.90}{29} = 161.62$$

Step 3: Square root.

$$s = \sqrt{161.62} = 12.70$$

Standard deviation = 12.70 years

EXERCISE 2.15 C): SOLUTION

Answer:

Data (n = 30):

23, 35, 14, 37, 38, 15, 45, 12, 40, 27, 13, 18, 19, 23, 37,
20, 29, 49, 40, 65, 53, 18, 17, 23, 27, 29, 31, 42, 35, 38, 22, 20, 15, 17, 21

c. Coefficient of variation

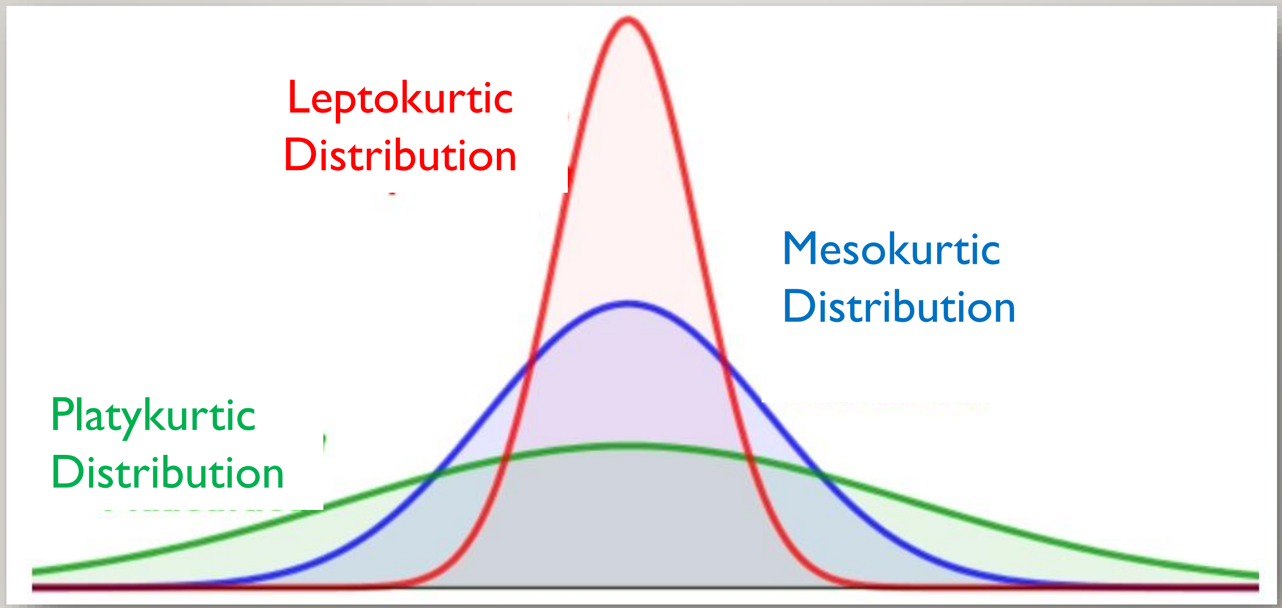
$$CV = \frac{s}{\bar{x}} \times 100$$

$$CV = \frac{12.70}{28.77} \times 100 = 44.15\%$$

Coefficient of variation = 44.15%

LECTURE 3: DESCRIPTIVE DATA ANALYSIS – MEASURES (CONTINUED)

KURTOSIS VISUALIZATION



KURTOSIS MEASURES: HOW TO IDENTIFY KURTOSIS WITH COEFFICIENTS

Measure	Formula	Interpretation
Quantile-based coefficient of kurtosis (Yule)	$K_Y = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$	$K_Y = 0.263$ → **Mesokurtic** (normal) $K_Y > 0.263$ → **Platykurtic** (flat) $K_Y < 0.263$ → **Leptokurtic** (peaked)
Moment coefficient of kurtosis (SPSS)	$K = \frac{\frac{1}{n}\sum(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum(x_i - \bar{x})^2\right)^2} - 3$	$K = 0$ → **Mesokurtic** $K > 0$ → **Leptokurtic** (heavy tails) $K < 0$ → **Platykurtic** (light tails)

Platykurtic
Ky > 0.263
K < 0

Mesokurtic
Ky = 0.263
K = 0

Leptokurtic
Ky < 0.263
K > 0

ADVANTAGES AND DISADVANTAGES OF KURTOSIS MEASURES

- **Quantile-based Coefficient of Kurtosis or Yule's Coefficient of Kurtosis (Q-Kurtosis)**
  ✅ Resistant to outliers, simple interpretation, non-parametric.
  ❌ Less precise for small samples, ignores detailed tail behavior.

- **Moment-based Coefficient of Kurtosis (SPSS / G2)**
  ✅ Uses all data points, standardized for statistical tests, widely available in software.
  ❌ Sensitive to outliers, more complex calculation, may exaggerate tail influence in small samples.

Yule's coefficient of kurtosis is considered nonparametric because it is based on quantiles rather than on the parameters or moments of a specific theoretical distribution.

EXERCISE: SKEWNESS AND KURTOSIS

A random sample of 12 daily sales (in hundreds of dollars) from a store is recorded as follows

Data: 6, 8, 10, 12, 14, 9, 11, 7, 13, 11, 15, 10

Questions:

a. Compute the **mean, median, and mode**.

b. Compute the **skewness** using:

- Pearson's coefficient: $(\text{Mean} - \text{Mode})/s$

c. Compute the **kurtosis (flattening)** using Quantile-based Coefficient

$$\text{Kurtosis} = \frac{Q3 - Q1}{2 \cdot (P90 - P10)}$$

EXERCISE: SOLUTION

Answer:

Step 1 – Organize the data (ascending)

6, 7, 8, 9, 10, 10, 11, 11, 12, 13, 14, 15

Step 2 – a. Mean, Median, Mode

- **Mean** $\bar{x} = 10.92$
- **Median** = 10.5
- **Mode** = 10 and 11 (bimodal)

Step 3 – b. Skewness

$$\text{Skewness} = \frac{10.92 - 10.5}{2.69} \approx 0.16$$

Step 4 – c. Kurtosis (quartile-based)

- Q1 = 8.5, Q3 = 12.5, P10 ≈ 7, P90 ≈ 14

$$\text{Kurtosis} = \frac{12.5 - 8.5}{2 \cdot (14 - 7)} = \frac{4}{14} \approx 0.29$$

- Interpretation: **platykurtic (flattened distribution)**

Summary:

- Slight positive skew (Skewness > 0)
- Platykurtic distribution (Kurtosis > 0.263)

MEASURES OF RELATIONSHIPS BETWEEN VARIABLES

Two measures of the relationship between variable are

- Covariance
  - a measure of the direction of a linear relationship between two variables
- Correlation Coefficient
  - a measure of both the direction and the strength of a linear relationship between two variables

– Only concerned with the strength of the relationship

– No causal effect is implied

Newbold et al (2013)

COVARIANCE

- The population covariance:

$$\text{Cov}(x,y) = \sigma_{xy} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N}$$

- The sample covariance:

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

  – Only concerned with the strength of the relationship
  – No causal effect is implied

Newbold et al (2013)

INTERPRETING COVARIANCE

- **Covariance** between two variables:

$\text{Cov}(x, y) > 0 \rightarrow$ *x* and *y* tend to move in the same direction

$\text{Cov}(x, y) < 0 \rightarrow$ *x* and *y* tend to move in opposite directions

$\text{Cov}(x, y) = 0 \rightarrow$ *x* and *y* are independent

Newbold et al (2013)

COEFFICIENT OF CORRELATION

- Measures the relative strength of the linear relationship between two variables
- Population correlation coefficient:

$$\rho = \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y}$$

- Sample correlation coefficient:

$$r = \frac{\mathrm{Cov}(x, y)}{s_x s_y}$$

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]}}$$

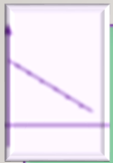
Newbold et al (2013)

FEATURES OF CORRELATION COEFFICIENT, R

- Unit free
- Ranges between −1 and 1
- The closer to −1, the stronger the negative linear relationship
- The closer to 1, the stronger the positive linear relationship
- The closer to 0, the weaker any positive linear relationship

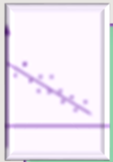
Newbold et al (2013)

INTERPRETING CORRELATION COEFFICIENT



r = -1

- Perfect negative linear correlation

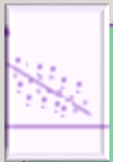


-1 < r ≤ -0,8

- Strong negative linear correlation

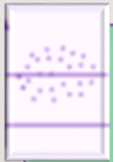
-0,8 < r ≤ -0,5

- Moderate negative linear correlation



-0,5 < r < 0

- Weak negative linear correlation



r = 0

- No linear correlation

0 < r < 0,5

- Weak positive linear correlation

0,5 ≤ r < 0,8

- Moderate positive linear correlation

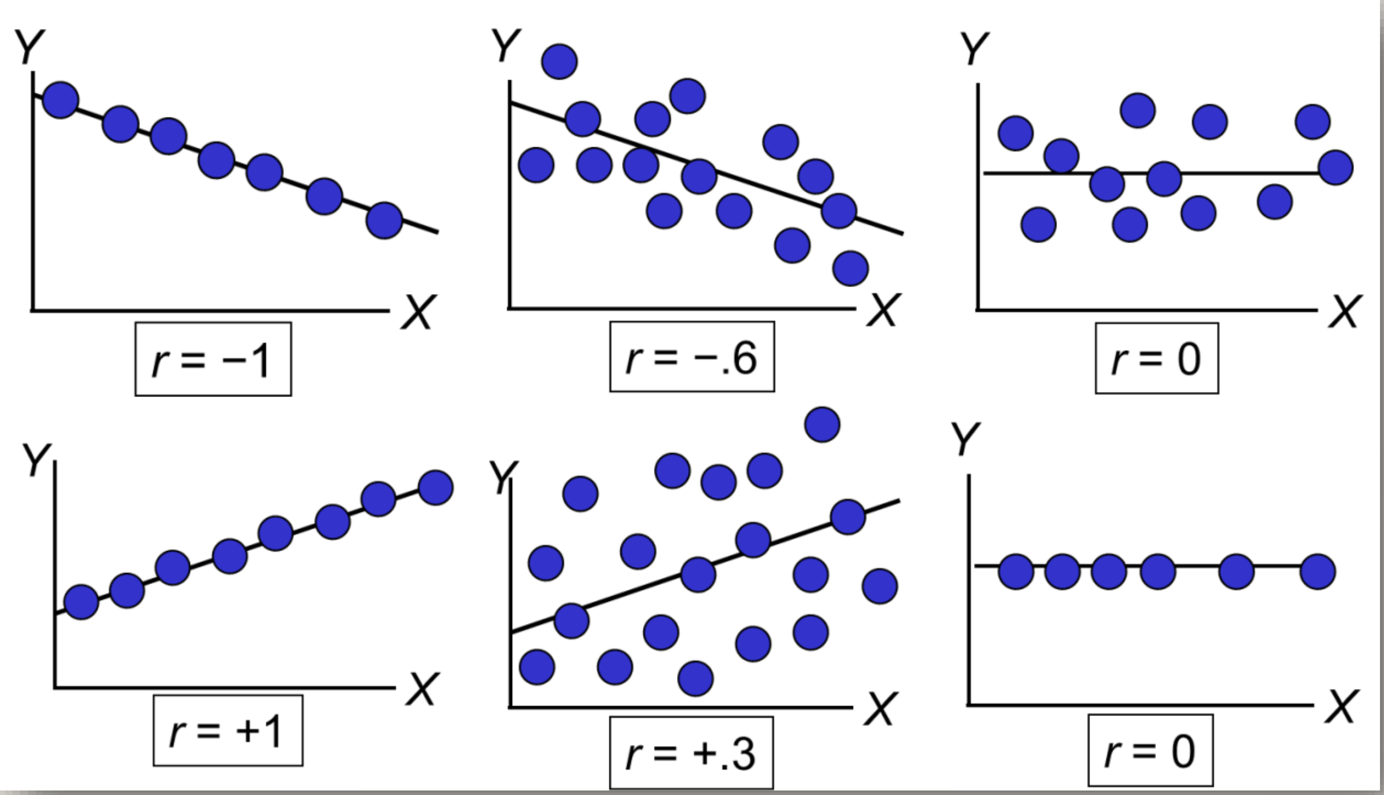
0,8 ≤ r < 1

- Strong positive linear correlation

r = 1

- Perfect positive linear correlation

SCATTER PLOTS OF DATA WITH VARIOUS CORRELATION COEFFICIENTS



Newbold et al (2013)

EXERCISE 2.55

2.55 A random sample for five exam scores produced the following (hours of study, grade) data values:

Hours Studied (x)	Test Grade (y)
3.5	88
2.4	76
4	92
5	85
1.1	60

a. Compute the covariance.
b. Compute the correlation coefficient

Newbold et al (2013)

EXERCISE 2.55: SOLUTION

✅ Answer:

a. Covariance

$$\mathrm{Cov}(x,y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

- Compute means: $\bar{x} = 3.2$, $\bar{y} = 80.2$
- Compute deviations and products, sum them → 74.6
- Divide by $n - 1 = 4$:

$$\mathrm{Cov}(x,y) \approx 18.65$$

---

b. Correlation Coefficient

$$r = \frac{\mathrm{Cov}(x,y)}{s_x s_y}$$

- Sample standard deviations: $s_x \approx 1.53$, $s_y \approx 13.4$

$$r \approx \frac{18.65}{1.53 \cdot 13.4} \approx 0.91$$

Summary:

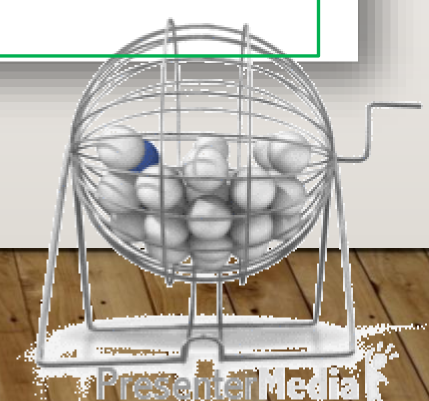
- Covariance ≈ 18.65
- Correlation = r ≈ 0.91 (strong positive linear relationship, 0.8 ≤ r < 1)

LECTURE 3: PROBABILITY

CONCEPTS OF PROBABILITIES

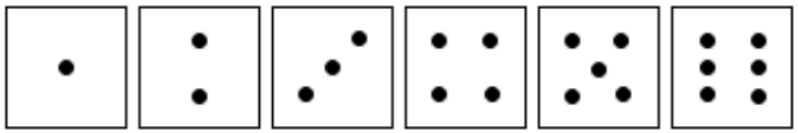
- Random Experiment – a process leading to an uncertain outcome
- Basic Outcome – a possible outcome of a random experiment
- Sample Space (*S*) – the collection of all possible outcomes of a random experiment
- Event (*E*) – any subset of basic outcomes from the sample space

Newbold et al (2013)



EXAMPLE: SAMPLE SPACE AND EVENTS

Let the Sample Space be the collection of all possible outcomes of rolling one die:



We can use either curly brackets { } or square brackets [ ] to define a set.

Sample Space S or Ω

$S = [1, 2, 3, 4, 5, 6]$

$S = \{1, 2, 3, 4, 5, 6\}$

Let *A* be the event "Number rolled is even"

Let *B* be the event "Number rolled is at least 4"

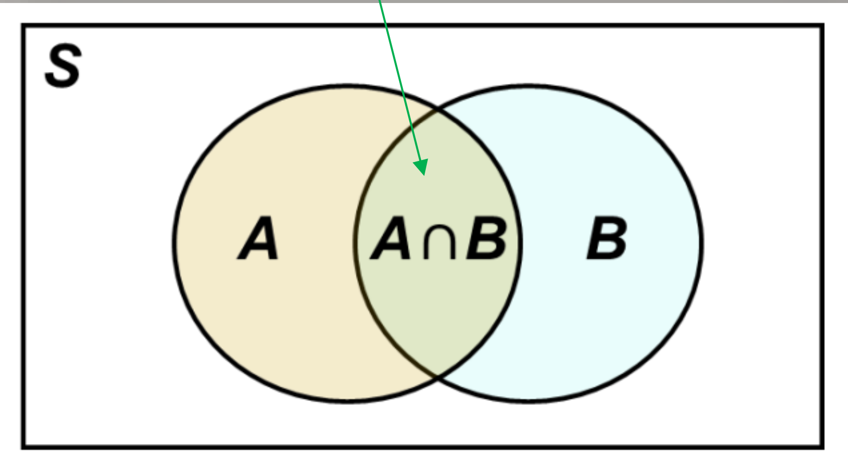
Then

Event A $A = [2, 4, 6]$ **and** $B = [4, 5, 6]$ Event B

Newbold et al (2013)

INTERSECTION OF EVENTS

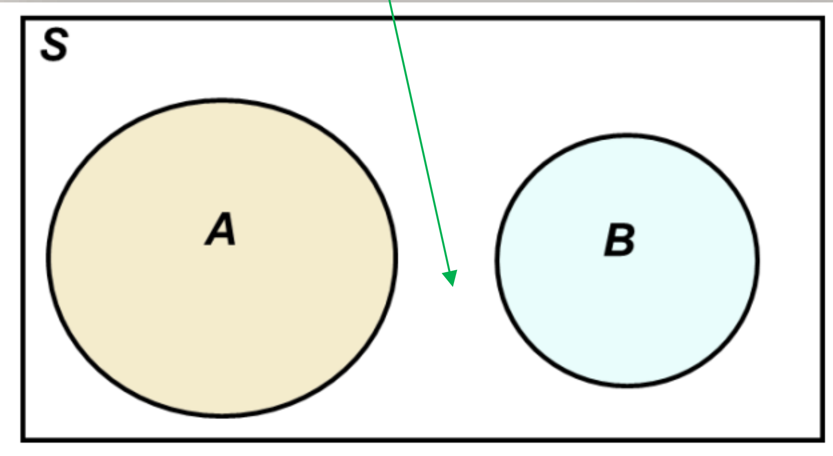
- Intersection of Events – If *A* and *B* are two events in a sample space *S*, then the intersection, $A \cap B$, is the set of all outcomes in *S* that belong to both *A* and *B*



Newbold et al (2013)

MUTUALLY EXCLUSIVE EVENTS

- *A* and *B* are Mutually Exclusive Events if they have no basic outcomes in common
  - i.e., the set $A \cap B$ is empty

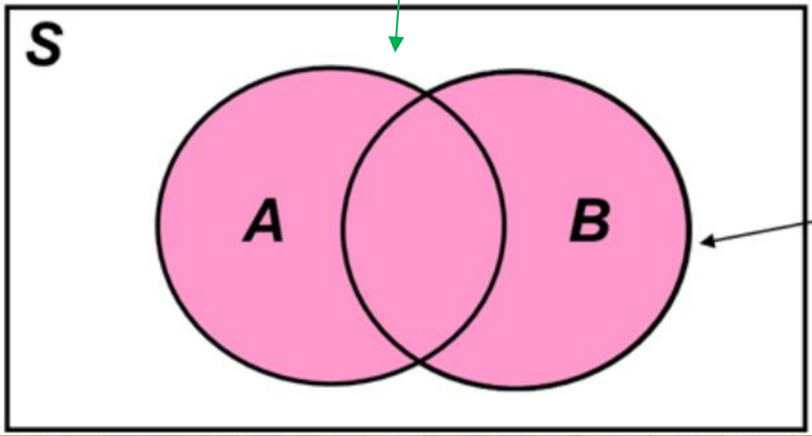


$$\boldsymbol{A \cap B = \{\ \}}$$

Newbold et al (2013)

UNION OF EVENTS

- Union of Events – If *A* and *B* are two events in a sample space *S*, then the union, $A \cup B$, is the set of all outcomes in *S* that belong to either *A* or *B*

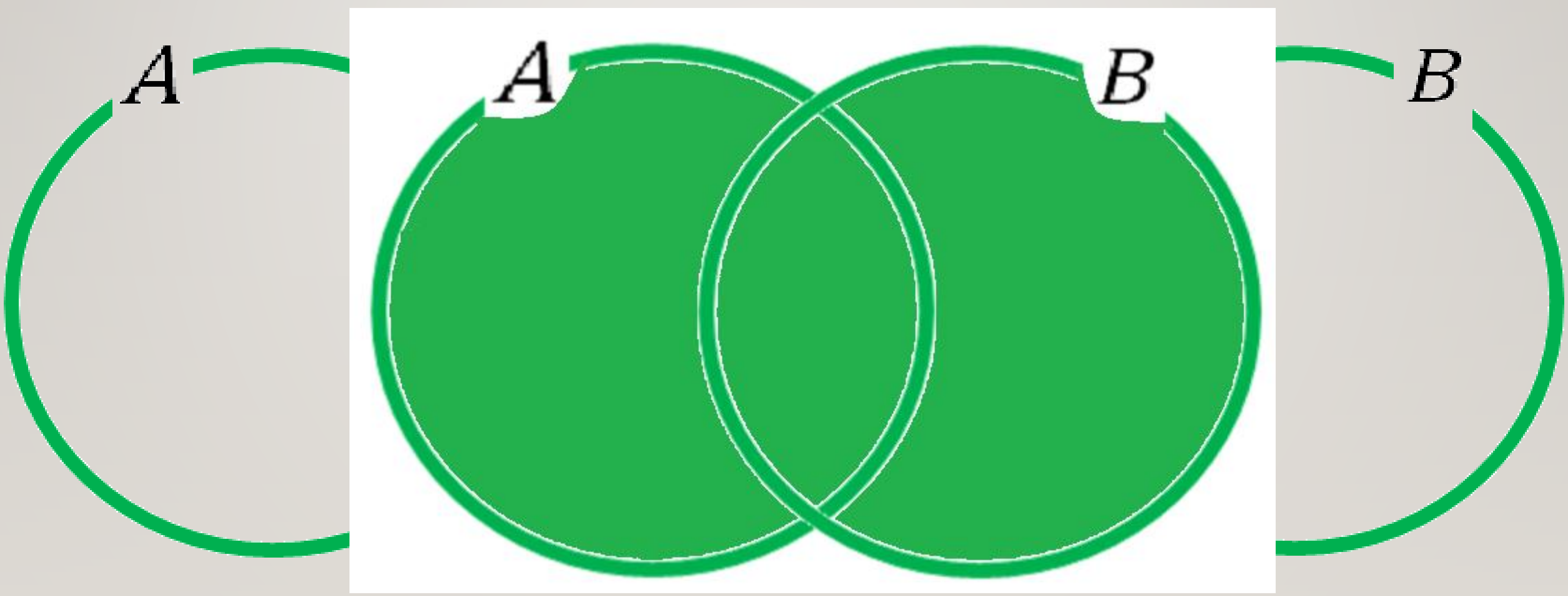


The entire shaded area represents $A \cup B$

Newbold et al (2013)

UNION OF EVENTS: VISUAL REPRESENTATION

$A \cup B$



Baseado em https://docente.ifrn.edu.br/julianaschivani/disciplinas/matematica-iii/probabilidade/slides-4-probabilidade

COLLECTIVELY EXHAUSTIVE EVENTS

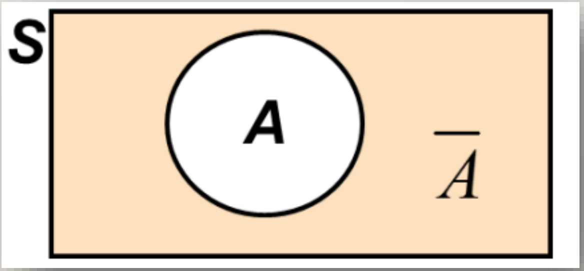
- Events $E_1, E_2, \ldots, E_k$ are Collectively Exhaustive events if $E_1 \cup E_2 \cup \ldots \cup E_k = S$
  - i.e., the events completely cover the sample space

Newbold et al (2013)

COMPLEMENT OF AN EVENT

- The Complement of an event *A* is the set of all basic outcomes in the sample space that do not belong to *A*. The complement is denoted $\overline{A}$

Newbold et al (2013)



$$A \cap \overline{A} = \{\ \}$$

$$A \cup \overline{A} = S$$

EXAMPLES: COMPLEMENTS, INTERSECTIONS, AND UNIONS

$S = [1, 2, 3, 4, 5, 6]$ $A = [2, 4, 6]$ $B = [4, 5, 6]$

Complements:

$\overline{A} =$ $\overline{B} =$

Intersections:

$A \cap B =$ $\overline{A} \cap B =$

Unions:

$A \cup B =$

$A \cup \overline{A} =$

Newbold et al (2013)

EXAMPLES: MUTUALLY EXCLUSIVE AND COLLECTIVELY EXHAUSTIVE

$$S = [1, 2, 3, 4, 5, 6] \qquad A = [2, 4, 6] \qquad B = [4, 5, 6]$$

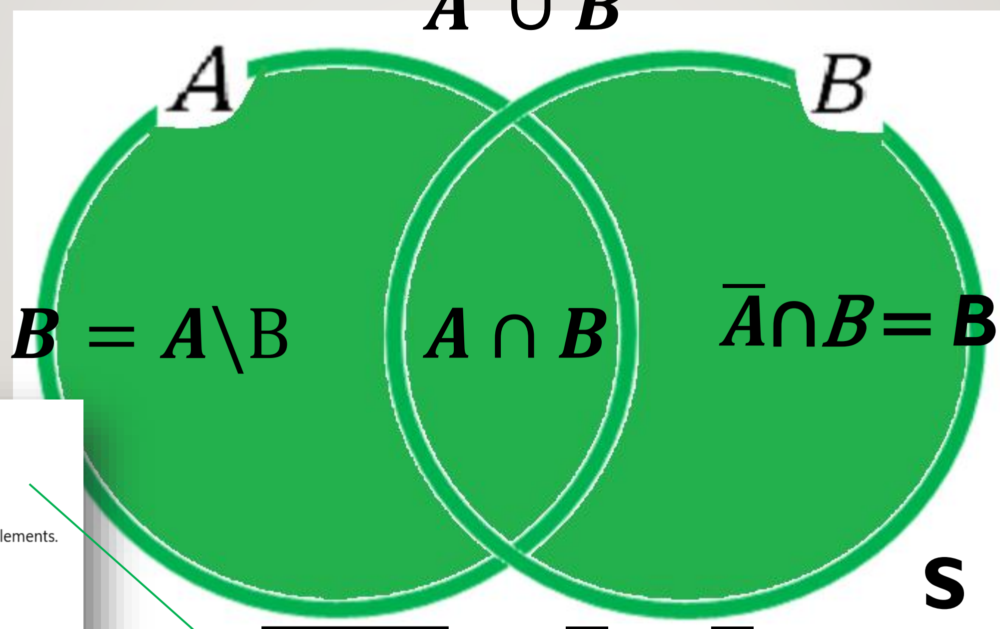
- Mutually exclusive:
  - *A* and *B* are not mutually exclusive
    - The outcomes 4 and 6 are common to both
- Collectively exhaustive:
  - *A* and *B* are not collectively exhaustive
    - $A \cup B$ does not contain 1 or 3

Newbold et al (2013)

VENN DIAGRAM

Venn Diagram:
Shows relationships between sets with overlapping circles.

$$A \cup B$$



$$A \cap \bar{B} = A - B = A \setminus B$$

$$A \cap B$$

$$\bar{A} \cap B = B - A = B \setminus A$$

De Morgan's Laws for Sets

1.

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

The complement of the union is the intersection of the complements.

2.

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

The complement of the intersection is the union of the complements.

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

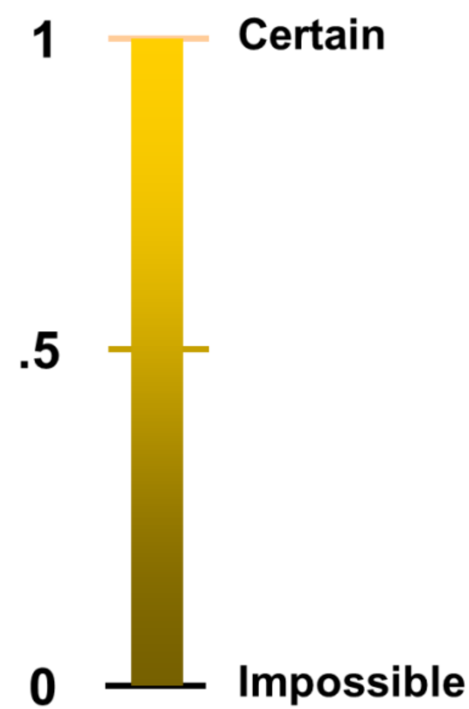
- "Not in A or B" means "outside A **and** outside B".
- "Not in both A and B" means "outside A **or** outside B".

Baseado em https://docente.ifrn.edu.br/julianaschivani/disciplinas/matematica-iii/probabilidade/slides-4-probabilidade

PROBABILITY

- Probability – the chance that an uncertain event will occur (always between 0 and 1)

$0 \le P(A) \le 1$ **For any event *A***



Newbold et al (2013)

ASSESSING PROBABILITY

- There are three approaches to assessing the probability of an uncertain event:

1. classical probability
2. relative frequency probability
3. subjective probability

Newbold et al (2013)

CLASSICAL PROBABILITY

- Assumes all outcomes in the sample space are equally likely to occur

Classical probability of event *A*:

$$P(A) = \frac{N_A}{N} = \frac{\text{number of outcomes that satisfy the event } A}{\text{total number of outcomes in the sample space}}$$

- Requires a count of the outcomes in the sample space

Newbold et al (2013)

RELATIVE FREQUENCY PROBABILITY AND SUBJECTIVE PROBABILITY

Three approaches (continued)

2. relative frequency probability

   - the limit of the proportion of times that an event *A* occurs in a large number of trials, *n*

$$P(A) = \frac{n_A}{n} = \frac{\text{number of events in the population that satisfy event } A}{\text{total number of events in the population}}$$

3. subjective probability

   an individual opinion or belief about the probability of occurrence

Newbold et al (2013)

PROBABILITY POSTULATES

1. If A is any event in the sample space S, then

$$0 \le P(A) \le 1$$

2. Let A be an event in S, and let O_i denote the basic outcomes. Then

$$P(A) = \sum_{A} P(O_i)$$

(the notation means that the summation is over all the basic outcomes in A)

3. $P(S) = 1$ $\qquad$ $P(\{\}) = 0$

Newbold et al (2013)

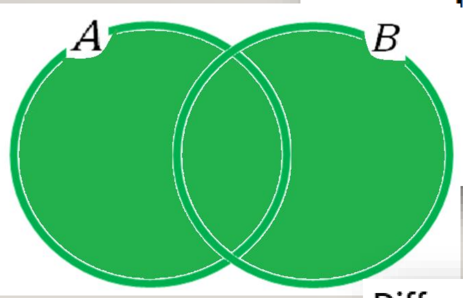
PROBABILITY RULES

- The Complement rule:

Newbold et al (2013)

$$P(\overline{A}) = 1 - P(A) \quad \text{i.e., } P(A) + P(\overline{A}) = 1$$

- The Addition rule:
  - The probability of the union of two events is



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

$$\boldsymbol{If\ A \cap B = \{\}, \text{then } P(A \cup B) = P(A) + P(B)}$$

Difference Rule *(Probability of A minus B)*

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

Note: $P(A \setminus B) \neq P(B \setminus A)$ in general.

Notation Note:
P(A-B) = P(A\B) in words. In Portuguese texts, P(A/B) is often used instead of P(A\B).

A PROBABILITY TABLE

Probabilities and joint probabilities for two events *A* and *B* are summarized in this table:

	B	$\overline{B}$	
A	$P(A \cap B)$	$P(A \cap \overline{B})$	$P(A)$
$\overline{A}$	$P(\overline{A} \cap B)$	$P(\overline{A} \cap \overline{B})$	$P(\overline{A})$
	$P(B)$	$P(\overline{B})$	$P(S) = 1.0$

Newbold et al (2013)

EXERCISE 3.17

3.17 A department store manager has monitored the number of complaints received per week about poor service. The probabilities for numbers of complaints in a week, established by this review, are shown in the following table. Let *A* be the event "there will be at least one complaint in a week" and *B* the event "there will be fewer than ten complaints in a week."

Number of complaints	0	1 to 3	4 to 6	7 to 9	10 to 12	More than 12
Probability	0.14	0.39	0.23	0.15	0.06	0.03

a. Find the probability of *A*.
b. Find the probability of *B*.
c. Find the probability of the complement of *A*.
d. Find the probability of the union of *A* and *B*.
e. Find the probability of the intersection of *A* and *B*.
f. Are *A* and *B* mutually exclusive?
g. Are *A* and *B* collectively exhaustive?

Newbold et al (2013)

EXERCISE 3.17 A): SOLUTION

Answer:

Given table:

Number of complaints	0	1–3	4–6	7–9	10–12	More than 12
Probability	0.14	0.39	0.23	0.15	0.06	0.03

Let:

- $A =$ "there will be **at least one complaint** in a week"
- $B =$ "there will be **fewer than ten complaints** in a week"

a. Probability of A

Event A = at least one complaint = all outcomes **except 0 complaints**.

$$P(A) = 1 - P(0) = 1 - 0.14 = 0.86$$

EXERCISE 3.17 B): SOLUTION

✅ Answer:

Given table:

Number of complaints	**0**	**1–3**	**4–6**	**7–9**	**10–12**	**More than 12**
Probability	0.14	0.39	0.23	0.15	0.06	0.03

Let:

- $A =$ "there will be **at least one complaint** in a week"
- $B =$ "there will be **fewer than ten complaints** in a week"

b. Probability of B

Event B = fewer than 10 complaints = sum of probabilities for 0, 1–3, 4–6, 7–9 complaints:

$$P(B) = 0.14 + 0.39 + 0.23 + 0.15 = 0.91$$

EXERCISE 3.17 C): SOLUTION

✅ Answer:

Given table:

Number of complaints	0	1–3	4–6	7–9	10–12	More than 12
Probability	0.14	0.39	0.23	0.15	0.06	0.03

Let:

- $A =$ "there will be **at least one complaint** in a week"
- $B =$ "there will be **fewer than ten complaints** in a week"

c. Probability of the complement of A

Complement of A = "no complaints in a week" = 0 complaints:

$$P(A^c) = P(0) = 0.14$$

HOMEWORK OF LECTURE 3: QUESTIONS

EXERCISE 2.56

2.56 A corporation administers an aptitude test to all new sales representatives. Management is interested in the extent to which this test is able to predict weekly sales of new representatives. Aptitude test scores range from 0 to 30 with greater scores indicating a higher aptitude. Weekly sales are recorded in hundreds of dollars for a random sample of 10 representatives. Test scores and weekly sales are as follows:

Test Score, x	12	30	15	24	14	18	28	26	19	27
Weekly Sales, y	20	60	27	50	21	30	61	54	32	57

a. Compute the covariance between test score and weekly sales.
b. Compute the correlation between test score and weekly sales.

Newbold et al (2013)

EXERCISE 3.13

3.13 A corporation has just received new machinery that must be installed and checked before it becomes operational. The accompanying table shows a manager's probability assessment for the number of days required before the machinery becomes operational.

Number of days	3	4	5	6	7
Probability	0.08	0.24	0.41	0.20	0.07

Let *A* be the event "it will be more than four days before the machinery becomes operational," and let *B* be the event "it will be less than six days before the machinery becomes available."

a. Find the probability of event *A*.
b. Find the probability of event *B*.
c. Find the probability of the complement of event *A*.
d. Find the probability of the intersection of events *A* and *B*.
e. Find the probability of the union of events *A* and *B*.

Newbold et al (2013)

EXERCISE 3.14

3.14 On a sample of 1,500 people in Sydney, Australia, 89 have no credit cards (event A), 750 have one (event B), 450 have two (event C) and the rest have more than two (event D). On the basis of the data, calculate each of the following.

a. The probability of event A
b. The probability of event D
c. The complement of event B
d. The complement of event C
e. The probability of event A or D

Newbold et al (2013)

THANKS!

Questions?