Statistical Laboratory

EXERCISES

1. Fundamental Concepts of Statistics

1.1 Distinguish the objectives of Descriptive Statistics from those of Inferential Statistics.

1.2 What is the importance of sampling studies compared to census studies? Illustrate your answer with the underlying concepts.

1.3 Distinguish between the concepts of *statistic* and *parameter*.

1.4 Distinguish between cross-sectional data, time series data, and panel data, providing examples.

1.5 How can variables be classified according to the type of measurement scale? Provide examples.

1.6 How can variables be classified according to the number of values they can take? Provide examples.

1.7 Provide examples of potential populations, variables, and statistical units of interest in the fields of Economics and Management.

1.8 Distinguish between the concepts of random variable and empirical variable

1.9 What is the general procedure underlying empirical research in the fields of Economics and Management?

1.10 Provide examples of databases available for conducting empirical research in the fields of Economics and Management.

2. Exploratory Data Anaysis

2.1 For quality-control purposes, the product is shipped in boxes of six units. A simple random sample without replacement of 40 boxes is drawn from the lot. For each box, the number of defective units is recorded. The observed sample is:

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

a) Define the statistical unit and the variable under analysis, and classify the variable.

b) Present the frequency table.

c) Provide a graphical representation of the data.

d) Compute the range and the interquartile range (IQR).

e) Assess the symmetry (skewness) of the distribution.

2.2 The following table shows the number of weekly study hours outside classes observed in a sample of MAEG students:

6.7	8.2	8.5	5.7	9	8	18.2	9.5	5.6	11.2	9.8	3.1	7.6	8.8	6.3

a) Define the statistical unit and the variable under study.

b) Compute the five-number summary for this dataset.

c) Analyse the symmetry of the data using a boxplot.

2.3 The statistical information on the monthly wages of workers in a textile company is presented below:

Salary (euros)	**Workers (%)**
[600 ; 900]	9
]900 ; 1000]	18
]1000 ; 1100]	35
]1100 ; 1200]	22
]1200 ; 1300]	16

a) Define and classify the variable under study.

b) Provide a graphical representation of the data.

c) Assess the symmetry of the data based on the graphical representation.

2.4 Consider the grades of 50 students in this course from the previous academic year:

93, 77, 67, 72, 52, 83, 66, 84, 59, 63,
75, 97, 84, 73, 81, 42, 61, 51, 91, 87,
34, 54, 71, 47, 79, 70, 65, 57, 90, 83,
58, 69, 82, 76, 71, 60, 38, 81, 74, 69,
68, 76, 85, 58, 45, 73, 75, 42, 93, 65

a) Represent the data in an ordered stem-and-leaf plot.

b) Compute the five-number summary from the plot.

c) Construct a boxplot and assess the symmetry (or asymmetry) of the data.

d) Group the data into 5 classes of equal width, with the lower bound of the first class set to 25 and construct a histogram.

2.5 The grades of 10 students in Statistics Laboratory (SL) and Linear Algebra (LA) are presented below:

SL	15	12	18	14	13	10	15	16	17	17
LA	14	15	17	13	15	10	14	15	18	17

a) Provide a graphical representation of these data.

b) Assess the possibility of a linear relationship between the grades in these two courses.

2.6 The statistical information on the heights of students in a school is presented below:

Height (cm)	Number of Students
[150 ; 158]	5
]158 ; 166]	18
]166 ; 174]	42
]174 ; 182]	27
]182 ; 190]	8

a) Identify the statistical unit and the variable, and classify the variable.

b) Briefly assess the data based on its graphical representation.

3. Organizing and Summarizing Data

3.1 The wages (in monetary units) of 40 factory workers are presented below:

50	55	85	62
35	70	75	68
90	38	72	70
45	120	50	75
80	50	48	83
38	35	45	92
42	125	150	85
110	115	70	140
100	95	75	145
60	90	155	90

a) Group the data into 4 classes of equal width, with the lower bound of the first class set at 35 monetary units, and construct the frequency table.

b) Provide a graphical representation of the data.

c) Present the cumulative frequency function for the class limits and draw the corresponding ogive.

d) Compute the measures of central tendency for these data.

3.2 Consider the information provided in exercise 2.3.

a) Assess the symmetry (or asymmetry) of the distribution using a boxplot.

b) Determine the new mean salary after a general wage increase of 10%.

c) Evaluate the concentration of the data.

3.3 A sports club measured the weight of its 40 athletes. The observed mean weight was 66.5 kg, the median was 64 kg, and the 1st and 3rd quartiles were 62 kg and 70 kg, respectively. The "extreme" values observed (in kg) were: 40, 51, 53, 53 (on the left tail) and 80, 85, 95, and 101 (on the right tail).

a) Explain the meaning of the 1st and 3rd quartiles in this context.

b) Identify and classify the outliers based on the extreme values.

c) Which measure of central tendency would you consider most appropriate to use? Justify your answer.

3.4 The information on the salaries of workers in a multinational company is presented below:

Classes	0 - 3	3 - 6	6 - 9	9 - 12
Number of Workers	500	1500	2500	500

a) Compute the mean and median salaries.

b) What percentage of workers earn more than the mean salary?

c) Determine the mean salary of the 30% of workers with the lowest salaries.

d) Determine the mean salary of the 30% of workers with the highest salaries.

3.5 The statistical information on the grades of a group of university students in Statistics and Economics is presented below:

Student	A	B	C	D	E	F	G	H	I	J
Statistics	10	15	13	16	15	12	11	14	10	11
Economics	a	13	13	15	16	13	12	18	12	15

a) Assess the variability of grades in each subject individually.

b) Assess the variability of grades in these subjects comparatively.

3.6 The data below show the diesel consumption (litres) of 10 cars of brands A and B over a 100 km trip:

Bard A: 7,8; 7,0; 8,2; 7,6; 6,9; 7,7; 7,2; 7,8; 7,3; 7,5

Bard B: 7,0; 7,1; 7,1; 7,3; 7,2; 7,0; 7,8; 8,0; 7,0; 7,0

a) Compare the average consumption using an appropriate measure of central tendency.

b) Using a back-to-back stem-and-leaf plot, assess the symmetry of the data and interpret it in context.

c) Using the cumulative frequency function, compute the proportion of brand A cars whose consumption exceeds 7.3 litres. Identify the corresponding frequency.

3.7 Show that the mean of the squares of the deviations of the values of a variable from a constant ccc is minimized when this constant is equal to the mean of the variable.

4. Association and Relationships Between Variables

4.1 The advertising expenses and sales (in monetary units) of a company are presented below:

Year	1	2	3	4	5	6	7	8	9	10
Sales	120	150	160	200	210	250	300	360	450	550
Expenses	10	12	14	20	22	25	30	38	50	60

a) Assess the relationship between sales and advertising expenses using an appropriate graphical representation.

b) Estimate the regression line of advertising expenses as a function of sales. Interpret the result.

c) Evaluate the quality of the estimated model using an appropriate measure.

4.2 The following data refer to the expenses (in monetary units) of an individual on food (F) and leisure (L) over 10 vacation days:

F	20	22	18	10	15	12	20	18	21	11
L	15	17	20	15	14	10	18	20	15	15

a) Compute the mean and variance of each variable.

b) Assess the existence of a possible linear relationship between the variables using an appropriate measure. Justify your answer.

4.3 Consider the following table on employment status by gender in a given region:

	Employed	**Unemployed**
Women	100	55
Men	120	35

a) Assess the existence of an association between gender and employment status using an appropriate measure.

b) Identify, in statistical terms, the frequencies corresponding to the following expressions:

b1) 100 / 155 **b2)** 35 / 310 **b3)** 155 / 310

4.4 The table below presents the election results in a given district:

Political Party	**A**	**B**	**C**
Employed	393	232	152
Unemployed	55	28	20

Using two appropriate measures, assess the existence of an association between employment status and voting choice. Justify your answer.

4.5 The table below presents information on the Humidity Index (H) and the Maximum Air Temperature (T) observed during the first ten days of a given month:

H	5,0	1,6	2,0	3,5	4,7	3,1	2,4	4,1	1,2	2,9
T	23,5	16,1	16,1	20,6	22,8	20,3	18,7	22,5	14,4	19,4

a) Assess the association between the variables using an appropriate graphical representation.

b) Estimate, using the least squares method, the linear regression equation to predict Maximum Air Temperature from the Humidity Index.

c) Evaluate the quality of the estimated model and compute the predicted value of Maximum Air Temperature when the Humidity Index is 3.3.

4.6 Show that the covariance can also be calculated using the following formula:

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

5. Index Numbers

5.1 The prices (in monetary units) and quantities (in kg) consumed of two raw materials, A and B, in a factory are presented below:

Year	1	2	3	4	5
$\mathbf{P_A}$	3	4	4,5	4	5,25
$\mathbf{Q_A}$	12	14	8	6	10
$\mathbf{P_B}$	5	6	6,5	5,5	3,9
$\mathbf{Q_B}$	9	11	8	10	12

a) Analyse the evolution of prices of raw material A using index numbers.

b) For year 2, relative to year 1, compute the Laspeyres quantity index and comment on the result.

5.2 The following data refer to the evolution of prices and quantities sold of three products:

	Year 1		***Year 2***		***Year 3***	
	Price	**Quantity**	**Price**	**Quantity**	**Price**	**Quantity**
Product 1	10	100	12	100	14	110
Product 2	12	150	16	120	15	120
Product 3	8	30	10	35	10	45

a) Assess the evolution of quantities sold using the Paasche Quantity Index, relative to year 1.

b) Suppose the price index (PI) of these products is given by:

	Year 1	**Year 2**	**Year3**
PI	100	110	115

The values for years 4 and 5 were calculated relative to year 3

	Year 4	**Year 5**
PI	120	135

From the data provided, compute and interpret the reconciled index, using year 3 as the base year.

5.3 The following data are used to calculate the Consumer Price Index (CPI) of a country:

Year	1		2	3	4
Product	**Price**	**Quantity**	**Price**	**Price**	**Price**
I	100	120	120	150	200
II	190	120	180	170	170
III	30	220	40	45	50
IV	80	150	100	90	110

Public expenditure on housing (current values, in monetary units):

Year	1	2	3	4
Expenses	74000	83600	84300	96200

a) Compute the Laspeyres Price Index using year 1 as the base year.

b) Determine the evolution of public expenditure on housing at constant prices, using year 1 as the base year.

5.4 The following table contains information on the prices and quantities of products A and B traded in years 1 and 2:

Year	Product A		Product B	
	Price	**Quantity**	**Price**	**Quantity**
1	*x*	10	12	5
2	7	*y*	15	10

a) For product B, assess the evolution of the price in comparison with the evolution of the quantity traded using index numbers.

b) From the collected data, the Laspeyres Price Index for year 1 relative to year 2 is 1.2, and the Paasche Quantity Index for year 2 relative to year 1 is 1.3. Under these conditions, determine x and y.

5.5 The data below refer to the number of online orders received by a company over the past 10 years:

Year	1	2	3	4	5	6	7	8	9	10
Orders	25	35	50	70	99	145	230	340	500	800

Using index numbers, assess the evolution of the variable under study relative to year 1.

6. Time Series

6.1 Consider the data from Exercise 5.5:

a) Analyse the evolution of the variable under study using an appropriate graphical representation.

b) Determine the trend values using 3-period moving averages.

6.2 The number of armed robberies (AR) in a given region over the past decade is presented below:

Year	1	2	3	4	5	6	7	8	9	10
AR	3	1	2	3	4	7	12	8	9	10

a) Provide an initial analysis of the data using a time series plot.

b) Using the least squares method, estimate the predicted trend value for year 12, assuming a linear trend.

6.3 The quarterly export values (in monetary units) of a product for years 1 and 2 are presented below:

Year / Quarter	**1**	**2**	**3**	**4**
1	22	25	24	33
2	37	43	42	52

a) Draw the time series plot of exports and analyse the trend of the series.

b) Assess the existence of additive seasonality using the method of quarterly averages.

6.4 The number of children observed in the ophthalmology clinic of a health center evolved as follows over the last two years:

Year / Quarter	**1**	**2**	**3**	**4**
1	110	125	133	128
2	130	140	145	120

a) Provide a graphical analysis of the evolution in the number of children observed in the ophthalmology clinic.

b) Determine the trend values using 3-period moving averages.

6.5 The number of mobile phones sold in a retail chain is shown below:

Year / Quarter	1	2	3	4
1	500	440	580	600
2	450	400	500	550

a) Graphically analyse the evolution of mobile phone sales.

b) Verify the existence of seasonality using an appropriate method and remove it (assuming additive seasonality).

c) Identify which components of the series are associated with the following events, briefly justifying: c1) the budget deficit stabilization agreement with the EU; c2) the Christmas season.

Short Answers

Chapter 2

2.1 a) Statistical Unit: Box of 6 units; Variable: number of defective products per box; **b)** x_i: 0, 1, 2, 3, 4, 5, 6 ; n_i: 14, 7, 5, 8, 1, 3, 2 ; f_i: 0.35, 0.175, 0.125, 0.2, 0.025, 0.075, 0.05; **d)** $D = [0, 6]$; $IQ = [0, 3]$; **e)** Positive skewness (left bias).

2.2 a) Statistical Unit: student; Variable: weekly study hours outside classes. **b)** $(3.1, 6.3, 8.2, 9.5, 18.2)$; **c)** Positive skewness.

2.3 a) Continuous variable: salary, in euros; **b)** I_j: $[600, 900]$, $]900, 1000]$, $]1000, 1100]$, $]1100, 1200]$, $]1200, 1300]$; f_j/h_j: 0.0003, 0.0018, 0.0035, 0.0022, 0.0016 ; **c)** Negative skewness.

2.4 a)

		n_i
3	4 8	2
4	2 2 5 7	4
5	1 2 4 7 8 8 9	7
6	0 1 3 5 5 6 7 8 9 9	10
7	0 1 1 2 3 3 4 5 5 6 6 7 9	13
8	1 1 2 3 3 4 4 5 7	9
9	0 1 3 3 7	5

b) $(34,\ 58.75, 71, 81.25, 97)$; **c)** Slight negative skewness; **d)** I_j: $[25, 40]$, $]40, 55]$, $]55, 70]$, $]70, 85]$, $]85, 100]$; f_j/h: 0.1333, 0.4667, 1, 1.3333, 0.4.

2.5 b) Relatively strong positive linear association.

2.6 a) Statistical unit: student ; Continuous variable: height, in cm; **b)** I_j: $[150, 158]$, $]158, 166]$, $]166, 174]$, $]174, 182]$, $]182, 190]$; F_j/h: 0.625, 2.25, 5.25, 3.375, 1; Negative skewness.

Chapter 3

3.1 a) I_j: $[35, 65]$, $]65, 95]$, $]95, 125]$, $]125, 155]$; x'_j: 50, 80, 110, 140 ; n_j: 14, 17, 5, 4 ; f_j: 0.35, 0.425, 0.125, 0.1 ; N_j: 14, 31, 36, 40 ; F_j^*: 0.35, 0.775, 0.9, 1 ; **b)** f_j/h: 0.4667, 0.5667, 0.1667, 0.1333 ; **c)** $F^*(35) = 0$, $F^*(65) = 0.35$, $F^*(95) = 0.775$, $F^*(125) = 0.9$, $F^*(155) = 1$; **d)** $\bar{x} = 79.25$; $m_e = 75.588$; $m_o = 71$ ou $m_o = 72.895$ (King´s formula).

3.2 a) $(600, 988.889, 1065.714, 1159.091, 1300)$; Negative skewness; **b)** 1164.9 ; **c)** $G = 0.081$.

3.3 a) $Q_1 = 62$; $Q_3 = 70$; **b)** Moderate outliers: 40, 85; Severe outliers: 95, 101; **c)** Median or Mode.

3.4 a) $\bar{x} = 6.3$; $m_e = 6.6$; **b)** 55% ; **c)** $\bar{x}_1 = 3.167$; **d)** $\bar{x}_2 = 9.1$.

3.5 a) $s^2_{EST} = 4.41$; $s^2_{ECO} = 3.69$; **b)** $CV_{EST} = 16.535$; $CV_{ECO} = 13.820$.

3.6 a) $\bar{x}_A = 7.5$; $\bar{x}_B = 7.25$;

b)

Folhas – A		Folhas – B
9	6*	
0 2 3	7	0 0 0 0 1 1 2 3
5 6 7 8 8	7*	8
2	8	0

c) $F^*(6.9) = 0$, $F^*(7.3) = 0.4$, $F^*(7.8) = 0.9$, $F^*(8.2) = 1$; 0.6.

Chapter 4

4.1 a) Relatively strong positive linear relationship; **b)** *y* = expenses, *x* = sale: $\hat{y} = -4.46 + 0.1184\,x$; **c)** $R^2 = 0.99674$.

4.2 a) $\bar{x}_A = 16.7$; $s_A^2 = 17.41$; $\bar{x}_L = 15.9$; $s_L^2 = 8.09$; **b)** $r_{AL} = -0.8957$.

4.3 a) $\chi^2 = 6.263$; $V = 0.142$: weak association.

4.4. $\chi^2 = 0.3631$; $T = 0.017$; $V = 0.02$: very weak association.

4.5 a) Relatively strong positive linear relationship; **b)** $\hat{T} = 12.19625 + 2.375\,H$; **c)** $R^2 = 0.96105$; 20.03375.

Chapter 5

5.1 a) $100, 133.333, 150, 133.333, 175$; **b)** 119.753.

5.2 a) $100, 88.974, 95.949$; **b)** $86.9565, 95.6522, 100, 120, 150$.

5.3 a) $100, 111.985, 115.730, 134.644$; **b)** $74000, 74652.855, 72841.960, 71447.669$.

5.4 a) Price Indices: $100, 125$; Quantity Indices: $100, 200$; **b)** $x = 19.309$, $y = 5.5$.

5.5. $100, 140, 200, 280, 396, 580, 920, 1360, 2000, 3200$.

Chapter 6

6.1 a) Exponential trend; **b)** $36.667, 51.667, 73, 104.667, 158, 238.333, 356.667, 546.667$.

6.2 a) Approximately linear trend; **b)** 12.953 (about 13).

6.3 a) Linear trend; **b)** $-5.25, -0.75, -1.75, 7.75$.

6.4 a) Approximately linear trend; **b)** $122.667, 128.667, 130.333, 132.667, 138.333, 135$.

6.5 a) Possible seasonality; **b)** Seasonal indices: $-27.5, -82.5, 37.5, 72.5$; Series without seasonality: $527.5, 522.5, 542.5, 527.5, 477.5, 482.5, 462.5, 427.5$; **c)** c_1) Cyclical component; c_2) seasonal component.