

DESCRIPTIVE STATISTICS

Exercise 1. The next table presents length of hospital stay in children under 15 years of age, who underwent adenoidectomy in two different hospitals. Data are reported as frequency distributions.

	Length of hospital stay (days)										Total
	1	2	3	4	5	6	7	8	9	10	
1st hospital	1	2	1	1	25						30
2nd hospital	102	8	7	6	15	12	9	2	1	1	163

1a) Compute mean, median and mode of length of hospital stay in the first and second hospitals.

1b) Does measures of central tendency differ between the two hospitals? Why?

1c) Which measures of central tendency and variability are more suited to describe these samples?

Exercise 2. Nine patients, who underwent surgery for pancreatic cancer, survive respectively 2, 2, 2, 3, 3, 4, 6, 9, 24 months. Which measure of central tendency is the most suited to describe survival in this series ?

- A) mean
- B) median
- C) mode
- D) geometric mean
- E) the 80th percentile

2b) Mode is equal to:

- A) 2
- B) 2.5
- C) 3
- D) 4
- E) 6.11
- F) 9

2c) Median is equal to:

- A) 2
- B) 2.5
- C) 3
- D) 4
- E) 6.11

F) 9

2d) Mean is equal to:

- A) 2
- B) 2.5
- C) 3
- D) 4
- E) 6.11
- F) 9

2e) The absolute frequency of the mode is equal to:

- A) 2
- B) 2.5
- C) 3
- D) 4
- E) 6.11
- F) 9

2f) The absolute frequency of the median is equal to:

- A) 2
- B) 2.5
- C) 3
- D) 4
- E) 6.11
- F) 9

2g) Range is equal to:

- A) 2
- B) 3
- C) 4
- D) 9
- E) 22
- F) 24

2h) Hence the distribution is:

- A) symmetric
- B) positively skewed
- C) negatively skewed
- D) bimodal
- E) multimodal

PROBABILITY

Exercise 3: In a cross-sectional investigation, the association between bronchial asthma and smoking habits was studied on a sample of the adult general population of Verona. The results of the survey are summarized in the following table:

		non-smokers	ex-smokers	Smokers	
Asthma attacks	NO	987	359	637	1983
In the last 12 months	YES	41	20	26	87
		1028	379	663	2070

- 3a)** What is the probability that an individual of this population has had asthma attacks in the last 12 months?
- 3b)** What is the probability that an individual of this population is a smoker?
- 3c)** What is the probability that one of these individuals has smoked during his life?
- 3d)** What is the conditional probability of having asthma attacks among non-smokers, between ex-smokers and smokers? Which of these conditional probabilities is higher? What is your interpretation?
- 3e)** In this sample chronic bronchitis was reported by 69 non-smokers, 45 ex-smokers and 117 smokers. Which category has the highest absolute frequency of chronic bronchitis? Which category presents the highest relative frequency of chronic bronchitis ?
- 3f)** In your opinion, can you affirm that there is a relationship between smoking and chronic bronchitis in this population? (Indicate the type of test to be performed, the null hypothesis and the alternative hypothesis, the level of significance adopted, the critical threshold and the value of the test).

DIAGNOSTIC TESTS

Exercise 4. Arterial hypertension is diagnosed when resting diastolic pressure is equal or greater than 90 mmHg and systolic pressure is equal or greater than 140 mmHg. However, arterial pressure should exceed the cut-offs at least twice to pose the diagnosis of arterial hypertension, as arterial pressure can easily change as a function of emotions, physical exercise, heat/cold exposure.

In an epidemiological study on hypertension, arterial pressure was measured in a sample of 1000 individuals and the first assessment was abnormally high in 75 patients, who were therefore classified as hypertensive. Subsequently with a more accurate screening the diagnosis of hypertension was confirmed in 60 patients and was formulated ex-novo in 30 other patients previously classified as normotensive.

- 4a) What is the probability that an individual in the population is hypertensive?
- 4b) If you use a single blood pressure measurement to diagnose hypertension, what is the sensitivity and specificity of this test?
- 4c) In this sample, what is the probability that an individual with a positive test is hypertensive? And the probability that an individual with a negative test is healthy?

INFERENCE STATISTICS

Exercise 5. Which is the 95% confidence interval of the mean weight in a population, if a sample of 9 subjects has a mean of 82 Kg and a standard deviation of 15 Kg ? Weight is normally distributed in the original population.

5b) Which is the confidence interval at a confidence level of 90% ? and at a confidence level of 99% ?

5c) Which would have been the 95% confidence interval with a sample size of 4 ? and with a sample size of 100 ?

5d) Which is the sample size needed to achieve a precision (confidence interval half-width) of 2 Kg?

Exercise 6. The following Table reports data on smoking habits and myocardial infarction in a group of 20 patients.

REC	SEX	AGEyears	Height_m	weight_Kg	Smoke	INFARCTION	BMI
1	M	42	1.70	58	S	N	
2	M	48	1.84	90	N	I	
3	M	51	1.66	70	S	I	
4	M	54	1.78	76	S	I	
5	M	58	1.74	72	N	N	
6	M	60	1.76	85	N	I	
7	M	62	1.64	62	S	I	
8	M	64	1.90	88	S	I	
9	M	65	1.72	69	N	N	
10	M	70	1.77	77	N	N	
11	M	75	1.68	73	S	I	
12	M	81	1.74	75	S	I	
13	F	45	1.68	59	S	N	
14	F	49	1.58	55	N	N	
15	F	51	1.62	68	N	N	
16	F	53	1.65	64	S	I	
17	F	60	1.72	70	N	N	
18	F	63	1.69	65	S	I	
19	F	68	1.70	73	N	I	
20	F	75	1.66	52	N	N	

6a) Compute BMI (Body Mass Index) (Indice di Massa Corporea), using the following equation: $BMI = \text{Weight in Kg} / (\text{Height in m})^2$. Report individual values in the last column of the previous Table.

6b) Compute mean, median and standard deviation of BMI

6c) Do height, weight, BMI significantly differ between genders?

6d) Does BMI significantly differ between people with and without myocardial infarction?