EXERCISES OF STATISTICS

Subject: Statistics Applied tho the Health Sciences
Course: 2^{nd}
Degree: Physiotherapy
Year: 2015-2016

Authors: Santiago Angulo Díaz-Parreño (sangulo@ceu.es)
José Miguel Cárdenas Rebollo (cardenas@ceu.es)
Anselmo Romero Limón (arlimon@ceu.es)
Alfredo Sánchez Alberca (asalber@ceu.es)

CEU
Universidad
San Pablo

Contents

1 Descriptive Statistics

1. Classify the following variables
   (a) Daily hours of exercise.
   (b) Nationality.
   (c) Blood pressure.
   (d) Severity of illness.
   (e) Number of sport injuries in a year.
   (f) Daily calorie intake.
   (g) Size of clothing.
   (h) Subjects passed in a course.

2. The number of injuries suffered by the members of a soccer team in a league were

0	1	2	1	3	0	1	0	1	2	0	1
1	1	2	0	1	3	2	1	2	1	0	1

   (a) Construct the frequency distribution table of the sample.
   (b) Draw the bar chart of the sample and the polygon.
   (c) Draw the cumulative frequency bar chart and the polygon.

3. A survey about the daily number of medicines consumed by people over 70 years, shows the following results:

3	1	2	2	0	1	4	2	3	5	1	3	2	3	1	4	2	4	3	2
3	5	0	1	2	0	2	3	0	1	1	5	3	4	2	3	0	1	2	3

   (a) Construct the frequency distribution table of the sample.
   (b) Draw the bar chart of the sample and the polygon.
   (c) Draw the cumulative relative frequency bar chart and the polygon.

4. In a survey about the dependency of older people, 23 persons over 75 years were asked about the help they need in daily life. The answers were

   B D A B C C B C D E A B C E A B C D B B A A B

   where the meanings of letters are:

   A No help.
   B Help climbing stairs.
   C Help climbing stairs and getting up from a chair or bed.
   D Help climbing stairs, getting up and dressing.
   E Help for almost everything.

   Construct the frequency distribution table and the suitable chart.

5. The number of people treated in the emergency service of a hospital every day of November was

15	23	12	10	28	7	12	17	20	21	18	13	11	12	26
30	6	16	19	22	14	17	21	28	9	16	13	11	16	20

   (a) Construct the frequency distribution table of the sample.
   (b) Draw a suitable chart for the frequency distribution.

(c) Draw a suitable chart for the cumulative frequency distribution.

6. The following frequency distribution table represents the distribution of time (in min) required by people attended in a medical dispensary.

Time	n_i	f_i	N_i	F_i
$[0, 5)$	2			
$[5, 10)$			8	
$[10, 15)$				0.7
$[15, 20)$	6			

(a) Complete the table.

(b) Draw the ogive.

7. Use the data of exercise 2 to calculate the following statistics and interpret them.

(a) Mean.

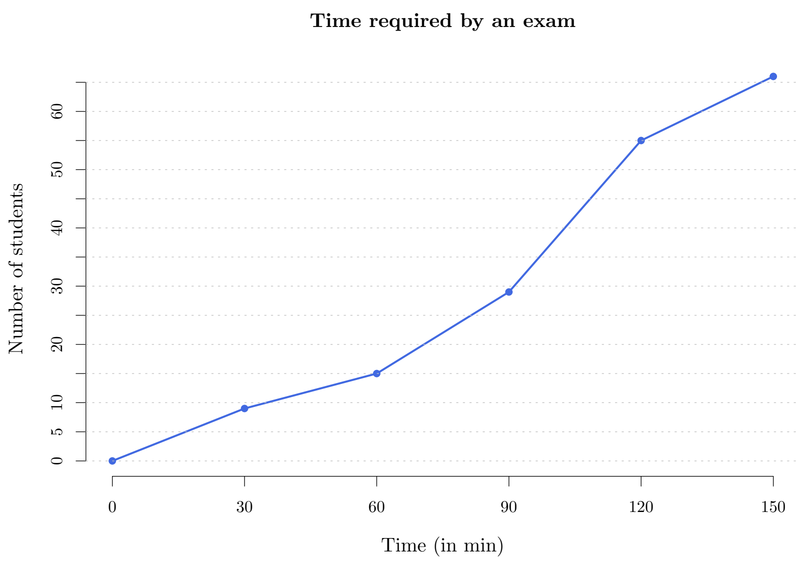
(b) Median.

(c) Mode.

(d) Quartiles.

(e) Percentile 32.

8. The chart below shows the cumulative distribution of the time (in min) required by 66 students to do an exam.



(a) A which time have finished half of the students? And 90% of students?

(b) Which percentage of students have finished after 100 minutes?

(c) Which is the time that best represent the time required by students in the sample to finish the exam? Is this value representative or not?

9. In a study about the children growth two samples where drawn, one for newborns and the other for one year old. The height in cm of children in both samples were

Newborn children:	51, 50, 51, 53, 49, 50, 53, 50, 47, 50
One year old children:	62, 65, 69, 71, 65, 66, 68, 69

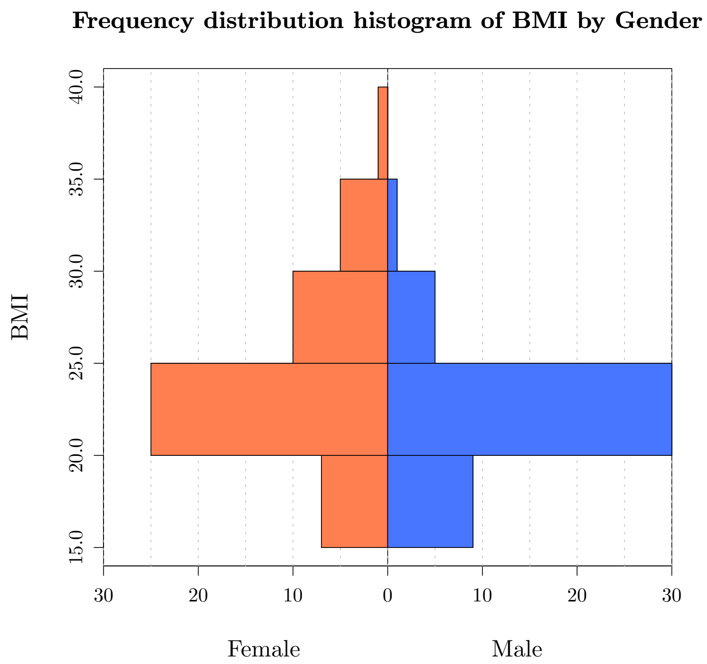
In which group is more representative the mean? Justify the answer.

10. To determine the accuracy of a method for measuring hematocrit in blood, the measurement was repeated 8 times on the same blood sample. The results in percentage of hematocrit in plasma were

$$42.2 \quad 42.1 \quad 41.9 \quad 41.8 \quad 42 \quad 42.1 \quad 41.9 \quad 42$$

What do you think about the accuracy of the method?

11. The histogram below shows the frequency distribution of the body mass index (BMI) of a group of people by gender.



(a) Draw the pie chart for the gender.

(b) In which group is more representative the mean of the BMI?

(c) Calculate the mean for the whole sample.

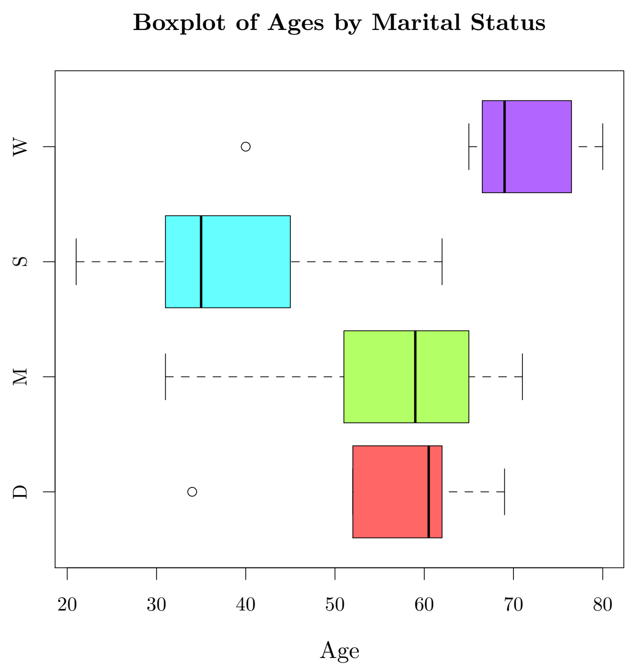
Use the following sums
Males: $\sum x_i = 1002$ kg/m^2 $\sum x_i^2 = 22781$ kg^2/m^4
Females: $\sum x_i = 1160$ kg/m^2 $\sum x_i^2 = 29050$ kg^2/m^4

12. The following table represents the frequency distribution of the yearly uses of a health insurance in a sample of clients of a insurance company.

Uses:	0	1	2	3	4	5	7
Clients:	4	8	6	3	2	1	1

Draw the box plot. How is the symmetry of the distribution?

13. The box plots below correspond to the age of a sample of people by marital status.



(a) Which group has higher ages?

(b) Which group has lower central dispersion?

(c) Which groups have outliers?

(d) Which group has a distribution of ages more asymmetric?

14. The following table represents the frequency distribution of ages at which a group of people suffered a heart attack.

Age	[40-50)	[50-60)	[60-70)	[70-80)	[80-90)
Persons	6	12	23	19	5

Could we assume that the sample comes from a normal population?

Use the following sums: $\sum x_i = 4275$ years, $\sum(x_i - \bar{x})^2 = 7462$ years^2, $\sum(x_i - \bar{x})^3 = -18249$ years^3, $\sum(x_i - \bar{x})^4 = 2099636$ years^4.

15. To compare two rehabilitation treatments A and B for an injury, every treatment was applied to a different group of people. The number of days required to cure the injury in every group is shown in the following table:

Days	A	B
20-40	5	8
40-60	20	15
60-80	18	20
80-100	7	7

(a) In which treatment is more representative the mean?

(b) In which treatment the distribution of days is more skew?

(c) In which treatment the distribution is more peaked?

Use the following sums:
A: $\sum x_i = 3040$ days, $\sum(x_i - \bar{x})^2 = 14568$ days^2, $\sum(x_i - \bar{x})^3 = 17011.2$ days^3, $\sum(x_i - \bar{x})^4 = 9989603$

days^4
B: $\sum x_i = 3020$ days, $\sum(x_i - \bar{x})^2 = 16992$ days^2, $\sum(x_i - \bar{x})^3 = -42393.6$ days^3, $\sum(x_i - \bar{x})^4 = 12551516$ days^4

16. The systolic blood pressure (in mmHg) of a sample of persons is

$$135 \quad 128 \quad 137 \quad 110 \quad 154 \quad 142 \quad 121 \quad 127 \quad 114 \quad 103$$

   (a) Calculate the central tendency statistics.
   (b) How is the relative dispersion with respect to the mean?
   (c) How is the skewness of the sample distribution?
   (d) How is the kurtosis of the sample distribution?
   (e) If we know that the method used for measuring the blood pressure is biased, and, in order to get the right values, we have to apply the linear transformation $y = 1.2x - 5$, which are values of the statistics required to answer the previous questions for the corrected values of the blood pressure?

   Use the following sums: $\sum x_i = 1271$ mmHg, $\sum(x_i - \bar{x})^2 = 2188.9$ mmHg^2, $\sum(x_i - \bar{x})^3 = 2764.32$ mmHg^3, $\sum(x_i - \bar{x})^4 = 1040080$ mmHg^4.

17. The table below contains the frequency of pregnancies, abortions and births of a sample of 999 women in a city.

Num	Pregnancies	Abortions	Births
0	61	751	67
1	64	183	80
2	328	51	400
3	301	10	300
4	122	2	90
5	81	2	62
6	29	0	0
7	11	0	0
8	2	0	0

   (a) How many birth outliers are in the sample?
   (b) Which variable has lower spread with respect to the mean?
   (c) Which value is relatively higher, 7 pregnancies or 4 abortions? Justify the answer.

   Use the following sums:
   Pregnancies: $\sum x_i = 2783$, $\sum x_i^2 = 9773$.
   Abortions: $\sum x_i = 333$, $\sum x_i^2 = 559$.
   Births: $\sum x_i = 2450$, $\sum x_i^2 = 7370$.

2 Regression and correlation

18. Give some examples of:
    (a) Non related variables.
    (b) Variables that are increasingly related.
    (c) Variables that are decreasingly related.

19. In an study about the effect of different doses of a medicament, 2 patients got 2 mg and took 5 days to cure, 4 patients got 2 mg and took 6 days to cure, 2 patients got 3 mg ant took 3 days to cure, 4 patients got 3 mg and took 5 days to cure, 1 patient got 3 mg and took 6 days to cure, 5 patients got 4 mg and took 3 days to cure and 2 patients got 4 mg and took 5 days to cure.
    (a) Construct the joint frequency table.
    (b) Get the marginal frequency distributions and compute the main statistics for every variable.
    (c) Compute the covariance and interpret it.

20. The table below shows the two-dimensional frequency distribution of a sample of 80 persons in a study about the relation between the blood cholesterol (X) in mg/dl and the high blood pressure (Y).

$X \setminus Y$	$[110,130)$	$[130,150)$	$[150,170)$	n_x
$[170,190)$		4		12
$[190,210)$	10	12	4	
$[210,230)$	7		8	
$[230,250)$	1			18
n_y		30	24	

    (a) Complete the table.
    (b) Construct the linear regression model of cholesterol on pressure.
    (c) Use the linear model to calculate the expected cholesterol for a person with pressure 160 mmHg.
    (d) According to the linear model, what is the expected pressure for a person with cholesterol 270 mg/dl?

    Use the following sums: $\sum x_i = 16960$ mg/dl, $\sum y_j = 11160$ mmHg, $\sum x_i^2 = 3627200$ (mg/dl)^2, $\sum y_j^2 = 1576800$ mmHg^2 y $\sum x_i y_j = 2378800$ mg/dl·mmHg.

21. A research study has been conducted to determine the loss of activity of a drug. The table below shows the results of the experiment.

Time (in years)	1	2	3	4	5
Activity (%)	96	84	70	58	52

    (a) Construct the linear regression model of activity on time.
    (b) According to the linear model, when will the activity be 80%? When will the drug have lost all activity?

22. A basketball team is testing a new stretching program to reduce the injuries during the league. The data below show the daily number of minutes doing stretching exercises and the number of injuries along the league.

Stretching minutes	0	30	10	15	5	25	35	40
Injuries	4	1	2	2	3	1	0	1

    (a) Construct the regression line of the number of injuries on the time of stretching.
    (b) What is the reduction of injuries for every minute of stretching?

(c) How many minutes of stretching are require for having no injuries? Is reliable this prediction?

Use the following sums (X=Number of minutes stretching, and Y=Number of injuries): $\sum x_i = 160$ min, $\sum y_j = 14$ injuries, $\sum x_i^2 = 4700$ min^2, $\sum y_j^2 = 36$ injuries^2 and $\sum x_i y_j = 160$ min·injuries.

23. For two variables X and Y we have

    - The regression line of Y on X is $y - x - 2 = 0$.
    - The regression line of X on Y is $y - 4x + 22 = 0$.

    Calculate:

    (a) The means $\bar{x}$ and $\bar{y}$.

    (b) The correlation coefficient.

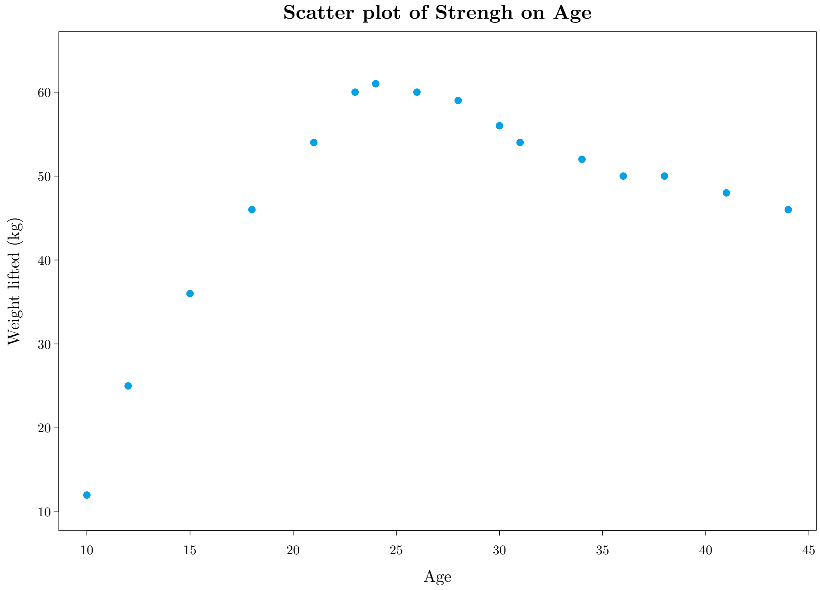
24. The means of two variables X and Y are $\bar{x} = 2$ and $\bar{y} = 1$, and the correlation coefficient is 0.

    (a) Predict the value of Y for $x = 10$.

    (b) Predict the value of X for $y = 5$.

    (c) Plot both regression lines.

25. A study to determine the relation between the age and the physical strength gave the scatter plot below.



    (a) Calculate the linear coefficient of determination for the whole sample.

    (b) Calculate the linear coefficient of determination for the sample of people younger than 25 years old.

    (c) Calculate the linear coefficient of determination for the sample of people older than 25 years old.

    (d) Which model explains better the relation between the age and the strength?

    Use the following sums (X=Age and Y=Weight lifted).

    - Whole sample: $\sum x_i = 431$ years, $\sum y_j = 769$ Kg, $\sum x_i^2 = 13173$ years^2, $\sum y_j^2 = 39675$ Kg^2 and $\sum x_i y_j = 21792$ years·Kg.
    - Young people: $\sum x_i = 123$ years, $\sum y_j = 294$ Kg, $\sum x_i^2 = 2339$ years^2, $\sum y_j^2 = 14418$ Kg^2 and $\sum x_i y_j = 5766$ years·Kg.

– Old people: $\sum x_i = 308$ years, $\sum y_j = 475$ Kg, $\sum x_i^2 = 10834$ years^2, $\sum y_j^2 = 25257$ Kg^2 and $\sum x_i y_j = 16026$ years·Kg.

26. A dietary center is testing a new diet in sample of 12 persons. The data below are the number of days of diet and the weight loss (in Kg) until them for every person.

(33 , 3.9), (51 , 5.9), (30 , 3.2), (55 , 6.0), (38 , 4.9), (62 , 6.2),
(35 , 4.5), (60 , 6.1), (44 , 5.6), (69 , 6.2), (47 , 5.8), (40 , 5.3)

(a) Draw the scatter plot. According to the point cloud, what type of regression model explains better the relation between the weight loss and the days of diet?

(b) Construct the linear regression model and the logarithmic regression model of the weight loss on the number of days of diet.

(c) Use the best model to predict the weight that will lose a person after 100 days of diet. Is this prediction reliable?

Use the following sums (X=days of diet and Y=weight loss): $\sum x_i = 564$ days, $\sum \log(x_i) = 45.8086$ log(days), $\sum y_j = 63.6$ Kg, $\sum x_i^2 = 28234$ days^2, $\sum \log(x_i)^2 = 175.6603$ log(days)^2, $\sum y_j^2 = 347.7$ Kg^2, $\sum x_i y_j = 3108.5$ days·Kg, $\sum \log(x_i) y_j = 245.4738$ log(days)·Kg.

27. The concentration of a drug in blood, in mg/dl, depends on time, in hours, according to the data below.

Drug concentration	2	3	4	5	6	7	8
Hours	25	36	48	64	86	114	168

(a) Construct the linear regression model of drug concentration on time.

(b) Construct the exponential regression model of drug concentration on time.

(c) Use the best regression model to predict the drug concentration after 4.8 hours? Is this prediction reliable? Justify the answer.

Use the following sums (C=Drug concentration and T=time): $\sum c_i = 35$ mg/dl, $\sum \log(c_i) = 10.6046$ log(mg/dl), $\sum t_j = 541$ hours, $\sum \log(t_j) = 29.147$ log(hours), $\sum c_i^2 = 203$ (mg/dl)^2, $\sum \log(c_i)^2 = 17.5206$ log(mg/dl)^2, $\sum t_j^2 = 56937$ hours^2, $\sum \log(t_j)^2 = 124.0131$ log(hours)^2, $\sum c_i t_j = 3328$ mg/dl·hours, $\sum c_i \log(t_j) = 154.3387$ mg/dl· log(hours), $\sum \log(c_i) t_j = 951.6961$ log(mg/dl)·hours, $\sum \log(c_i)\log(t_j) = 46.08046$ log(mg/dl) · log(hours).

28. A researcher is studying the relation between the obesity and the response to pain. Te obesity is measured as the percentage over the ideal weight, and the response to pain as the nociceptive flexion pain threshold. The results of the study appears in the table below.

Obesity	89	90	75	30	51	75	62	45	90	20
Pain threshold	10	12	4	4.5	5.5	7	9	8	15	3

(a) According to the scatter plot, what model explains better the relation of the response to pain on the obesity?

(b) According to the best regression model, what is the response to pain expected for a person with an obesity of 50%? Is this prection reliable?

(c) According to the best regression model, what is the expected obesity for a person with a pain threshold of 10? Is this prediction reliable?

Use the following sums (X=Obesity and Y=Pain threshold): $\sum x_i = 627$, $\sum \log(x_i) = 40.3858$, $\sum y_j = 78$, $\sum \log(y_j) = 19.4119$, $\sum x_i^2 = 45141$, $\sum \log(x_i)^2 = 165.4516$, $\sum y_j^2 = 738.5$, $\sum \log(y_j)^2 = 40.0458$, $\sum x_i y_j = 5538.5$, $\sum x_i \log(y_j) = 1306.051$, $\sum \log(x_i) y_j = 327.3887$, $\sum \log(x_i)\log(y_j) = 80.1831$.

3 Probability

29. Construct the sample space of the following random experiments:
    (a) Pick a random person and measure the gender and whether she or he is smoker or not.
    (b) Pick a random person and measure the blood type and whether she or he is smoker or not.
    (c) Pick a random person and measure the gender, the blood type and whether she or he is smoker or not.

30. There are two boxes with balls of different colors. The first box contains 3 white balls and 2 black balls, and the second one contains 2 blue balls, 1 red ball and 1 green ball. Construct the sample space of the following random experiments:
    (a) Pick a random ball from every box.
    (b) Pick two random balls from every box.

31. The Morgan's laws state that given two events A and B from the same sample space, $\overline{A \cup B} = \bar{A} \cap \bar{B}$ and $\overline{A \cap B} = \bar{A} \cup \bar{B}$. Proof both statements graphically using Venn diagrams.

32. Calculate the probability of the following events of the random experiment consisting in tossing 3 coins:
    (a) Get exactly 1 heads.
    (b) Get exactly 2 tails.
    (c) Get two or more heads.
    (d) Get some tails.

33. In a laboratory there are 4 flasks with sulfuric acid and 2 with nitric acid, and in another laboratory there are 1 flask with sulfuric acid and 3 with nitric acid. A random experiment consist in picking two flask, one from every laboratory. Calculate the probability of the following events:
    (a) The two picked flasks are of sulfuric acid.
    (b) The two picked flasks are of nitric acid.
    (c) The tow picked flasks contains different acids.

    Calculate the same probabilities if the flask picked in the first laboratory is put in the second laboratory before picking the flask from it.

34. Let A and B be events of the same sample space, such that $P(A) = 3/8$, $P(B) = 1/2$, $P(A \cap B) = 1/4$. Calculate the following probabilities:
    (a) $P(A \cup B)$.
    (b) $P(\bar{A})$ y $P(\bar{B})$.
    (c) $P(\bar{A} \cap \bar{B})$.
    (d) $P(A \cap \bar{B})$.
    (e) $P(A|B)$.
    (f) $P(A|\bar{B})$.

35. In a hospital the probability of getting hepatitis in a blood transfusion from a unit of blood is 0.01. A patient gets two units of blood while staying at the hospital. What is the probability of getting hepatitis?

36. Let A and B be two events of the same sample space, such that $P(A) = 0.6$ and $P(A \cup B) = 0.9$. Calculate $P(B)$ with the following assumptions:
    (a) A and B are incompatible.
    (b) A and B are independent.

37. A study about smoking has published that 40% of smokers have a smoker father, 25% have a smoker mother and 52% have al least one of the parents smoker. We pick a random person from this population. Answer the following questions:

    (a) What is the probability of having a smoker mother if the father smokes?

    (b) What is the probability of having a smoker mother if the father doesn't smoke?

    (c) Are independent the events having a smoker father and having a smoker mother?

38. The probability that an injury A is repeated is $4/5$, the probability that another injury B is repeated is $1/2$, and the probability that both injuries are repeated is $1/3$. Calculate the probability of the following events:

    (a) Only injury B is repeated.

    (b) At least one injury is repeated.

    (c) Injury B is repeated if injury A has been repeated.

    (d) Injury B is repeated if injury A hasn't been repeated.

39. In a digestive clinic from every 1000 patients that arrive with stomach pain, 700 have gastritis, 200 have an ulcer and 100 have cancer. After analyzing the gastric symptoms, it is known that the probability of having vomiting is 0.3 in case of gastritis, 0.6 in case of ulcer and 0.9 in case of cancer. What is the diagnosis for a new patient with stomach pain that has vomiting?

    Note: Assume that the only diseases are gastritis, ulcer and cancer and that are incompatible among them.

40. To evaluate the effectiveness of a diagnosis test, the test was applied to a sample of people with the following results:

	Test +	Test −
Sick	2020	80
Healthy	140	7760

    Calculate for this test:

    (a) The sensibility and the specificity.

    (b) The positive and negative predictive value.

    (c) The probability of a correct diagnosis.

41. We know, from a research study, that 10% of people over 50 years suffer a particular type or arthritis. We have developed a new method to detect the disease and after clinical trials we observe that if we apply the method to people with arthritis we get a positive result in 85% of cases, while if we apply the method to people without arthritis, we get a positive result in 4% of cases. Answer the following questions:

    (a) What is the probability of getting a positive result after applying the method to a random person?

    (b) If the result of applying the method to one person has been positive, what is the probability of having arthritis?

42. A severe pain without effusion in a particular zone of the knee joint is a symptom of sprained lateral collateral ligament (SLCL). If the sprains in that ligament are classified into grade 1, when there is only distension (60% of cases), grade 2 when there is a partial tearing (30% of cases) and grade 3 when there is a complete tearing (10% of cases). Taking into account that the symptom appears in 80% of cases of grade 1 sprains, 90% of cases of grade 2 and 100% of cases of grade 3, answer the following questions:

    (a) If a person has SLCL what is the probability that he or she present severe pain without effusion?

(b) What is the diagnosis for a person with severe pain without effusion?

(c) From a total of 10000 people with severe pain without effusion, how many are expected to have a grade 1 sprain? How many are expected to have a grade 2 sprain? And a grade 3 sprain?

43. A physiotherapist uses two techniques A and B to cure an injury. It is known that the injury is 3 times more frequent in people over 30 than in people under 30. It is also known that in people over 30 technique A works in 30% of cases and technique B in 60%, while in people under 30 technique A works in 50% of cases and technique B in 70%. If both techniques are applied with the same probability, no matter the age,

    (a) What is the probability of a random person under 30 to cure? And for a people over 30?

    (b) What is the probability of a random person to cure?

    (c) If after applying a technique to a person over 30, the person doesn't cure, what is the probability that the technique applied was A?

44. We have two different test A and B to diagnose a disease. Test A have a sensitivity of 98% and a specificity of 80%, while test B have a sensitivity of 75% and a specificity of 99%.

    (a) What test is better to confirm the disease?

    (b) What test is better to rule out the disease?

    (c) Often a test is used to discard the presence of the disease in a large amount of people apparently healthy. This type of test is known as *screening test*. What test will work better as a screening test? What is the positive predictive value (PPV) of this test if the prevalence of the disease is 0.01? And if the prevalence of de disease is 0.2?

    (d) The positive predictive value of a screening test used to be not too high. How can combine tests A and B to have a higher confidence in the diagnosis of the disease? Calculate the post-test probability of having the disease with the combination of both tests if the outcome of both test is positive for a prevalence of 0.01.

4 Discrete random variables

45. Let X be a discrete random variable with the following probability distribution

X	4	5	6	7	8
$f(x)$	0.15	0.35	0.10	0.25	0.15

   (a) Calculate and represent graphically the distribution function.
   (b) Calculate the following probabilities
      i. $P(X < 7.5)$.
      ii. $P(X > 8)$.
      iii. $P(4 \leq X \leq 6.5)$.
      iv. $P(5 < X < 6)$.

46. Let X be a discrete random variable with the following probability distribution

$$F(x) = \begin{cases} 0 & \text{si } x < 1, \\ 1/5 & \text{si } 1 \leq x < 4, \\ 3/4 & \text{si } 4 \leq x < 6, \\ 1 & \text{si } 6 \leq x. \end{cases}$$

   (a) Calculate the probability function.
   (b) Calculate the following probabilities
      i. $P(X = 6)$.
      ii. $P(X = 5)$.
      iii. $P(2 < X < 5.5)$.
      iv. $P(0 \leq X < 4)$.
   (c) Calculate the mean.
   (d) Calculate the standard deviation.

47. An experiment consist in injecting a virus to three different types of rats and observing if they survive or not. It is known that the probability of surviving is 0.5 for the first type of rats, 0.4 for the second type and 0.3 for the third type.
   (a) Calculate the probability function of the variable X that measures the number of surviving rats.
   (b) Calculate the distribution function.
   (c) Calculate $P(X \leq 1)$, $P(X \geq 2)$ y $P(X = 1.5)$.
   (d) Calculate the mean and the standard deviation. Is representative the mean?

48. The chance of being cured with a treatment is 0.85. If we apply the treatment to 6 patients,
   (a) What is the probability that half of them are cured?
   (b) What is the probability that a least 4 of them are cured?

49. Ten persons came into contact with a person infected by tuberculosis. The probability of being infected after contacting a person with tuberculosis is 0.10.
   (a) What is the probability that nobody are infected?
   (b) What is the probability that at least 2 persons are infected?
   (c) What is the expected number of infected persons?

50. The probability of having an adverse reaction to a vaccine is 0.001. If 2000 persons are vaccinated, what is the probability of having some adverse reaction?

51. The average number of calls per minute that arrive to a telephone switchboard is 120.

(a) What is the probability of receiving less than 4 calls in 2 seconds?

(b) What is the probability of receiving at least 3 calls in 3 seconds?

52. A test contains 10 questions with 3 possible options each. For every question you win a point if you give the right answer and loss half a point if the answer is wrong. A student knows the right answer for 3 of the 10 questions and answers the rest randomly. What is the probability of passing the exam?

53. It has been observed experimentally that 1 of every 20 trillions of cells exposed to radiation mutates becoming carcinogenic. We know that the human body has approximately 1 trillion of cells by kilogram ot tissue. Calculate the probability that a 60 kg person exposed to radiation develops cancer. If the radiation affects 3 persons weighing 60 kg, what is the probability that a least one of them develops cancer?

54. A diagnostic test for a disease returns 1% of positive outcomes, and the positivie and negative predictive values are 0.95 and 0.98 respectively.

    (a) Calculate the prevalence of the disease.

    (b) Calculate the sensitivity and the specificity of the test.

    (c) If the test is applied to 12 sick persons, what is the probability of getting at least a wrong diagnosis?

    (d) If the test is applied to 12 persons, what is the probability of getting a right diagnosis for all of them?

55. In a study about a parasite that attack the kidney of rats it is known that the average number of parasites per kidney is 3.

    (a) Calculate the probability that a rat have more than 3 parasites.

    (b) Calculate the probability of having at least 9 rats infected in a sample of 10 rats.

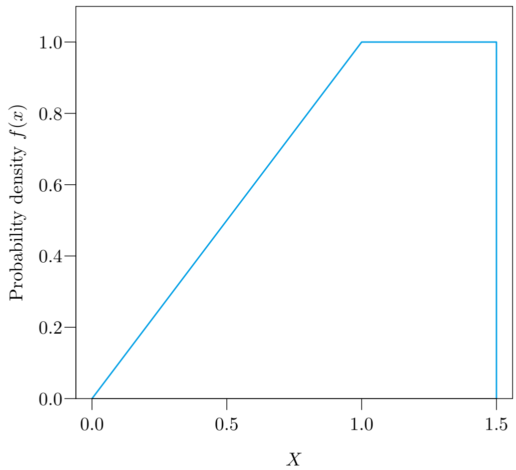
56. In a physiotherapy course there are 60% of females and 40% of males.

    (a) If 6 random students have to go to a hospital for making practices, what is the probability of going more males than females?

    (b) In 5 samples of 6 students, what is the probability of having some sample without males?

5 Continuous random variables

57. Given the continuous random variable X with the following probability density function chart,



(a) Check that $f(x)$ is a probability density function.
(b) Calculate the following probabilities
   i. $P(X < 1)$
   ii. $P(X > 0)$
   iii. $P(X = 1/4)$
   iv. $P(1/2 \leq X \leq 3/2)$
(c) Calculate the distribution function.

58. A worker can arrive to the workplace in any instant between 6 and 7 in the morning with the same likelihood.

(a) Compute and plot the probability density function of the variable that measures the arrival time.
(b) compute and plot the distribution function.
(c) Compute the probability of arriving quarter past six and half past six.
(d) What is the expected arrival time?

59. Let Z be a random variable following a standard normal distribution model. Calculate the following probabilities using the table of the distribution function:

(a) $P(Z < 1.24)$
(b) $P(Z > -0.68)$
(c) $P(-1.35 \leq Z \leq 0.44)$

60. Let Z be a random variable following a standard normal distribution model. Determine the value of x in the following cases using the table of the distribution function:

(a) $P(Z < x) = 0.6406$.
(b) $P(Z > x) = 0.0606$.
(c) $P(0 \leq Z \leq x) = 0.4783$.
(d) $P(-1.5 \leq Z \leq x) = 0.2313$.
(e) $P(-x \leq Z \leq x) = 0.5467$.

61. Let X be a random variable following a normal distribution model $N(10, 2)$.

(a) Calculate $P(X \leq 10)$.
(b) Calculate $P(8 \leq X \leq 14)$.
(c) Calculate the interquartile range.
(d) Calculate the third decile.

62. It is known that the glucose level in blood of diabetic persons follows a normal distribution model with mean 106 mg/100 ml and standard deviation 8 mg/100 ml.

    (a) Calculate the probability of a random diabetic person having a glucose level less than 120 mg/100 ml.
    (b) What percentage of persons have a glucose level between 90 and 120 mg/100 ml?
    (c) Calculate and interpret the first quartile of the glucose level.

63. It is known that the cholesterol level in males 30 years old follows a normal distribution with mean 220 mg/dl and standard deviation 30 mg/dl. If there are 20000 males 30 years old in the population,

    (a) how many of them have a cholesterol level between 210 and 240 mg/dl?
    (b) If a cholesterol level greater than 250 mg/dl can provoke a thrombosis, how many of them are in risk of thrombosis?
    (c) Calculate the cholesterol level above which 20% of the males are?

64. In an exam done by 100 students, the average degree is 4.2 and only 32 students pass. Assuming that the grade follows a normal distribution model, how many students got a grade greater than 7?

65. In a population with 40000 persons, 2276 have between 0.80 and 0.84 milligrams of bilirubin per deciliter of blood, and 11508 have more than 0.84. Assuming that the level of bilirubin in blood follows a normal distribution model,

    (a) Calculate the mean and the standard deviation.
    (b) How many persons have more than 1 mg of bilirubin per dl of blood?

66. It is known that the blood pressure of people in a population with 20000 persons follows a normal distribution model with mean 13 mm Hg and interquartile range 4 mm Hg.

    (a) How many persons have a blood pressure above 16 mm Hg?
    (b) How much have to decrease the blood pressure of a person with 16 mm Hg in order to be below the 40% of people with lowest blood pressure?

67. A study tries to determine the effect of a low fat diet in the lifetime of rats. The rats where divided into two groups, one with a normal diet and another with a low fat diet. It is assumed that the lifetime of both groups follows normal distribution model with the same variance but different mean. If 20% of rats with normal diet lived more than 12 months, 5% less than 8 moth, and 85% of rats with low fat diet lived more than 11 moths,

    (a) what is the mean and the standard deviation of lifetime of rats following a low fat diet?
    (b) If there was 40% of rats with normal diet, and 60% of rats with low fat diet, what is the probability that a random rat die before 9 months?

68. A diagnostic test to determine doping of athletes returns a positive outcome when the concentration of a substance in blood is greater than 4 μg/ml. If the distribution of the substance concentration in doped athletes follows a normal distribution model with mean 4.5 μg/ml and standard deviation 0.2 μg/ml, and in non-doped athletes follow a normal distribution model with mean 3 μg/ml and standard deviation 0.3 μg/ml,

    (a) what is the sensitivity and specificity of the test?
    (b) If there are a 10% of doped athletes in a competition, what is the positive predicted value?

69. According to the central limit theorem, for big samples ($n \geq 30$) the sample mean $\bar{x}$ follows a normal distribution model $N(\mu, \sigma/\sqrt{n})$, where μ is the population mean and σ the population standard deviation.

    It is known that in a population the sural triceps elongation follows has mean 60 cm and standard deviation 15 cm. If you draw a sample of 30 individuals from this population, what is the probability of having a sample mean greater than 62 cm? If a sample is atypical if its mean is below the 5th percentile, is atypical a sample of 60 individuals with $\bar{x} = 57$?

70. The curing time of a knee injury in soccer players follows a normal distribution model with mean 50 days and standard deviation 10 days. If there is a final match in 65 days, what is the probability that a player that has just injured his knee will miss the final? If the semifinal match is in 40 days, and 4 players has just injured the knee, what is the probability that some of them can play the semifinal?