Bachelor Degree in Computer Engineering

Statistics group E (English)

FIRST PARTIAL EXAM

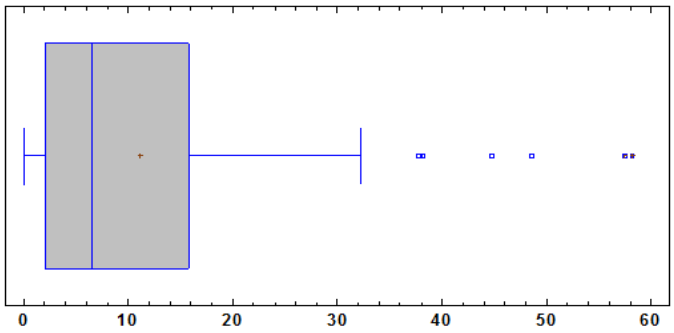
March 21st 2016

Surname, name	
Signature	

Instructions

1. Write your name and sign in this page.
2. Answer each question in the corresponding page.
3. All answers must be justified.
4. Personal notes in the formula tables will not be allowed.
5. Mobile phones are not permitted over the table. It is only permitted to have the DNI (identification document), calculator, pen, and the formula tables. Mobile phones cannot be used as calculators.
6. Do not unstaple any page of the exam (do not remove the staple).
7. All questions score the same (over 10).
8. At the end, it is compulsory to sign in the list on the professor's table in order to justify that the exam has been handed in.
9. Time available: **2 hours**.

1. A set of 100 values have been collected corresponding to the time (in months) of operation until failure of electronic components used in computer systems. These values have been represented in the following plot:



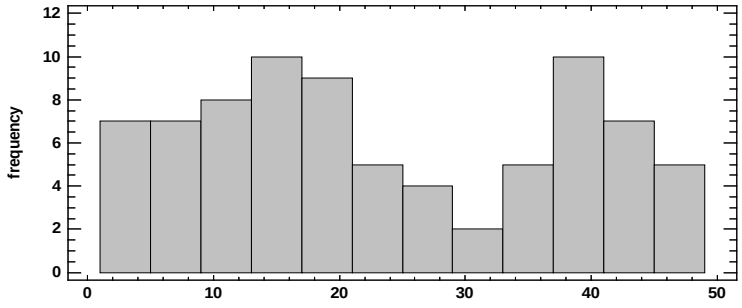
a) Do you think that the results correspond to any known distribution model? If so, what would it be? Justify your answer. *(2 points)*

b) What would be the most appropriate parameter of position in this case? Why? Calculate its approximate value. *(2 points)*

c) What would be the most appropriate parameter of dispersion in this case? Why? Calculate its approximate value. *(2 points)*

d) What can you say in this case about the value of the standardized skewness coefficient? *(2 points)*

e) Certain report indicates that, based on the same data, another graphical representation was obtained, which is shown below. What is the name of this plot? Is it compatible with the box-whisker shown above or, on the contrary, do you think that it must correspond to an error in the report? Justify conveniently your reply. *(2 points)*



2. Certain school of computer science has performed a survey about the brand of laptops owned by the students. The following results were obtained:

	TOSHIBA	**ACER**	**HP**
Men	120	80	20
Women	60	40	10

Considering that the set of surveyed students constitutes the population under study and assuming that each student only owns one laptop, answer the following questions by clearly defining the events considered, and justifying conveniently your replies.

a) What is the percentage of female students in the survey? *(2 points)*

b) What is the probability of a surveyed student to be male and to own an ACER laptop? *(2 points)*

c) If a given laptop belongs to a female student, what is the probability to be HP? *(2 points)*

d) The following events are considered: A = "to be male" and B = "to own a laptop of brand TOSHIBA". Are these events independent? *(2 points)*

e) Knowing that a laptop is not of the brand ACER, what is the probability of belonging to a female student? *(2 points)*

3. A given company manufactures certain type of network card, which are sold in large lots. The company has found that 1% of the manufactured cards in each lot are defective on average. In order to assess the quality of each lot, a random sample of 20 cards is taken, and the number of defective units is determined in this sample.

a) Define the random variable and indicate its distribution. *(2 points)*

b) Calculate the probability of finding a maximum of 2 defective cards. *(4 points)*

c) Calculate the number of network cards, **x**, so that the probability to find **x** or less defective cards should be less than 95%. *(4 points)*

4. In certain computer terminal, the time of occupation by each user follows an exponential model, with a median of 5 minutes.

a) In order to properly optimize the system, what should be the time of occupation ***t*** so that 90% of users have finished their inquiries? *(3 points)*

b) A new user arrives and finds that the terminal is being occupied by someone who has been working during 5 minutes. What is the probability to have to wait during 5 additional minutes until the terminal is available? Discuss the result obtained indicating what is the property that allows the calculation of the requested probability. *(4 points)*

c) Repeat the calculation of probability in the previous section (b) by assuming that the occupation time of the terminal by each user is a uniform variable ranging from zero to 15 minutes. *(3 points)*

SOLUTION

1a) Data could be modelled by means of the exponential distribution because the box-and-whisker plot indicates a positively skewed distribution and the first quartile is close to zero.

1b) The median is a parameter of position more appropriate than the mean in strongly skewed distributions, like this case, because it is not affected by the presence of extreme values (shown in the plot). It corresponds to the vertical line inside the box (approximate value: **6.5**).

1c) The interquartile range (IQR) is a parameter of dispersion more appropriate than the variance or the range because it is not affected by extreme values.
IQR = Q_3 - Q_1 = 16 - 2 = **14** months

1d) When the standardized skewness coefficient (SSC) is small (between -2 and 2), the data can be regarded as a random sample taken from a symmetric distribution. In this case, the box-and-whisker plot indicates a clearly positively skewed distribution because the right whisker is much longer than the left one, and the median is displaced to the left. Thus, the SSC will be for sure greater than 2.

1e) This plot is a histogram (also called "frequency histogram"). The shape of this plot is not skewed and, hence, it does not match with the box-whisker plot. According to the histogram, there are 27 values greater than 32, but only 6 based on the box-whisker. By summing the height of all bars, the histogram was built with 79 values, while the box-whisker was obtained with 100 data. Both plots were clearly obtained with different values and, hence, it must correspond to an error in the report.

2a) Event F: the student is female (woman). There are 110 women, 220 men (330 in total). By applying the rule of Laplace, there are 110 students who satisfy this event: P(F)=110/330 = 1/3 = **33.3%**

2b) Event M: the student is male (man). Event A: the student owns an ACER laptop. By applying the rule of Laplace, there are 80 students who satisfy both events, so that: $P(M \cap A) = 80/330 = \mathbf{0.242}$

Note: The expression below can only be considered as correct if calculations are provided to demonstrate that M and A are independent events.

$$P(M \cap A) = P(M) \cdot P(A) = \frac{220}{330} \cdot \frac{120}{330} = 0.242$$

2c) Event F: the student is female. Event HP: the student owns a HP laptop. There are 10 women who own a HP laptop in the subset of 110 women. Thus, by applying the rule of Laplace: P(HP/F) = 10/110 = **0.091**

Note: The expression below can only be considered as correct if calculations are provided to demonstrate that F and HP are independent events.

$$P(HP/F) = \frac{P(HP \cap F)}{P(F)} = \frac{P(HP) \cdot P(F)}{P(F)} = P(HP) = \frac{30}{330} = 0.091$$

2d) Two events are independent if they satisfy:
P(A/B) = P(A); P(A/B) = 120/180 = 2/3; P(A) = 220/330 = 2/3
P(B/A) = P(B); P(B/A) = 120/220 = 0.545; P(B) = 180/330 = 0.545
$P(A \cap B) = P(A) \cdot P(B)$; $120/330 = (220/330) \cdot (180/330) = 0.364$
These conditions are satisfied and, hence, A and B are independent events.

2e) Event F: the student is female. Event A: the student owns an ACER laptop. Among the 210 laptops not ACER (180 Toshiba and 30 HP), 70 of them belong to women. Thus, by applying the rule of Laplace:

$$P\left(F/\overline{A}\right) = \frac{60+10}{180+30} = \frac{70}{210} = \frac{1}{3} = \mathbf{0.33}$$

Note: The expression below can only be considered as correct if calculations are provided to demonstrate that F and not-Acer are independent events.

$$P\left(F/\overline{A}\right) = \frac{P\left(F \cap \overline{A}\right)}{P\left(\overline{A}\right)} = \frac{P(F) \cdot P\left(\overline{A}\right)}{P\left(\overline{A}\right)} = P(F) = \frac{110}{330} = 0.33$$

3a) Random variable X: number of defective units found in the sample of 20 network cards.
X follows a binomial distribution: B (n=20, p=0.01) because the minimum value is 0 and the maximum is 20.

3b) $P(X \le 2) = P(X=0) + P(X=1) + P(X=2) = \binom{20}{0} \cdot 0.01^0 \cdot 0.99^{20} +$

$$+ \binom{20}{1} \cdot 0.01^1 \cdot 0.99^{19} + \binom{20}{2} \cdot 0.01^2 \cdot 0.99^{18} = 0.81791 + 0.16523 + 0.01586 = 0.9990$$

The same result is obtained by approximating the Binomial by means of a Poisson distribution with λ = 0.2 and reading on the curve 2 in the abacus.

3c) P(X≤**x**) < 0.95. According to the results obtained in 3b):
P(X=0) = 0.82; P(X≤1) = 0.82+0.16 = 0.98 which is > 0.95. Then, the solution has to be zero. It is not possible to find any value **x** different from zero that satisfies the requested condition.

4a) $P(T > 5) = 0.5 = e^{-\alpha \cdot 5}$; $\alpha = -(\ln 0.5)/5 = 0.1386$
$P(T < t) = 0.9$; $P(T > t) = 0.1 = e^{-0.1386 \cdot t}$; $t = -(\ln 0.1)/0.1386 = \mathbf{16.61\ minutes}$

4b) T: time of occupation of the computer terminal. The probability to wait more than 10 minutes is equivalent to the probability that the terminal is occupied more than 10 min. Based on the lack-of-memory property of the exponential distribution, and taking into account that 5 is the value of the median, the requested probability can be calculated as:
$P\left[(T>10)/(T>5)\right] = P(T>5) = \mathbf{0.5}$

4c) In this case, X ≈ U(0; 15) ; P(T<t) = (t-0)/(15-0)

$$P\left[(T>10)/(T>5)\right] = \frac{P\left[(T>10) \cap (T>5)\right]}{P(T>5)} = \frac{P(T>10)}{P(T>5)} = \frac{1-P(T<10)}{1-P(T<5)} = \frac{1-10/15}{1-5/15} = 0.5$$