

**Bachelor Degree in Computer Engineering**

**Statistics**

**FIRST PARTIAL EXAM**

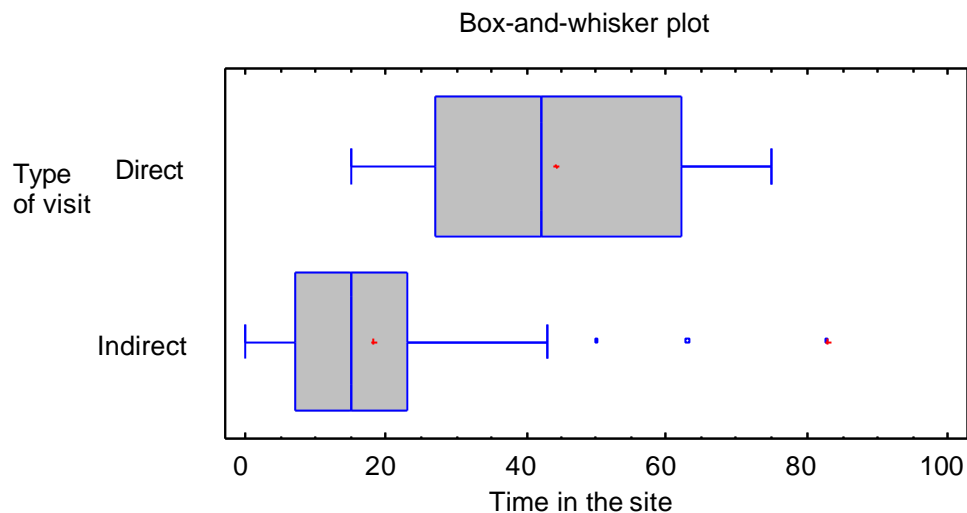
**March 7<sup>th</sup> 2011**

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. Over the table it is only permitted to have the DNI (identification document), calculator, pen, and the formula tables.
5. Do not unstaple any page of the exam (do not remove the staple).
6. All questions score the same (over 10).
7. 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.
8. Time available: 2 hours.

**1.** One company of web design wants to study the profile of those who search in a certain web page as a function of the type of visit. One visit is direct if the user writes the web address of the site that wants to visit or accesses through the favorite markers of the navigator, while one indirect visit comes from search engines (Google, MSN Search, Yahoo!) or an external link in other web sites. The Box-and-Whiskers plot of the time registered as a function of the type of visit is the following:



According to this plot obtained, answer the following questions:

**a)** What would be the most suitable parameters of position and dispersion in order to characterize the variable of time, according to the type of visit?

**b)** From a descriptive point of view, can you detect differences regarding the dispersion of the time, according to the type of visit of the user?

**c)** Is there any indication of asymmetry (skewness) in the times recorded according to one or the other type of visit?

**d)** If possible, estimate the number of data that have been taken to represent the figure.

**e)** Indicate if the following sentence is correct: “Approximately 50% of the persons who access by indirect visit pass less than 15 minutes in the web page while this percentage is 25% for those who access by direct visit”.

2. One software has been created based on pattern recognition techniques, for the automatic classification of waste product in 3 different categories: Organic (O), Containers (C) and Paper (P). Sometimes the software makes errors, and classifies one waste product that is actually organic (Or) in the category of containers (Cc) or in the category of Paper (Pc). Similarly, one waste product corresponding in fact to the category of Containers (Cr) is classified in the category of organic (Oc) or in the category of paper (Pc). The same occurs with the category of Paper. The following table shows what is called confusion matrix of this classification software, which is in fact a cross-tabulation table:

		CLASSIFICATION SOFTWARE		
		Oc	Cc	Pc
REAL	Or	350	10	11
	Cr	15	200	12
	Pr	3	5	150

Being:

Oc={The classification software says that the object is Organic}

Ec={The classification software says that the object is Container}

Pc={The classification software says that the object is Paper}

Or={The object is in fact Organic}

Er={The object is in fact Container}

Pr={The object is in fact Paper}

According to this information, answer the following questions:

a) What percentage of objects are actually organic?

b) What percentage of objects belong to the category of Paper and are classified as Container?

**c) What percentage of objects classified as Organic are actually Paper?**

**d) What percentage of objects have been classified as Container?**

**e) What is the percentage of correct classification of this software?**

**3.** One productive process manufactures integrated circuits in two different lines. The first line has double production capacity than the second one. The first line produces 0.5% of defective circuits. The second line produces 0.3% of defective circuits.

In a certain moment, the process detects a defective circuit. What is the probability that this circuit comes from the second line?

**4.** One company devoted to the marketing of computer equipment produces certain units used in the assembly of network boosters. The company wants to guarantee that the proportion of defective units in the batches that sells does not exceed 0.7 per cent. For this purpose,  $N$  units are randomly selected from each batch, and the batch is accepted if all of them are correct. Calculate the minimum value of  $N$  so that the probability of accepting as good one batch with 0.7 per cent or more of defective units does not exceed 10%.

**5.** One on-line consulting center receives in average 6 consultations per hour. It can be considered that the number of consultations per hour follows a Poisson distribution.

**a)** Calculate the probability to receive exactly 3 consultations in a certain hour.

**b)** Calculate the probability to receive more than 20 consultations in a period of 3 hours.