



UNIVERSIDADE DE COIMBRA
FACULDADE DE CIÊNCIAS E TECNOLOGIA
Departamento de Engenharia Informática
Pólo II - Pinhal de Marrocos, 3030 Coimbra - Portugal

Exam of Experimental Methods in Computer Science

January 12th, 2018

Maximum time allowed: 2h:15m

Name: _____ No.: _____

Note: The students are allowed to use their own material in paper format during the exam. The only electronic device allowed is an electronic calculator, provided that the calculator cannot connect to the Internet.

- 1) Assume you are the system manager of a large organization and, in such position, you are responsible for a quite large and complex data centre and network infrastructure. In recent months, you have noticed a strange behaviour in one of the data storage servers, which consists of an abrupt drop in the performance of the server, followed by an also inexplicable resume of normal performance after some time (tens of minutes). The company that has provided the storage server has already analysed the system (hardware and software) and has declared that the storage server is in perfect conditions, claiming that the strange behaviour could be caused by an interaction with the other systems that use the services of the storage server. As the problem persists, you decided to increase the amount of data logged in the storage server and in other related machines (to have more information that may help identifying the problem) and after some analysis you believe you have found an explanation for the phenomenon. Describe the formal steps you would take to confirm that your explanation is right. (2.0/20)

- 2) Representativeness of an experiment is one of the most important properties to be considered in experiment design. Assuming the general context of experiments involving computer science and informatics engineering, describe the key aspects to take into account to assure good representativeness of an experiment. Make sure your answer is concise yet accurate and complete. (1.5/20)

- 3) A panel of 15 individuals, representing typical users, participating in the evaluation of the improved version of the interface of a mobile application for shopping in a big supermarket chain. The evaluation consists of a set of tasks to be performed and the result is the time taken by the user to complete the tasks. The new version of the interface (V2) is supposed to be simpler than the previous version (V1), so the time needed to execute the tasks in the interface V2 is expected to be shorter than in V1. The experiment was organized in the following way: 7 individuals were given the original interface application (V1) and 8 were given the application with the new interface (V2). The set of tasks performed by the two sub-groups of the panel in each version of the application are the same. The following table shows the results:

	Measurements								Average	Std. Dev
V1	6,5	4,2	10,9	5,3	7,8	5,7	9,7		7,16	2,44
V2	4,8	8,4	4,0	6,9	5,0	6,3	7,8	4,9	6,01	1,58

- a) Based on these results, do you think V1 should be replaced by V2? Consider 95% of confidence. (4.0/20)

- b) Could you propose a better design for this experiment? Explain briefly and precisely how you would conduct your improved experiment. (1.5/20)

- 4) An engineer is asked to improve a system for predicting fraud in financial transactions in a large bank. The system analyses large amounts of data recorded in the database of the bank in order to find financial transactions or groups of transactions that are marked as suspicious. Then a set of procedures and checks are performed, which will confirm if it is a case of fraud or if it is a false positive. As these checks involve manual steps, their cost is very high. It is therefore important that the fraud detection system has the minimum possible number of false positives. This is precisely the focus of the engineer's work: reducing the number of false positives. In any case, the reduction in the false positives must not have significant impact on the percentage of real frauds the system can detect.

To do this, the engineer tests three fraud predicting models (named here as FM1, FM2 and FM3) that can be used by the system. Each model is tried with various model settings and submitted to some sets of historical data with a variety of known fraud attempts. As the financial frauds in the historical data are known, it is possible to determine if each model detects the frauds correctly, as well as if the model signals frauds that are not there (false positives). The result of all the three predicting models when analysing a set of financial transactions is expressed as a binary result indicated if there is a fraud in a set of data or not (i.e., the output is not probabilistic; it is a yes or no).

Given the circumstances described above, answer to the following questions:

- a) What kind of statistical test would you use to test hypotheses in the scenario described? Justify your answer. (1.0/20)

- b) Dependent variable(s) (1.0/20)

- c) Independent variables (1.0/20)

- d) Examples of possible levels for the independent variables in the experiments (1.0/20)

- e) Two examples of hypothesis that could be tested (e.g. in different phases of the experiment) (1.5/20)

- 5) A company intends to purchase a large server for high-performance processing and, given the various offerings from the different vendors, the company has decided to ask each vendor for a server performance test in the configuration proposed by the vendor. If there are no performance differences in the servers, it means that the purchase decision will be based on other criteria (i.e., not the performance), in particular criterion such as the cost and conditions of the maintenance contract.

Knowing that the execution time depends on the type of program, the company defined a workload composed of several programs representative of the type of processing that the company needs, having provided this workload to each vendor for the tests, and also indicated that the workload should be executed several times so that the execution time was calculated as the average of the execution times of each run. The results received from each vendor are described in the following table:

	Serv. F1	Serv. F2	Serv. F3	Serv. F4
Number of runs	7	6	6	7
Average execution time (sec.)	20,2	15,8	21.9	24,0
Standard deviation	5,13	5,59	4,78	3,66

- a) In view of these results, the company wants to know if there are significant performance differences in the servers, considering a confidence level of 95%. Show all the steps and calculations. (4.0/20)

- b) Independently of the conclusion found in a), suppose you want to know whether the server F2, which seems to be faster than the rest, is in fact the fastest server or not. Explain the approach you would use to find if F2 is faster than the other servers (just explain; no need for calculations). (1.5/20)