Experimental Methods
in Computer Science (and Informatics Engineering)
2017-2018

Training exercises 2

This set of training exercises is meant to help students in the preparation for the written exams. If, for any reason, you conclude that it is not possible to answer a given question explain what is wrong and indicate the reason why the question cannot be answered.

1) The following summary table shows the results of the execution time of a package of benchmark programs compiled with four different compilers, named as A, B, C and D for experiment purposes. Show whether the compilers used have any effect on the execution time of the compiled programs or not considering 95% of confidence.

Compiler	Number of runs	Exec. time mean	Exec. time standard deviation
A	8	4.35	2.25
B	11	3.79	1.63
C	10	4.91	1.89
D	9	6.33	2.21

2) A Bank has developed a new database system with the goal of improving the speed of typical web banking transactions done by the Bank customers. There are several types of transactions (e.g., viewing account balances, funds transfers between the customer's accounts, viewing transactions in a given period indicated by the customer, etc.) The response time of each transaction depends on several factors related to both the queries issued by the customer and the load of the system. In order to assess the new database response, the Bank asked its Information Systems Department to perform an experiment to test the speed of the new database and compare the speed with the current database. The experiment includes a set of representative accounts with typical transactions profiles to assure that the measurements of the speed in both databases (the new one and the already existing) are meaningful.

   a) State the different elements of the experiment (problem definition, variables, levels, hypothesis,…) and justify your choices.

   b) In very few cases it was observed the response time of the transactions was several times larger than the average response time. Explain the choices you have to deal with this problem.

3) A company has developed a new word-searching algorithm in files in smartphones and wants to test the speed of the new algorithm in three operating systems (Android, iOS, and Ubuntu Touch) and in two smartphones (named A and B for the experiment purposes). After a set of measurements, the data collected originated the following ANOVA table.

	SS	df	MS	F	P
Factor OS	23.333	2	11.6667	17.5	1.384e-06
Factor Smartphone	1.667	1	1.667	2.5	0.1197
Interaction	23.333	2	11.6667	17.5	1.384e-06
Within (error)	36.000	54	0.6667		

Based on this table, what conclusions can you take considering a significance level of 5%.

4) In the design of a large-scale experiment involving virtualized infrastructures such as the ones that support the Cloud you have identified a very large number of variables. You are aware of the fact that a large number of variables complicates the experiments and increases the whole experimental cost. Because of that, you are reanalysing the elements of your experiment to simplify (and to focus) the experiment as much as possible. In particular, should the following two variables (A and B) be considered in the experiment (as independent variables) or can you use just one of them? Justify your answer.

Variable A	Variable B
2.34	3.36
3.98	2.23
7.89	-2.98
12.13	-7.56
13.45	-7.98

5) Your company decided to try a training course meant to improve the productivity of code developers. The course is provided by an international software-consulting firm and is quite expensive, thus your company decided to run a pilot study with a group of 8 programmers randomly selected among the employees, to assess the effectiveness of the course before extending the course to the nearly 300 programmers of the company. In order to measure the productivity of code developers your company use "function points", which provide a more reliable and realistic measure of programmers productivity than the "classic lines of code" metrics. You have been assigned to the task of planning and managing the pilot study. In order to have a reference point you measure the productivity of the 8 programmers selected for the pilot study before and after the training course. To assure that the function point measurements are comparable, you used the specification of a set of benchmark programs especially design to assess code development productivity. The following table shows the productivity (in function points per unit of time) of each programmer before and after the training.

Developer ID	1	2	3	4	5	6	7	8
Function points (before)	12	11	9	16	12	11	7	13
Function points (after)	11	15	12	14	16	12	11	9

Can you state with 95% of confidence that the training has improved the code development productivity of the programmers, so you can confidently recommend the extension of the course to all the programmers of your company? Justify your answer.

6) A company is deciding between two programming languages (named here as languages A and B) for a given project. Due to the nature of the project, the cost of the software maintenance is very important, as it is expected a long period of changes and adjustments in the code after the deployment of the first version of the product. The company decided to perform a preliminary experiment to evaluate whether there is a difference between the languages A and B concerning the time needed to introduce modifications in the program modules. Describe the following basic elements of the experiment:

   c) Problem statement

   d) Variables

   e) Examples of levels for the different independent variables

   f) Hypothesis

   g) Assumptions and hypothesis testing technique that should be used in this experiment.

7) After a series of security incidents that compromised the web pages of the University of Conimbriga, campus of Condeixa, the administration of the university has decided to install a new intrusion detection system (IDS) and redefined in full the security procedures and configurations in the data centre and in the university network. In order to verify if the new IDS and security procedures represent an improvement concerning the previous scenario, the university decided to hire a Red Team from a specialized security company to simulate attacks (using penetration tests and multi-layered attack simulation) and compare the level of protection of the University of Conimbriga network and data infrastructure in the two scenarios (i.e., before the installation of the new IDS and new set of security policies and after the new IDS and new security policies). The report produced by the Red Team is quite comprehensive and detailed but the results can be summarized as follows:

   - None of the 100 simulated attacks requiring "low effort" have succeeded in any of the two security configurations (i.e., the original scenario and the new IDS and security configuration).
   - The simulated attacks requiring "very high effort" produced the following results:
     - Original scenario: 5 of the 75 simulated attacks compromised the University of Conimbriga network and data infrastructure.
     - New IDS and security configuration: 3 of the 79 simulated attacks compromised the University of Conimbriga network and data infrastructure.

   Is it possible to state that the new IDS and security configuration has improved the security of University of Conimbriga network and data infrastructure with 95% of confidence? Explain your answer.

8) In an experiment you suspected that two variables are correlated. You applied a Pearson's test to samples of the two variables and the result for the Pearson's coefficient was r = 0.989. Based on this result you concluded that there is a strong positive correlation between the two variables. But you are interested in knowing if this correlation does in fact imply causation.

   a) Why is it important to know that correlation also implies causation? In another works, what can you do if you are sure that the high positive correlation also implies causation for the above-mentioned variables?

   b) Explain the type of analysis you would perform to confirm whether the correlation also implies causation or not.

9) The University of Coimbra (UC) has decided to reshape the interface of Inforestudante and Infordocente to improve the general quality user interaction and improve the effectiveness of the same tasks (especially in the Infordocente side) that take a long time to the user. After launching the versions of Inforestudante and Infordocente with the new interface the UC wants to assess whether the new interface is better than the previous one or not. Suppose you are in charge of performing such assessment; what type of experimental technique would you use: 1) Observation and data analysis, 2) Controlled experiment, or 3) Interviews and surveys? Explain your answer and if you consider that the situation would recommend more than one of these techniques explain your point of view.

10) In experiments with people you can obtain the subjective participant's opinion using interviews and surveys but you can also assess the subjective participant's opinion when you perform controlled experiments with people. Explain the differences between these two situations.

11) What is the problem of treating two dependent samples as two independent samples in a hypothesis testing? Explain the situation clearly.

12) You are trying to improve the average execution time of a long aggregation query in a decision support database (normally know as data warehouses). These types of queries typically scan big portions of the main table of the data warehouse (called the fact table) and filter the registers using several satellite dimension tables. As the execution time of such queries can take minutes or even hours, all the improvements are very welcome. After generating a new set of indexes and materialized views (the typical "tricks" used to speed up such queries) you tried several variants of the query and you observed some reduction in the average execution time. However, when comparing the samples (i.e., execution time measurements) obtained before and after the introduction of the new indexes and materialized views, you concluded that you could not reject the null hypotheses that the average execution time is the same in both scenarios with 90% of confidence. In any case, based on your calculations, you would be able to reject the null hypothesis with 75% of confidence. In these circumstances, do you think you should leave the data warehouse as it was or should you consider the new set of indexes and materialized views as the next production version of the data warehouse?