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

Training exercises 1

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) Explain the differences between laboratory experiments and pilot studies, providing simple and concise examples to clarify the key differences between them.

2) What is the main difference between measuring and benchmarking?

3) One of the first steps in the design of an experiment is to define the problem statement (or research question). Give a concrete example of problem statement (i.e., provide the actual sentence that express the problem statement) and briefly explain the experiment context related to each problem statement.

4) For the example of problem statement provided in your answer to the previous question, indicate the following:

a) Dependent variable(s)

b) Independent variables

c) Examples of possible levels for the independent variables in the experiments

d) Hypothesis that could be tested (indicate if the hypothesis is directional or non-directional).

5) An engineer conducts a hypothesis test and concludes that his hypothesis is correct. Explain why this conclusion is never an appropriate decision in hypothesis testing.

6) The definition of the levels taken by the independent variables is one of the most important decisions for the success of an experiment. Explain the key points/goals that should be taken into account in the choice of the levels for the independent variables.

7) Outliers obtained in the measurements should be reported but, in general, are removed from the analysis. Explain what should be taken into account in the decision of ignoring or not the outliers.

8) Explain what should be done to deal with the two types of measurement uncertainties: random and systematic uncertainties.

9) Suppose you are the data center administrator of a big organization and you are about to decide if your organization should accept the new cooling system that was recently installed in the data center. As the cooling system is very expensive, the contract defines precisely the conditions that should be met, concerning the temperature inside the server racks. Quoting the contract: "The cooling system assures that the temperature in the racks

is always in the range of 16.00 ± 0.80 Celsius degrees, with a confidence of 95%". In order to be sure that the cooling system is operating under the values defined in the contract, you decided to measure the temperature in the racks using a high precision electronic thermometer. To assure representative measurements, you took 100 measurements, including a variety of server loading scenarios and covering the 24 hours of the day. The results obtained show an average temperature of 16.245 Celsius with a standard deviation of 2.234. Do you think the conditions defined by the contract are met and your organization should accept the cooling system as is?

10) Consider the scenario described in the previous question but now suppose that you could only take 15 measurements. Explain what is different in this case. Consider both the arguments of the data center administrator (your perspective) and the vending representative of the cooling system.

11) Consider the following problem statement: the number of software bugs found in the tests of program units developed by programmers is dependent on the average number of sleeping hours of the programmers. Assume that you have the detailed specifications of a set of program units to be developed, and consider that the program units include units of high, medium and low complexity. Additionally, you have comprehensive unit test suits to test each program unit.

In these circumstances, describe how you would organize an experiment to answer the proposed problem statement. Your answer should be as complete as possible, focusing on the experiment design steps (obviously, it does not make sense to speculate about the experiment results and conclusions), and indicate the dependent and independent variables, the levels you would consider for the independent variables, the hypothesis under evaluation and the hypothesis testing technique you would use. Also describe, very briefly, the experimental setup and take into account in your answer to the whole question that the experiment deals with people (the programmers).

12) You want to specify the response time of a Web service you have developed for the Lisbon Stock Exchange, showing the average response time in milliseconds and the confidence interval. The Web service receives as input parameters the identifier of a company (VAT) and a date, and provides as output the total amount in euros for the buying and selling transactions of the papers of the company in the specified date. You have performed a set of experiments with a representative load (set of invocations of the service using companies from the PSI20 and a variety of dates) and measured the response time of the Web service. The Web service was used inside the same network where the servers of the Lisbon Stock Exchange are connected to, in order to avoid the unpredictable delays of remote networks. The results obtained are the following (in milliseconds):

- Average response time: 45.28
- Number of tests of the web service: 180
- Standard deviation: 8.23

a) Calculate the confidence interval with a confidence level of 95% and indicate the response time of the Web service in a technically correct and complete way.

b) In the experiments performed you noticed that in some executions the response time of the Web service was considerably longer than the average. For example, in the first time the web service is invoked with the VAT of a given company, the response time could be three or four times longer than the response time obtained in the subsequent invocations of the service for that company. Do you think these cases in which the response time was much longer should be considered as outliers and excluded from the calculation of the confidence interval or not? Justify your answer.

13) In hypothesis testing there are two types of errors: type II errors (false negatives) and type I errors (false positive). Explain the differences between these two errors and compare the consequences of each one.

14) Code inspections (a specific form of software inspections proposed by Michael Fagan more than four decades ago) is a technique to find bugs in the software under development in order to produce high quality software. Very briefly, in a code inspection a group of inspectors (experienced programmers) examine the code of a given code module (in general, short blocs of code with less than 100 lines of code) in order to find possible bugs. Obviously, the inspectors are not the authors of the code under inspection.

The goal is to perform an experiment to evaluate the effectiveness of code inspection in finding bugs. The idea is to use a set of code modules that have been very well tested and are considered totally correct (i.e., with no bugs) and insert a small number of bugs (3 or 4 for bugs per module). The inspectors will perform the code inspection on these modules and the goal is to check if they can detect the bugs that have been inserted in the modules.

In each individual code inspection, the inspector indicates the number of bugs found in the module. As the person that is managing the experiment knows exactly the bugs that have been inserted in each module, it is possible to confirm the quality of the code inspection done by each inspector. In practice, the result of each individual inspection is the number of bugs found and the number of mistakes done by the inspector. The mistakes are of two types: bugs that are wrongly identified (i.e., the code is correct but the inspector thinks there is a bug; this is a false positive) and real bugs that are not identified by the inspector (i.e., he looked at the code but did not find anything wrong; this is a false negative). These are the results that will be used to evaluate the effectiveness of the code inspections.

Assume that the modules (blocks of code) available for the experiment are written in C, C++ and Java, and that the authors of the modules have inserted the bugs in the code, to assure that the bugs are realistic (i.e., not obvious bugs, very easy to find). Additionally, consider that the group of inspectors available for the experiment fill a questioner with some demographic data (age, sex, etc.) and state the number of years of experience as programmer/code inspector.

In these circumstances, and considering the problem statement of knowing if code inspections are effective in finding bugs, indicate the following elements of the experiment:

a) Hypothesis

b) Variables (dependent and independent) and levels used for the independent variables, justifying your choices