Homework #3
Empirical Methods in Software Engineering and Computer Science

Delivered by: Ekrem Kocaguneli
Assigned: February 18, 2010
Due: February 25, 2010

Briefly discuss the following:

1. **What are the reasons for choosing 2^k factorial design?**

The characteristics of factorial design are:

- It uses resources efficiently. It enables researcher to receive maximum information with minimum effort.
- Unlike traditional one-factor-at-a-time (OFAT) approach, which suggests to change one factor and keep all the others constant, thereby ignoring the interaction between various factors, factorial design helps us to see the the interactions clearly.

The 2^k f factorial design considers *k* factors each having 2 levels (or values). Furthermore, in their study the researchers have 5 factors (independent variables) which correspond to a k value of 5 in a 2^k f factorial design and the level of their factors (the value that a factor could have) is 2. Therefore, due to characteristics of factorial design that enables researcher to observe interactions between factors in a very efficient manner as well as due to the fact that their factors conform well to 2^k factorial design, the researchers have chosen to use 2^k factorial design in their research.

2. List the independent and dependent (i.e., response) variables.

Independent variables (they are called in the paper as the factors) are the variables that affect the response (dependent) variables. In their research, the researchers have given the following identification of 5 independent and 2 dependent variables, which are listed as follows:

- **Independent variables:**
  1. Network density
  2. Node mobility
  3. Traffic load
  4. Network size
  5. Medium access control scheme

- **Dependent (response variables):**
  1. Packet delivery ratio
  2. End-to-end delay

2. Identify several threads to validity of this experiment. Elaborate on your answers.
The types of validity to this empirical study can be of 4 types: Construct validity, internal validity, conclusion validity and external validity. Their brief definitions and how they were addressed in the study are as follows:

- **Construct Validity:** Construct validity is basically asking the question whether we are testing what we intend to test. This is a very important validity, because the failure of construct validity would make all the study invalid. Fortunately, there is no problem with the construct validity of the paper. The authors follow the statistical design of experiment guidelines very closely and evaluate their results according to various performance criteria subject to statistical significance tests.
- **Internal Validity:** Internal validity questions whether the results elicited in a certain study are solely due to experiment manipulations. One way to ensure internal validity is to employ appropriate statistical tests. In the case of this paper, the authors have used analysis of variance (ANOVA) test to ensure that the findings are statistically significant and they have used the results of statistical significance tests in their model.
- **Conclusion Validity:** The conclusion validity asks whether the conclusions made in a particular study are justified. The authors have used both statistical tests and various performance criteria to justify their findings and they have also given the reasons why particular tests or performance criteria were chosen. Therefore, I think the conclusions have validity.
- **External Validity:** The aim of external validity is to figure out how and in what circumstances are the results applicable. I think this is the weakest part of the paper, in the sense that there is not a very clear indicator up to which point the results are scalable and since the data that was used in this study is a simulated dataset, it is hard to predict the actual performance of the model in a real life setting. Therefore, I would say the study lacks proper external validity.

2. What did you like most about this paper? What did you like least?
Liked the most:

- They have followed the steps of experimental design very closely. Furthermore, they have defined each step very clearly and have stated how their research conforms to the experimental design choice they choose. This greatly increases the readability of paper, because at each step you are aware of what is being done and why.
- Another thing that I found very nice about this paper is that rather than providing information regarding to a single community of mobile ad hoc networks, it provides guidelines to every researcher, who needs to follow a complicated 2^k factorial design in his\her research of interest.

Liked the least:

- They have used a simulation data and they have only justified the use of 5 factors to the fact that their study aims illustrating the effectiveness of statistical DOE strategy. Although their use of only 5 of many important factors and using a limited simulation dataset can be justifiable, I think it raises questions about the scalability of their findings in their study.

II. Assume this paper was submitted to a journal on Wireless and Mobile Systems and you are asked to review the paper. Fill in the review form given below. Note that the written comments (both to the editor and to the authors) are very important part of your review. As it is stated in the review form "Without clear, detailed support, it is hard for authors and editors to make good use of the review."

Section I. Overview

A. Reader Interest

How relevant is this manuscript to the readers of this periodical? Please explain under the Public Comments section below.

Very Relevant - **X**
Relevant
Interesting - but not very relevant
Irrelevant

B. Content

Is the manuscript technically sound? Please explain under the Public Comments section below.

Yes
Appears to be, but did not check completely **- X**
Partially
No

C. Presentation

1. Are the title, abstract, and keywords appropriate? Please explain under the Public Comments section below.

Yes **- X**
No

2. Does the manuscript contain sufficient and appropriate references? Please explain under the Public

Comments section below.

References are sufficient and appropriate **- X**
Important references are missing; more references are needed
Number of references is excessive

3. Please rate the organization and readability of this manuscript. Please explain under the Public Comments section below.

Easy to read **- X**
Readable - but requires some effort to understand
Difficult to read and understand
Unreadable

Section II. Summary and Recommendation

Evaluation

Please rate the manuscript. Explain your rating under the Public Comments section below.

Award Quality
Excellent **- X**
Good
Fair
Poor

Recommendation

Accept with no changes
Author should prepare a minor revision **- X**
Author should prepare a major revision for a second review
Revise and resubmit as "new"
Reject

Comments

Confidential Comments (authors will not see these comments)

Public Comments (these will be made available to the author)

Please give detailed justifications and explanations for your assessments, including positive and negative aspects of the manuscript. Without clear, detailed support, it is hard for authors and editors to make good use of the review.

1. **Reader Interest**
   - This paper has been submitted to MSWIM (Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems) and it follows a rigorous statistical experimental design while analyzing the interaction between dependent and independent variables on their simulated data of mobile ad hoc networks. Furthermore, they do not merely say that the observed results are true. Instead they use a statistical significance test (ANOVA) to see whether their observed effects of factors on the independent variables are also statistically significant and for each independent variable they clearly state which factors had a statistically significant effect on them. Therefore, I think both the topic and the adopted strategy in this study conforms to the requirements and interests of this conference.
2. **Content**
   - Manuscript seems technically sound to me, however I did not do all the calculations one by one. The technical details regarding the manuscript are as follows:
     - The writers define dependent and independent variables, then simulate a dataset to observe which independent variables have a statistical impact on the independent variables.
     - While doing that they use a 2^k factorial design and they elaborate on the definitions and application of this design technique. Therefore, I am assuming that their application of 2^k factorial design was correct.
     - Furthermore, to observe the statistical significance of the observed results out of 2^k factorial design, they utilize a statistical significance test: ANOVA. Again, I suffice to assume that ANOVA application was correct, because the results elicited from ANOVA that are given in Table 6 and also the predictive empirical model derived on the statistically significant independent features seems to have generated sensible results.

- For two independent variables that researchers have questioned in their study they provide R-Square, RMSE (root mean square error) and CV (coefficient of variation) to observe how well their model performs when compared to the actual data. Since the model was built on a solid statistical basis and since the statistics confirm that model performs well, I do not think that researchers had any technical problems with R-Square, RMSE or CV.

3. Presentation

i. **Are the title, abstract, and keywords appropriate?**

- Authors have described a 2^k factorial design for their research in which they have questioned the main and two way effect of 5 independent factors. They have then analyzed their findings statistically and built up two empirical linear first order regression models. The whole procedure and the reasons for such a story is succinctly summarized in the abstract and I believe title also describes the main aspects of this paper well.

ii. **Does the manuscript contain sufficient and appropriate references?**

- This manuscript is a conference paper and it aims to provide an answer to a research question rather than making a whole literature review. Therefore, I would say 18 references are fair enough. However, since I do not have an in depth knowledge about the field, I cannot say whether all the references are relevant or appropriate or whether some important literature work is missed.

iii. **Please rate the organization and readability of this manuscript.**

- The paper was very easy to read from the abstract on I think. Because in abstract as well as in introduction, they lay out the grounds of their research and you know that you are expecting to see a study that followed the statistical design of experiment rules very closely. After introduction, they continue with background of the related work, which I think makes sense in their situation, because especially for readers, who are not familiar to their field, having an early background section helps to understand the problem better.
- After the background, they continue with the description and terminology of statistical design of experiment approach and while doing that they also tell where their parameters fit in this model. I think that makes statistical DOE both more concrete and helps the reader to understand the problem better.
- Following the statistical DOE, they start the thorough statistical analysis of their results and apply statistical tests. Assuming that their calculations are correct, such a thorough analysis increases the credibility of their results.
- Lastly they use the results to build up two linear regression models and evaluate the performance of their models according to various criteria such as R-square, RMSE and CV.
- Therefore, I would say that the authors have made a nice organization for such a paper which -as they say in their conclusion paragraph- aims to have 3 different contributions.

4. Evaluation

- I would say that the manuscript is of very high quality. It tries to address three different goals in the same paper:
  - Systematic statistical design of experiment and its use in analyzing mobile networks

- How to use a factorial design to efficiently analyze and quantify main and two way interaction of factors in their domain
- How to use the main and two way interaction factors to build up an empirical first order regression model

While trying to give three different messages in the same paper it is really difficult to organize all the ideas and paragraphs so as to keep everything coherent, but the authors have done a good job on this.

My only comment for the improvement of the paper would be regarding the external validity. I believe there is not enough clarification regarding their choice of independent variables as well as how their simulated data could scale to real life situations. The chosen independent variables could be one of the most important ones according to some certain criteria and their data could be much more complex than a real life dataset, but the reader may not know these information and brief explanations on these aspects would make the work in the manuscript clearer.