PA2554: RESEARCH METHODOLOGY IN SOFTWARE

ENGINEERING AND COMPUTER SCIENCE

Student Name	P. Number	Contribution to the assignment (25% for equal contribution)
Vadrevu Krithi Sameera	961031-6086	25%
Kanneganti Joshika	970429-1781	25%
Mogili Anusha	970626-8167	25%
Bejawada Sravani	970525-9563	25%

SCENARIO – 1

Answer the following questions related to the design of this experiment:

1. What are the objects, subjects, treatment and factors used in the four designs listed above for this experiment?
Answer:
The objects, subjects, treatment and factors used in the four designs are:

Objects: Objects are the things that are used to perform a specific operation in a study.

Subjects: The participants who are involved in the experiment are known as subjects.

Treatment: In an experiment the different objects and procedures, which are to be compared are known as treatment.

Factors: A variable which is used to define a category, is known as a factor. The levels of the variable are set by an experimenter.

Design A:

Objects: prog 1

Subjects:12 students

Treatment: IDE-A and IDE-B

Factors: It comes under fixed factor and the level of the variable is 2.

-In Design A prog 1 is implemented by 12 students randomly in IDE-A and IDE-B to analyse.

Design B:

Objects: prog 1

Subjects:12 students

Treatment: IDE-A and IDE-B

Factors: It comes under fixed factor and the level of the variable is 2.

-In Design B prog 1 is implemented by 12 students orderly in both IDE's (A&B)

Design C:

Objects: prog 1 and prog 2

Subjects:12 students

Treatment: IDE-A and IDE-B

Factors: It comes under fixed factor and the level of the variable is 2.

-In Design C the prog 1 and prog 2 are implemented by 12 students evenly using IDE-A and IDE-B

Design D:

Objects: prog 1 and prog 2

Subjects:12 students

Treatment: IDE-A and IDE-B

Factors: It comes under fixed factor and the level of the variable is 2.

-In Design D the prog 1 and prog 2 are implemented by 12 students. Half of the students in IDE-A and remaining students in IDE-B.

2. How would you describe Design-A and Design-D in terms of a standard design type e.g. one factor, two treatments?

Answer:

-In Design-A the subject consisting of 12 students are used to implement the object of the study that is prog 1 by using two different treatments, which consists IDE-A and IDE-B with one factor.

-In design-D the subjects are classified into four possible combinations to perform the implementation of prog 1 and prog 2 using two treatments IDE-A and IDE-B with two factors.

3. What are the benefits and limitations of using Design-B instead of Design-A?

BENEFITS: Every student implements prog 1 in IDE-A and IDE-B randomly. It gives the exact result to analyze whereas Design -A randomly half of the students perform prog 1 in IDE-A or IDE-B and vice versa this type of experiment in not valid for analyzing.

LIMITATIONS: It consumes lot of time and memory for processing, but Design -A saves time and space because, half of the students perform using IDE-A and remaining students by using IDE-B on implementing prog 1.

4. What problems/mistakes can you identify in Design-C?

PROBLEMS:

-In Design C the students are performing prog 1 and prog 2 in IDE-A and IDE-B. There is no proper sequence for implementing prog, which results in confusion while analyzing the obtained data.

-There is no schedule for processing IDE's, which results conflicts in analyzing the study.

5. Does Design-D solve the problems you have identified in Design-B and Design-C?

Answer:
-Design-B and Design C consists of same problem. They both deals with improper sequence which leads to confusing state while evaluating results.

-Whereas in design B, the prog 1 is only implemented using IDE's. In Design C, both the programs 1 and 2 are implemented to analyze.

-In Design D the 12 students are used to implement prog 1 and prog 2 in both the IDE-A and IDE-B.

- The main advantage in Design-D is the students are divided into 4 possible combinations and they are used to perform the operations, which helps in saving time and space by avoiding repetitions that is the same work done twice can be avoided as the students performing in IDE-A in prog 1 are not performing in any other operation.

-It reduces the main conflict in design B and design C that is improper sequence as there is no sequence needed in design D.

-In Design D the schedule is perfectly managed as when one team in performing in IDE-A and prog 1 other can perform prog 2 in ide B and vice versa which reduce the time conflicts as in design B and design C.

6. What are the benefits and limitations of the designs Design-A and Design-D? e.g. one is easier to analyze.

Answer:
Design A:

BENEFITS: The students are equally divided to perform prog 1 in IDE-A and IDE-B. without repetitions.

LIMITATIONS: The students working for IDE-A are not working on IDE-B and vice versa. This leads to inaccurate results. The main study in this scenario depends on both prog 1 and prog 2. whereas in design A prog 2 is not implemented.

-The Design A is easy to understand and hard to analyze the result obtained after implementation, but it is incomplete process.

Design D:

BENEFITS: In design D the students are divided in four combinations randomly and are used to implement prog 1 and prog 2 in IDE-A and IDE-B. and which provides accurate results for analyzing.

LIMITATIONS: Here every student doesn't get a chance to implement both the programs in both the IDE'S.

-The Design D is easy to analyze and to understand and helps in avoiding repetitions.

7. What variables must be controlled in Design-A to increase the validity of the experiment? e.g. previous experience/familiarity of subjects to IDE's?

Answer:

-To increase the validity of the experiment in design A the IDE'S that are used to implement prog 1 needs to be controlled. The prog1 should be implemented in both IDE-A and IDE-B by all the students orderly. Which helps in getting exact results in the experiment.

Answer the following questions related to analysis of an experiment with Design-A (as shown in Table 3) and the results in Table 7:

1. State the null and alternative hypothesis for this investigation.
Answer:
Null hypothesis(H0): IDE-A and IDE-B are equally efficient and give accurate results.

H0: μIDE-A = μIDE-B

Alternative hypothesis(H1): IDE-B gives efficient and accurate results than IDE-A *vice versa.*

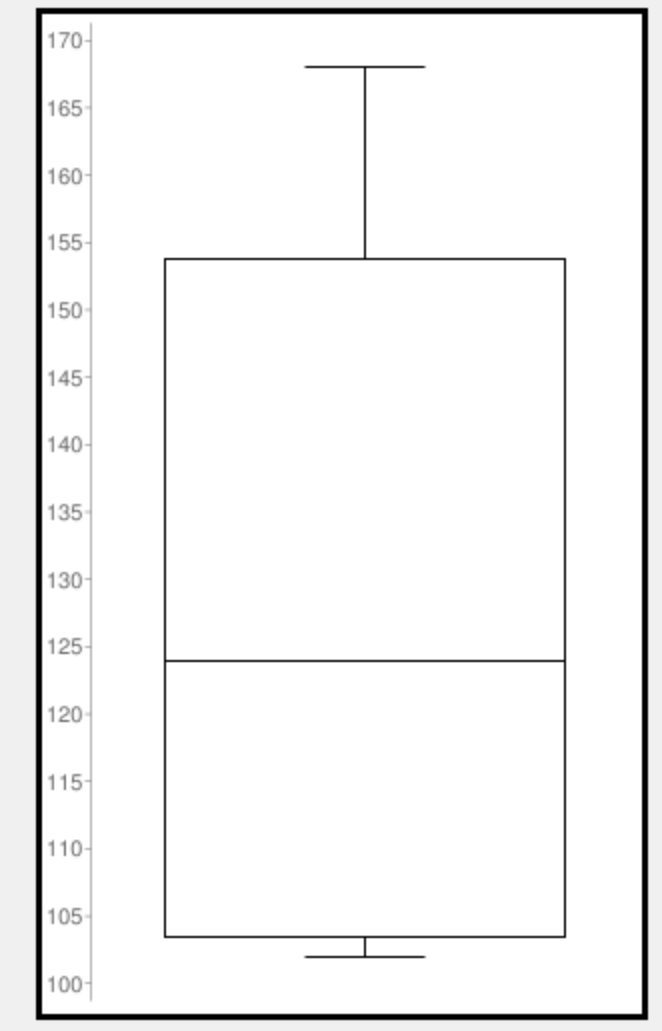
H1: μIDE-A ≠μIDE-B, μIDE-A <μIDE-B or μIDE-A > μIDE-B

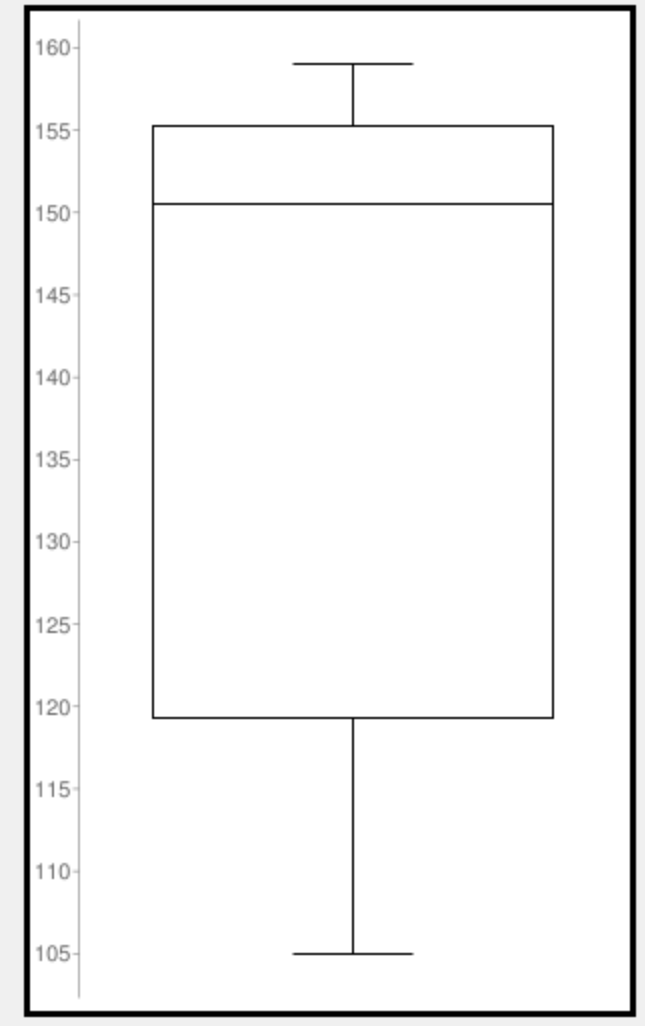
2. Use descriptive statistics and visualize the data in Table 7 use e.g. box plot, histograms and scatter plot. Which visualization tool helped you develop some insights into the data? What were the insights e.g. any interesting patterns or trends in the data, a clear difference in efficiency between two IDEs, outliers.

Answer:

In table 7, every student has different time of execution. Using box plot data can be visualized clearly and accurate time of execution can be marked. Box plots help in concluding that IDE-B is preferable than IDE-A. Here we arrange the given data in ascending order and make a note of range values and interquartile range and median values to plot the graph.

Median	IDE-A= 124	IDE-B= 150.5
Mean	IDE-A= 128.5	IDE-B= 140.5
– Range	IDE-A= (102,168)	IDE-B= (105,159)
– Interquartile range(IQR)	IDE-A= 45	IDE-B= 30
– Standard deviation	IDE-A= 27.06	IDE-B= 21.24


170
165
160
155
150
145
140
135
130
125
120
115
110
105
100



3. Choose and justify your choice of a parametric/non-parametric test for analyzing the given data (document the steps you undertook and the results).
Answer:
We have chosen the parametric test for a statistical data, gives significant results.
By using t-test we have analyzed the data

t =0.8545, p-value = 0.4128
levelofsignificance-0.05
standard error of difference = 14.044

4. Run the statistical method and report if you can reject the null hypothesis? Please interpret your results, what does this imply for the objective of the study?
Answer:

P value = 0.4
The p value is less than 0.05(level of significance) so we reject the null hypothesis.
So, Null Hypothesis (H0) is rejected.

5. Based on the results would you be confident to recommend an IDE either IDE-A or IDE-B for use in your company. Why or why not?
Answer: Based on the results, we would be confident to recommend IDE-A.
The results prove that IDE-A is more efficient compared to IDE-B. Using efficient IDE benefits the programmers.

SCENARIO - 2

A. Describe the approach that you will follow to analyze the given data. Also, describe your mechanism for coding the data. Also, explain why you chose the approach over other alternatives.

Answer: Data can be analyzed constructively using Thematic Coding Analysis approach.

The steps for coding the data

1. Acquiring knowledge - reading and transcribing data
2. Generating initial codes - codes are drawn based on interviews and observations
3. Themes formulation - Searching, reviewing, defining and naming the themes
4. Documentation – writing a final report

The main objective of selecting Thematic coding approach is, it provides flexibility to researchers by allowing them to add more than one theory, unlike Quasi-static and Grounded theory approach. This approach can be used by researchers who have little knowledge of qualitative research since it is an easy and quick method to learn [3].

B. Please describe the coding procedure that you followed. For each step, please provide an example of how you coded the information in the papers.

Answer:

Steps	**Procedure**	**Description & Example**
1	Resource Accumulating	Gather information sources, build a clear understanding and making short notes. Example: we highlighted important points and created a draft document.
2	Developing codes	Analyzing related parts of the data into data sets and generating potential codes and sub-codes. Example: “Challenges” in the papers given is a code and industry, research, general and academia challenges as sub-codes.
3	Defining themes	Sorting different codes into subsets to form themes.

		For example, taking challenges from industry and research we form a theme as practitioner challenges
4	Refining themes	1) Reworking of the existing themes by recombining or condense to smaller units. 2) After finalizing theme, we name them. Example: Naming practitioners challenges as 'practitioners attitude'
5	Final report	Documenting final analysis of the report.

C. Answer the following questions by citing examples from your analysis of the three studies

1. Which challenges or impediments for industry-academia collaborations have been raised by the papers?

Answer:

(i) Fluent and direct communication is needed among the partners [2][3].
(ii) Identification of goal [3].
(iii) Identify realistic sources of data [3].
(iv) Understand the real context in the industry [3].
(v) Changes of people in the collaboration team [2][3].
(vi) Need for the practitioner to balance to dedicate time [2][3].
(vii) Local opinion-based knowledge is trusted more than empirically-based knowledge in industry [2].
(viii) Requirements obtained in large volumes from multiple sources should be handled [4].

2. What patterns have been proposed for industry-academia collaborations?

Answer:

1) The approach used for collection of data is focus group, as it is considered a proven and tested technique. It acquires perceptions from selected people of the group by enabling free flowing discussions and providing an opportunity for everyone to participate [3].
2) Long-term collaborations establish a balance between a company 'pull' and academia 'push', creating more mutual benefits [2].

3. What should be avoided during industry-academia collaborations?

Answer: Practices that should be avoided are

1) Low levels of maturity of research [3].

2) Limited presence of the researcher in the company [2][4].
3) Research projects should not be unfunded [2].

REFERENCES

[1] C. Robson and K. McCartan, *Real World Research*. John Wiley & Sons, 2016.

[2] P. Runeson, "It Takes Two to Tango – An Experience Report on Industry-Academia Collaboration," in *Verification and Validation 2012 IEEE Fifth International Conference on Software Testing*, 2012, pp. 872–877.

[3] S. Martínez-Fernández and H. M. Marques, "Practical Experiences in Designing and Conducting Empirical Studies in Industry-academia Collaboration," in *Proceedings of the 2Nd International Workshop on Conducting Empirical Studies in Industry*, New York, NY, USA, 2014, pp. 15–20.

[4] T. Gorschek, P. Garre, S. Larsson, and C. Wohlin, "A Model for Technology Transfer in Practice," *IEEE Software.*, vol. 23, no. 6, pp. 88–95, Nov. 2006.

[5] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering*. Berlin Heidelberg: Springer-Verlag, 2012.