

April 2022

2IMP40 2IMP40 Applications of Data Science to Software Engineering 21/22 · 3 exercises · 30.0 points

1 Part 1

20.0 points · 20 questions

The first part of the exam consists of **twenty** multiple choice questions. Each question has four answer options and exactly one is correct.

Text

a Ewusi-Mensah claims that “regardless of the particular process model ...every software project will feature: (1) the requirements-definition and functional-specification phase; (2) the design phase; ...(3) the implementation; ...and (4) the installation, operation, and maintenance phase.” This claim can be seen as reflecting

1.0 point · Multiple choice · 4 alternatives

- | | |
|---|-----|
| <input type="radio"/> empiricism | 0.0 |
| <input checked="" type="radio"/> rationalism | 1.0 |
| <input type="radio"/> neither empiricism nor rationalism | 0.0 |
| <input type="radio"/> there is not enough information to answer this question | 0.0 |

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

b Recall from the lecture that Jane is a PhD student working on fisheye visualisations. Her theory is that fisheye views correspond well to the way that people see and navigate in the world, by offering more detail of a specific area of focus, together with a less detailed overview of the peripheral regions, and a smooth way of moving the focus of attention. The theory suggests that less time spent scrolling and fewer clicks should reduce navigation time. Based on this theory Jane hypothesized that fisheye views cause more efficient file navigation than traditional file tree explorer views. From the below list of approaches please select the best approach that Jane can use to test her hypothesis.

1.0 point · Multiple choice · 4 alternatives

- conduct a survey of developers using fisheye views and using traditional file tree explorer views. 0.0
- conduct a controlled experiment with some subjects using fisheye views and others using traditional file tree explorer views. 1.0
- mine Git repositories of developers using fisheye views and those using traditional file tree explorer views. 0.0
- apply regression discontinuity design (RDD). 0.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

c Wu et al. present a case study of an industrial real-time avionics operating system. The analysis focuses on the variation in quality due to testing and associated characteristics in a concrete real-time operating system (RTOS). What research strategy did Wu et al. use?

1.0 point · Multiple choice · 4 alternatives

- a laboratory experiment 0.0
- a sample study 0.0
- an experimental simulation 0.0
- a field study 1.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

d Thongtanunam et al. study factors that influence review participation, using a dataset of over 196,000 reviews from 3 OSS projects. The study develops a number of models with generalizable statements, and does not focus specifically on the contexts of the 3 projects. What research strategy did Thongtanunam et al. use?

1.0 point · Multiple choice · 4 alternatives

- a laboratory experiment 0.0
- a sample study 1.0
- an experimental simulation 0.0
- a field study 0.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

e Martinez et al. investigate the effectiveness of automatic test-suite-based repair on Defects4J, which is a large dataset of real-world Java bugs. Using state-of-the-art repair methods, it was demonstrated that patches could be generated for 47 out of 224 bugs. What research strategy did Martinez et al. use?

1.0 point · Multiple choice · 4 alternatives

- | | |
|--|-----|
| <input checked="" type="radio"/> a laboratory experiment | 1.0 |
| <input type="radio"/> a sample study | 0.0 |
| <input type="radio"/> an experimental simulation | 0.0 |
| <input type="radio"/> a field study | 0.0 |

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

f Researchers conduct a literature study of what kind of statistical techniques are used in empirical software engineering research. To this end they they focused on four leading software engineering venues, and used a true random number generator to randomly select five papers published at each venue in 2021. What kind of sampling strategy did they use?

1.0 point · Multiple choice · 4 alternatives

- | | |
|---|-----|
| <input type="radio"/> simple random sampling | 0.0 |
| <input type="radio"/> snowball sampling | 0.0 |
| <input checked="" type="radio"/> stratified random sampling | 1.0 |
| <input type="radio"/> cluster sampling | 0.0 |

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

g To conduct a series of interviews with visually-impaired software developers researchers have decided to perform a respondent-driven sampling. As an expert on empirical methods you would recommend them

1.0 point · Multiple choice · 4 alternatives

- to start with well-connected (in terms professional contacts) developers from different continents that are to ensure that the initial set of participants is diverse enough 1.0
- to ask the interviewees to identify prospective participants rather than recruit them: such that the researchers can select the most promising participants for the next interview round 0.0
- Feedback
no, this would introduce selection bias
- to encourage the interviewees to provide as many names as possible such that the researchers can reach many people in short amount of time 0.0
- Feedback
no, highly-connected participants will bias the sample
- use traditional statistical methods to analyse the data obtained after conducting interviews.0.0
- Feedback
no, traditional methods assume that the sample is random and here the sample is clearly not random.

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

h Which one of the following sampling strategies guarantees that the sample is representative?

1.0 point · Multiple choice · 4 alternatives

random selection (each item in sampling frame has an equal probability of inclusion) 0.0

Feedback

Assumes large sample size, good sampling frame and no coincidences.

size (sample is so large that missing a subpopulation is unlikely) 0.0

Feedback

Bias toward one or more subpopulations or on important parameters.

breadth (sample captures a large variance on important parameters) 0.0

Feedback

Point estimates unreliable; groups may not be proportional.

none of the above 1.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

i While working at university X Professor Smith has conducted a study of agile practices in small companies. After completing the interviews, Professor Smith got a new position at a different university and left their previous position at university X. Their new research will focus on emotions expressed by software developers. In order to kick-start the activities of their new research group they decided to reuse the interview transcripts of the agile practices study since some of the interviewees got quite emotional while discussing their experiences. Professor Smith intends to combine the interview data with findings from a literature study to obtain new insights.

Would you consider behaviour of Professor Smith to be ethical?

1.0 point · Multiple choice · 4 alternatives

- Yes, Professor Smith collected this data themselves and are entitled to use the data for their research. 0.0
- No, unless the study participants have explicitly provided their consent for their interview data to be used for purposes other than the study of agile practices. 1.0
- Yes, if the ethical review board of the new university approves the reuse of the interview data. 0.0
- No, unless the data is stored on Professor Smith's private computer as opposed to official data storage provided by university X. 0.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

j A researcher conducts an exploratory study of log comparison practices in a large company counting 32 000 employees. The researcher would like to understand how developers compare logs of two subsequent versions of a software system to identify bugs. The researcher should

1.0 point · Multiple choice · 4 alternatives

- Send a survey to 380 randomly chosen employees corresponding to 95% confidence level and 5% confidence interval (simple random sampling). 0.0
- Identify organisational units employing developers that analyse logs and interview several experienced developers from each unit (multi-stage sampling) 1.0
- Collect data from version control repositories owned by the company (repository mining) on how developers introduce and remove log statements. 0.0
- Conduct a controlled experiment with developers employed by the company to learn how developers use log statements. 0.0

Feedback

An experiment is not a good approach to conduct exploratory research.

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

k Bird et al. have used IP addresses in the mail headers to identify geographic location of the mailing list participants. Which other element of the mail header can be used for the same goal?

1.0 point · Multiple choice · 4 alternatives

- Date 1.0
- Delivered-To 0.0
- Message-ID 0.0
- MIME-version 0.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

l Which of the following is not a bug (defect) tracking system?

1.0 point · Multiple choice · 4 alternatives

- | | |
|---|-----|
| <input type="radio"/> Bugzilla | 0.0 |
| <input type="radio"/> JIRA | 0.0 |
| <input checked="" type="radio"/> Gerrit | 1.0 |
| <input type="radio"/> Launchpad | 0.0 |

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

m A researcher wants to study whether the number of lines of code of C functions is smaller than the number of lines of code of Java methods. They have collected a large corpus of C functions and Java methods and recorded the number of lines of code for each one of them. Based on the previous literature the researcher knows that the distributions cannot be assumed to be normal, and this is why they decide to use

1.0 point · Multiple choice · 4 alternatives

- | | |
|---|-----|
| <input type="radio"/> Student's t test | 0.0 |
| <input type="radio"/> Kolmogorov-Smirnov's test | 0.0 |
| <input type="radio"/> Fisher's exact test | 0.0 |
| <input checked="" type="radio"/> Mann-Whitney-Wilcoxon test | 1.0 |

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

n Which one of the following statements related to p-values is **false**?

1.0 point · Multiple choice · 4 alternatives

- In null-hypothesis significance testing, the p-value is the probability of obtaining test results at least as extreme as the results actually observed, under the assumption that the null hypothesis is correct. 0.0
- techniques such as Benjamini-Hochberg designed to adjust the p-values due to multiple comparisons decrease the p-values 1.0
- applying the chi-square test and Fisher's exact test to the same data is likely to result in different p-values 0.0
- when the tool reports " $p < 2.2 \times 10^{-16}$ " this means that the p-value is too small to be computed exactly. 0.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

o Not every statistically significant difference is meaningful in practice: to capture this intuition several measures have been introduced to describe how "big" the differences are. Which one of the following measures would be appropriate to this end?

1.0 point · Multiple choice · 4 alternatives

- Cliff's delta 1.0
- Pearson's r 0.0
- Nakagawa's R2 0.0
- chi-square 0.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

p Which one of the following variables is commonly modeled as a random effect in mixed-effects modeling?

1.0 point · Multiple choice · 4 alternatives

- project ID 1.0
- sentiment of a code review comment 0.0
- number of lines of code 0.0
- number of developers 0.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

q Which one of the following statements about Grounded Theory is **true**?

1.0 point · Multiple choice · 4 alternatives

- Application of Grounded Theory requires statistical methods such as ANOVA/Kruskal-Wallis test. 0.0
- Grounded Theory assumes that the labels are known in advance. 0.0
- Grounded Theory requires constant comparison of the emergent categories. 1.0
- To ensure intercoder reliability, application of Grounded Theory requires as many coders as possible. 0.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

r Choose one option to complete the following sentence: Machine learning approaches to code smell prediction ...

1.0 point · Multiple choice · 4 alternatives

- consistently outperform heuristic approaches 0.0
- suffer from the threshold problem 0.0
- should only be used in combination with data balancing 0.0
- require manually validated data 1.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

s

When people take part in an experiment they might try to figure out what the purpose and intended result of the experiment is. Then they are likely to base their behavior on their guesses about the hypotheses, either positively or negatively, depending on their attitude to the anticipated hypothesis. Which type of validity would this threaten?

1.0 point · Multiple choice · 4 alternatives

- construct validity 1.0
- conclusion validity 0.0
- internal validity 0.0
- external validity 0.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

t When conducting an experimental study it is possible that the subjects react differently as time passes. Examples are when the subjects are affected negatively (tired or bored) during the experiment, or positively (learning) during the course of the experiment. Which type of validity would this threaten?

1.0 point · Multiple choice · 4 alternatives

- | | | |
|----------------------------------|---------------------|-----|
| <input type="radio"/> | construct validity | 0.0 |
| <input type="radio"/> | conclusion validity | 0.0 |
| <input checked="" type="radio"/> | internal validity | 1.0 |
| <input type="radio"/> | external validity | 0.0 |

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

2 Part 2

10.0 points · 4 questions

To answer this question you need to read a scientific paper and identify issues that might have threatened validity of the conclusions of this study.

As appendix to this exam you should have received a copy of the paper by Shimida et al. in which the threats to validity section has been redacted. Please read the paper and come-up with two viable and distinct threats to validity. For each threat to validity please use the answer fields below to describe and classify it. The threats you describe should describe **two different types of threats to validity** (So if the first threat to validity you describe belongs to the category external, the second threat you describe cannot belong to the category external).

You should describe your first threat to validity in 2.a, and classify it in 2.b, and you should describe your second threat to validity in 2.c and classify it in 2.d.

Text

a Describe your first threat to validity

3.0 points · Open · 1/2 Page

+3 points

The issue described can indeed threaten validity of the conclusion.

0 points

The issue described cannot threaten validity of the conclusion, OR the issue is not clearly described.

b Classify your first described threat as either Internal, External, Construct, or Conclusion.

2.0 points · Open · 1/20 Page

+2 points

The issue described in 2a has been correctly classified.

0 points

The issue described in 2a has not been correctly classified.

c Describe your second threat to validity.

3.0 points · Open · 1/2 Page

+3 points

The issue described can indeed threaten validity of the conclusion.

0 points

The issue described cannot threaten validity of the conclusion, OR the issue is not clearly described.

d Classify your second described threat to validity as either Internal, External, Construct or Conclusion.

2.0 points · Open · 1/100 Page

+2 points

The issue described in 2c has been correctly classified.

0 points

The issue described in 2c has not been correctly classified.