ECE750: Usable Security and Privacy
Assignment 2

Kami Vaniea

July 7, 2025

Due: July 29th

Submission: Crowdmark for Assignment 2. Upload a PDF or Word file with the answers as well as the `tahaei-data.csv` file with filled-in values.

1 Introduction

In this assignment you will be analyzing the dataset from Tahaei et al. [4] who asked people who claimed to have prior programming experience to fill out four survey scales: one on programming skill and three on security/privacy.

In this assignment you are going to be computing the scales from the dataset and then using them to answer research questions.

In this assignment you will be doing the following:

1. Compute two scales

2. Compute some simple descriptive statistics

3. Answer statistical questions

2 Context

This section contains information about the study that produced the dataset you are working with. All the information you need about the study is provided here, there is no need to read Tahaei *et al.* to answer any of the questions.

2.1 Dataset

You can find the dataset files on the course website under Assignment 2. The data can be found in `tahaei-data.csv`. You are also provided with descriptions of all the questions asked and associated column names: `tahaei-header-descriptions.csv`. Please use these files as they have been cleaned and have empty columns you will be asked to fill. The paper links to public data which has the same values, but doesn't have the same column names or the blank columns.

The original dataset was collected using the following approach:

- Recruitment: Advertisements to take a pre-screening survey were sent out to users of Appen, Clickworker, and Mechanical Turk (mturk). Participants who responded to the ad were asked several questions, including "Do you have computer programming skill?" If they said yes, they were added to the *recruitment pool*.

- Prolific requires all their workers to fill out a survey in advance. People posting tasks can limit task visibility based on the survey questions. All Prolific users who answered 'yes' to Prolific's question about if they knew how to write programs were considered part of the *recruitment pool*.

- Advertisements for the full survey were posted on the 4 crowd worker platforms visible only to workers in the *recruitment pool*. A separate advertisement was emailed to active students in a Computer Science program at an University in the UK.

- After participants took the survey, the researchers removed all participants from the resulting survey dataset who had incomplete answers or who answered any of the attention check questions incorrectly.

2.2 Scales

The research used 4 survey scales which are listed below.

- REALCODE [1] - The scale measures if someone has basic level development skills. It is based on research into what survey questions best predict actual development skill without requiring a survey respondent to read long code blocks.

- IUIPC [3] - Internet Users' Information Privacy Concerns

- SeBIS [2] - Security Behavior Intentions Scale

- SECDEV [5] - Scale for Secure Software Development Self-Efficacy (SSD-SES)[1]. The dataset contains the scale, but we will not be using the SECDEV scale in this assignment.

3 Assignment questions

3.1 Compute descriptive statistics

Descriptive statistics provide a rough summary of a dataset. They are normally computed using common mathematical functions like average, min, max, median and standard divination. Most research papers that conduct research on a dataset will provide descriptive statistics about it in the methodology or results sections. For this assignment, compute the descriptive statistics stated below:

1. How many participants were in the study assuming each participant only took the survey once? Provide a number. The question is only asking about the provided dataset, you do not need to lookup the original number of participants before exclusions happened. (2)

2. What percentage of the participants self-identified as: male, female, and other genders? (2)

3. How many participants were successfully recruited from each channel? (2)

   The researchers advertised across several different services (channels) asking people to come take the survey.

4. How many years of any programming experience did participants have on average? And how many years of software development experience did they have? Briefly write why you think these numbers are different. (3)

5. What percentage of the participants know how to write programs? (4)

   To answer this question you need to use the REALCODE scale. The column `really.correct` provides a sum between 0 and 5 of how many REALCODE questions the participant answered correctly. Participants who answer all 5 questions correctly are considered able to write programs according to the scale.

6. Using descriptive statistics answer the main question of the research paper: *Which platforms are best to recruit from to get the highest percentage of participants who can actually write programs?* Use descriptive statistics to justify your answer. (10)

7. REALCODE is made up of 5 questions. Read the questions. In your experience, would students who struggle with programming likely struggle with these four questions? Explain your answer. (4)

---

[1]We call the SSD-SES scale SECDEV in this assignment because that is how it is referred to in the dataset.

3.2   Run a simple t-test

Start by trying to answer a relatively simple research question: *Do participants from MTurk and from Prolific have similar or different numbers of years of software development experience?* Most of the participants are from MTurk and Prolific, so we will look at these two first.

More formally a t-test helps us determine if the means of two groups are statistically distinct. In other words, if they are likely drawn from the same distribution and observed differences are just random chance, or if the differences are likely due to the distributions being truly different. To perform such a test we need to phrase our research question in terms of a hypothesis that can be proven or disproved:

H1: The number of years of experience MTurk and Prolific participants have in software development is different.

8. Test if H1 holds using a t-test. You may use your choice of programming language to run the test. You may even use Excel which calls it a Student t-test. The following are details you will need to run the test: (10)

- The $\alpha = .05$
- Use a two-tailed or two-sided test (this is because H1 does not care if prolific > mturk or prolific < mturk, only that they are different).
- Two sample t-test, not paired

After running the test, answer the following questions:

(a) What p-value did you get, state the number.

(b) Mean years of Prolific participants' software experience.

(c) Mean years of MTurk participants' software experience.

(d) What does the p-value mean? Are the means for Prolific and MTurk statistically different?

9. Mechanical Turk (MTurk) and Prolific are general crowdworker sites often used by researchers who want to run surveys. MTurk is based out of the USA and Prolific is based out of Europe. In your opinion, is the result you got in question 8 expected? Do they make sense? Explain your reasoning. (4)

3.3   Compute IUIPC, and SeBIS scales

The `tahaei-data.csv` file has empty columns at the end with the following headers:

- iuipc.control
- iuipc.awareness
- iuipc.collection
- sebis.device
- sebis.password
- sebis.awareness
- sebis.update

Compute these values and submit the `tahaei-data.csv` file with the appropriate columns filled in. I will not be grading these files directly, but I may use them to assign partial grades if the answers in the PDF are wrong.

To compute the blank columns you will need to lookup the scales described in Section 2.2. The IUIPC scale we did in class is computed by converting the Likert scale words into a numeric scale from 1 (Strongly disagree) to 7 (Strongly agree) and then adding the questions associated with each subscale. IUIPC has three subscales: control, awareness, and collection.

Some scales have *inverted* questions where the Likert value needs to be reversed before being added. Imagine the two statements: "Security is important" and "Security is a complete waste of time". I would expect someone who marked "strongly agree" to the first question would mark "strongly disagree" to the second question. These questions are inverted, so adding them does not work, first we have to reverse one of them. So the first question would be computed with 1 (Strongly disagree) to 7 (Strongly agree) and the second question would be computed 7 (Strongly disagree) to 1 (Strongly agree). To compute the scales you will need to pull open the research papers and find which, if any, questions are inverted.

10. Create three box-and-whisker plots of the IUIPC subscales against the channel the participants were recruited from. There are three IUIPC subscales, so you should either have three plots, or if you know how, you can plot all three at once. (5)

    Figure 1 below shows an example of a box-and-whisker plot for the participants ages against the channel they were recruited from. Provide a similar diagram, but with the IUIPC subscale instead of the age.
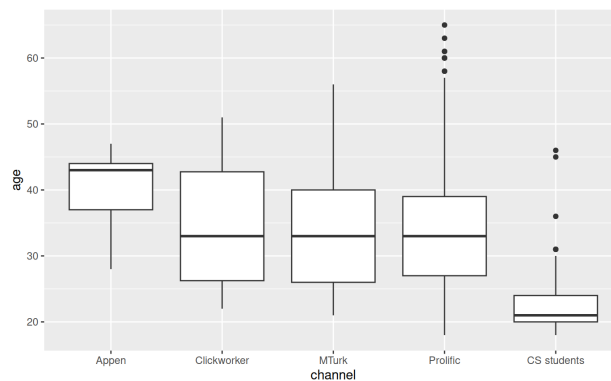

Figure 1: Example Box and Whisker plot of the ages of participants by the channel they were recruited from.

11. Create four box-and-whisker plots of the SeBIS subscales against the channel the participants were recruited from. There are four SeBIS subscales, so you should either have four plots, or if you know how, you can plot all four at once. (6)

3.4 Do North Americans have different levels of privacy concern than Europeans?

Its a common view in the privacy community that North Americans (predominantly the USA) are behind in terms of privacy. North American laws around privacy are less strict than European ones, and it is assumed that Europeans generally care more about privacy than North Americans. You will be using the provided data set to test if there is a difference in their privacy concerns.

H2: The privacy concern level of North Americans is different from Europeans as measured by the IUIPC scale.

12. Use the IUIPC scale to answer the research question. Start by summing the three subscales so: (10) `iuipc.total = iuipc.awareness + iuipc.collection + iuipc.control`. Then use a t-test the same way you did in Question 8 with the same provided setting values.

    After running the test, answer the following questions:

    (a) What p-value did you get, state the number.

    (b) Mean European IUIPC score.

    (c) Mean North American IUIPC score.

(d) What does that p-value mean? Are the means for people from North American and Europe statistically different or the same? If they are diffarnet, who has higher privacy concerns?

13. Explain the main finding from Question 12 in terms of the three IUIPC subscales which measure the components of privacy concern. Take a look at how European and North American participants answered the subscales. You may want to run more t-tests. (10)

3.5 Behavioral Intentions: Europeans vs North Americans

In the section above we explored the privacy concerns of the two geographical groups. But does concern transition into their behavior?

14. Similar to the calculation you did for IUIPC, sum the four SeBIS subscales together and compare North Americans and Europeans in terms of behavioral intention. (10)

After running the test, answer the following questions:

(a) What p-value did you get, state the number.

(b) Mean European SeBIS score.

(c) Mean North American SeBIS score.

(d) What does that p-value mean? Are the means for people from North American and Europe statistically different? If they are different, who behaves more securely? If they are the same, how should we interpret the number the scale produces?

15. Both IUIPC and SeBIS have a subscale called "awareness". Use a t-test to compare the answers to the two subscales for all participants (not just North Americans and Europeans). The scales have different max values, so you will need to normalize the values first. The two values are also both coming from the same person, so you will need to run a "paired t-test" instead of the "two sample" you were using earlier. (12)

After running the test, answer the following questions:

(a) What p-value did you get, state the number.

(b) Mean IUIPC awareness normalized score.

(c) Mean SeBIS awareness normalized score.

(d) What does the p-value mean? Are the means for the IUIPC and SeBIS awareness scales the same or different?

(e) In your opinion, what is the cause of the finding in (d) above? You may need to look at the exact wording of the questions used in the scales to answer the question.

3.6 Self-reflection

16. What did you find most surprising or unexpected when doing the assignment? (5)

This is a self-reflection question, it is graded lightly, but I expect to see an answer that shows you were engaged with the assignment.

4 Turn in:

Use Crowdmark to turn in the final PDF.

- PDF containing the answers to the questions above.

Please keep the data files and the code till grading is returned. If I see weird answers I may ask you for these files to try and understand what might happen.

- tahaei-data.csv file with all the column numbers turned in. Should be in csv format. This file will not be graded directly, but it will be used to give partial points if the provided answer is wrong.

- Code you used to compute the scales and the statistica tests. The code will not be graded directly, but it may be referenced to add context to your answers.

References

[1] Anastasia Danilova, Alena Naiakshina, Stefan Horstmann, and Matthew Smith. Do you really code? designing and evaluating screening questions for online surveys with programmers. In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, pages 537–548. IEEE, 2021.

[2] Serge Egelman and Eyal Peer. Scaling the security wall: Developing a security behavior intentions scale (sebis). In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, page 2873–2882, New York, NY, USA, 2015. Association for Computing Machinery.

[3] Naresh K Malhotra, Sung S Kim, and James Agarwal. Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model. *Information systems research*, 15(4):336–355, 2004.

[4] Mohammad Tahaei and Kami Vaniea. Recruiting participants with programming skills: A comparison of four crowdsourcing platforms and a cs student mailing list. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1–15. ACM, May 2022.

[5] Daniel Votipka, Desiree Abrokwa, and Michelle L. Mazurek. Building and validating a scale for secure software development self-efficacy. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–20, New York, NY, USA, 2020. Association for Computing Machinery.