ELEC-E7890 - USER RESEARCH
AALTO UNIVESITY, DEPT. COMMUNICATIONS AND ENGINEERING

Exercises, Week 1

Jussi P.P. Jokinen

September 12, 2019

1 IDENTIFY USER RESEARCH TOUCHPOINTS FOR A PRODUCT (8P)

The goal of this task is to start thinking in terms of user research. You are asked to pick a product of your choice (some ideas are provided below), and analyse it critically from the perspective of usability and user experience (so, please pick a product that has identifiable problems in this area). After identifying usability and user experience issues with the product, state the methods that could be used in collecting user data that would help designers to improve the product.

You should read one of the materials in the supplementary reading and refer to it in your response.

- Dorst, K., & Cross, N. (2001). Creativity in the design process: co-evolution of problem-solution. Design studies, 22(5), 425-437.
- Simon, H.A. (1996/1969): The Sciences of the Artificial. Chapter 1.
- Saariluoma, P., & Leikas, J. (2010). Life-based design - An approach to design for life. Global Journal of Management and Business Research, 10(5), 17-23.

Some ideas for the product (that has identifiable problems):

- a clearly confusing and badly designed website, e.g., `http://art.yale.edu/`
- a bad app you have been using, or that has received a lot of negative publicity
- perhaps as part of your studies, you have had to use a bad interface?

Please answer following questions, either question-by-question, or as a complete essay. A sufficient answer, containing at least a couple of complete sentences, yields two points.

- What are the key usability and/or user experience problems with the product?
- What kinds of user information would be required to improve the product?
- What user research methods can provide the requested information? Here, you can provide your own ideas on this, because we have not yet covered different user research methodologies.
- How would the kinds of methods listed help a designer in improving the product? Please explicitly refer to one of the articles listed above.

2 GET FAMILIAR WITH DATASETS (12P)

The goal of this task is to make sure that you know how to manage and view data on your choice of statistical software (R with `tidyverse` recommended). In particular, you should be able to

- load any structured dataset,
- get an overview of the data (e.g., how many variables and observations there are),
- summarise the data (e.g., in terms of the number of different factors, or arithmetic mean of variables,
- make basic visualisations, such as boxplots or bargraphs.

If you are unsure about the basic statistical concepts, such as mean and standard deviation, please rehearse them (the internet is full of text and video tutorials, and these are not super difficult concepts to grasp).

You are given two datasets.

DOGS AT NURSING HOMES

The first dataset (`crowley-1996.csv`) is a recreation of the data from the article that was shown on the lecture: Crowley-Robinson et al. (1996). "A long-term study of elderly people in nursing homes with visiting and resident dogs." The study introduced a resident dog to an elderly nursing home, and for the duration of six months, one per month, asked the elderly inhabitants about their mood, such as depression. The original dataset is not available, but it has been created from Fig. 2 in the article with an assumption that most of the time depression goes down, not up, for an individual. However, please don't treat these reconstructed data as actual empirical data, these are just for practice purposes.

Although the dataset is not from an HCI study, it is a prototypical example, suitable for HCI. Often in HCI, there is treatment (here, a visiting dog), and control conditions. Some response variable is measured (here, depression) for a length of time, or during multiple tasks, and the final interest is in the effectiveness of the treatment. The question might for example be, if the new layout improvement actually increases user experience or task efficiency. The dataset holds following variables.

- `id`: participant specific identificator
- `condition`: whether the participant was in the resident dog house, or in the control condition
- `m`x: self-report on depression on a given month, such that 3 = high, 2 = moderate, and 1 = low

Please answer the following questions. You don't need to provide code used to get the results, but please answer using whole sentences.

1. Get to know the data. (1p)
    a) Identify the between and within subjects variables (or factors, if a factor is spread among multiple variables).
    b) How many participants were there in total? How many participants in each home?
2. Describe the main response variable of the dataset, depression. How many responses, in total, are there in the three levels of depression? (1p)
3. Visualise the effect of the treatment. Make a bargraph similar to that of Fig 2, shown on the slides of the first lecture (p. 29). The "control" and "resident" conditions should be shown next to each other, or one on top of the other. Bargraphs should show, for each month and depression level, how many participants reported that level of depression. (3p)
4. Is there an internal validity threat in the starting (m1) depression between the groups? Investigate how m1 differs between the groups, and provide your thoughts. (1p)

Tips (R specific).

- The data points are separated by a comma (","), so you need to specify this to `read.table`.
- The data are in a wide format. In order to create the visualisation, it is easier to change the data into long format using `gather`. Please consult the R-clinic instructions to see how this works.
- When using `geom_bar()` in plotting, the bars are stacked by default. Use `position = ''dodge''` to place the bars side by side.
- You can make two separate graphs for the two conditions, but make sure that their y-axis scales match (otherwise comparison will be difficult). However, it is easier to use `facet_grid` to make separate plots within a plot.
- In order to analyse how m1 behaves between the two groups, you can for instance calculate how many respondents belong to each of the three depression groups within the conditions. In R, this can be done with a correct `group_by()`, followed by summarising using `n()` function.

TYPING ON A SMARTPHONE

The second dataset (`smartphone-typing.csv`) is from a study conducted at our research group. Young and old adults were asked to transcribe sentences with a normal smartphone Qwerty keyboard. The dataset contains following

- `id`: unique identificator for the participant
- `age.group`: whether the participant is young or old adult
- `sentence`: unique identificator for the sentence
- `n.letters`: how many letters that sentence contains

- `trialtime`: how much time from the first keypress to pressing enter the participant took, in milliseconds
- `errors`: how many errors, as percent of the sentence length, in terms of edit distance, were there in the final sentence (not counting corrections) (note that this variable can have values over 1.0 in some extreme cases)
- `backspaces`: how many times backspace was pressed for error correction purposes

This exercise will not go through all the ways these data could be analysed. We'll return to this dataset in later exercises.

Please answer the following questions. You don't need to provide code used to get the results, but please answer using whole sentences.

1. Get to know the data. (1p)
   a) Identify the between and within subjects variables (or factors, if a factor is spread among multiple variables).
   b) How many participants are there? How many old and young adults?
2. Create a new variable, called `cps`, which indicates, on average, how many characters the participant typed in a second. In other words, for each sentence, divide the number of letters by trial time. Note that trialtime is in milliseconds, but we wish `cps` to be *characters per second*. Please report the summary statistics of this new variable: its minimum and maximum values, its mean, and its standard deviation. Also provide a histogram showing the distribution of the variable. (1p)
3. How did the old and young adults, in general, differ in their typing speed (`cps`)? We'll examine this question in more detail in the future. For now, it's enough that you provide a boxplot of participant-specific average `cps`. In other words, for each participant, calculate their own mean cps, and use this dataset to make a boxplot, that shows differences between young and old adults. (4p)

Tips (R specific).

- The data points are separated by a comma (","), so you need to specify this to `read.table`.
- The data are in a long format, and for the purposes of this exercise, it's already the correct one.
- Use `mutate` command to calculate `cps`, but remember that `trialtime` is in milliseconds.
- If you wish to have a better intuition on the observed typing speeds, there are different rules at calculating *WPM* (words per minute) from *CPS*. For example, you can define word to be 5 chars, and so $WPM = CPS \times 60/5$. If you wish, you can present your results in words per minute instead of characters per second.
- A histogram of a variable is created by supplying just that variable to the `aes()` function, and adding `geom_histogram()` to the plot. You can get histograms of different

detail by specifying `bins` parameter to it.

- For id-specific mean `cps`, you need to group the data by `id`, and then mutate `cps` as its own mean. However, to retain the `age.group` in the resulting dataset, also group by it. It does not affect the mean `cps` calculation, because every participant belongs to just one age group.