Report Template coursework assignment A - 2021

CS4125 Seminar Research Methodology for Data Science

Student names (student numbers)

20/04/2021

Contents

1 Part 1 - Design and set-up of true experiment

1.1 The motivation for the planned research

(Max 250 words)

1.2 The theory underlying the research

(Max 250 words) Preferable based on theories reported in literature

##Research questions The research question that will be examined in the experiment (or alternatively the hypothesis that will be tested in the experiment)

##The related conceptual model This model should include: *Independent variable(s)* Dependent variable *Mediating variable (at least 1)* Moderating variable (at least 1)

##Experimental Design Note that the study should have a true experimental design

1.3 Experimental procedure

Describe how the experiment will be executed step by step

1.4 Measures

Describe the measure that will be used

1.5 Participants

Describe which participants will recruit in the study and how they will be recruited

1.6 Suggested statistical analyses

Describe the statistical test you suggest to care out on the collected data

2 Part 2 - Generalized linear models

2.1 Question 1 Twitter sentiment analysis (Between groups - single factor)

2.1.1 Conceptual model

Make a conceptual model for the following research question: Is there a difference in the sentiment of the tweets related to the different celebrities?

2.1.2 Collecting tweets, and data preparation

Include the annotated R script (excluding your personal Keys and Access Tokens information), but put echo=FALSE, so code is not included in the output pdf file.

2.1.3 Homogeneity of variance analysis

Analyze the homogeneity of variance of sentiments of the tweets of the different celebrities, and provide interpretation

```
#include your code and output in the document
```

2.1.4 Visual inspection Mean and distribution sentiments

Graphically examine the mean and distribution sentiments of tweets for each celebrity, and provide interpretation

```
#include your code and output in the document
```

2.1.5 Frequentist approach

2.1.5.1 Linear model

Use a linear model to analyze whether the knowledge to which celebrity a tweet relates has a significant impact on explaining the sentiments of the tweets. Provide interpretation of results

```
#include your code and output in the document
```

2.1.5.2 Post Hoc analysis

If a model that includes the celebrity is better in explaining the sentiments of tweets than a model without such predictor, conduct a post-hoc analysis with e.g. Bonferroni correction, to examine which of celebrity tweets differ from the other celebrity tweets. Provide interpretation of the results

```
#include your code and output in the document
```

2.1.5.3 Report section for a scientific publication

Write a small section for a scientific publication, in which you report the results of the analyses, and explain the conclusions that can be drawn.

2.1.6 Bayesian Approach

2.1.6.1 Model description

Describe the mathematical model fitted on the most extensive model. (hint, look at the mark down file of the lectures to see example on formulate mathematical models in markdown). Justify the priors.

2.1.6.2 Model comparison

Conduct model analysis and provide brief interpretation of the results

```
#include your code and output in the document
```

2.1.6.3 Comparison celebrity pair

Compare sentiments of celebrity pairs and provide a brief interpretation (e.g. CIs)

2.2 Question 2 - Website visits (between groups - Two factors)

2.2.1 Conceptual model

Make a conceptual model underlying this research question

2.2.2 Visual inspection

Graphically examine the variation in page visits for different factors levels (e.g. histogram, density plot etc.)

```
#include your code and output in the document
```

2.2.3 Normality check

Visually inspect if variable page visits deviates from a Gaussian distribution, and discuss implication for general linear model analysis.

```
#include your code and output in the document
```

2.2.4 Frequentist Approach

2.2.4.1 Model analysis

Conduct a model analysis, to examine the added values of adding 2 factors and interaction between the factors in the model to predict page visits, and include brief interpretation of the results.

```
#include your code and output in the document
```

2.2.4.2 Simple effect analysis

If the analysis shows a significant two-way interaction effect, conduct a Simple Effect analysis to explore this interaction effect in more detail.It helps first to look at the means of different conditions in a figure. Provide brief interpretation of the results.

```
#include your code and output in the document
```

2.2.4.3 Report section for a scientific publication

Write a small section for a scientific publication, in which you report the results of the analyses, and explain the conclusions that can be drawn.

2.2.5 Bayesian Approach

2.2.5.1 Model description

Describe the mathematical model fitted on the most extensive model. (hint, look at the mark down file of the lectures to see example on formulate mathematical models in markdown). Justify the priors.

2.2.5.2 Model comparison

Conduct model analysis and provide brief interpretation of the results

```
#include your code and output in the document
```

3 Part 3 - Multilevel model

3.1 Visual inspection

Use graphics to inspect the distribution of the score, and relationship between session and score

```
#include your code and output in the document
```

3.2 Frequentist approach

3.2.1 Multilevel analysis

Conduct multilevel analysis and calculate 95% confidence intervals, determine:

- If session has an impact on people score
- If there is significant variance between the participants in their score

```
#include your code and output in the document
```

3.2.2 Report section for a scientific publication

Write a small section for a scientific publication, in which you report the results of the analyses, and explain the conclusions that can be drawn.

3.3 Bayesian approach

3.3.1 Model description

Describe the mathematical model fitted on the most extensive model. (hint, look at the mark down file of the lectures to see example on formulate mathematical models in markdown). Justify the priors.

3.3.2 Model comparison

Select the first 100 participants from the data set. (hint to overcome the Stan problem with a zero index, increase subject id number with 1). Compare models with with increasing complexity.

```
#include your code and output in the document
```

3.3.3 Estimates examination

Examine the estimate of parameters of the model with best fitt, and provide a brief interpretation.

```
#include your code and output in the document
```