Title of your report

Leave author name blank

Your submission date

Preliminaries

This is an R Markdown document. For more details on using R Markdown see http://rmarkdown.rstudio.com.

This document serves a multiple purpose:

- It provides you with some initial code which is needed in order to read the data which are required to complete this assignment;
- It contains instructions on what is expected from your assignment;
- You produce your actual assignment by adding text and code to this document, and eventually knitting it when finished.

Your final submitted document will be an report, written in R Markdown, which may contain text, equations, code, and images, and which should be 3000 words long. It is admitted that "word count" may be an unclearly defined concept for such a heterogeneous document. We therefore recommend to use the `wordcountaddin` to establish your wordcount (see code at the end of this document). To avoid the document becoming excessively long, we think of an image as representing 50 words, and of R code contributing another 20 (or so) words per chunk. So, assuming that you have a reasonable number of such elements included into your report, then a word count of **2000** according to the `wordcountaddin` will typically be appropriate. You *don't* need to evidence how this would add up to 3000 when accounting for code and/or figures.

The final document needs to be a knitted PDF version of this Markdown document, with the `Preliminaries` and `Presentation` sections removed, and of course your content added. This is non-negotiable. MS Office documents will not be accepted. The PDF file should be submitted using Gradescope.

Your report needs to have four sections: Introduction, Methods, Analysis and Discussion of results. The framework and mark scheme of these is already set up below, with some indications of envisaged content. Within and between these four sections, you have some flexibility to move content around (in particular; these indications should not be understood as prescribing or limiting the scope of your work), but you should provide an overall sound story which addresses the overall question of this assignment.

An individual submission is required (that is, it should be your own work). In writing your essay, you may refer to any of the teaching materials provided as part of this module. You may also refer to any other reference, provided it is cited and listed as part of a bibliography.

R Packages The following are some R packages that you are likely to need for your analysis. It is recommended to load these now.

```
require(lme4)
require(lmerTest)
require(ggplot2)
require(sjPlot)
```

Data We now introduce the data to be used in this assignment. You can modify this part and move it into one of the subsections of your assignment as you deem appropriate.

We consider data resulting from a longitudinal multisite trial on stress of nurses working in Accident and Emergency (A&E) departments in hospitals. For each of 20 hospitals, nurses working in the Accident and Emergency department were randomly assigned to an experimental or a control condition. In the experimental condition, all nurses receive a training program to cope with job-related stress. After the program is completed, a sample of 10 nurses from each hospital's A&E department are given a test that measures job-related stress, at intervals of 1 month, 2 months and 3 months after the date of the training program.

You should work with your own unique data set, which is labelled according to your first name (in lower case and as appearing in the class register of our course Ultra page). For example, my data can be read into a data frame `MST` as follows:

```
MST <- read.csv("https://andygolightly.github.io/teaching/MATH43515/summative/andy.csv",
                header=TRUE)
#MST<- read.csv("andy.csv", header=TRUE) #Use this if the file is saved locally
head(MST)
```

```
##   ID Hospital Responset1 Responset2 Responset3 Trt Experience Gender Size
## 1  1        1         36         38         38   1        6.8      1    0
## 2  2        1         35         39         39   1        9.1      1    0
## 3  3        1         46         41         41   0        6.0      1    0
## 4  4        1         31         31         40   1        3.7      0    0
## 5  5        1         36         36         39   1       12.1      1    0
## 6  6        1         29         33         36   1       15.8      0    0
```

```
dim(MST)
```

```
## [1] 200   9
```

We see that the data set possesses 200 rows and 9 columns. These columns represent the following variables:

- `ID`: anonymized nurse identifer;
- `Hospital`: hospital identifier $\{1, 2, \ldots, 20\}$;

- `Trt`: treatment / experiment indicator (0=control; 1=training program given (treatment));
- `Experience`: nurse experience (continuous scale, units of years);
- `Gender`: gender indicator (0=male; 1=female);
- `Size`: indicator of A&E department size (0=small, 1=large);
- `Responset1`: A post-test stress score for each nurse at time 1;
- `Responset2`: A post-test stress score for each nurse at follow-up time 2;
- `Responset3`: A post-test stress score for each nurse at follow-up time 3;

All test scores were recorded on an integer valued scale on $[0, 100]$ ranging from "no stress" to "maximum stress".

Question of interest Does the experimental 'Intervention' have a significant impact on nurse stress (post-test)? We are also interested in investigating how this impact evolves over time.

Note: There are several ways to look at this data. You could just consider the first post-test score as response (and ignore the other two post-tests). One could then consider fitting separate such models to the

other post-test scores. Alternatively, a longitudinal 3-level model could be set up. This requires re-shaping the data into long format first, and needs more skill and expertise. While it is not unreasonable to attempt this given the tools acquired in this course, it is noted that a strong mark can be obtained without doing this.

Here your actual essay will start. Remove all unnecessary prior parts before submitting.

Introduction (20 marks)

Give some background on multisite trials, and their relevance for assessing the effectiveness of an intervention. This subsection could also introduce the data, identify the variables at each level, and give some simple illustrative exploratory analysis, without fitting any models.

Methods (20 marks)

Explain what multilevel models do and why they are relevant for the modelling of this data set. Explain the type of models that you are considering for your report (2-level or 3-level, longitudinal, etc). This subsection should also give details on how to decompose variance (ICC etc) and a brief description of the methods used for model building.

Analysis (35 marks)

This section should contain the actual modelling and data analysis. You are welcome to fit several potential models, and apply suitable tests or diagnostic tools to compare them. Possible questions to guide your model building are:

1. What is the intervention effect and its confidence interval estimate; is the intervention effect significant?
2. Is there evidence of heterogeneity of the intervention effect between hospitals?
3. Is there evidence of a change of the intervention effect over time?
4. What other covariates are important in explaining stress?
5. For the final fitted model(s), are the model assumptions reasonable?

Be sure to make clear the models you are fitting at each stage of your analysis.

Discussion of results (15 marks)

In this section you will discuss, interpret and summarize the results from your analysis in the light of the original problem: do these data give evidence that the intervention works? You can also discuss other aspects of your analysis and identify any limitations of the approach taken.

Presentation (10 marks)

Marks for correct structure (including title), mathematical typesetting (where required), writing style, presentation and clarity of figures etc.

Remove this section from your final report.

Word count

```
#install.packages("devtools")
#devtools::install_github("benmarwick/wordcountaddin", type = "source", dependencies = TRUE)
require(wordcountaddin)
```

```
## Loading required package: wordcountaddin
```

```
word_count()
```

```
## For information on available language packages for 'koRpus', run
##
##    available.koRpus.lang()
##
## and see ?install.koRpus.lang()


##
## Attaching package: 'koRpus'


## The following object is masked from 'package:wordcountaddin':
##
##      readability


## [1] 1160
```

```
text_stats()
```

Method	koRpus	stringi
Word count	1160	1126
Character count	7182	7172
Sentence count	94	Not available
Reading time	5.8 minutes	5.6 minutes