Data analysis and statistics course

End of semester assignment

==================================

Answer all questions. The length limit is **strictly** 12 pages. Create an R project, copy the data file into the folder, create an .Rmd Notebook file, set it up, do the analysis, knit to an html or pdf or docx report, zip the entire folder, upload it to VULA when finished.

**Question 1 - 35 marks**

A Psychology PhD student is interested in interventions that try to promote growth mindsets in children learning to read (if you don’t know what ‘growth mindsets’ are, take a look at some of Carol Dweck’s research).

The student in question mounts an intervention with two groups of children in Grade 2, namely a control group (control), and an intervention group (growth, or growth\_mindset). The control group receives the usual classroom activities, whereas the growth mindset group spends an hour each week of the year doing activities aimed at promoting a growth mindset (if you want to know more about what these might be take a look at the various mindset interventions Dweck and colleagues have tested). Each child is tested at the beginning of the program (January), and then halfway through the program (June), and again at the end of the program (December).

The data for the control group is in the R RDS object *control.RDS* and the data for the growth\_mindset group is in the R RDS object *growth.RDS*. Start by reading the objects into R, and then merging them (hint: you can use a full join, or a row bind command; additional hint: you need to create a vector in the new dataframe that records which group they are from). The variables in the RDS objects are: participant\_id (unique number assigned to each child), time (January, June or December), and reading\_score (performance on a standardized reading test, maximum score of 35).

Explore the data descriptively, creating appropriate tables and figures, where needed.

There are two research hypotheses to investigate: the researcher thinks that the growth group will improve more than the control group does over time; the researcher thinks that both groups will improve over time. Run suitable Analysis of Variance or Mixed Linear Model analyses to test these hypotheses. Report the results, including necessary tables and figures.

**PISA data – background information for Questions 2 and 3**

The global consortium PISA (Programme for International Student Assessment) conducts extensive surveys of student abilities and attitudes, among other things, every three years. See <https://www.oecd.org/pisa/>.

The data file pisa\_2018.csv contains de-identified data for a random sample of 500 New Zealand students, drawn from the various PISA databases for that year.

The Excel file pisa\_codes.xlsx provides some detail for each of the items in the data file. Additional details can be found at <https://www.oecd-ilibrary.org/sites/0a428b07-en/index.html?itemId=/content/component/0a428b07-en>, and <https://www.oecd.org/pisa/data/pisa2018technicalreport/PISA2018_Technical-Report-Chapter-16-Background-Questionnaires.pdf>. Note – you don’t need to consult those sources, the Excel file should provide enough detail.

**Question 2 - 20 marks**

Two scales that were measured in the PISA 2018 data were those of *resilience*, and *fear of failure*. These are referenced in the codebook referred to above.

Combine these items in a factor analysis and assess whether there is good evidence for a two-factor structure, each factor corresponding to one of the original scales. Compute measures of internal consistency for each of the scales.

**Question 3 - 45 marks**

A key outcome measure in the PISA 2018 dataset, for our purposes, is reading fluency. Build, test, and report a regression model that takes reading fluency as the outcome variable, and models it with predictors available to you in the dataset. Select variables for inclusion in the model on the basis of whether they seem likely to you to be useful predictors. Remember to think about possible control variables (and how you would treat them as controls). Provide useful tables and figures summarizing your analysis, and findings.