Multivariate Statistics and Methodology in R  
Coursework report

2 Overview

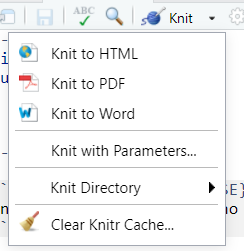
You need to produce a report answering the assignment questions detailed on the following pages. You will also need to provide the R code which reproduces the answers you give in the report.

There are two sections to the coursework, based on the two five-week blocks of the course. Try to write each section of the report as if you are writing a paper, or your dissertation. Try to approach each whole section as a single study, rather than separate independent questions (i.e., avoid answering a list of research questions, but providing a comprehensive analysis of the data in light of the question(s) of interest).

We would recommend that for both sections of the coursework you write an analysis section and a results section. The analyses section should detail the appropriate analyses you undertook and how they will provide answers to the research questions. The results section should present and discuss your findings, utilising graphics where necessary to illustrate your points. Analyses will draw on the methodologies we have discussed in lectures and weekly exercises.

3 What you need to submit

You are required to submit 2 documents.

1. Your final compiled report, detailing your analyses, results, interpretation and conclusions. This can be **.docx** (or equivalent), **.pdf**, or **.html**.  
   [https://rstudio.ppls.ed.ac.uk](https://rstudio.ppls.ed.ac.uk/) now supports knitting to **.pdf** and word (**.docx**); choose by clicking the drop-down arrow next to “Knit”.  
   
2. A **.Rmd** document which reproduces the results you give in the report.

4 Tips for writing

For the task below, the *compiled* report (final .pdf, .html, or word document) is expected to include:

1. **Clear written details of the analysis conducted** in order to answer the research questions, including transparency with regards to decisions made about the data prior to and during analysis.
2. **Results**, in appropriate detail (for instance, a test statistic, standard error and p-value, not just one of these).
3. **Presentation** of results where appropriate (in the form of tables or plots).
4. **Interpretation** (in the form of a written text referencing relevant parts of your results) leading to a conclusion regarding the research questions.

The code you write in your submitted .Rmd file should successfully undertake the analysis described in A), which returns B). You should also include the code to produce C).

As the compiled document will not contain visible R code, a large part of the challenge comes in clearly describing all aspects of the analysis procedure.  
A reader of your compiled document should be able to more or less replicate your analyses **without** referring to your R code.

**IMPORTANT:** Some functions (such as fa.parallel()) involve randomly generating numbers, and so results will vary slightly each time you run them. To ensure that your results are the same each time, at the top of your code, use set.seed() to set the random seed. Choose a number (any length) and pass it to set.seed(). Then, every time you run random number generations, it will produce the same results.  
For example:

set.seed(8675309) *# This is an example, choose your own!*

5 Specifics

5.1 Report Formatting

Beyond assessing your ability to create reproducible results, the focus of this assignment is not on the formatting of RMarkdown files; it is on implementing and interpreting analyses in R.

This means that you are more than welcome, for instance, to knit to word, and then to subsequently make edits to your text and formatting in the word document itself. *The important thing to remember is that the data analysis and modelling results should match those produced in your RMarkdown file.*

If you do wish to do your formatting in RMarkdown, then we suggest the following readings for help:

* [RMarkdown CheatSheet](https://rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf)
* [Writing Math in Rmd](https://rpruim.github.io/s341/S19/from-class/MathinRmd.html#:~:text=Math%20inside%20RMarkdown,10n%3D1n2.)
* For linear regression tables to .html and .pdf, the [**stargazer** package](https://cran.r-project.org/web/packages/stargazer/vignettes/stargazer.pdf)
* For linear regression tables to .html, the [tab\_model() from the](https://strengejacke.github.io/sjPlot/articles/tab_model_estimates.html)**[sjPlot](https://strengejacke.github.io/sjPlot/articles/tab_model_estimates.html)**[package](https://strengejacke.github.io/sjPlot/articles/tab_model_estimates.html)

Feel free also to post formatting questions on the discussion forum on Piazza.

5.2 Page limit

**Your report should be no longer than 5 pages, you may place tables and figures in an appendix should you wish** (if you are knitting to html, please open the html in a browser and print to pdf in order to check how many pages your submission takes up).

5.3 Grading

Marks will be awarded for providing evidence of the ability to understand and execute appropriate statistical methods to answer the research questions, and for clear presentation of results, interpretation and conclusions.

* Ten points will be deducted from your final grade if your marker cannot determine how the results in your report were generated in your .Rmd document (for example, if the .Rmd document produces errors during knitting, or produces values different from those reported). This means that an 82 out of 100 becomes 72 out of 100.
* **Up to** ten points will be added for good use of R and RMarkdown (for example, where code is elegant, or when inline constructs are used to report results in the text). For example, a 72 might be raised up to 82.

5.4 Submission

**Check that your code runs!**

Before submitting, we strongly advise you to check that your code runs. The easiest way to check this is by checking that your .Rmd compiles (i.e., can you knit your Rmarkdown document into .html/.pdf/.docx without error?)

**If you use RStudio on the PPLS Server, you will need to export the file to your computer in order to upload it to Turnitin.**

**Filenames**

For both files which you submit, the filename should be your Exam Number with the appropriate extension, and nothing else.  
For example, a student with exam number **B047847** would submit two files:

* **B047847.Rmd**
* one of: **B047847.docx** / **B047847.pdf** / **B047847.html**

**Where to submit**

Go to the Assessments page on Learn, and look for “Assessment Submission”. There you will find an own-work declaration which requires marking as reviewed, before two submission boxes will be visible (one for each file).

**7 COURSEWORK TASK**

7.1 Section 1

A study on individual differences in metaphor comprehension tested 76 participants on comprehension of 40 metaphors which were categorised as either ‘literary’ or ‘non-literary’. For example:

* “Literary”
  + A body is a prison for the soul.
  + The tongue is a bayonet.
  + The mind is a mountainous landscape.
* “Non Literary”
  + The nose is the antenna of scent.
  + Invention is the child of an inventor.
  + Purgatory is the lobby of heaven.

Participants also completed a measure of fluid intelligence / non-verbal problem solving (Raven’s Progressive Matrices, RPM) and crystalized intelligence / semantic knowledge (Semantic Similarities Test, SST).

**Research Questions**  
1. Are “literary” metaphors harder to understand than “non-literary” metaphors? (Is comprehension accuracy lower?)  
2. Is metaphor comprehension affected by fluid intelligence? Is fluid intelligence more important for comprehending literary or non-literary metaphors?  
3. Is metaphor comprehension affected by semantic knowledge (crystalized intelligence)? Is semantic knowledge more important for comprehending literary or non-literary metaphors?

The data in **.RData** format is available at <https://uoepsy.github.io/data/metaphor.RData>, and a data dictionary can be found in Table [7.1](https://www.learn.ed.ac.uk/bbcswebdav/pid-5853120-dt-content-rid-20183043_1/courses/PSYL110542020-1SV1SEM2/msmr_2021_cwork.html#tab:tab2).

| Table 7.1: Metaphor Comprehension: Data Dictionary | |
| --- | --- |
| **variable** | **description** |
| SubjectCode | Unique participant identifier |
| Age | Participant age at time of testing |
| RPM | Score on Raven’s Progressive Matrices (RPM) test of fluid intelligence (non-verbal problem solving). Number of matrices solved correctly (0-12) |
| SST | Score on Semantic Similarities Test (SST) of crystallized intelligence (semantic knowledge). Number of correct responses (19-37) |
| Gender | Participant gender (Male, Female, or not specified) |
| Item | Unique metaphor identifier |
| Correct | Participant’s comprehension of the metaphor statement (1 = Correct, 0 = Incorrect) |
| Condition | Metaphor type (Literary or Non-Literary) |

7.2 Section 2

A researcher is interested in the role of parenting styles in mediating the association between maternal depression and child behavioural problems. n=1200n=1200 mothers were recruited when their child was three years of age. They provided data on their own depression symptoms and parenting style at age three and their child’s conduct problems at age five. Each construct (maternal depression, erratic parenting, conduct problems) was measured using between 4 and 6-items. In previous exploratory factor analyses, all three scales showed unidimensionality, i.e., the optimal factor structure was a single factor model.

Using the data described in Table [7.2](https://www.learn.ed.ac.uk/bbcswebdav/pid-5853120-dt-content-rid-20183043_1/courses/PSYL110542020-1SV1SEM2/msmr_2021_cwork.html#tab:tab1) and available (in **.csv** format) at <https://uoepsy.github.io/data/msmr_2021_assignment.csv>, conduct, interpret and write up a set of analyses that answer the researcher’s research question. Include a brief discussion of the limits of what can be inferred from the analyses.

**Research Question**  
Does erratic parenting (a parenting style characterised by failing to be consistent in discipline) mediate the association between maternal depression and child conduct (behavioural) problems at age 5? Does the gender of the child make any difference?

| Table 7.2: Child Behavioural Problems - Data Dictionary | |
| --- | --- |
| **variable** | **description** |
| ID | Unique participant identifier |
| Child\_gend | Child gender with 0=female, 1=male |
| Dep1 | ‘Anhedonia’ measured on a 5-point Likert scale, higher scores mean higher levels of depression |
| Dep2 | ‘Lack of hope’ measured on a 5-point Likert scale, higher scores mean higher levels of depression |
| Dep3 | ‘Lack of energy’ measured on a 5-point Likert scale, higher scores mean higher levels of depression |
| Dep4 | ‘Sad without reason’ measured on a 5-point Likert scale, higher scores mean higher levels of depression |
| Dep5 | ‘Tired all the time’ measured on a 5-point Likert scale, higher scores mean higher levels of depression |
| Dep6 | ‘Feeling worthless’ measured on a 5-point Likert scale |
| Errat1 | ‘Failure to follow through on punishments’ measured on a 5-point Likert scale, higher scores mean higher levels of erratic parenting |
| Errat2 | ‘Severity of punishment depends on mood’ measured on a 5-point Likert scale, higher scores mean higher levels of erratic parenting |
| Errat3 | ‘Child talks parent out of punishment’ measured on a 5-point Likert scale, higher scores mean higher levels of erratic parenting |
| Errat4 | ‘Restrictions are lifted early because they are too much effort’ measured on a 5-point Likert scale, higher scores mean higher levels of erratic parenting |
| CP1 | ‘Child has regular long-lasting temper tantrums’ measured on a 5-point Likert scale, higher scores mean higher levels of conduct problems |
| CP2 | ‘Child bites, kicks, or hits other children’ measured on a 5-point Likert scale, higher scores mean higher levels of conduct problems |
| CP3 | ‘Child refuses to do as told’ measured on a 5-point Likert scale, higher scores mean higher levels of conduct problems |
| CP4 | ‘Child deliberately breaks things’ measured on a 5-point Likert scale, higher scores mean higher levels of conduct problems |
| CP5 | ‘Child steals things’ measured on a 5-point Likert scale, higher scores mean higher levels of conduct problems |
| CP6 | ‘Child lies frequently’ measured on a 5-point Likert scale, higher scores mean higher levels of conduct problems |