

QML – Summative Assessment 2

YOUR EXAM NUMBER

2023-11-23

Contents

1	Overview	1
2	Read the data	1
3	Plot the data	1
4	Model the data	2
5	Results	2
6	Discuss the results	2

1 Overview

PLEASE READ CAREFULLY

DUE Week 12 - Thu 07 December at noon

- You must include your **exam number as the author** in the document preamble above.
- Make sure you can render the Rmd file to PDF frequently while completing the assessment to avoid last minute panic!
- Be careful with formatting. Make sure there is an empty line after each paragraph and/or code chunk.

2 Read the data

Instructions:

- Read the chosen data.
- Write below a brief description about the data:
 - Include the number of observations and the number of columns.
 - Identify the variables/columns in the data that will be relevant to answer the research question, their type (categorical or numeric continuous/discrete) and, if categorical, report the levels of the variable.
 - Report counts and/or summary measures of the relevant variables plus any other you think might be important to mention (for example, info about participants).

3 Plot the data

Instructions:

- Now plot the data as you see fit. Remember the research question of the chosen data to help you decide which plots to include.
- For each plot, write a brief description as text (include information on the components of the plot like axes, colour, panels etc and briefly describe the patterns you see).
- Plotting the data is a good way of familiarising yourself with it, so that when it comes to modelling you have a better sense of what the data looks like!
- **Include at least 3 plots.**

4 Model the data

Instructions

- Finally, model the data and write a model report.
- Pick a model formula that will help you answer the research question.
- Follow these steps to help you model the data:
 - Identify the outcome variable and the predictors.
 - Do you need to transform the variables and/or reorder levels?
 - Identify the appropriate distribution for the outcome variable.
 - Is an interaction term necessary?
 - Do you need to centre any predictor?

5 Results

Instructions

- You should include at least two paragraphs.
 1. One reporting the model specifications (outcome, distribution of the outcome, predictors, coding, ...)
 2. One reporting the results from the model. To simplify things, here is some guidance. **Include at least the following:**
 - For each model coefficient (the “population-level effects” that get printed with `summary()`), report the 95% CrI and the mean (`Estimate`) with the estimate error (i.e. the SD, `Est.error`) of the posterior probability of the coefficient.
 - For each combination of levels of categorical predictors, report the 95% CrI of the conditional posterior probability.
 - If your model includes a numeric predictor, you don’t need to report conditional posterior probabilities depending on that predictor, but you can just report the 95% CrI of the conditional posterior probabilities depending on levels of the categorical predictor, *when the continuous predictor is at its mean*.
 3. **Optional:**
 - If you wish, you can report the 95% CrI of relevant differences between levels/values of categorical/numeric predictors.
- Include plots of the posterior distributions of each coefficient and the conditional posterior distributions.

6 Discuss the results

Instructions:

- Based on the results that you reported above, try and answer the research question.
- To give you an example of what we are looking for, here is a short paragraph on interpreting the results from the Week 8 lecture on morphological parsing.

As suggested by the 95% CrI of the interaction term (in log-odds [-1.72, 0.42]), there is quite a lot of uncertainty as to whether the difference in probability of correct response in unrelated vs constituent in right-branching pairs differs from that in left-branching pairs, since the interval covers both negative and positive values. Moreover, the conditional posterior probabilities of unrelated and right-branching on the one hand and constituent and right branching on the other are very similar, as can be seen in the plot above (and as suggested by the respective 95% CrIs: 90-97% vs 91-97% respectively).