

# QML - Formative Assessment 2

YOUR EXAM NUMBER

2023-11-10

## 1 Overview

### PLEASE READ CAREFULLY

#### DUE Week 10 - Thu 23 November at noon

You must include your **exam number as the author** in the document preamble above.

You'll notice that line 10 of the preamble says `mainfont: DejaVu Sans`. We would appreciate it if you'd use this font (or at least some other sans-serif font), but it's fine to use any other sans-serif font, like Arial. **Try to render this Rmd file now, before making any changes, to see if you have this font installed.**

- If you get an error message saying that DejaVu Sans could not be found, you can download it here: <http://sourceforge.net/projects/dejavu/files/dejavu/2.37/dejavu-fonts-ttf-2.37.zip>.
- Then, install it appropriately for your operating system.
  - Here's a guide for Windows: <https://www.digitaltrends.com/computing/how-to-install-fonts-in-windows-10/>.
  - Here's a guide for Mac: <https://support.apple.com/en-gb/HT201749>.
- If you are having issues with this, feel free to use any other sans-serif font you have installed on your machine, like Arial.

This assessment covers **Weeks 1 to 8**.

You are asked to **read, wrangle, summarise, plot and model data** from Lorson, Alex, Cummins, Chris & Rohde, Hannah. 2021. 'Strategic use of (un)certainity expressions'. *Frontiers in Communication*. 6. DOI: <https://doi.org/10.3389/fcomm.2021.635156>.

Read the paper abstract to get a sense of what the study is about.

You will analyse the data from the second study in the paper. You can skim Section 3.2 of the paper for an overview of the study. You will work on the Participant, Response, Scenario, Gender and Age columns.

#### When you are ready to submit:

1. Render the Rmd file to **PDF**.
2. **Rename** the PDF to your exam number only.
3. **Upload** the PDF file to Learn.

## 2 Read the data

### Instructions:

- Read the data in `data/lorson2021-Study2d.csv`.

- Write a paragraph describing the data frame. Include the following information (some of these will require you to write R code):
  - Number of observations and columns.
  - Number of participants (`Participant`).
  - Minimum and maximum number of observations per participant.
  - Levels of `Response`.
  - Levels of `Gender`.
  - Levels of `Scenario`.

### 3 Calculate summary measures

#### Instructions:

- Since all the relevant variables except `Age` are categorical, it might be useful to just report for these the number of each response (“know” vs “believe”) in the different combinations of `Scenario` and `Gender`.
- Write R code to obtain those counts. If you want you can write a markdown table with the counts. You can use this online tool to create and copy-paste a markdown table: [https://www.tablesgenerator.com/markdown\\_tables](https://www.tablesgenerator.com/markdown_tables).
- Although we will not include `Age` in our model later, it is customary to provide readers with general socio-demographic information on the participants and this usually includes summary measures for the participants’ age.
- Go ahead and calculate the appropriate summary measures for `Age`. Write a sentence reporting the measures.

### 4 Plot the data

#### Instructions:

- Now plot the data as you see fit. Remember that the focus of the study is how participants use “know” vs “believe” (`Response`) so you might want your plots to focus on that variable, but feel free to also plot combinations of any of `Participant`, `Response`, `Scenario`, `Gender` and `Age`.
- For each plot, write a brief description (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!

### 5 Model the data

#### Instructions

- Finally, you can model the data and write a model report.
- You are trying to answer the following research question: *Do participants show a preference for “know” vs “believe” depending on their gender and the scenario they are in?*
- 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?

## 6 Results

### Instructions

- You should include at least two paragraphs, (a) one reporting the model specifications (outcome, distribution of the outcome, predictors, coding, ...) and (b) one reporting the results from the model (estimates and estimates' SD, Credible Intervals, conditional posteriors and differences across different levels, etc...)
- Include plots of the posterior distributions of each coefficient (listed in the model summary) and the conditional posterior distributions. You can also include the posterior distributions of differences among levels not covered by the coefficients, if any.

We fitted a Bayesian model...

According to the results, ...

## 7 Discuss the results

### Instructions:

- Based on the results which you reported above, try and answer the research question: *Do participants show a preference for "know" vs "believe" depending on their gender and the scenario they are in?*
- 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).