HAVEN Work Experience Placement Exercise

This is an exercise to familiarise yourself with the work that we are doing with HAVEN (at least the brain-based stuff) and to provide guidance for your data analysis and presentation.

**What is the HAVEN project?**

To help with the analysis that we will be working on, it makes sense to understand what the study (HAVEN) is trying to understand (at least in part).

Essentially, we are looking to understand how the brain changes over time, and how memory function is affected by these changes. We look to understand these changes by using functional magnetic resonance imaging (fMRI) whilst older adults complete an episodic memory task.

Your presentation should start by giving a brief overview of this part of the HAVEN project.

Here are some points which you can briefly cover in your intro:

1. What are we interested with understanding? Here are some hints:

* What is episodic memory? Why are we looking at that specifically?
* How does episodic memory change as we age? Why?
* What are we making people do in the scanner? Why did we choose for them to do this specifically?
* How can we measure how ‘well’ people do at the task?
* What factors do you think will affect this? (age being the obvious one)

1. What brain regions are we interested in and why?

* What brain regions are involved with episodic memory?
* How do these brain regions change over time?

1. What are the methods that we are using? Why have we chosen these particular methods?
   1. What is MRI? What is fMRI? What is the difference?
2. What is the importance of our research in the long term?

* Think about what we can do in a longitudinal study

You should have access to several files, including:

Text, application, table

Description automatically generated

The HAVEN information sheet and PowerPoint provides information for you to work with, but a simple Google will probably be the best method. The HAVEN webpage may also be useful: <https://research.reading.ac.uk/cinn/take-part/haven/>

* If you decide to have a read through these, just focus on the brain/psychology part (you don't need to worry about platelets!)

**Basic research/data analysis in R**

Once you have understood what HAVEN is, you can work on some of the data! I have attached a file containing mainly demographic data (e.g., Age, Sex, Height, Weight) but also whole-brain lesion data as well. This is contained within an R project which is a collection of files associated with some analysis/project that you are running.

I will help to set up the R project, but here are a few tasks to start with:

* Load in the “participant\_data.csv” file
* Assign the “participant\_data.csv” file to a variable
* Print out the current directory in the R console
* Run some basic commands such as *head* and *summary* on the data

Before you work on the data, ask yourself:

1. What questions would you be interested in understanding? Why? What are the expected results that you would (hopefully) see?

We are creating some basic hypotheses before you do your analysis. This is a basic principle of science, as we use our existing knowledge to predict the results and experimentally validate them using our research.

* Some of these may be straight forward (e.g., height and weight), but make some hypotheses about the lesion data. Do some research into the effects of lesions on …

Once you have thought of some questions, think about how you are going to test them using the data? Maybe you can visualise them? And then perform some statistical tests? Which graphs work best for each of your questions? Which statistical tests are you going to use?

Once you have an idea of what your research questions are, let me know and we can work on them together!

Using ChatGPT to write code in R

You are most likely to have never programmed before. Up until a few months ago, that would have made writing your own code to do analyses in R very difficult. But, in December 2022 OpenAI released ChatGPT. ChatGPT is incredibly useful for a number of things, but importantly it will write code based upon what you tell it using words. This is quite incredible.

The thing with ChatGPT is that the more specific you are with what you give it, the better results it will churn out. For example, let’s say you loaded in the ‘participant\_data.csv’ file, assigned it to the ‘participant\_data’ variable, and wanted to plot height vs age. This is the ‘prompt’ that I would give ChatGPT:

Graphical user interface, application, Teams

Description automatically generated

This is a simple example, but you can use the same principle to create any graph in R.

I will be available to help, but if you want a hint towards things that you can look at, and potentially to create the code yourself, you can open the “haven\_data\_workbook.html” which is a file containing some of the things that I would look at, and the code to run them.

Table

Description automatically generated

This should take up most of the day, but if you have time at the end, also consider some things that we could include in our study which we have missed:

1. What other untested relationships could we look at in the future? What are some of the limitations of our study?

**Bonus material: Reaction times and accuracy**

If we get this far in the tutorial, then I will go through some additional data gathered during the task that people completed whilst in the scanner. Essentially, we had people watch a video (a 30 minute episode of a sitcom) at about 8am, and then at about 10am they were shown in the MRI scanner two scenes from the episode, to which they had to select which scene came earlier in the video.

Graphical user interface

Description automatically generated

You will already understand what this task is trying to measure (episodic memory) and whilst we are not expecting differences in accuracy within the group, it could be something of interest.

* Why are we not expecting significant differences with accuracy across the group?

To analyse this data, you can load the ‘rt\_data.csv’ file into R. This gives information regarding the percentage accuracy of participants, as well as the time taken (on average).

Here are some questions to work on:

* Are there differences in the accuracy/reaction time across age/gender?
* Is there an influence of the lesion data on either of these?

What might reaction time tell us about people’s brain activity? But, what might be a reason to not ‘trust’ this particular measurement?

**Simulating a dataset**

There are many scenarios in which we need data which we don’t currently have. Take for example the case where you are wanting to predict what types of outcomes you expect to see from a planned experiment. Without any data, you can’t understand these outcomes.

That’s where we can essentially ‘make up’ the data, again using ChatGPT.

Take the following example, where I wanted to simulate a metacognition wager task on a group of people with high and low metacognition. You can see the code in the file wager\_meta\_task.R.

A screenshot of a computer program

Description automatically generated with medium confidence

The prompt which I gave ChatGPT isn’t included (that prompt is me asking ChatGPT what the prompt was) but it was quite long. This is because I need to be very specific in:

* What the experiment consists of
* What the characteristics of my population are and how they behave with regards to the task
* What outcomes I would like recorded

It also is an iterative process, by no means did I end up with this in the first go!

But once you have an idea of these three rules above, you can make an effective simulated dataset.

Think about an experiment that you would like to run and create a simulated dataset.

* Create some code analysing some of the data