**DSA8022 Behaviour Analytics Practical Skills Assessment**

This is a practical skills assessment associated with the DSA Behaviour Analytics section of the Frontiers in Analytics. It is not an exam, but it has similar time-limited conditions that are associated with an exam. This will be a test of knowledge and of practical skills that are useful for you to learn, skills that complement the lecture materials presented to you during the two and a half days of behaviour analytics teaching.

The assessment concentrates on practical skills. These are the skills that are complementary to the lectures and are very similar to the exercises completed in class. The scripts will be quite short and instructions and some example code will be provided to allow the completion of the scripts. Each of these four questions will be approximately evenly weighted to make up the remaining marks with 14, 14, 14, and 13 marks available for each of these questions. Four skill areas will be examined and you will have to answer four questions from the skill areas for the assessment.

As part of learning these skills it is also useful to be aware of some of the commands in the Tidyverse set of packages. In particular being able to do data wrangling and change the form of data using the dplyr package and being able to load in data and install packages in Rstudio is are important general skills required to answer these questions. I also encourage you to put lots of comments into the files to explain what you are doing at each stage of the code you produce so that I can see that you understand what you are doing. Marks will be allocated on the basis of your ability to complete the task and on the ability to explain how you have done this, the more you comment your code the more easy it is for me to understand and reward what you have done.

A first skill is the ability to run and complete an inter-rater reliability calculation. Two of the most important of these are Krippendorf’s alpha and the intra-class correlation. It would be useful if you could practice these skills and be prepared to do one or both of these calculations on a dataset that will be provided during the assessment, some advice will be given on the nature of the dataset and the research goals that will allow you to make decisions. The Hallgren paper contains the best account of the decisions that need to be made for an intra class correlation. Also, it would be useful to practice and see if you can run a Krippendorf’s alpha test.

A second skill is being able to develop and use R functions that are related to physiological measurement. We spent some time going through some R skills and you were provided with a full R Shiny application code that allows you to examine and develop these skills – the Physio\_dash.R shiny web app script. There are functions in this script that allow you to create simulated physiological code and to print them out as dynamic plots. The complementary practical skill that it is desirable that you develop is the ability to use this code outside of the web app in which it was initially placed. In other words, to be able to adapt the code to be used as functions in isolation.

A third skill is to be able to run models on these kinds of codes. As was discussed in class the models are almost always non-linear in these circumstances – therefore generalized additive models and generalized additive mixed models are very useful and a practical skill is knowing how to implement these. In addition, it is useful knowing how to implement traditional linear models in these respects.

A fourth practical skill that would be useful would be the analysis of sentiment in relation to a textual corpus. We did an exercise of this nature in class where we looked at the sentiment from a variety of news outlets and classified them according to the Warriner sentiment scores for words along the Valence, Arousal, Dominance levels.

If you are to take the questions and this document as a guide to the practical skills we wish you to acquire to complement the lectures then you should be very well placed to revisit those skills in the assessment and perform well in the assessment. Some example questions are provided below.

**DSA8022 Behaviour Analytics Questions – Example Questions**

Complete each of these questions and return the requested materials through QOL. The marks are distributed approximately evenly for each of the questions with 14, 13 and 13 marks available for each of the three questions respectively. You may use your notes and the slides and the papers to refer to as required.

1. **Behavioural Measurement Data Wrangling and reliability.**

Some psychologists have used a coding scheme to assess the level of expressivity in people and they want to know if raters can reliably follow the scheme. Load in the Example Data and decide whether Krippendorf’s alpha or Intra-class Correlation is the more appropriate statistic to use. The irr package help file, the Hallgren (2012) and the Shrout and Fleiss (1979) journal articles should help you in making these decisions. Then wrangle the data into the appropriate form and report either Krippendorf’s alpha or your chosen icc statistics summary – if you choose icc say why you made the choices you did. Upload the R script you used to wrangle the data and the output from your chosen statistic to QOL.

The data can be found in the ExampleData1.csv file.

1. **Body measurement**

Adapt some code from the Physio\_dash.R shiny app to create a function and code that allows you to create some breathing rate simulated data. Then adapt some more code again from the Physio\_dash.R shiny app, using the dygraphs and xts packages, to create a script that plots a dygraph plot and print the dygraph plot of the simulated data.

Write the simulated data to a datafile so that it can be used to check the graph.

Submit an R script with the code you used, a pdf of the plot and the datafile of your simulated data.

1. **Statistical modelling**

Using the mgcv package take the simulated breathing rate data and create a linear model and a GAM model using the lm() command and the gam() command respectively (the gam command is part of the mgcv package). Report the summary output of the lm and gam models

Submit an R script with the code you used and with the reported statistics in the comments at the end of the script.

1. **Sentiment Analysis**

The TweetCorpusPolarity.csv file contains nearly 6000 tweets. About half of these tweets have been given a polarity rating of positive and the other half negative. Using the Warriner ratings in the Ratings\_WarrinerVAD.csv file and the TweetCorpusPolarity.csv file based on the in class exercise produce a ggplot of each of the valence, arousal and dominance dimensions from the WarrinerVAD file. How do these relate to the polarity? Submit the code you created to achieve this as well as a image file of the plot.