Title: Linear Mixed Effects and Multiple Regression Analysis of Fitbit Data and PANAS survey data from LifeSnaps Study

Student Name: Fiona Hegarty
Supervisors: Professor Thomas J. Naughton

Abstract:
This presentation gives an overview of the findings of our analysis on the LifeSnaps dataset [1] – specifically the PANAS survey and Fitbit data collected in the LifeSnaps study. LifeSnaps was a study funded by the European Union's Horizon 2020 research programme and ran for two separate 9 week (63 day) periods in 2021 and 2022, with 71 participants drawn from four separate European countries. During this study participants wore a Fitbit watch that tracked activity, sleep, heart rate and other biological signals, they also completed a PANAS mood survey once per week.

An analysis was performed on the data both at the group and individual levels using PCA and the resulting components were use in mixed effects and multiple regression models, as appropriate. At the group level this dataset is a repeated measures dataset where multiple data points are collected over time from an individual. Standard multiple regression assumes that the data is independent. This is not the case for data collected over time from the same individual. Additionally there may be grouping effects on the individual. Mixed effect regression is best suited to data of this type as it can handle repeated measures from different individuals. PCA was used as tool to uncover the variance in the dataset and to relate this to mood. It was difficult to uncover relationships between the dataset and mood. This was not a straightforward analysis. At the group level most of the variance in the dataset was related to activity, however there were weak correlations with mood. To find meaning in this dataset we explored two types of models – one that used polynomial time and one without time that used previous mood features as predictors, both models had other features in the model. Although features relating to activity are the only ones that were found to be significant in the models the resulting models are a very poor fit for the data.

The individual level analysis focused on the top eight participants with more than 50 datapoints over the study. Each was analysed separately. The data in this case are multiple timeseries and as such temporal autocorrelations in the data need to be addressed in the resulting models. PCA was also used to understand the variance in each dataset – in most cases the most variance in the dataset was related to activity however this could not always be tied directly to mood. A cross-correlation analysis was performed at the individual level to find potential features that could be used in a multiple regression model. Given the small dataset size at the individual level, PCA was essential in this analysis. Current and lagged features were run through PCA and the resulting components modelled in a multiple linear regression once the temporal autocorrelations in the data were understood and modelled. The results will be discussed and presented at the postgraduate conference.

One aim of this analysis was to confirm previously published research findings into the relationship between mood and activity, another was to confirm that mood can be modelled at the group level using only data from an over-the-counter wearable. Additionally, the results of this analysis will be used in the next stages of this Phd research.

[1] Yfantidou, S., Karagianni, C. & Efstathiou, S. e. a. LifeSnaps, a 4-month multi-modal dataset capturing unobtrusive snapshots of our lives in the wild. Sci Data 9, 663 (2022).