# Drug Use Based on Different Personality Traits

Amna *ejam@itu.dk*
Chrisanna Kate Cornish *ccor@itu.dk*
Christian Margo Hansen *chmh@itu.dk*
Zainab Khudoir *zakh@itu.dk*

## I. Data

*The data set and early thoughts*

We will be working with the Drug consumption (quantified) Data set. This data was collected by online survey for this paper in 2015. It has 1885 responses.

There are seven personality traits for which participants are assigned a score, which have been standardised by the authors. There is also some demographic data such as age, education level, etc. This is presented as categories/binned data. We note that the age distribution of the participants is skewed towards the younger end, generally educated, from the UK/US and white. There is an even balance of male and female participants

There are 18 drugs, both legal and illegal, plus one fictitious drug, and participants were asked when they last used each one. There are seven time periods: "*Never Used*", "*Used over a Decade Ago*", "*Used in Last Decade*", "*Used in Last Year*", "*Used in Last Month*", "*Used in Last Week*", and "*Used in Last Day*".

We propose using this to create either binary classes, i.e used in the last year, or more than a year ago, or possibly 3 classes. We need to look at how age and drug use time periods are linked as this could be a confounding variable (fx. someone in 18-24 age group is unlikely to select the "over a decade ago" option).

The fictitious drug only has 3 people who claim to have used it in the last year. The authors claim this was included to identify people who were 'over-claimers'.

*Potential Research Questions:*

- Evaluation of the risk associated with substance use for each drug, offering valuable insights into the factors contributing to an individual's likelihood of being a drug consumer.
- Can personality data be used to predict whether someone is likely to be actively (within the last year) use certain drugs?
- Is there a different personality profile to the users of different drugs and can this be used to predict drug usage (historic or current).
- Is the personality profile of people who have never used illegal drugs different from those who have at some point in their lives. Does this also hold for legal drugs (very very few here though, as coffee, chocolate etc are included)
- Can we predict whether a younger user is likely to continue using particular drugs in the future (might not have the data for this...)

## II. Methods

### A. Bayesian Inference Modelling

We propose using a Bayesian inference model. We will calculate prior probabilities for drug use based on the dataset, and then estimate posterior distribution.

### B. Multi-hypothesis testing

We will test multiple hypotheses by trying to predict each drug separately based on personality scores (and possibly also age, educational level etc).

$H_0$ : Data doesn't predict drug use

$H_1$ : Data predicts usage of *DRUG*$_1$ in the last year

$H_2$ : Data predicts usage of *DRUG*$_2$ in the last year

$H_3$ : Data predicts usage of *DRUG*$_3$ in the last year...

### C. Cross Validation/Confidence Intervals

We can conduct cross validation to evaluate the result of our models.

### D. Gaussian Mixtures modelling?

## III. Questions to discuss at supervision:

- Does this make sense?
- Does it cover the requirements, do we need 4 methods, or is 3 sufficient?