ASA DataFest 2021 Summary Paper

The Data Eaters
Yi Heng Wang, Yaqi Shi, Laiba Zaman, Nina Tan

Our team analyzed the information based on the Germany dataset. We worked on the question "What are the contributing factors of prescription drug abuse?" and we focused on the following two sub-questions:

- Which drugs are used, and which ones are used in combination?
- What are the demographic profiles that contribute most to drug usage?

Which drugs are used, and which ones are used in combination?

Our analysis of drug categories focused on the data related to pain relievers listed from question 11 to 37 (from Fentanyl to THC/Cannabinoid) in the questionnaire. Using a bar chart of the frequency of usage of each drug type, we found out that although drug usage is not very frequent for most drugs, Codeine is surprisingly widely used, with 42.84% of the survey takers using that drug. By conducting a confidence interval estimation, we are confident in a 95% level that the proportion of Codeine users from the population targeted by the survey is in the range [42.05%, 43.6%].

We also used a correlation matrix on the drug types to determine whether any pair of drugs tends to be taken together. Overall, we observed low positive correlation among all pairs with maximum being 0.34. This indicates that there is little association between the usage of any two drugs and the fact that a survey taker has used a particular drug is independent to the usage of another drug.

What demographic profiles contribute most to drug usage?

To measure drug usage, we relied on the metric DAST_CAT in the DAST-10 test. It is a discrete random variable measuring of drug involvement that ranges from 1 ("No drug involvement reported") to 5 ("Severe level") of drug usage. It is a summary value for the number of "Yes" responses to ten questions about drug usage in the same test.

As people get older, it is less likely misuse the drugs. We plotted the variable over different age groups and we saw an increasing trend of the proportion "No Drug" group. We also tested the hypothesis that drug involvement (measured by DAST_CAT) and age group are independent. We divided the survey takers into those whose DAST_CAT score is 1 and those whose DAST_CAT score is greater than 1. Using a Pearson's Chi-squared test for independence, we obtained a p-value < 2.2e-16 that suggests that the association between young age and drug involvement is in fact significant.

We also consider the variable of education. We plotted the average of the DAST_CAT score over different education groups. We noticed that bars look similar over different education group. To further confirm the preliminary result, we fit a logistic regression model and we noticed that all the parameters of the categorical variables are insignificant. Based on the result, we conclude education has little impact on drug involvement.