

# **THE IMPACT OF PERSONAL QUALITIES ON DRUG CONSUMPTION**

Lubomir Dlhy  
Anastasiia Solomiia Hrytsyna

Course: Data Science Programming

Semester: WS 2021/2022

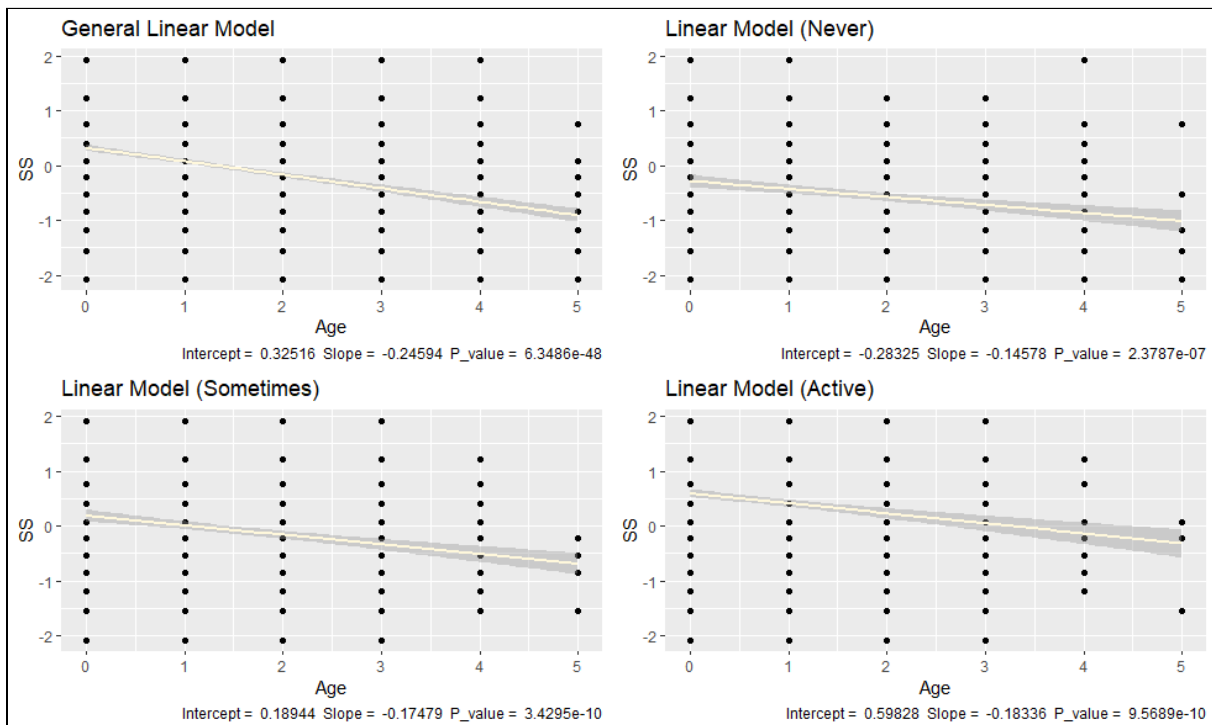
Supervisor: doc. Mgr. Michal Kovac, MSc., PhD.

Year of study: 3

Addiction is a serious problem that affects millions of people in the modern world. **Drug addiction**, also called as a substance use disorder (SUD), is a disease that may badly affect a person's brain and behavior. It is extremely important to know how SUD can impact your life, what harm it may cause to know how to cope with the abuse in case it is necessary.

There are a lot of **different reasons why** people start doing drugs. Someone can start with experimental narcotic use in public, someone take them for stress relieving or as a medicine and for someone the drug use may become even more frequent. In our study we decided to examine not only the simple explanations that we mentioned above, but also **whether there is any dependency between the personality features and the possibility of doing hard drugs during the lifetime**.

In our survey we used the information from **1884 participants** of several social groups. The dataset contains information about some **personal characteristics** (level of neuroticism, extraversion, openness to experience, agreeableness, conscientiousness, impulsiveness, sensation seeking (SS) measured by NEO-FFI-R test) and **periodicity of using** different kinds of soft and hard **drugs**. 34% of all members were **youth** (18-24 y.o.), which actually might have an impact on the future results. Almost 40% of people from our survey have **left school at an early age, have not finished any university or obtained a professional certificate**. In addition, 55% of all participants have **UK origin** and the other 30% - **USA**, which makes these two groups the dominant ones.



**Figure 1. Linear Models For 3 Main Groups And The General One**

Afterwards, we decided to remove from our survey all **members who used to take a fictitious drug Semeron**, as in case they were suspected for being deceitful - their answers may be uncertain. Moreover, information about **soft drugs and some personal qualities was also erased**, as we are not going to analyse them.

We assume that the level of SS has a significant influence on a person's behaviour. Literally thrill-seeking is the tendency to hunt new sensations, feelings, and experiences (including drug consumption). **The higher SS is, the higher possibility to start doing something new, start doing drugs.** That is why we divided the whole dataset into 3 groups:

1. People who had **never** done hard drugs during their lifetime (the coefficient in the dataset was 0);
2. People who are still doing drugs (the coefficient in the dataset was 4, 5, or 6; we suppose that there is a small possibility that such members do drugs for the first time and at the same period have participated in the test. That is why we consider such individuals as **currently addicted - the 'active' ones**);
3. People who used to do drugs long time ago (the remaining group of people - **the 'sometimes' group**)

In the next step we calculated a **linear model** for each group and described the dependency of SS on the age of each group member. All of the models are shown on Figure 1 above. It is obvious that the graph is quite inaccurate, as both **SS and Age are categorical** (discrete) values and points (that represent the people) are distributed on the whole graph. However, it is noticeable that the **P-value is extremely small** (less than 5%), so  **$H_0$  (that  $K = 0$ ) has to be rejected**. Therefore, both **SS and Age are dependent**. We also might point out that the average SS value in an active group **(0,45) is a lot higher** than the one in other groups (sometimes\_group: -0,09, all\_data: -0,007). On the contrary, the average SS value among people who never used drugs **(-0,54) is much smaller** than the others.

Afterwards, we decided to check our model with a **repeated K-fold cross-validation method** for how precise our models are. Unfortunately, it did not show as good results, as we expected, but the outcome is next:

- RMSE (root-mean-square deviation) is **80-90,9%**
- MAE (mean absolute error) is **64,4-72,7%**
- $R^2$  error is **6-11%**

Both RMSE and MAE values must be less in order to make a better model. Instead a proper model should have a high value of R-squared error. Our results are quite inappropriate as both SS and Age are categorical values, so it is complicated to create one correct linear model and function that fits all distributed values.

To sum up, sensation seeking has also an impact on drug consumption as it is an adrenaline-filled activity. The higher level of SS, the higher is the probability of becoming addicted. It is expected that the SS level decreases the older a person gets. Linear model graph is a great representation of all of these facts. Cross-validation method does not show the good results of evaluating our models as we have used categorical values. That is why these methods of data analysis can not be used.