

Review of a final project for team A (Timur Abragimovič, Jakub Benetin, Čestmír Vejmola)

By team B - Vladyslav Babyč, Tomáš Klouček,
Ondřej Pelikán, Dejan Ribarovski

January 2025

1 Summary

Our team found the presented topic of the semestral work highly interesting and believes that it holds significant potential for practical applications in the future. Due to a highly specified topic, our team had to reread some of the parts of the work several times to become more familiar with the topic. We also needed to search for additional information about the subject presented.

After becoming more familiar with the topic, we were able to evaluate the presented report and all the statistical methods used in this work. From the introduction, we were able to get basic information about the presented dataset and the intent of the work. We greatly appreciate the presence of the section *Validation of Cross-over design*, as it shows the quality and independence of the collected data. Also we were pleased to see very detailed descriptions of all the bands and reflection on the given factors such as arena, week, etc. The section which focuses on the creation of the model is highly detailed and provides interesting insights into the influence which different kinds of drugs and their volumes have on the results in the EEG bands. The next section on clustering methods was well described and had interesting conclusions in which we could see a major improvement in the performance after performing the Ledoit-Wolf shrinkage, and the amount of improvement in the given method kindly surprised us.

In conclusion, the work is well organized and provides an analysis of an interesting topic. The statistical methods used are presented in high quality and in great detail. We would like to state that we consider this work to be of high quality. However, we discovered some possible downsides that we would like to point out. These points are presented below.

2 Positive sides

This project shows a higher level of citation than the usual projects of this course. It also holds very well on academic language and structure. Overall, the project sets a high benchmark for future work in this course. The source of the data set is very well documented. We especially want to recognize the description of data collection. The dataset itself, collected using EEG implants, provides a very interesting insight into the influences of drugs, compared to techniques such as video collection, blood tests, etc. The use of EEG data is also very well done, with detailed analysis of the entire spectrum.

Sections 3, 4 and 5 bring significant results. Section 4 shows interesting effects of drugs (even though we would appreciate a deeper analysis of the created model), Section 3 demonstrates the influences of different variables in the dataset on each other very well with a comprehensive review, and the last LDA model in Section 5 differentiates reasonably well between the different drugs.

3 Negative sides

We did not fully understand the study design as it is described in only two sentences from which the design is unclear (at least to us):

"In each run, one substance was selected for each group and injected in a cross-over design to \pm eighteen rats over ten weeks. Each rat was given each of the ten substances just once within ten weeks, in randomized order." Section 2.3

In Section 4 the authors describe choosing linear mixed modeling due to the size of the dataset. Although this decision yields good, interpretable results (the graphs on page 13 are very informative), the reasoning seems odd, as well as the decision not to discuss the resulting p-values. While the discriminant analysis reviews different models, we would like to see this discussion in the Section 4 as well.

4 Questions

1. Is it possible that the strong time dependence on brain activity (discussed in Section 3.6) was caused by the effect of drugs itself?
2. Why was there such a big spread in the number of recordings for each drug (only 11 for MDPV while 35 for cocaine)? Does this mean that the data used for discriminant analysis was also unequally distributed across the groups? If so, do you think it affected the results?
3. With ethanol needed for solubility, did you account for the potential effect of ethanol on EEG readings?
4. Even though it is far outside the scope of this project, do you have insight into how well these results translate to humans?