Group Contract – #Group 28

1) Goals

1. What are our team goals for this project?

Our team will be working with dataset 6: Diabetes. An important component of our project is to distinguish between models created for explanation versus those for prediction, focusing on criteria for choosing, evaluating, and interpreting models in each context.

We will select suitable methods, such as regression, generalized linear models (GLMs), sample size estimation, and variable selection, and discuss the advantages and disadvantages of each. One of the goals is to correctly interpret computer output concerning the posed statistical questions, ensuring our analyses are appropriate for the intended audience.

Additionally, we will identify and verify the assumptions required for each method to yield reliable results, addressing the consequences of mismatching methods to questions or data types. By adopting this comprehensive approach, we aim to derive meaningful insights from the data and enhance our understanding and application of various statistical techniques.

2. What do we want to accomplish?

Ultimately, we want to derive meaningful insights from the dataset, enhancing our understanding and application of various statistical techniques to address complex data science problems.

3. What skills do we want to develop or refine?

With this project, our group hopes to refine our skill in interpreting various types of plots and to apply these interpretations to real life data sets. We also hope to improve in group work in R, which has not been done in class yet. To round it all out, we hope to refine our skills in resolving conflict regarding the project should the situation arise.

4. What do we expect of one another regarding attendance at meetings, participation, frequency of communication, quality of work, etc.?

We expect all group members to respect each other's efforts. This means attending every meeting or giving ample notice if unable to attend, and coming prepared with any pre-work completed to fully participate. Communication should primarily occur during meetings, which will be held at least once a week, with increased frequency as the project deadline approaches. We also expect timely completion of work to the best of our abilities, meeting pre-set standards.

Everyone is expected to put in the same amount of effort into the group project, however this can be distributed throughout the project. For example, if one were to take a lighter role in one phase of the project, we expect them to contribute more in another phase of the project.

We expect group members to seek help when needed on material relevant to the project, whether through other group members, the Internet, or office hours. However, this should only be done after one has tried to answer their question alone or has tried different approaches to the problem—in short, other group members are not a crutch, but a resource.

We expect group members to adhere to academic honesty guidelines, citation requirements and other conventions of the sort.

(2) <u>Policies & Procedures</u>

- What rules and procedures can we agree on to help us meet our goals and expectations and deal with members who do not fulfill their part in the contract?

Inform the group if you're unable to complete your work or attend a meeting. We understand if exams or other commitments prevent attendance, but members must attend at least 80% of meetings. If meetings need to be rescheduled, we expect members to give the rest of the group 24h notice and find alternative means.

The group should discuss and set expectations for assignments beforehand, ensuring the work aligns with the overall project goals.

(3) **Consequences:**

- How will we address non-performance regarding these goals, expectations, policies and procedures?

Our group has agreed to implement a 3-strike system in the unlikely case that non-performance arises. Before a group member gets a strike, discussion among the group should happen and reason if a group member's non-performance was preventable. The strikes and their respective actions are as follows:

Strike 1: Issue a warning to the group member and tell them to do their part, giving them assistance or guidance if needed.

Strike 2: Hold a meeting to discuss ways we can get the non-performing group member back on track, so that work on the project can progress.

Strike 3: Inform the professor or a TA about the group member's non-performance.

Through communication, we hope that we will not have to issue any strikes at all.

(3) **Data:**

https://www.kaggle.com/datasets/whenamancodes/predict-diabities?resource=download

The Predict Diabetes dataset, sourced from the National Institute of Diabetes and Digestive and Kidney Diseases, includes 9 variables (Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age and Outcome) and 768 data points. It is designed to predict diabetes diagnoses. The dataset focuses on all patients here who are females at least 21 years old and of Pima Indian heritage. The response variable is Outcome, a binary indicator showing whether or not the individual has diabetes.

Sign:
Bowen Zhao
Xuanya Liu
Yulu Duan
Anne Nguyen