**Plan for PSQF 6270 HW5: Generalized Linear Models on Your Own Data (2 points)
Due Monday 3/31/2025 by 11:59 PM under “assignments” in ICON

Please submit this document in an editable format (e.g., .docx or .rtf extension)
using this file-naming convention: PSQF6270\_Lastname\_Firstname\_Plan**

The goal of HW5 is for you to practice conducting and reporting analyses using generalized linear models on real data. **Ideally these analyses would focus on data you already care about** in the context of your research or employment (i.e., in which these analyses could serve as the basis for a future conference presentation or manuscript to be submitted for publication). If needed, however, you can also use publicly available data, such as from these example sources:

Kaggle: <https://www.kaggle.com/datasets>

ICPSR: <https://www.icpsr.umich.edu/web/pages/ICPSR/index.html>

Berkely archives: <https://sda.berkeley.edu/archive.htm>

Harvard archives: <https://dataverse.harvard.edu/>

Early Childhood Longitudinal Studies: <https://nces.ed.gov/ecls/dataproducts.asp>

Competitive E-sports: <https://doi.org/10.1038/s41597-024-04364-z>

Healthy Minds Data (must submit request and be approved first): <https://healthymindsnetwork.org/research/data-for-researchers/>

Your analyses should include **4–6 variables in total**, in which 1–2 would be treated as outcomes and the rest would be treated as predictors. Please note that **you will need to have complete data for all variables** to be included in your models (unless you already know how to use software in which missing data are allowed after bringing predictor variables into the likelihood, which we will not cover until unit 6). Please select your variables with this restriction in mind.

The type of outcome(s) you have will logically dictate which type of generalized linear model(s) you use—I would recommend choosing them based on the particular variants we have covered in class, but I will allow exceptions if sufficiently motivated. Use of standard general linear models (i.e., regression, ANOVA) is not allowed. Ideally your data will include only a single dimension of sampling (i.e., not longitudinal or clustered data), but I will allow exceptions to this as well if you have had other coursework addressing these (multilevel) extensions.

Please answer the three questions below directly in this document so that I can provide feedback you with on your analysis plan:

1. Briefly describe your sample and its salient characteristics. Your description should include where your data came from, the total sample size available for your variables of interest, and any additional sources of nesting or crossing for multilevel sampling designs.

Answer:
2. Briefly describe each of the 4–6 variables of interest that will be available for your full sample. Your description should include what construct each variable measures (i.e., what each variable is supposed to represent conceptually), how it was measured (e.g., scale score, performance index like response time or accuracy), and its format (e.g., binary, ordinal, nominal, quantitative), including the number of categories or range of possible values. You may do so in a table if that is more convenient.

Answer:
3. What do you want to know with respect to these variables? Briefly describe your research questions as best you can, as well as the models you plan to estimate to answer them (words and/or equations are fine for this). I will then try to help you figure out this part as needed.

Answer: