Weekly Status Submission Template

Week Number: 6-7

Date: 10/5/2025

Student Name: Abdullah Siddiqui

Original Plan for This Week: (5 points)

The plan for Weeks 6–7 was to expand the data analysis by incorporating interaction effects between education and income levels, conduct stratified analyses by demographic variables such as sex and age group, and generate more advanced visualizations for regression results. Additional goals included evaluating model performance through ROC/AUC metrics, confusion matrices, and saving all outputs (summaries, tables, and plots) for documentation.

Tasks Accomplished This Week: (10 points)

This week, I successfully completed the advanced statistical modeling phase of my BRFSS 2023 project. I fitted both logistic and linear regression models with an education × income interaction to explore how socioeconomic factors jointly influence mental health outcomes. I then performed stratified analyses by sex and age group to understand subgroup-specific trends. All models were cleaned, summarized, and visualized using ggplot2. I also evaluated model performance through ROC/AUC analysis (AUC = 0.731) and created a confusion matrix to assess classification accuracy. All results, including coefficients, summaries, and visualizations, were exported to organized folders in my working directory.

Comparison of Planned vs Actual: (10 points)

Planned Tasks	Actual Tasks Completed
Include interaction terms in regression models	Implemented education × income interaction in both logistic and linear models
Conduct stratified analyses by demographics	Completed analyses stratified by sex and age groups
Generate polished coefficient and relationship plots	Created 5 professional visualizations including coefficient plots, grouped bar charts, and a heatmap
Evaluate model performance (ROC/AUC, confusion matrix)	Successfully computed AUC = 0.731 and saved ROC plot and confusion matrix
Save and document all outputs	Exported all model summaries, coefficient tables, and figures into dedicated folders (outputs, tables, plots)

Self-Rating of Progress: (5 points)

- [Done] Met the Planned Tasks
- [ ] Did Not Meet the Planned Tasks

All planned objectives were successfully completed. The code ran smoothly with no major errors, and the model performance results were strong. I am on schedule with the project timeline and will begin preparing interpretation notes and documentation in Week 8.

Plan for Next Week: (5 points)

- Begin drafting a detailed results and discussion section summarizing regression outcomes.
- Interpret the interaction effects and subgroup findings in context with public health literature.
- Refine figures and tables for inclusion in the final report.
- Update GitHub repository with markdown explanations, references, and structured folders for reproducibility.

Evidence of Tasks Completed: (15 points)

All code, outputs, and plots are stored in my GitHub repository:
https://github.com/abdullahs1357/SeniorThesis
Key Evidence:
- Code: BRFSS2023_interaction_week6_7.R
- Tables:

tables/logit_interaction_coeffs.csv
tables/lm_interaction_coeffs.csv
tables/stratified_by_sex_logit_coeffs.csv
tables/stratified_by_age_logit_coeffs.csv

- Plots:

plots/logit_coefficients.png
plots/lm_coefficients.png
plots/grouped_exercise_education_sex.png
plots/mh_edu_income_heatmap.png
plots/logit_interaction_ROC.png

- Outputs:

outputs/logit_interaction_summary.txt
outputs/lm_interaction_summary.txt
outputs/confusion_matrix_logit.txt
outputs/summary_run_week6_7.txt

The logistic regression model achieved an AUC of 0.731, indicating moderate predictive accuracy. Visualizations effectively highlight relationships between education, income, exercise, and mental health outcomes. This completes the advanced modeling and evaluation stages of my project.