Sandra Osei
GVA-Benzie 103116
42nd Ave, Apt 6
Allendale, MI 49401
April 16, 2025

Prof. David Zeitler
Department of Statistics
Grand Valley State University
Allendale, MI 49401

Dear Prof. Zeitler,

Self-Evaluation Letter

This semester, your statistical modeling course has been an eye-opening and rewarding journey. Each week built on the last, helping me develop a deeper understanding of data, statistical models, and the power of thoughtful interpretation. I began the course with a basic grasp of linear regression, but through the lectures, hands-on assignments, and team projects, I've gained confidence in applying a wide range of models. My original goals were to strengthen my proficiency in statistical modeling, improve my collaboration skills, and manage my time more effectively. Reflecting on the semester, I've made meaningful progress in all three areas, faced some challenges, and discovered strengths I didn't know I had. This letter is my reflection on that growth, framed through the work I did with the Auto and K-12 datasets.

In the first part of the course, I focused on simple and multiple linear regression. Learning to interpret R-squared values, p-values, and residual plots gave me the tools to assess model fit and assumptions. One particularly satisfying moment was when I modeled miles per gallon using horsepower in the Auto dataset and discovered that horsepower alone explained about 60% of the variation in mpg. It was like solving a mystery. Techniques like forward and backward selection helped me build more efficient models. In a model predicting acceleration, I saw how predictors like weight and displacement interacted, showing me the importance of balance between complexity and clarity.

As we moved into logistic regression, I learned to think beyond just prediction, focusing instead on interpreting odds ratios and understanding model quality. Concepts like ROC curves and pseudo R-squared values became familiar, and I began to see how careful variable selection and interaction terms could shift the story a model tells. One assignment involving categorical variables and interaction terms really brought this to life and emphasized the value of checking assumptions and context in every model.

The K-12 project was a significant milestone. My teammate and I analyzed data from the National Household Education Surveys to model student academic performance (mostly A's, B's, or C's). We used multinomial logistic regression and selected predictors like absenteeism,

homework hours, family help, and parent education level. We applied forward selection and tested the model's accuracy through cross-validation. Although the final accuracy wasn't high, the process gave me a clearer picture of how these variables contribute to educational outcomes. It was also my first time working with multiclass classification metrics — confusion matrices, Brier scores, and ROC AUC — which were intimidating at first but ultimately became a major learning success.

In weeks seven and eight, we explored generalized linear models. Working with Poisson regression for count data required a mindset shift from linear regression. Concepts like overdispersion were challenging, but they taught me how to adapt models to different types of data. This flexibility is something I now value deeply.

A big takeaway for me was learning how to evaluate models properly — not just by how well they fit the training data, but how well they generalize. K-fold cross-validation became a key part of my process, helping prevent overfitting. The YouTube video about "Cross Validation the Wrong Way" by Robert Tibshirani and Jonathan Taylor, made a big impression on me. It showed how easy it is to misapply validation techniques and why model evaluation should be just as deliberate as model selection.

I was also surprised by how much I enjoyed learning about regularization methods like LASSO. Initially, I was intimidated by the idea of penalizing coefficients, but I quickly saw the value in maintaining interpretability while managing complex models. This was one of the areas where I grew the most unexpectedly.

in the course, we explored nonlinear modeling. While I didn't use splines in the K-12 project, I came to appreciate the importance of visualization in identifying nonlinear patterns. Using tools like ggpairs helped me recognize correlations visually, especially in the Auto dataset. This strengthened my data storytelling skills and made my analyses clearer and more engaging.

Throughout the semester, I made it a point to follow tidy modeling practices, organizing my work, documenting each step, and ensuring reproducibility. While not one of my original goals, this became a defining characteristic of my assignments and projects. It made my work more professional and allowed me to reflect on the process, not just the results.

The K-12 project also tested my collaborative skills. I worked closely with my teammate to divide tasks and explain concepts like multinomial logistic regression and Linear discriminant analysis. I met my goal of communicating effectively, but I also realized I sometimes hesitated to take the lead. The experience helped me build confidence, and I plan to push myself further in future group projects by taking on more leadership responsibilities.

Time management was another important focus for me. Early in the semester, I submitted one assignment late it was a tough but important wake-up call. I learned to manage my schedule more proactively, using tools like Google Calendar and building in buffer time before deadlines.

This improved my consistency, though I'm still learning how to better handle weeks with heavier workloads or difficult concepts, like GLMs.

There was also a moment when my group and I mistakenly submitted the wrong file for Project Two. It was frustrating at first, but it taught me a valuable lesson about double-checking submissions, a lesson that applies well beyond the classroom and into the workplace. I've learned to take extra care when handling collaborative work, especially when representing a team's efforts.

What I didn't expect to find was a passion for visualization and data storytelling. It emerged naturally and has added so much value to how I communicate findings. I also didn't anticipate becoming so comfortable with LASSO and regularization techniques, which seemed daunting at first. These unexpected strengths have broadened my confidence and built a solid foundation for more advanced modeling in the future.

Looking ahead, I want to improve my ability to interpret complex interactions and better manage categorical variables. These are areas where I still feel uncertain, and I plan to explore them further through independent projects and additional resources. Long-term, I'm excited to apply statistical modeling to fields like education and public health, where data can drive meaningful change. I am also interested in branching into machine learning to complement the regression techniques I have learned this semester.

This course has changed how I approach data. I now prioritize curiosity, structure, and careful reasoning in everything I do. I've developed both technical skills and personal resilience, and I am incredibly grateful for the supportive learning environment you have created. Thank you for guiding us through this journey. I'm excited to carry these lessons with me into future academic and professional experiences.

Sincerely,
Sandra Osei