Food Expenditures

Homework #2

Statistics 469: Analysis of Correlated Data

Problem Background

The question of how much money people spend on “eating out” as their income rises or falls has been a question that has long interested economists. If restaurant food is a “normal” good, one would expect consumption (and spending) to increase as income increases. This relationship has implications for projections of food demand and restaurant ownership as a viable career as incomes grow or as countries develop and become wealthier. However, just because people will spend more at restaurants as their income increases, on average, the relationship is likely more complicated due to high degree of variability for high income individuals. Also of interest is how demand for “quality” and “convenience” changes as income changes, which relates to whether people spend more money on food at home vs. away from home as income increases.

The dataset *FoodExpenses.txt* contains data from the Food Demand Survey (FooDS) on $n = 523$ households income and associated expenses on “eating out”. Specifically, *FoodExpenses.txt* contains the following variables:

Variable Name	Description
Income	Annual household income (in thousands)
EatingOut	Average weekly expenditure on food not cooked at home.

To analyze this dataset, do the following:

Analysis Questions:

1. Create exploratory plots and calculate summary statistics from the data. Comment on any *potential* relationships you see between `Income` and `EatingOut`.
2. Using a homoskedastic linear model, fit a regression model to `EatingOut` using `Income` as the explanatory variable. Determine if the equal variance assumption is met. If it not met, discuss what impact the violation of this assumption could have on an analysis on the relationship between income and food expenditure.
3. Write down a heteroskedastic linear regression model (in matrix and vector form) in terms of population parameters including your specification for the variance function with `EatingOut` as the response and `Income` as the explanatory variable. Explain the meaning of any parameters in your model. Explain how statistical inference for your model can be used to answer the effect of income on food expenditure.
4. Fit your model from #3 to `EatingOut`. Validate the model L-I-N-E assumptions so you will be confident that the statistical inference you perform below will be correct.
5. Validate your predictions based on your model in #3 via cross-validation (any of leave-one-out, Monte Carlo or K-fold). Report your model RPMSE and coverage. Additionally, show your predictions and 95% prediction interval bounds on a scatterplot of income vs. food expenditure.

6. Report $\widehat{\beta}_{\text{inc}}$ along with a 95% confidence interval for the model in #4. Report any variance parameters (including the variance function parameters) along with appropriate 95% confidence intervals. Correctly interpret all intervals in context.

7. Economists with the National Restaurant Association (which, perhaps unfortunately, shares its acronym with another institution), hypothesize that a "healthy" restaurant economy should see increases of about $0.50 or more per week for each $1000 increase in income. Using your heteroskedastic model, test if the economy is NOT "healthy" for restaurant owners. State your hypotheses, p-value and an appropriate conclusion.

8. Predict how much you will be spending at restaurants for your desired income level upon graduation (meaning at your first job). Report a 95% prediction interval and interpret the interval in context.