Problem Set I

Saskia Jacob & Martin Klarmann

Winter Term 2025/26

Data For all the problems here, please use the data for the Fulton Fish Market in New York City, which is apparently one of the largest fish markets in the world. Data on sales of a specific fish ("Whiting") was collected by Kathryn Graddy, who is an economist at Brandeis University in the US.

The data were collected in April and May of 1992. Two relevant datasets are available, and the task descriptions specify which one should be used for each task. Both datasets can be downloaded here. Under "Data: Fish", you will find the "Daily Fulton Fish Market Data" and the "Detailed Fulton Fish Market Data", as well as a data dictionary ("dictionary") that provides additional documentation, variable descriptions, and measurement details.

It is also worth noting that Kathryn Graddy published some scientific papers based on this dataset. You may want to look at them, but they are of limited relevance with regard to this problem set. They may become more relevant later in this class. If so, I will then refer to them explicitly.

Document This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

Relevant R packages All tasks from this week can be completed using the R base package. You may want to additionally use other packages for visualization and data handling (e.g., ggplot2, dplyr, psych).

Deliverables and Deadline

1. A Markdown Notebook with all the code and your argumentation. We may apply this code to the original dataset, so please make sure that all changes to the dataset are programmed in the notebook (and not done in Excel etc.).
2. A PDF document with your responses to all tasks (max. 20 pages). We recommend using your Markdown Notebook document to create your PDF file.

Results need to be uploaded to ILIAS in the relevant section on Friday, 21st November 2025, at 5pm.

Task 1 "Descriptive analysis" - 5 out of 25 points Please describe the market situation for Whiting in April and May 1992 using appropriate descriptive statistics and simple plots. Please illustrate five interesting insights using appropriate graphs, etc. and provide additional interpretation. What did surprise you the most? Explain.

Please note that this task is not related to any specific part of the lecture. The key goal of this task is that you familiarize yourselves with the programming environment while drawing on your existing statistical knowledge, e.g., from the class "Market Research". A number of short videos on R functionality are available in this course to assist you in completing this task.

For this task, please use the "Detailed Fulton Fish Market Data."

Task 2 "Nonlinear regression" - 12 out of 25 points A relationship that is of focal interest for many retailers is the association between price and demand. Please identify and remove outliers, if applicable, using the methods introduced in class. The outlier analysis should be conducted from three perspectives. First, examine potential outliers in the dependent variable (demand/quantity). Next, identify outliers in the independent variable (price). Finally, consider both variables simultaneously to detect cases where extreme values occur in combination.

Next, compare three possible functional forms of the price–demand relationship: one linear and two non-linear (polynomial and logarithmic). Please provide a plausible justification for the structure of each model. Please compare the accuracy of the different models using relevant model selection criteria (such as adjusted R-squared, AIC, or BIC), and select the model that best predicts the demand for fish. Clearly explain your choice and provide reasons based on the criteria used. Follow the procedures suggested in class. Interpret and discuss the implications of your findings.

For this task, please use the "Daily Fulton Fish Market Data". This task explicitly refers to chapters 1.1, 1.2, 1.4, and 1.5 of the lecture.

Task 3 "Moderated regression" - 8 out of 25 points Does price sensitivity of individual customers depend on context characteristics? Please use moderated regression to test for multiple context variables from the "Detailed Fulton Fish Market Data" whether they have an effect on a (linear) effect of price on demand. For this reason, please develop three hypotheses based on plausibility considerations and apply the process to analyzing moderated effects outlined in the lecture. Based on your analyses: Can you identify a target group for Whiting that is particularly attractive from a seller's perspective?

Note that in the "Detailed Fulton Fish Market Data" set, quantities may differ between customers due to different sizes of customer businesses. For this reason, you may want to find a way of standardizing quantities within customers. For instance, you could look at quantity deviations from the customer average. Hence, in a first step, you will need to come up with a solid (linear) specification of your price-demand model for this data.

For this task, please use the "Detailed Fulton Fish Market Data." Make sure that you go through all the steps of moderated regression described in the course. This task explicitly refers to chapters 1.1 and 1.3 of the lecture. You will also need some parts of chapter 1.2 (especially regarding the F-Test for R-squared differences).