Assignment 04

Evidence and Model Selection

This assignment is intended to have you build your understanding of using information criteria for model selection. Turn in a printed document that includes your responses to each of the questions on this assignment. This assignment is worth 13 points.

Please adhere to the following guidelines for further formatting your assignment:

- All graphics should be resized so that they do not take up more room than necessary and should have an appropriate caption.
- Any typed mathematics (equations, matrices, vectors, etc.) should be appropriately typeset within the document.
- If you are using Markdown, all syntax should be hidden (i.e., not displayed) unless specifically asked for. Any messages or warnings produced from loading packages should also be hidden.

For this assignment, you will use the file *wine.csv*. This file contains data on 200 different wines. These data are a subset of a larger database ($n = 6,613$) from Wine.com, one of the biggest wine e-commerce retailers in the U.S. It allows customers to buy wine according to any price range, grape variety, country of origin, etc. The data were made available at http://insightmine.com/.The variables are:

- `wine`: Wine name
- `vintage`: Year the wine was produced (centered so that 0 = 2008, 1 = 2009, etc.)
- `region`: Region of the world where the wine was produced
- `varietal`: Grape varietal (e.g., Cabernet Sauvignon)
- `rating`: Wine rating on a 100 pt. scale (these are from sources such as *Wine Spectator*, the *Wine Advocate*, and the *Wine Enthusiast*)
- `price`: Price in U.S. dollars

You will be using these data to examine several different predictors of wine rating (a measure of the wine's quality). The literature has suggested that price of wine is quite predictive of a wine's quality. You will be carrying out a replication study (using a different data set) of a study published by Snipes and Taylor (2014).

Preparation

Read the article Model selection and Akaike Information Criteria: An example from wine ratings and prices.

Model Selection 1

1. Fit the same nine candidate models that Snipes and Taylor fitted in their analysis, but use the *wine.csv* data. In these models use wine rating (`rating`) as the outcome. Create a table of model evidence that includes the following information for each of the nine candidate models. **(2pts.)**
    - Model
    - Log-likelihood
    - K
    - AICc
    - ΔAICc
    - Evidence Ratio
    - Model Probability

Use this table of model evidence to answer Questions 2–5.

2. Use the AICc values to select the model with the most empirical evidence. Write the fitted equation for this model.
3. Interpret the model probability for the model with the most empirical evidence.
4. Interpret the evidence ratio that compares the two models with the most empirical evidence.
5. Examine the tenability of the regression assumptions (linearity, independence, normality, homoscedasticity) for the model with the most empirical evidence. Comment on the tenability of these asumptions and provide any evidence (graphical or numerical) you use in this endeavor.

Model Selection 2

6. Re-fit all nine candidate models using the natural logarithm of wine rating as the outcome. Also log-transform price (using the natural logarithm) in all of the models. Create the same table of model evidence you did in Question 1, but for the re-fitted models. **(2pts.)**

Use this table of model evidence to answer Questions 7–11.

7. Using the model that has the most empirical evidence, examine the tenability of the regression assumptions (linearity, independence, normality, homoscedasticity). Comment on the tenability of these asumptions and provide any evidence (graphical or numerical) you use in this endeavor.
8. Based on previous literature, Snipes and Taylor hypothesized that price was an important predictor of wine quality. Based on your analyses, is price an important predictor of wine quality? Justify your response by referring to values reported in the table of model evidence.
9. Does the empirical evidence support adopting more than one candidate model? Justify your response by referring to values reported in the table of model evidence.
10. Does the empirical evidence from the Snipes and Taylor analyses support adopting more than one candidate model? Justify your response by by referring to values reported in their table of model evidence.
11. Based on your responses to the last two questions, which set of analyses (yours or Snipes and Taylor) has more model selection uncertainty? Explain.