Assignment 02

Evidence and Model Selection

This assignment is intended to have you build your understanding of using information criteria for model selection. Use R Markdown to produce an HTML (or PDF) file that includes your responses to each of the questions on this assignment. 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. Learn how to do this in a code chunk using knitr syntax.
- Any typed mathematics (equations, matrices, vectors, etc.) should be appropriately typeset within the document using Markdown's display equations. See here for some examples of how mathematics can be typeset in R 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 each question, specify the question number using a section header. Submit your HTML (or PDF) file and your RMD file via email that you send to both Jonathan and Andy.

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 using wine rating as the outcome. Create a table that includes the following information for each of the nine candidate models.

   - Model
   - Log-likelihood
   - K
   - AICc
   - ΔAICc
   - Evidence Ratio
   - Model Probability

2. By referring to the evidence ratio of the second-best model, which candidate model(s) you will adopt? Explain. (Hint: Back-transform the log of the evidence ratio.)

3. Interpret the model probability for your adopted model.

4. Write the fitted equation for your adopted model.

Model Selection 2

5. Re-fit all nine candidate models using the natural logarithm of rating as the outcome. Also log-transform price in all of the models. Create the same table you did in Question 1, but for the re-fitted models.

6. Use the AICc value to select the best log-transformed model. Write the fitted equation for your adopted model.

7. Examine the tenability of the regression assumptions (linearity, independence, normality, homoscedasticity) for the model you adopted in Question 2 and for the model adopted in Question are satisfied for this model. Based on these analyses, which of the two adopted candidate models (the un-transformed, or log-transformed model) should be adopted and interpreted. Explain. Provide any evidence (graphical or numerical) you use in this endeavor.

Discussion of Results

8. Using all the evidence from the table of log-transformed models you created in Question 5, does the empirical evidence support adopting more than one candidate model? Explain.

9. Based on previous literature, Snipes and Taylor hypothesized that price was an important predictor of wine quality. Using evidence from the table of candidate models you identified in Question 5, is price an important predictor of wine quality? Explain.

10. Did you select the same best model as Snipes and Taylor? Explain. Also comment on whether the uncertainty in this selection was the same as that in the Snipes and Taylor paper.