

Final Project AIC Version

MBA 694

Due: July 29, 2022 at 5:00 pm

AIC Version

This version is if I indicated your research question fit an AIC Model Selection path rather than a hypothesis testing approach.

Instructions

The final project should be written up just like a homework assignment. The difference is that I will not tell you what the question is, or provide more than just a rough structure for the paper. You are responsible for showing me that you have learned to write a paper (in my preferred format) through all 8 homework assignments.

You are also responsible for deciding how to split up the work between you and your partner. However you decide to do it, you are *BOTH* responsible for the *whole* paper/grade. So make sure to edit each others work! It is best practice in the business world to have someone else edit documents, so this is not a criticism it is a fresh set of eyes.

The project will be out of 150 points.

1. (10 points) Using the recs2015 dataset determine a research question of interest to you, with the following limitations:
 - a. It **MUST** be a question not previously addressed in a homework assignment or quiz. I will have you submit your variables and research question to me for a quiz score by July 22, at 8:00 am. Easiest way to change the question is to change the response variable.
 - b. It **MUST** have at least three explanatory variables (sorry no Chi-Squared research questions allowed.)
 - c. You **don't** have to restrict the data to the subset used for the homework assignments. But if you choose to use the full dataset, know that the categorical variables are coded numerically and you will need to refer to the codebook Excel file to understand what the codes represent.
2. The paper will have the following sections:
 - a. (10 points) Introduction
 - b. Data Description/Summaries

- i. (10 points) Describe variables
 - ii. (10 points) Summarize variables in words and with appropriate tables.
 - iii. (10 points) Include data visualizations
 - c. Statistical Analysis
 - i. (10 points) Statistical Full Model
 - ii. (10 points) Describe the model selection approach and the criteria you will be using to decide on the “best” model.
 - iii. (10 points) Validity Conditions for full Model
 - d. Statistical Results
 - i. (10 points) Describe the best model according to your model selection criteria. Are there any models that are indistinguishable on the AIC scale?
 - ii. (10 points) Scope of Inference
 - iii. (5 points) Reassess validity conditions for the “best” model if the best model is not the full model. If the full model was the “best” model, if there was an indistinguishable model on the AIC scale, check the diagnostics for that model. If there wasn’t an indistinguishable model, state that and why you don’t need to reassess validity conditions.
 - iv. (5 points) Discuss if the model you started with is considered a “good model” and if it is “trustworthy” or not.
 - e. Conclusion
 - i. (10 points) Next steps; limitations of the data, model or analysis. If you didn’t start with a good model, what might you try if you were to perform the analysis again?
 - ii. (10 points) Paper conclusion to mirror introductory paragraph. Did your thesis/research question get answered?
 - f. Appendix
 - i. (10 points) Include all code and output
- 3. Professionalism
 - a. (10 points) Structured in paper form, no spelling mistakes, good grammar, readable to someone outside of the class.

IMPORTANT!

Just because this data comes from a national survey of households, doesn't mean you can't create your own story about why you are looking at this data. There is plenty of room for creativity. In fact, the more creative you are about the reasons for this analysis the easier it might be to write the introduction and conclusion and provide motivation for the analysis. However, please come up with something that makes sense! I.e. **not** this: "I'm a prisoner of statistical torture island and my jailers are forcing me to do this pointless analysis for no good reason..."

Appendix A (Code and Output):

Code and output used in this report

The code chunks below will automatically include all code and output used in your homework assignment. This is necessary for me to see what you have done. Do not modify (other than to change `eval=FALSE` to `eval=TRUE`) or remove the below code chunks.

You need to change `eval=FALSE` to `eval=TRUE` before compiling your report.

Appendix B (Packages):

The R package `rmarkdown` (Allaire et al. 2022) was used to create this report document using the R language (R Core Team 2022). In addition, the following were packages used for the analysis and/or formatting of this document:

Allaire, JJ, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. 2022. *Rmarkdown: Dynamic Documents for r*. <https://CRAN.R-project.org/package=rmarkdown>.

Pruim, Randall, Daniel T. Kaplan, and Nicholas J. Horton. 2021. *Mosaic: Project MOSAIC Statistics and Mathematics Teaching Utilities*. <https://CRAN.R-project.org/package=mosaic>.

Pruim, Randall, Daniel T Kaplan, and Nicholas J Horton. 2017. "The Mosaic Package: Helping Students to 'Think with Data' Using r." *The R Journal* 9 (1): 77–102. <https://journal.r-project.org/archive/2017/RJ-2017-024/index.html>.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Wickham, Hadley. 2021. *Tidyverse: Easily Install and Load the Tidyverse*. <https://CRAN.R-project.org/package=tidyverse>.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich

Leisch, and Roger D. Peng. Chapman; Hall/CRC. <http://www.crcpress.com/product/isbn/9781466561595>.

———. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.

———. 2022. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.

Xie, Yihui, J. J. Allaire, and Garrett Grolemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown>.

Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook*. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown-cookbook>.