

Biostat 202: Final Project Guidance/Hints

Your final project should have the following components:

Background – Provide some information that helps to identify why you have chosen your particular research question. This section does not have to be long, but needs to show that you have thought about why you are looking at the research question. The research question does not have to be novel or profound. You may not need references.

Research question – State specifically your research question of interest, preferably in the form of: Does xxx (predictors) predict/associated with (yyy) outcome among your particular population of interest (for supervised learning)? Or Determine which xxx characteristics are associated with yyy outcome among your population of interest (for unsupervised learning).

Predictors – Please identify your predictors of your model, especially any predictors that you have had to derive from the dataset. Remember to include any definitions that you have used to define derived predictors especially ones that aren't commonly used (commonly used being BMI, for example).

Outcome (for supervised learning algorithms) – Please list your primary outcome(s) and secondary outcome(s), if any

Methods – Identify whether you intend to perform a supervised or unsupervised learning algorithm and list the algorithm(s) that you will use to perform your analysis in SPSS (you should be using multiple algorithms given that it is easy to do so in SPSS. Also identify any particular challenges or concerns you had faced when importing and cleaning your data in SPSS. If you performed the data cleaning procedures in an alternative program, please state which program you used.

Results

- Comment on your “Table 1” or baseline characteristics of your population of interest, especially any particular points of interest
- List your overall model results including results from various models (if you utilized several different models)
- State any sensitivity analysis results that you may have

Discussion

- Comment upon your results and which model you felt performed the best to look at your specific question of interest
- Comment upon the accuracy of the model and how you evaluated the accuracy of the model
- Comment upon any limitations that you can think of with respect to analysis of your data and to the interpretation of the results
- Conclude with some comments specifically addressing whether or not your research question was adequately answered