Final Project Log

1. EDA, checking graphs of all variables to see which ones should be converted to factors. Converted inv to a factor, and grade to an ordered factor.
2. Model 1 with all main effects. Cumulative proportional odds with logit link.
3. Check VIFs for all predictors. No issues with multicollinearity.
4. Checked Cook's Distance. No observations were influential.
5. Fit Model 2 with all main effects and two way interactions. Gives warning due to non convergence. Tried a LR test to check significance of predictors. This failed because of the same reasons. Looked at the Wald p-values to see which predictors were leading to these issues. These predictors contained NA values for the z value.
6. Fit Model 3. Removed predictors from model 2 that led to warnings (psa:age, psa:cap, and age:cap). Obtained p-values for each predictor using an LRT.

   Note: will not remove any variables that are significant under 10% false discovery rate
7. Removed predictors from model 3 that had very large p-values from LRT (vol:wt, vol:bph, vol:in, vol:cap, age:bph, and bph:cap). Fit Model 4.
8. Removed predictors from model 4 that had very large p-values from LRT (psa:wt, psa:bph, psa:inv, wt:bph, wt:inv, and age:inv). Fit model 5.
9. Removed least significant interaction, inv:cap, based on LRT. Fit model 6.
10. Removed least significant interaction, bph:inv, based on LRT. Fit model 7.
11. Removed least significant predictor, bph, based on LRT. Fit model 8.
12. Removed least significant predictor, inv, based on LRT. Fit model 9.
13. Removed least significant interaction, psa:vol, based on LRT. Fit model 10.
14. Removed least significant interaction, vol:age, based on LRT. Fit model 11.
15. Removed least significant predictor, vol, based on LRT. Fit model 12.

16. Removed least significant interaction, wt:age, based on LRT. Fit model 13. All predictors are now significant so this a final candidate model. Contains predictors psa+wt+age+cap+wt:cap.
17. Check AIC of model 13.
18. Fitting model 14. Contains same predictors as model 13 but uses probit link instead of logit. AIC is larger, not as good of a model. Also receive some warnings that predicted probs are very close to 0 (same warnings as full model 2)
19. Fitting model 15. Contains same predictors as model 13 but uses cloglog link. AIC is larger, not as good of a model. Receive same warnings as mentioned before.
20. Fitting model 16. A non proportional odds model with logit link. Model has failed and did not converge. Not going to use this model.
21. Calculate overdispersion ratio using the Pearson Chi Square Stat and df formula. No overdispersion present.
22. Check VIFs for model 13. The wt:cap interaction VIF is large, 12, gives evidence of multicollinearity.
23. Remove the interaction and check multicollinearity. All predictors are good.
24. Fit model 17 with the collinear interaction removed. Contains predictors psa, age, wt, and cap. Wt is not significant based on LRT.
25. Fit model 18. Has wt removed, contains psa, age, and cap. All predictors significant.
26. Check model 18 AIC. Larger than model 13, but only by a small amount. This model is better than model 13 due to model 13's multicollinearity issues.
27. Check for multicollinearity. All good.
28. Check for influential observations. None present
29. Check for overdispersion. None present.
30. LRT comparing model 18 (age, psa, and cap) to model 2 (model with all main effects and two way interactions). No evidence that the full model is better than model 18. Model 18 is the final model.
31. Created prediction table of all unique combos of the min, median, and max of each variable present in the final model.
32. Transformations of the coefficients to get them in the form needed for interpretation.