## **Review for Exam 2**

You are not required to write any SAS code for this exam, however, you will be answering questions based on the SAS output (with certain key values possibly removed). The missing values will be able to be calculated from the values provided. Think about the questions that were asked to interpret the results on the homework problem sets.

## **Topic 4: Multiple Regression II**

- 1. Determine which general linear tests will test specific problems.
  - a. Be able to state the appropriate general linear test to determine if specific terms (more than one) should be included in the model and interpret the results from that test.
- 2. For a general linear test, be able to write the null and alternative hypothesis, the test statistic and degrees of freedom, p-value, decision and write the conclusion in the context of the problem.
- 3. Be able to write down the equation for the extra sum of squares for both Type I and Type II.
  - a. Be able to calculate SSM from the appropriate extra sum of squares.
  - b. Be able to use the equations to calculate other sum of squares.
- 4. Be able to use the extra sum of squares to be able to determine if an explanatory variable is necessary in the model.
- 5. Be able to use the extra SS to determine if there is a problem with multicollinearity.
- 6. If there is a problem with multicollinearity, determine which values are questionable and which values are acceptable to use using extra SS.
- 7. Be able to state the method that will possibly correct the multicollinearity problem when you have quadratic and interaction terms.
- 8. Be able to determine when a quadratic regression should be performed.
- 9. When using a quadratic regression, be able to make an estimation of when the results should be used and when they should not be used.
- 10. Be able to interpret the results of a quadratic regression (including confidence intervals).
- 11. Be able to state how qualitative predictors are coded in regression.
- 12. Be able to interpret the results from regression using a qualitative predictor or dummy variable (including confidence intervals).
  - a. This includes writing the two fitted regression equations and interpreting the scatterplot (with two lines).
- 13. When there is a qualitative predictor, be able to state and interpret the hypothesis test (and/or confidence intervals) to determine whether the lines are the same or the slopes are the same for the different values of the qualitative variable.
- 14. Be able explain the procedure on how to interpret the results of two continuous variables with an interaction term.
- 15. Given the best models for each number of explanatory variables and the  $R^2$  values, adjusted  $R^2$  values and  $C_P$ , be able to determine which models would be the best to choose.
- 16. Be able to explain briefly how the automatic search methods, forward selection, backward elimination and forward stepwise regression work and which one normally provides the best model.
- 17. Be able to state what methods should be used to choose the best model for a particular situation given the results of objective 15 and the results from the automatic procedures.

## **Topic 5: Diagnostics and Remedial Measures (Regression)**

- 18. Be able to state the assumptions involved in linear regression and be able to tell using the appropriate plots whether the assumptions are met or not (repeat from Midterm 1).
- 19. Be able to interpret a partial regression plot (should the predictor be included in the model, should a quadratic term for the predictor be added to the model)
- 20. Be able to use the studentized deleted residuals to detect outliers for Y observations.
  - a. You will need to know the formula of what to compare the value to, however, you will not need to calculate the actual value.
- 21. Be able to use the diagonal elements of the hat matrix to detect outliers for the X observations.
- 22. Be able to use DFFITS and the Cook's distance to determine if there are any influential points for  $\hat{Y}$ .
- 23. Be able to use the DFBETAS to determine if there are any influential points for the parameters.
- 24. Be able to use the Covariance Ratio to determine if there are any influential points for the variance.
- 25. Be able to use the variance inflation factor and tolerance to determine if there is a problem in multicollinearity.
  - a. Be able to convert between the variance inflation factor and the tolerance.
  - b. Be able to calculate the squared multiple correlation from the variance inflation factor or the tolerance.
- 26. Be able to determine if either the weighted least squares method or ridge regression is required for remedial measures.
- 27. Be able to state the advantages and disadvantages of using the weighted least squares method.
- 28. Know the general procedure of weighted least squares. That is, be able to write down the steps involved.
- 29. Be able to interpret the results from weighted least squares procedure.
- 30. Be able to state what is required to determine if the weighted least squares procedure corrected the problem.
- 31. Be able to state the 7 remedial methods for multicollinearity problems including when each method can be used and the advantages and disadvantages for each.
- 32. Be able to determine what value of 'c' to use for the ridge regression using the ridge trace and variance inflation factors methods.
- 33. Be able to interpret the results of the ridge regression.

## Topic 6: One-Way ANOVA

- 34. Be able to state the model and interpret the results from the cell means model.
- 35. Be able to complete the ANOVA table including how to determine the degrees of freedom, MS and F test statistic.
- 36. Be able to perform a F-test (significance test) for ANOVA for the cell means model given the ANOVA table. That is, state the null and alternative hypothesis, the test statistic with the degrees of freedom, the p-value, decision and the conclusion in words.
- 37. Be able to state the model and interpret the results from the factor effects model.
  - a. This includes the constraints for the two common constraint methods that we are using, conceptual and SAS.
- 38. Be able to state why constraints are necessary in the factor effects model.

- 39. Be able to perform a F-test (significance test) for ANOVA for the factor effects model. That is, state the null and alternative hypothesis, the test statistic with the degrees of freedom, the p-value, decision and the conclusion in words.
- 40. Be able to convert between the parameters of the cell means model and the factor effects model (both constraint systems) and between factor effects models with different constraints.
- 41. Be able to interpret the confidence intervals for the individual means using either proc means or proc glm and using the Bonferroni correction.
- 42. Be able to interpret the difference between the means using the Tukey method (from both cldiff and lines output)
  - a. This includes the interpretation of the confidence interval provided in the cldiff output when the '\*\*\*' is NOT included in the output and determining which level should be chosen.
- 43. Be able to state and explain in which circumstances each of the following procedures should be used, Tukey, Scheffé, Bonferroni.
- 44. Be able to interpret the SAS results from specific contrasts.
- 45. Be able to state why you would use qualitative variables in ANOVA.
- 46. Be able to state the assumptions required for ANOVA and how to diagnose whether they are met by using scatterplots, residual plots, histograms and qqplots.
- 47. Given the results of the Levene test, be able to state whether the constant variance assumption is met or not. That is, state the null and alternative hypothesis, the test statistic with the degrees of freedom, the p-value, decision and the conclusion in words.
- 48. Given a problem with the assumptions for ANOVA, be able to state how to remedy the situation.
  - a. This includes the weighted least squares with the modification that ANOVA uses and knowing the simple guide for Y-transformations in addition to how to use the boxcox procedure.
- 49. Be able to state what are the consequences are when the assumptions in ANOVA are not met.