STAT 705 Test–Spring 2020

You may use SAS help at `sas.support.com` as well as **R** online help; either package is appropriate for any question. You may also review your PROC IMPORT statements for either Excel workbooks or tab-delimited text files with headers. The data set used in this exam (Test 2020.xlsx) is available in my SAS Studio STAT 705 course folder, in Blackboard, and on the course website as an Excel workbook. This data set obtained through Zillow includes the variables Price–home values (in $1000), Square_Footage, Bedrooms–number of bedrooms, and Neighborhood for houses in expensive neighborhoods in Columbia (Arsenal Hill, Congaree Park, Heathwood, Wheeler Hill).

1. We will first consider a one-way ANOVA model with Price as the response and Neighborhood as the factor.

   (a) Construct side-by-side boxplots of Price by Neighborhood. Comment on any patterns.

   (b) Does the Box-Cox procedure suggest any transformations? If so, transform the data, generate a new set of side-by-side boxplots and discuss.

   (c) Conduct a one-way analysis of the model, and test for equality of variances. Regardless of the result, what test(s) might you have run if the variances were unequal?

   (d) Test all possible pairwise comparisons at $\alpha = 0.10$ and discuss the results.

   (e) The Heathwood neighborhood is in a suburban setting with larger lots. Write a suitable contrast in the Neighborhood means and test for significance.

2. We will now consider an ANCOVA model using the same response and factor as in Question 1(b) with Square_Footage added as the concomitant variable, (I checked, and a transformation of Square_Footage appears unnecessary).

   (a) Plot your response from Question 1(b) against Square_Footage, grouping by Neighborhood; include regression lines. Comment on the appropriateness of the no-interaction ANCOVA model.

   (b) Compute pairwise LSMEANS for the no-interaction ANCOVA model at $\alpha = 0.10$ and compare your results to Question 1(d).

   (c) Write an expression for $E(Y_{ij})$ for the interaction model. Now test for a difference in means between Heathwood and Arsenal Hill for 3658.35 square-foot houses for the interaction model using an ESTIMATE statement.

   (d) Compute pairwise LSMEANS for the interaction ANCOVA model and compare the pairwise difference between Heathwood and Arsenal Hill to your answer in 2(c). Explain.

3. Consider a two-way ANOVA model with factors Bedrooms and Neighborhood; continue to use the response you chose in Question 1(b).

(a) Construct an interaction plot and summarize any departures from an additive model.

(b) Conduct a formal test for interaction and compare results to your graph.

(c) Fit an additive model and save predicted values in an output data set.

(d) Compute LSMEANS for Bedrooms for the additive model. Use the table of predicted values for individual cell means from 3(c) and the formula for LSMEANS to verify the LSMEANS value you obtained for 4-bedroom homes.