**Module Assignment**

**Module 8**

**QMB-6304 Analytical Methods for Business**



Write a simple R script to execute the following data preprocessing and statistical analysis. Where required show analytical output and interpretations.

**Preprocessing**

1. Load the file “6304 Module 8 Assignment Data.xlsx” into R. This file contains information on 46,484 vehicles listed for sale on Craig’s List in the United States.
2. Create a single data frame for your analysis which meets the following characteristics:
   1. Only includes cars with 4, 6, or 8 cylinder engines.
   2. Only includes cars using gasoline or diesel as fuel.
   3. Includes all variables appearing in the master (N=46,484) data set.
   4. Has a random sample of n=150 cars from each of three states: Illinois, North Carolina, and Texas. The lists at the end of this assignment specify the regions which you will aggregate to represent the three states. Remember to apply the numerical portion of your U number as the random number seed. This type of sample is referred to as a stratified sample.
   5. Includes a new variable identifying the state from which a car has been drawn. This will be a factor variable with the levels "Illinois", "Texas", and "North Carolina".

There are several ways this can be accomplished. Carefully plan your method for creating this 3-state n=450 data set.

**Analysis**

1. Within your n=450 stratified sample, determine if asking.price has an equal variance across the three states. Briefly interpret your results. If you determine there is a difference in variances across the three factor levels state where the difference(s) is/are.
2. Using your sample (n=450) data set conduct a one-way analysis of variance with asking.price as the dependent variable and state as the independent variable. Plot the results of a Tukey HSD test to show whether/where differences in asking.price among the states exist. Briefly explain the results shown in the plot, stating which pairs of states which do and do not appear to show significant mean differences in asking.price. Make sure state names can be clearly and completely read on the appropriate axis of your plot.
3. Repeat Steps 1 and 2 above using odometer as the dependent variable and state as the independent. Again, briefly explain your analysis results and make sure state names can be clearly and completely read on the appropriate axis of your plot.
4. Drawing on the n=150 sample, use only the vehicles for sale in the state of Texas to conduct a one-way ANOVA using asking.price as the dependent variable and region as the independent variable. Plot the results of a Tukey HSD test to show whether/where there are differences in asking.price among the regions of Texas. Briefly explain the results shown in the plot, stating which regions do appear to show significant mean differences in asking.price. Make sure region names can be clearly and completely read on the appropriate axis of your plot.
5. Using your n=450 sample conduct an ANOVA using asking.price as the dependent variable and fuel and condition as independent variables. Plot the results of a Tukey HSD test to show whether/where there are differences in asking.price by independent variables. Make certain both Tukey plots are visible on the same graphic as demonstrated in class. Make sure names of levels of independent variables can be clearly and completely read on the appropriate axis of your plots.

Your deliverable will be a single MS-Word file created using R Markdown. Your file will show 1) the R script which executes the above instructions and 2) the results of those instructions. The first two lines of your deliverable will state this is “Module 3 Assignment” of our course and your name as it appears in Canvas. Your code chunks and analysis results should be presented in the order in which they are listed here. Deliverable due time will be announced in class and on Canvas. **This is an individual assignment to be completed before you leave the classroom. No collaboration of any sort is allowed on this assignment.**

Regions to be aggregated to represent targeted states.

|  |  |  |
| --- | --- | --- |
| **Texas** | **Illinois** | **North Carolina** |
| amarillo, TX | champaign urbana | asheville, NC |
| austin, TX | chicago | boone, NC |
| brownsville, TX | danville | charlotte, NC |
| college station, TX | peoria, IL | eastern NC |
| corpus christi, TX | quad cities, IA/IL | fayetteville, NC |
| dallas / fort worth | rockford, IL | greensboro, NC |
| el paso, TX | southern illinois | wilmington, NC |
| galveston, TX | springfield, IL | winston-salem, NC |
| houston, TX |  |  |
| lubbock, TX |  |  |
| odessa / midland |  |  |
| tyler / east TX |  |  |
| waco, TX |  |  |