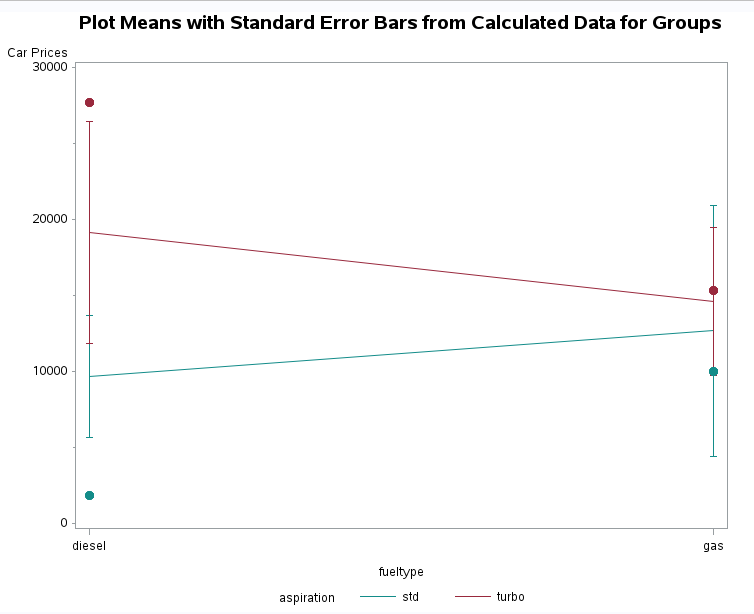
Addressing Objective 2 – Fueltype & Aspiration

State what route you are going to take 2way ANOVA or Time series and summarize the goal. Required

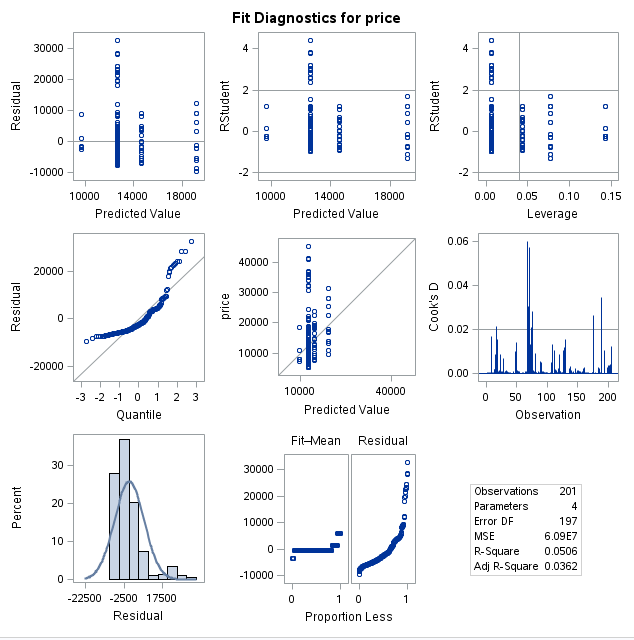
Main Analysis Content Required

* What situation are we in?
  + One Factor of interest with a Block: Randomized Complete Block Design
  + Two Factors: Both Factors are of interest
* Plot data (visualize through mean profile plotting)

A 2 way anova of the 2 categorical variables of fueltype and aspiration. The factors for aspiration are std and tubro while the factors for fueltype are diesel and diesel and gas. It is an unbalanced design because there are unequal number of observations.



* Fit full saturated model with both factors and the interaction (nonadditive) term
* Diagnostics
  + Residuals



* + Normality, Independence, Constant Variance

While there is some departure from normality, in both the histogram and q-q plot, for this exercise, we will assume that the central limit theorem will kick in.

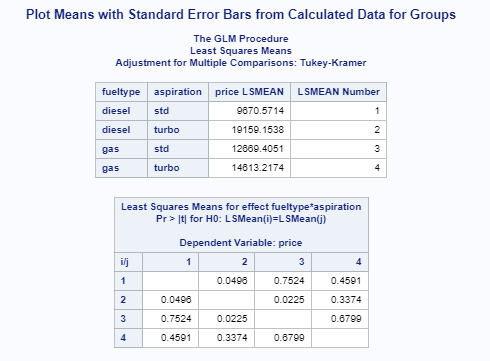
There are no repeated measures in this dataset so independence is maintained.

* + Outliers (Don’t concern with leverage as these apply more to continuous explanatory variables)
* Testing
* High level (ANOVA)
* Contrasts



There is strong evidence from the ANOVA that at least one aspiration and/or at least one fueltype group has a different mean price than the others (p-value of .0165)

With a p-value of.7029, fueltype is not statistically significant in the model. With a p-value of .0053, aspiration is statistically significant. While the interaction term of fueltype\*aspiration has a p-value of .064, it is borderline statistically insignificant with an alpha of .05. The plot means with standard error bars chart also gives some indication of an interaction which makes this an additive model.



Level of Measurement: The criterion variable should be assessed on an interval or ratio level of measurement (i.e. continuous). Both predictor variables should be nominal-level variables (i.e. categorical variables).

Independent observations (No repeated measures)

Random sampling

Normal distributions: Each cell (factor level combination) should be drawn from a normally distributed population. If each cell contains more than 30 participants, the test is robust against moderate departures from normality (CLT kicks in…)

Homogeneity of variance: Cell populations (factor level combinations) should have equal variances. If the number of participants in the largest cell is no more than 1.5 times greater than the number of participants in the smallest cell, then the test is robust against violations of the homogeneity assumption.

Conclusion/Discussion Required

Summary: Factorial ANOVA Test

1. Is the interaction between the factors statistically significant?
   1. Use: plots, examination of data, p-value
   2. If yes, answer the following:
      1. How strong is the interaction?
      2. What is its nature?
   3. If no, test the main effects.
2. Is the main effect for either factor significant?
   1. How strong is the effect?

What is the nature of the differences in means across the levels of the factor?