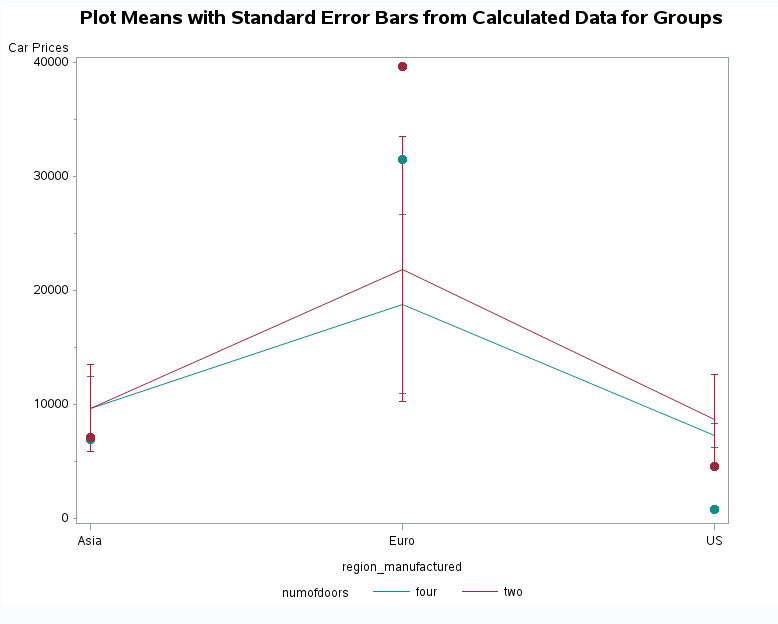
Addressing Objective 2 – region\_manufactured and numofdoors

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

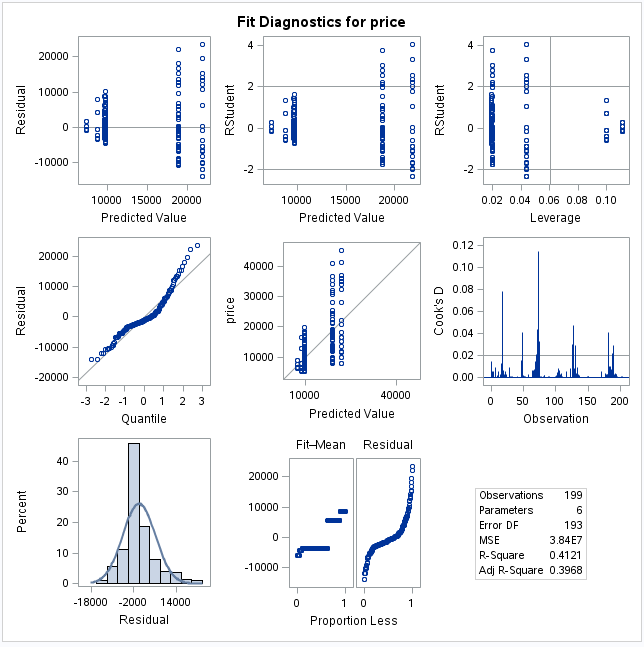
Main Analysis Content Required

* 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 region\_manufactured and numofdoors. The factors for region\_manufactured are US, Asia, and Europe while the factors for numofdoors are two and four. 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



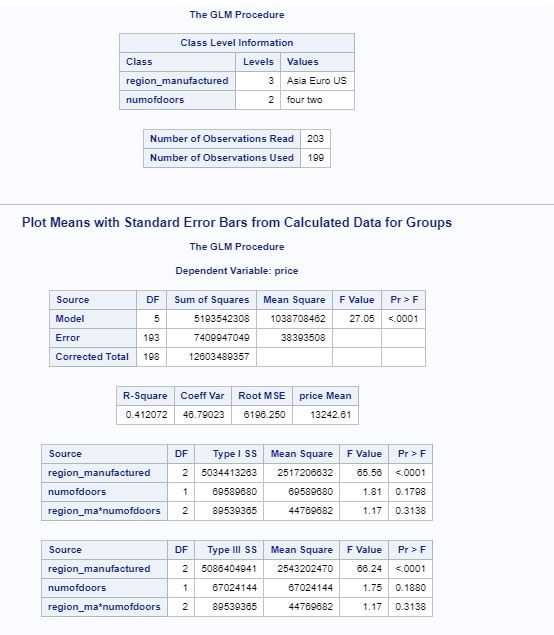
* + Normality, Independence, Constant Variance

While the histogram shows little evidence against normality, there is some curvature towards the top in the QQ plot. We will assume that the CLT will kick in and proceed with caution.

While there is some clustering of the residuals, the Q-Q plot for predicted value residuals looks pretty good. Per our class discussion, our options are limited to addressing residuals. We tried a log transformation on pricing which is our response variable to see if it would generate more of a random cloud, but that did not appear to improve the constant variance.

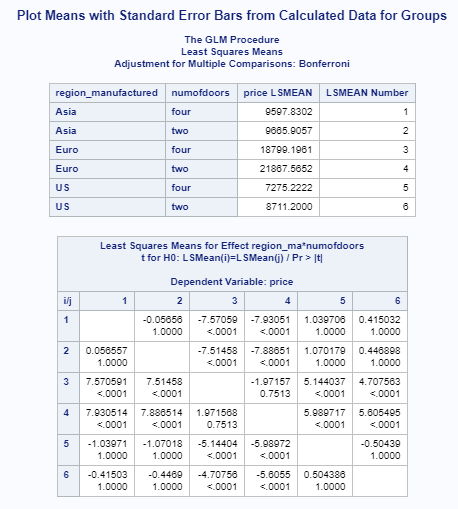
There are no repeated measures in this dataset so independence is maintained.

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



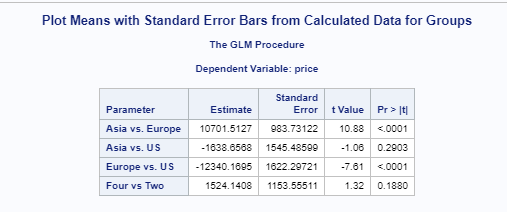
There is strong evidence from the ANOVA that at least one region\_manufactured and numofdoors group has a different mean price than the others (p-value of < .0001 )

With a p-value of < .0001, region\_manufactured is statistically significant. With a p-value of .188, numofdoors is not statistically significant. The interaction term of region\_manufactured \* numofdoors has a p-value of .3138 which is not statistically significant. The plot means with standard error bars chart also gives indication of no interaction which makes this a non-additive model.



The PDIFF has given us the following statistically significant values that indicate which factor levels are different:

* Asia/four-door and Euro/four-door (p-value < .0001) (1/3)
* Asia/four-door and Euro/two-door (p-value < .0001) (1/4)
* Asia/two-door and Europe/four-door (p-value < .0001) (2/3)
* Asia/two-door and Europe/two-door (p-value < .0001) (2/4)
* Europe/four-door and US/four-door (p-value < .0001) (3/5)
* Europe/four-door and US/two-door (p-value < .0001) (3/6)
* Europe/two-door and US/four-door (p-value < .0001) (4/5)
* Europe/two-door and US/two-door (p-value < .0001) (4/6)



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?