1. Linear Models Exercise

In this exercise we will use linear least squares models together with a subset of the EPA mileage data to explore the factors that impact gas mileage.

Although any statistical package (MINITAB, R, SPSS, SAS, S, S-plus, STATPACK, etc., etc.) could be used, we will make use of MINITAB because it is installed in the lab.

Your job is to build two linear models, one for city and one for highway miles per gallon. The objective is to make the residual variation as measured by the error mean square as small as possible (i.e., the model should explain as much of the variation about the mean in the data as possible).

You must decide which variables to include in your model, and, possibly, whether to include the variable as a continuous (a "covariate" in MINITAB) or discrete variable ("Model:" in MINITAB)

Your model should not include any predictors that are not significant at the $\alpha = .05$ level, as indicated by the "P" column in the ANOVA for line corresponding to that predictor.

For this exercise please use the GLM (General Linear Model) procedure under the $Stat- > ANOVA- > GeneralLinearModel$ pulldown unless you have only continuous predictors. In that case, use the $Regression- > Regression$ pulldown.

For GLM, click the "Results" button and select the "In addition, coefficients for all terms" to get the least squares solutions for the parameters in your model. If you use Regression, they will appear automatically.

Your writeup should include the error mean square values, the predictors in your model, and the least squares estimates of their coefficients.

MA225 General Linear Models exercise

Names:
a)   b)
c)   d)

Model: (in the form $Y = \mu + \beta_1 X_1 + \cdots + \beta_n X_n$)

Error mean square:

Least squares solution for parameters: