ST 512 - Lab 7 - Model Selection in MLR

1. Open the `Cheese.sas` file again (from Lab 6). Here you will see data from a study designed to investigate people's taste preference of cheese. The response is a taste score (a subjective taste test score, obtained by combining the scores of several tasters) and the explanatory variables are the natural log of acetic acid, the natural log of hydrogen sulfide, and the concentration of lactic acid.

2. In Lab 6 we fit the MLR model using all three predictors. Fit this model again using the code provided.

```
proc glm data = cheese;
    model taste = acetic h2s lactic;
run;
```

   (a) Are the tests given in the parameter estimates table Type I or Type III tests?

   (b) What are the hypotheses being tested by the Type I SS? What is the difference between those tests and the type III tests?

   (c) Change the order of the variables entered into the model to h2s, lactic, acetic. What model seems sufficient to use? Give a test statistic and a p-value to back up your claim.

3. Let's use an automated procedure to help us determine a good model. Run the following code to use forward selection:

```
proc glmselect data = cheese;
   model taste = acetic h2s lactic / selection = forward(select = sl);
run;
```

   (a) What is the default significance level for a variable to enter the model?

   (b) What is the final model selected by forward selection?

4. Run the following code to use backward selection with removal criteria 0.2:

```
proc glmselect data = cheese;
   model taste = acetic h2s lactic / selection = backward(select = sl)
                                     SLstay = 0.20;
run;
```

   (a) What is the final model selected by backward selection?

5. Run the following code to use stepwise selection with entry criteria 0.2 and removal criteria 0.3:

```
proc glmselect data = cheese;
   model taste = acetic h2s lactic / selection = backward(select = sl)
                                     SLentry = 0.20 SLstay = 0.30;
run;
```

   (a) What is the final model selected by stepwise selection?

6. Run the following code to use subset selection using the Mallow's C_p criterion:

```
proc glmselect data = cheese;
   model taste = acetic h2s lactic / select = cp;
run;
```

   (a) What is the final model selected by subset selection?

Note: These procedures will not usually select the same model. However, when they do that gives good evidence for your model being the one to use from your set of candidate models.

Note: The model you select should ***not*** be used for any final inferences (e.g. for publication). A separate **confirmatory study** would be needed to make inferences about your model!