ST 512 - Lab 7 - Multiple Linear Regression

1. Open the `Cheese.sas` file. 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. Use PROC CORR to obtain a correlation matrix and scatter plot matrix of all four variables. Which single variable do you think would do the best job predicting Taste?

3. All of these variable may affect the taste and so ideally we would look at them together. This is something that MLR allows us to do. Use the following code to fit the additive MLR model with all 3 predictors:

```
ods graphics on;
proc glm data = cheese;
    model taste = acetic h2s lactic;
run;
ods graphics off;
```

   (a) What is the estimated least squares prediction line?

   (b) What are the interpretations of each β parameter? What is the estimate of σ^2 and its interpretation?

   (c) Identify the null and alternative hypotheses tested by the ANOVA p-value and the p-values in the parameter estimates table.

4. We can get SAS to print out CI's for the slope parameters using the following model statement: `model taste = acetic h2s lactic / clparm;`. How do these CIs relate the the p-values of the hypothesis tests we just looked at?

5. In addition to the `clparm` option to provide CI for the model parameters, we can print out the (augmented) $\mathbf{X^TX}$ matrix and it's inverse. Use the options `xpx` and `I` in your model statement to request these matrices: `model taste = acetic h2s lactic / clparm xpx i;`.

   (a) What should the dimension of $\mathbf{X^TX}$ be? What is the other information that SAS is printing in this matrix?

   (b) What is the relationship between the X-transpose-X matrix and the covariance matrix of our regression estimates?

   (c) How can we use this output to form a 90% CI for the mean taste score with acetic = 10, h2s = 5.5, and lactic=4? No need to carry out the calculation, just write out what formula we should be using and be able to obtain each of the components!

6. Note that with multiple predictors, SAS will not create a plot with confidence and prediction bands on it since it requires more than two dimensions. However, we can still use SAS to find the find a 90% confidence interval for the mean taste score with acetic = 10, h2s = 5.5, and lactic=4 and also find a 90% prediction interval for the same observation. Recall this can be done using the missing y trick and a couple of data steps! (Refer back to older code if you need to!) Once you've added the missing value, get the CI using:

```
ods graphics off;
proc glm data = cheese;
    model taste = acetic h2s lactic/clm alpha=0.1;
run;
```