# STAT 425 Assignment 1

Due Monday, February 8, 11:59pm. Submit through Moodle.

## Name: (insert your name here)

**Netid: (insert)**

Submit your computational work both as an R markdown (*.Rmd) document and as a pdf, along with any files needed to run the code. Embed your answers to each problem in the document below after the question statement. If you have hand-written work, please scan or take pictures of it and include in a pdf file, ideally combined with your pdf output file from R Markdown.

## Problem 1

Thirty samples of cheddar cheese were analyzed for their content of acetic acid, hydrogen sulfide and lactic acid. A panel of judges tasted each sample and scored them, and the average taste score for each sample was recorded. The data are available as the data frame 'cheddar' in the **faraway** library. After loading the library enter 'help(cheddar)' for more information.

**a)** Make a scatter plot of 'taste' versus 'Lactic' and include the least squares regression line on the graph. Comment on whether the graph appears consistent with data that follow a linear model.

**Answer:**

**b)** Obtain and display the summary of the least square fitted model, including coefficient estimates, standard errors, t-values and p-values. Is there is a statistically significant association between lactic acid content and the average taste score, using a significance level of $\alpha = 0.05$? Explain based on your results, making clear what information from the results you are using.

**Answer:**

**c)** In R, the 'cor' function can compute the sample correlation coefficient between two variables in a data set. Compute the **squared** correlation between 'taste' and 'Lactic'. Verify that this is numerically equal to $R^2$ for the model. (Note: to refer to a variable within a data frame use the dataframe$variable syntax.)

**Answer:**

**d)** Compute a 95% confidence interval for the coefficient of 'Lactic' in the model.

**Answer:**

**e)** Compute a 95% confidence interval for the mean taste value expected for a cheddar cheese sample with lactic acid concentration of 2.0.

**Answer:**

## Problem 2

The simple regression through the origin model has the form

$$y_i = \beta_1 x_i + e_i, \quad i = 1, 2, \ldots, n,$$

where the standard assumptions are that $E(e_i) = 0$, $\text{var}(e_i) = \sigma^2$, and $\text{cov}(e_i, e_j) = 0$ if $i \neq j$. The least squares estimate of $\beta_1$ minimizes the residual sum of squares,

$$RSS(\beta_1) = \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2$$

as a function of $\beta_1$.

**a)** Take the derivative of $RSS(\beta_1)$ with respect to $\beta_1$, set the derivative to zero. Solve the resulting equation algebraically to obtain the formula for the estimate, $\hat{\beta}_1$.

**Answer:** (here or indicate where it is in the attached pdf file)

**b)** Use your formula form Part a) to show that $\hat{\beta}_1$ is an unbiased estimator of $\beta_1$ under the standard assumptions.

**Answer:**

**c)** Show that

$$\text{var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$$

**Answer:**

**d)** Under the standard assumptions find $E(y_1)$ and $\text{var}(y_1)$.

**Answer:**

**e)** Under the standard assumptions find expressions for $E(\hat{y}_1)$ and $\text{var}(\hat{y}_1)$.

**Answer:**

## Problem 3:

This problem refers again to the 'cheddar' data described in Problem 1.

**a)** Make a 'pairs' plot of the data, i.e., a matrix of all the pairwise scatter plots between variables.

**Answer:**

**b)** Fit a multiple linear regression model with 'taste' as the response and the three chemical constituent concentrations as the predictors. Display a summary of your fitted model. Note: the 'lm' function can fit a multiple linear regression model using a formula of the form 'y ~ x1 + x2 + ... + xp'.

**Answer:**

**c)** Report the values of the regression coefficients for associated with the predictors.

**Answer:**

**d)** Which of the predictor variables have statistically significant coefficients, rejecting the null hypothesis that the coefficient is zero, at the 5% level of significance? Explain.

**Answer:**

**e)** Compute an estimate of the average taste score for a cheddar sample with Acetic= 5.5, H2S=5.0, Lactic=1.5.

**Answer:**