Homework 5

Due, Friday Oct 20 by 3:00

1. Load the stackloss data:

```
data(stackloss)
names(stackloss)
help(stackloss)
```

(a) Plot the data.

(b) Fit a multiple regression model to predict `stackloss` from the three other variables. The model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

where Y is stackloss, X_1 is airflow, X_2 is water temperature and X_3 is acid. Summarize the results.

(c) Construct 90 percent confidence intervals for the coefficients of the linear regresion model.

(d) Construct a 99 percent prediction interval for a new observation when Airflow = 58, Water temperature = 20 and Acid = 86.

(e) Test the null hypothesis $H_0 : \beta_3 = 0$. What is the p-value? What is the conclusion (at $\alpha = 0.10$)?

2. Suppose that

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i, \quad i = 1, \ldots, n$$

where $\epsilon_i \sim N(0, \sigma^2)$. Notice that there is no intercept. Suppose that

$$\sum_i X_{1i} X_{2i} = 0.$$

Show that the least squares estimators $\widehat{\beta}_1$ and $\widehat{\beta}_2$ from the multiple regression are the same as if we were to fit separate, simple regressions on X_1 and X_2.

3. Consider these data:

X_1	4	3	10	7
X_2	5	4	9	10
Y	25	20	57	51

(a) Fit the multiple regression in R and summarize the results.

(b) Construct $\mathbf{X}^T\mathbf{X}$ and $(\mathbf{X}^T\mathbf{X})^{-1}$.

(c) Construct $\widehat{\beta}$ directly (show your work) and confirm that you get the same answer as you got from R.

(d) Construct the hat matrix $\mathbf{H}$.

(e) Compute $\text{Var}(\widehat{\beta})$ using your calculations.

4. Let $\mathbf{H}$ be the hat matrix from multiple regression. Show that $\text{tr}(\mathbf{H}) = p+1$ where p is the number of covariates.

5. Recall that two vectors v and w or orthogonal if $v^T w = 0$. Let $\mathbf{e}$ be the vector of residuals and let $\widehat{\mathbf{Y}}$ be the vector of fitted values. Use the properties of the hat matrix to show that $\mathbf{e}$ and $\widehat{\mathbf{Y}}$ are orthogonal.