STAT505 Assessment #1

1. For the following matrix, calculate the sample mean vector $\overline{\mathbf{x}}$, the sample covariance matrix $\mathbf{S}$, the sample correlation matrix $\mathbf{R}$, and the generalized sample variance $|\mathbf{S}|$.

$$\begin{bmatrix} X_1 & X_2 \\ 3 & 4 \\ 6 & -2 \\ 3 & 1 \end{bmatrix}$$

2. (*adapted from J&W Exercise 3.15*) Consider the data matrix

$$\begin{bmatrix} X_1 & X_2 & X_3 \\ 1 & 4 & 3 \\ 6 & 2 & 6 \\ 8 & 3 & 3 \end{bmatrix}.$$

We have $n = 3$ observations on $p = 3$ variables X_1, X_2, and X_3. Form the linear combinations

$$\mathbf{b}'\mathbf{X} = [1 \quad 1 \quad 1] \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = X_1 + X_2 + X_3$$

$$\mathbf{c}'\mathbf{X} = [1 \quad 2 \quad -3] \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = X_1 + 2X_2 - 3X_3$$

Find the sample means, variances, and covariance of $\mathbf{b}'\mathbf{X}$ and $\mathbf{c}'\mathbf{X}$

3. Consider the data "corporations.dat" from 10 U.S. corporations. The variables are Sales, Profits, and Assets. All figures are in millions of dollars. These data may be input into SAS using the following code:

```
data corp;
infile 'v:\corporations.dat' delimiter='09'x;
input name :$20. sales profits assets;
run;
```

The "`delimiter`" command tells SAS the columns of data are delimited by tabs, and the "`:$20.`" symbols tell SAS the corporation names may have a length of up to 20 characters.

Consider the following linear combinations:

$$\text{Overhead} = \text{Sales} - \text{Profits}$$
$$\text{Assets after Sales} = \text{Assets} + \text{Sales}$$

(a) Find the sample mean of Overhead and the sample mean of Assets after Sales.

(b) Find the sample variance of Overhead and the sample variance of Assets after Sales.

(c) Find the sample covariance and correlation between Overhead and Assets after sales. Describe the relationship between them.