Assignment 2

Due: Thursday September 13

1. Please do the following exercises from 'Applied Multivariate Statistical Analysis' by Johnson and Wichern. Next to each exercise there is an **H** and/or **R**, indicating wether the the exercise should be done **by hand** [**H**] and/or **by R** [**R**].

   - 2.3:
     - (a,b,c) [**H,R**]
     - (d) [**H**]
   - 2.4:
     - (a,b) [**H**]
   - 2.5 [**H,R**]
   - 2.7:
     - (a) [**H,R**]
     - (b) [**H**]
     - (c,d) [**H,R**]
   - 2.12 [**H**]
   - 2.13 [**H**]

2. In class we determined the least squares solution for a linear regression model ($\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$). Recall the linear regression model is $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Where the $Cov(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}$.

   (a) Outline the least squares problem and solution.

   (b) Using the 'mathmarks' data set and matrix routines in R (not using lm()), solve for $\hat{\boldsymbol{\beta}}$ when we regress the 'statistics' variable against all the other variables.

   (c) Show by hand (proof) that the covariance of $\hat{\boldsymbol{\beta}}$ denoted by $Cov(\hat{\boldsymbol{\beta}})$ is equal to $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.
   Where the estimate of σ^2 is $\hat{\sigma^2} = \frac{(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})}{n-p}$, n is the sample size, and p is the number of parameters that are estimated in the model not counting σ^2.

      i. Using matrix routines in R (not using lm()), calculate the estimated $Cov(\hat{\boldsymbol{\beta}})$.

      ii. Show that the estimated covariance matrix is positive definite.

iii. Using the det() command in R, evaluate the determinant of the estimated $Cov(\hat{\boldsymbol{\beta}})$. Can the result from 2.12 be applied to the estimated covariance matrix and covariance matrices in general? If so, use the result as another means of calculating the determinant for the estimated covariance matrix.