Statistics 312/612, fall 2016
Homework # 2
Due: Monday 19 September

[1]

(i) (10 points) Suppose $X = (x_1, \ldots, x_p)$ is an $n \times p$ matrix and y is $n \times 1$. Let $X_1 = (x_1, \ldots, x_k)$ and $X_2 = (x_{k+1}, \ldots, x_p)$. Define a new $n \times p$ matrix Z by $z_i = x_i$ for $i \leq k$ and

```
z_i <- lm(x_i \sim -1 + X_1)$res
```

for $k + 1 \leq i \leq p$. (Of course **R** would object to the subscripts.) Calculate two least squares fits.

```
outX <- lm(y ~ -1+X); outZ <- lm(y ~ -1+Z)
```

Explain geometrically (using the language of projections and orthonormal bases, not the $(X'X)^{-1}X'y$ stuff) why the fitted vectors are the same in both cases. Hint: Why do the columns of Z span the same subspace of $\mathbb{R}^n$ as the columns of X?

(ii) (extra credit) Suppose X is of full rank (that is, its columns are linearly independent). Derive a formula relating the coefficients in `outX` to the coefficients in `outZ`. Then use your formula to explain why some coefficients are the same and some are different.

(iii) (5 points) Illustrate your answers to (i) (and (ii)?) by including with your solution an annotated (highlighted, with comments added) copy of the following summaries of the calculations from the handout Longley.pdf (or Longley.Rmd):

```
L1 <- longley
for (nn in names(longley)[1:5]){
  L1[[nn]] <- lm(longley[[nn]] ~ longley[["Year"]])$residual
}
summary(lm(Employed ~ . , data = L1))
summary(lm(Employed ~ . , data = longley))
```

Be sure to explain how this example is a special case of (i).

[2] Suppose Q_1 is an $n \times k$ matrix whose columns provide an orthonormal basis for the subspace $\mathcal{X}$ of $\mathbb{R}^n$ spanned by the columns of an $n \times p$ matrix $X = (x_1, \ldots, x_p)$. The hat matrix $H = Q_1 Q_1^T$ projects vectors orthogonally onto $\mathcal{X}$.

(i) (5 points) Explain why $H = H^T = H^2$.

(ii) (5 points) Explain why the diagonal elements H_{ii} all take values in the range $[0, 1]$. Hint: Consider $\|He_i\|$ where e_i denotes the ith column of the identity matrix I_n.

(iii) (5 points) Explain why $\sum_i H_{ii} = k$. Hint: This sum is called the trace of the matrix H. If A and B are two matrices for which both AB and BA are well defined then $\text{trace}(AB) = \text{trace}(BA)$.

(iv) (harder; 5 points) Suppose the first two rows of X are the same. Explain why the first two rows of H are the same.

[3] Suppose X is an $n \times p$ matrix with singular-value decomposition $X = \sum_{i \leq p} \lambda_i u_i v_i^T$. Suppose also that $y = \sum_{i \leq n} s_i u_i$ is an $n \times 1$ vector. Let $\gamma > 0$ be a constant.

(i) (10 points) Show that the vector

$$\widehat{b}_\gamma = \sum_{i \leq p} \frac{s_i \lambda_i}{\lambda_i^2 + \gamma} v_i$$

minimizes the penalized sum of squared residuals, $S_\gamma(b) = \|y - Xb\|^2 + \gamma \|b\|^2$.

(ii) (extra credit) As usual, let $\widehat{y}$ denote the orthogonal projection of y onto the subspace spanned by the columns of X. Show that

$$\left\| \widehat{y} - X\widehat{b}_\gamma \right\| \leq \left(\frac{\gamma}{\lambda_k^2 + \gamma} \right) \|\widehat{y}\|,$$

where λ_k is the smallest nonzero singular value of X.