BST 140.751
Problem Set 1

1 Regression and background

1. If your normalize (mean 0 and variance 1) your Y and X vectors, argue that the regression slope estimate is the correlation.

2. Let Y and X be one dimensional vectors of length n. Give the relationship between the slope from regressing Y on X and X on Y.

3. Consider the residuals after mean only regression. Argue that they sum to 0.

4. Consider the residuals after regression through the origin. Argue that they are orthogonal to the regressor.

5. Consider the residuals from ordinary linear regression. Argue that the residuals are orthogonal to both J_n and X.

2 Least squares

1. Show that $I - H$ is an idempotent matrix where H is idempotent.

2. Let $X = [X_1 X_2]$ be an $n \times 2$ design matrix and consider

$$||Y - X\beta||^2$$

   where $\beta = (\beta_1 \beta 2)'$. Show that $\hat{\beta}_2$ can be obtained by taking the residuals after regressing X_1 out of Y and X_2 then doing regression through the origin on the residuals.

3. Argue that X, X', $X'X$ and XX' all have the same matrix rank.

4. Suppose that X is such that $X'X = I$. Find the associated least squares estimate of β.

5. Suppose that the design matrix is of the form $J_A \otimes I$ where J_A is a vector of length A and I is a $B \times B$ identity matrix. Let Y be a $AB \times 1$ length vector. Find the least squares estimates associated with this design matrix.

6. Let $X = [X1X2]$ where X_1 is $n \times p_1$ and X_2 is $n \times p_2$. Consider minimizing $||Y - X\beta||$ where $\beta = (\beta_1' \beta_2')'$. Argue that the least squares estimate of β_1 can be obtained by regressing $e_{Y|X_2} = (I - X_2(X_2'X_2)^{-1}X_2')Y$ as the outcome and $e_{X_1|X_2} = (I - X_1(X_1'X_1)^{-1}X_1')X_2$. as the predictor.

7. Show that if $X = [X_1 \ldots X_p]$ is such that $X'X = I$. Then the least squares estimate of β is $X'Y$ and further $\hat{Y}$ is $\sum_{j=1}^p X_i < Y, X_i >$.

8. Assume that the columns of X have been mean centered so that $J_n' X = (0 \dots 0)'$. Suppose further that $X = UDV'$ where $U'U = V'V = I$ and D is a diagonal matrix of singular values. Argue that the matrix $U = XV'D^{-1}$ results in an orthonormal basis for the same space as the column space of X. Thus, the $\hat{Y}$ matrix treating X as the outcome and treating U as the outcome are the same and further $\hat{Y} = \sum_{j=1}^{p} U_i < U_i, Y >$ where U_i are the columns of U.

9. If $U = XW$ where W is an invertibble matrix, relate the estimated coefficients obtained when using U as the design matrix and W as the design matrix.

3 Computing and analysis

1. Write an R function that takes a Y vector and X matrix and obtains the least squares fit for the associated linear model.

2. Write and R function that takes an $n \times p$ data matrix, X, and "whitens" it via subtracting out a mean and multiplying by a matrix so that the resulting matrix has (p) sample column means of 0 and $p \times p$ sample covariance matrix of I.