18.S096 Problem Set 3 Fall 2013
Regression Analysis
Due Date: 10/8/2013

The Projection('Hat') Matrix and Case Influence/Leverage

Recall the setup for a linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{y}$ and $\boldsymbol{\epsilon}$ are $n-$vectors, $\boldsymbol{X}$ is an $n \times p$ matrix (of full rank $p \leq n$) and $\boldsymbol{\beta}$ is the p-vector regression parameter.

The Ordinary-Least-Squares (OLS) estimate of the regression parameter is:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

The vector of fitted values of the dependent variable is given by:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} = \boldsymbol{H}\boldsymbol{y},$$

where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ is the $n \times n$ "Hat Matrix"

and the vector of residuals is given by:

$$\hat{\boldsymbol{\epsilon}} = (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y},$$

1 (a) Prove that H is a projection matrix, i.e., H has the following properties:

- Symmetric: $H^T = H$
- Idempotent: $H \times H = H$

1 (b) The ith diagonal element of H, $H_{i,i}$ is called the *leverage* of case i. Show that

$$\frac{d\hat{y}_i}{dy_i} = H_{i,i}$$

1 (c) If $\boldsymbol{X}$ has full column rank p,

$$Average(H_{i,i}) = \frac{p}{n}$$

Hint: Use the property: $tr(AB) = tr(BA)$ for conformal matrices A and B.

1 (d) Prove that the Hat matrix H is unchanged if we replace the $(n \times p)$ matrix $\boldsymbol{X}$ by $\boldsymbol{X}' = \boldsymbol{X}G$ for any non-singular $(p \times p)$ matrix $\boldsymbol{G}$.

1 (e) Consider the case where $\boldsymbol{X}$ is $n \times (p+1)$ with a constant term and p independent variables defining the regression model, i.e.,

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{p,n} \end{bmatrix}$$

Define G as follows:

$$\boldsymbol{G} = \begin{bmatrix} 1 & -\bar{x}_1 & -\bar{x}_2 & \cdots & -\bar{x}_p \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & & 1 \end{bmatrix}$$

where $\bar{x}_j = \sum_{i=1}^{n} x_{i,j}/n$, for $j = 1, 2, \ldots, p$.

By (d), the regression model with $\boldsymbol{X'} = \boldsymbol{X}G$ is equivalent to the original regression model in terms of having the same fitted values $\hat{\boldsymbol{y}}$ and residuals $\hat{\boldsymbol{\epsilon}}$

- If $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ is the regression parameter for $\boldsymbol{X}$, show that

  $$\boldsymbol{\beta'} = G^{-1}\beta \text{ is the regression parameter for } \boldsymbol{X'}.$$

  Solve for G^{-1} and provide explicit formulas for the elements of $\boldsymbol{\beta'}$.

- Show that:

  $$[\boldsymbol{X'}^T\boldsymbol{X'}] = \begin{bmatrix} n & \boldsymbol{0}_p^T \\ \boldsymbol{0}_p & \mathcal{X}^T\mathcal{X} \end{bmatrix}$$

  where $\mathcal{X} = \begin{bmatrix} x_{1,1} - \bar{x}_1 & x_{1,2} - \bar{x}_2 & \cdots & x_{1,p} - \bar{x}_p \\ x_{2,1} - \bar{x}_1 & x_{2,2} - \bar{x}_2 & \cdots & x_{2,p} - \bar{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} - \bar{x}_1 & x_{n,2} - \bar{x}_2 & \cdots & x_{p,n} - \bar{x}_p \end{bmatrix}$

- Prove the following formula for elements of the projection/hat matrix:

  $$H_{i,j} = \frac{1}{n} + (x_i - \bar{x})^T[\mathcal{X}^T\mathcal{X}]^{-1}(x_j - \bar{x})$$

  where $x_i = (x_{i,1}, x_{i,2}, \ldots x_{i,p})^T$ is the vector of independent variable values for case i, and $\bar{x} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)^T$.

  The *leverage* of case i, $H_{i,i}$, increases with the second term, the squared *Mahalanobis* distance between x_i and the mean vector $\bar{x}$.

Case Deletion Influence Measures

2 (a) Sherman-Morrison-Woodbury (S-M-W) Theorem: Suppose that A is a $p \times p$ symmetric matrix of rank p, and a and b are each $q \times p$ matrices of rank $q < p$. Then provided inverses exist

$$(A + a^T b)^{-1} = A^{-1} - A^{-1}a^T(I_q + bA^{-1}a^T)^{-1}bA^{-1}.$$

Prove the theorem.

2 (b) Case deletion impact on $\hat{\beta}$: Apply the S-M-W Theorem to show that the least squares estimate of β when the ith case is deleted from the data is

$$\hat{\beta}_{(i)} = \hat{\beta} - \frac{(\boldsymbol{X}^T\boldsymbol{X})^{-1}x_i\hat{\epsilon}_i}{1-H_{i,i}},$$

where x_i^T is the ith row of $\boldsymbol{X}$ and $\hat{\epsilon}_i = y_i - \hat{y}_i = y_i - x_i^T\hat{\beta}$.

2 (c) A popular influence measure for a case i is the ith Cook's distance

$$CD_i = \left(\frac{1}{p\hat{\sigma}^2}\right)|\hat{y} - \hat{y}_{(i)}|^2$$

where $\hat{y}_{(i)} = \boldsymbol{X}\hat{\beta}_{(i)}$. Show that

$$CD_i = \frac{\hat{\epsilon}_i^2}{p\hat{\sigma}^2} \cdot \frac{H_{i,i}}{(1-H_{i,i})^2}$$

2 (d) Case deletion impact on $\hat{\sigma}^2$: Let $\hat{\sigma}^2_{(i)}$ be the unbiased estimate of the residual variance σ^2 when case i is deleted from the data. Show that:

$$\hat{\sigma}^2_{(i)} = \hat{\sigma}^2 + \left(\frac{1}{n-p-1}\right)\left(\hat{\sigma}^2 - \frac{\hat{\epsilon}_i^2}{1-H_{i,i}}\right)$$

Sequential ANOVA in Normal Linear Regression Models via the QR Decomposition

Recall from the lecture notes that the QR-decomposition, $\boldsymbol{X} = \boldsymbol{QR}$ is a factorization of the $n \times p$ matrix $\boldsymbol{X}$ into $\boldsymbol{Q}$, an $n \times p$ column-orthonormal matrix ($\boldsymbol{Q}^T\boldsymbol{Q} = I_p$, the $p \times p$ identify matrix) times $\boldsymbol{R}$, a $p \times p$ upper-triangular matrix.

Denoting the jth column of $\boldsymbol{X}$ and of $\boldsymbol{Q}$ by $X_{[j]}$ and $Q_{[j]}$, respectively, we can write out the QR-decomposition for X, column-wise:

$$\begin{aligned}
X_{[1]} &= Q_{[1]}R_{1,1} \\
X_{[2]} &= Q_{[1]}R_{1,2} + Q_{[2]}R_{2,2} \\
X_{[3]} &= Q_{[1]}R_{1,3} + Q_{[2]}R_{2,3} + Q_{[3]}R_{3,3} \\
&\vdots \\
X_{[p]} &= Q_{[1]}R_{1,p} + Q_{[2]}R_{2,p} + Q_{[3]}R_{3,p} + \cdots + Q_{[p]}R_{p,p}
\end{aligned}$$

A common issue arising in a regression analysis with p explanatory variables is whether just the first k $(< p)$ explanatory variables (given by the first k columns of $\boldsymbol{X}$) enter in the regression model. This can be expressed as an hypothesis about the regression parameter $\boldsymbol{\beta}$,

$$H_0\colon\ \beta_{k+1} = \beta_{k+2} = \cdots = \beta_p \equiv 0.$$

3 (a) Consider the estimate $\hat{\beta}_0 = \begin{pmatrix} \hat{\beta}_I \\ 0_{p-k} \end{pmatrix}$ where

$$\begin{aligned} \hat{\beta}_I &= (X_I^T X_I)^{-1} X_I^T y \\ X_I &= \left[X_{[1]} X_{[2]} \cdots X_{[k]}\right] \end{aligned}$$

Show that $\hat{\beta}_0$ is the constrained least-squares estimate of β corresponding to the hypothesis H_0, i.e.,

$$\hat{\beta}_0 \text{ minimizes: } SS(\beta) = (y - X\beta)^T (y - X\beta)$$

$$\text{subject to}$$

$$\hat{\beta}_j = 0,\ j = k+1, k+2, \ldots, p.$$

3 (b) Show that the QR-decomposition of X_I is $X_I = Q_I R_I$, where Q_I is the matrix of the first k columns of Q and R_I is the upper-left $k \times k$ block of R. Furthermore, verify that:

$$\hat{\beta}_I = R_I^{-1} Q_I^T y, \text{ and}$$

$$\hat{y}_I = H_I y,$$

where $H_I = Q_I Q_I^T$, the $n \times n$ projection/Hat matrix under the null hypothesis.

3 (c) From the lecture notes, recall the definition of

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{Q}^T \\ \boldsymbol{W}^T \end{bmatrix}, \text{ where}$$

- $\boldsymbol{A}$ is an $(n \times n)$ orthogonal matrix (i.e. $\boldsymbol{A}^T = A^{-1}$)
- $\boldsymbol{Q}$ is the column-orthonormal matrix in a Q-R decomposition of $\boldsymbol{X}$

Note: $\boldsymbol{W}$ can be constructed by continuing the *Gram-Schmidt Orthonormalization* process (which was used to construct $\boldsymbol{Q}$ from $\boldsymbol{X}$) with $\boldsymbol{X}^* = [\ \boldsymbol{X} \mid \boldsymbol{I}_n\ ]$.

Then, consider

$$\boldsymbol{z} = \boldsymbol{A}\boldsymbol{y} = \begin{bmatrix} \boldsymbol{Q}^T\boldsymbol{y} \\ \boldsymbol{W}^T\boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{z}_{\boldsymbol{Q}} \\ \boldsymbol{z}_{\boldsymbol{W}} \end{bmatrix} \quad \begin{matrix} (p \times 1) \\ (n-p) \times 1 \end{matrix}$$

Prove the following relationships for the unconstrained regression model:

- $$\begin{aligned} y^T y &= y_1^2 + y_2^2 + \cdots + y_n^2 \\ &= z_1^2 + z_2^2 + \cdots + z_n^2 \\ &= z^T z \end{aligned}$$
- $$\begin{aligned} \hat{y}^T \hat{y} &= \hat{y}_1^2 + \hat{y}_2^2 + \cdots + \hat{y}_n^2 \\ &= z_1^2 + z_2^2 + \cdots + z_k^2 + \cdots + z_p^2 \end{aligned}$$
- $$\begin{aligned} \hat{\epsilon}^T \hat{\epsilon} &= \hat{\epsilon}_1^2 + \hat{\epsilon}_2^2 + \cdots + \hat{\epsilon}_n^2 \\ &= z_{p+1}^2 + z_{p+2}^2 + \cdots + z_n^2 \end{aligned}$$

Prove the following relationships for the constrained regression model:

- $$\begin{aligned} y^T y &= y_1^2 + y_2^2 + \cdots + y_n^2 \\ &= z_1^2 + z_2^2 + \cdots + z_n^2 \\ &= z^T z \end{aligned}$$
- $$\begin{aligned} \hat{y}_I^T \hat{y}_I &= (\hat{y}_I)_1^2 + (\hat{y}_I)_2^2 + \cdots + (\hat{y}_I)_n^2 \\ &= z_1^2 + z_2^2 + \cdots + z_k^2 \end{aligned}$$
- $$\begin{aligned} \hat{\epsilon}_I^T \hat{\epsilon}_I &= \hat{(\epsilon_I)}_1^2 + \hat{(\epsilon_I)}_2^2 + \cdots + \hat{(\epsilon_I)}_n^2 \\ &= z_{k+1}^2 + \cdots + z_p^2 + z_{p+1}^2 + z_{p+2}^2 + \cdots + z_n^2 \end{aligned}$$

3 (d) Under the assumption of a normal linear regression model, the lecture notes detail how the distribution of $\boldsymbol{z} = \boldsymbol{Ay}$ is

$$z = \begin{pmatrix} \boldsymbol{z_Q} \\ \boldsymbol{z_W} \end{pmatrix} \sim N_n \left[\begin{pmatrix} \boldsymbol{R\beta} \\ \boldsymbol{O}_{n-p} \end{pmatrix}, \sigma^2 \boldsymbol{I}_n \right]$$

$\Longrightarrow$

$$\boldsymbol{z_Q} \sim N_p[(\boldsymbol{R\beta}), \sigma^2 \boldsymbol{I}_p]$$
$$\boldsymbol{z_W} \sim N_{(n-p)}[(\boldsymbol{O}_{(n-p)}, \sigma^2 \boldsymbol{I}_{(n-p)}]$$

and $\boldsymbol{z_Q}$ and $\boldsymbol{z_W}$ are independent.

- For the unconstrained (and the constrained) model, deduce that:
  $$SS_{ERROR} = \hat{\epsilon}^T \hat{\epsilon} \sim \sigma^2 \times \chi^2_{n-p}$$
  a Chi-Square r.v. with $(n-p)$ degrees of freedom scaled by σ^2.
- For the constrained model under H_0, deduce that:
  $$\begin{aligned} SS_{REG(k+1,\ldots,p|1,2,\ldots,k)} &= \hat{y}^T \hat{y} - \hat{y}_I^T \hat{y}_I \\ &= \hat{\epsilon}_I^T \hat{\epsilon}_I - \hat{\epsilon}^T \hat{\epsilon} \\ &= z_{k+1}^2 + \cdots z_p^2 \\ &\sim \sigma^2 \times \chi^2_{p-k}, \end{aligned}$$
  a σ^2 multiple of a Chi-Square r.v. with $(p-k)$ degrees of freedom which is independent of SS_{ERROR}.
- Under H_0, deduce that the statistic:

$$\hat{F} = \frac{SS_{REG(k+1,\ldots,p|1,2,\ldots,k)}/(p-k)}{SS_{ERROR}/(n-p)}$$

has an F distribution with $(p-k)$ degrees of freedom 'for the numerator' and $(n-p)$ degrees of freedom 'for the denominator.'

It is common practice to summarize in a table the calculations of the F-statistics for testing the null hypothesis that the last $(p-k)$ components of the regression parameter are zero:

ANOVA Table

Source	Sum of Squares	Degrees of Freedom	Mean Square	F-Statistic
Regression on '$1, 2, \ldots, k$'	$\hat{y}_I^T \hat{y}_I$	k	—	
Regression on '$k+1, \ldots, p$' Adjusting for '$1, 2, \ldots, k$'	$\hat{y}^T \hat{y} - \hat{y}_I^T \hat{y}_I$	(p-k)	$MS_0 = \frac{\hat{y}^T \hat{y} - \hat{y}_I^T \hat{y}_I}{(p-k)}$	$\hat{F} = \frac{MS_0}{MS_{Error}}$
Error	$\hat{\epsilon}^T \hat{\epsilon}$	$(n-p)$	$MS_{Error} = \frac{\hat{\epsilon}^T \hat{\epsilon}}{(n-p)}$	
Total	$y^T y$	n		

MIT OpenCourseWare
http://ocw.mit.edu

18.S096 Mathematical Applications in Financial Industry
Fall 2013

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.