Statistics 312/612, fall 2016
Homework # 10
Due: Friday 9 December

Please attempt this homework by yourself, with no help from others. Please cite explicitly any sources that you use.

Consider again the usual least squares fit obtained by choosing the vector $b \in \mathbb{R}^p$ to minimize $\|y - Xb\|^2$, where $y \in \mathbb{R}^n$ and X is a given $n \times p$ matrix, not necessarily of full rank.

Suppose we are worried about the observation y_i. For concreteness take $i = 1$ and partition the matrices as

$$y = \begin{pmatrix} y_1 \\ Y \end{pmatrix} \qquad \text{and} \qquad X = \begin{pmatrix} w_1^T \\ W \end{pmatrix} \qquad \text{where } W := \begin{pmatrix} w_2^T \\ \vdots \\ w_n^T \end{pmatrix},$$

where Y is the $(n-1) \times 1$ vector $[y_2, \dots, y_n]'$ and W is the $(n-1) \times p$ matrix obtained by deleting the first row, w_1^T, from X.

Let $\mathcal{X}$ denote the subspace of $\mathbb{R}^n$ spanned by the columns of X. Let H denote the hat matrix for orthogonal projection onto $\mathcal{X}$. Let $\mathcal{W}$ denote the subspace of $\mathbb{R}^p$ spanned by $\{w_i : 2 \le i \le n\}$.

The least squares estimator $\widehat{b}$ is defined to minimize $\|y - Xb\|^2$. If $\text{rank}(X) = k < p$ then $\widehat{b}$ is not unique, but all solutions give the same fitted value $X\widehat{b} = \widehat{y} = Hy$, where H denotes the hat matrix for orthogonal projection onto $\mathcal{X}$. Similarly, the $\widehat{B}$ that minimizes $\|Y - Wb\|^2$ need not be unique but all solutions give the same $\widehat{Y} = W\widehat{B}$.

There are various diagnostic procedures that try to detect bad violations of the normality assumption. This Homework describes three seemingly different diagnostics that turn out to be almost equivalent.

[1] Write e_1 for the unit vector with 1 in the first position.

(i) (10 points) Show that

$$\|y - Xb - e_1 c\|^2 = (y_1 - w_1^T b - c)^2 + \|Y - Wb\|^2$$

and that the left-hand side is minimized by choosing b equal to any $\widehat{B}$ that minimizes $\|Y - Wb\|^2$ and then choosing $\widehat{c}$ appropriately. Find $\widehat{c}$.

(ii) (5 points) Explain why $\widehat{c}$ takes the same value for all choices of $\widehat{B}$ in (i) if and only if w_1 lies in $\mathcal{W}$. Hint: overparametrized handout.

(iii) (5 points) If $e_1 \in \mathcal{X}$ show that $H_{11} = 1$. *(Here H_{11} denotes the element $H[1,1]$.)*

(iv) (5 points) If $e_1 \in \mathcal{X}$ show that $w_1 \notin \mathcal{W}$. Hint: $w_1 = X^T e_1$.

(v) (5 points) If $e_1 \notin \mathcal{X}$ show that $H_{11} < 1$ and $w_1 \in \mathcal{W}$. Hint: $(I - H)e_1 \perp \mathcal{X}$.

From now on assume $e_1 \notin \mathcal{X}$. Write $\widetilde{\mathcal{X}}$ for the subspace spanned by the e_1 and the columns of X. Write κ for $\sqrt{1 - H_{11}}$, the length of the vector $z := (I - H)e_1$.

[2] Define $q_0 := z/\kappa$. Let $\{q_j : 1 \le j \le k\}$ be an onb for $\mathcal{X}$.

(i) (5 points) Explain why $\{q_j : 0 \le j \le k\}$ is an onb for $\widetilde{\mathcal{X}}$.

(ii) (5 points) Show that $\widetilde{H} = H + q_0 q_0^T$ is the hat matrix for orthogonal projection onto $\widetilde{\mathcal{X}}$.

(iii) (10 points) Show that the component of y in the q_0 direction equals $\widehat{c}z$.

(iv) (10 points) If $y \sim N(\mu, \sigma^2 I_n)$ with $\mu \in \mathcal{X}$, show that $\widehat{c} \sim N(0, \sigma^2/\kappa^2)$.

[3] Define $\widehat{\sigma}^2 = \|y - X\widehat{b}\|^2/(n-k)$ and $\widehat{S}^2 = \|Y - W\widehat{B}\|^2/(n-k-1)$. Suppose $y \sim N(\mu, \sigma^2 I_n)$ with $\mu \in \mathcal{X}$.

(i) (10 points) Show that the statistic

$$\mathrm{ESR}_1 := \kappa\widehat{c}/\widehat{S} = q'y/\widehat{S}$$

has a t_{n-k-1} distribution.

(ii) (extra credit) Define

$$\mathrm{ISR}_1 := \kappa\widehat{c}/\widehat{\sigma} = q'y/\widehat{\sigma}.$$

Show that ISR_1 is a monotonely increasing function of ESR_1.

[4] (extra credit) Define

$$\mathcal{D}_1 = \frac{\|X\widehat{b} - X\widehat{B}\|^2}{k\widehat{\sigma}^2}.$$

Show that $\mathcal{D}_1$ is a monotonely increasing function of $|\mathrm{ISR}_1|$.