Stat 151a Linear Models
Homework 2 Solutions

October 5, 2015

1. Defining $a = (a_1, \ldots, a_n)$, we have from the question $\mathbb{E}(a^T y) = \beta_1$. Since $\mathbb{E}(y) = X\beta$, we have

$$a^T X\beta = \beta_1$$

or another way of writing the same thing,

$$(X^T a)^T \beta = \beta_1$$

Defining a new vector c as $X^T a$ we notice two things, Firstly, $c \in \mathcal{R}(X)$ (since it is a linear combination of rows of X). Secondly, $c^T \beta = \beta_1$. By the definition of estimability, this means that β_1 is estimable.

2. (a) Since the model is $y_i = \beta_0 + \beta_1 + e_i$, the design matrix X looks like,

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix}_{n\times 2}$$

(b) The normal equations are

$$(X^T X)\beta = X^T y$$
$$\text{Or, } \begin{pmatrix} n & n \\ n & n \end{pmatrix} \beta = \begin{pmatrix} \sum_i y_i \\ \sum_i y_i \end{pmatrix}$$
$$\text{Dividing both sides by } n, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \beta = \begin{pmatrix} \bar{y} \\ \bar{y} \end{pmatrix}$$

So we only have one equation in 2 unknowns,

$$\beta_0 + \beta_1 = \bar{y} \tag{1}$$

Any values of β_0, β_1 which satisfy this equation 1 satisfies the normal equations. For example,

$$\beta_0 = 0, \beta_1 = \bar{y} \qquad\qquad \beta_0 = \frac{\bar{y}}{3}, \beta_1 = \frac{2\bar{y}}{3}$$

are both (non-unique) solutions to normal equations.

(c) Recall from Lecture 6 notes (just after Equation 1), that the least squares estimates of $v_i^T \beta$ are

$$\widehat{v_i^T \beta}_{ls} = \frac{u_i^T y}{\sigma_i}$$

Since v_i are orthonormal, this implies that the least squares estimate of $c^T\beta$, where c is assumed to be in row space of X, would be

$$\widehat{c^T\beta}_{ls} = \sum_i (c, v_i)\widehat{v_i^T\beta}_{ls} = \sum_i (c, v_i)\frac{u_i^T y}{\sigma_i}$$

But this looks a little cumbersome, going back to Lecture 6, right after equation 2, we see that

$$\hat{\beta}_{ls} = \sum_i \frac{u_i^T y}{\sigma_i} v_i == (X^T X)^- X^T y \tag{2}$$

which, by the magical properties of sets of orthonormal vectors, gives a more manageable expression,

$$\widehat{c^T\beta}_{ls} = \sum_i (c, v_i)(\hat{\beta}_{ls}, v_i) = (c, \hat{\beta}_{ls})$$

where in the last expression we have again used the fact that $c \in \mathcal{R}(X)$. A subtlety here is that as long as $c \in \mathcal{R}(X)$ we do not need β, the full vector, to be estimable. In which case, $\hat{\beta}_{ls}$ is simply formally defined as in equation 2. This is a derivation we will only do once and store away for posterity in many problems. Coming back to the precise problem we have at hand, we only need $\hat{\beta}_{ls}$. Remember that,

$$\hat{\beta}_{ls} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}^- \begin{pmatrix} \bar{y} \\ \bar{y} \end{pmatrix}$$

where the singular value decomposition of $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = 2 \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

giving

$$\hat{\beta}_{ls} = \begin{pmatrix} \bar{y}/2 \\ \bar{y}/2 \end{pmatrix}$$

Finally, the least square estimate of $\beta_0 + \beta_1$ is $\bar{y}$.

Alternatively, Note that, we get the same answer by using any of the solutions we found in the previous part. For example,

$$0 + \bar{y} = \bar{y}$$

or,

$$\frac{\bar{y}}{3} + \frac{2\bar{y}}{3} = \bar{y}$$

(d) In this case, the design matrix gains an additional row

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 2 \end{pmatrix}_{(n+1)\times 2}$$

Now, $X^T X$,

$$X^T X = \begin{pmatrix} n+1 & n+2 \\ n+2 & n+4 \end{pmatrix}$$

is invertible with inverse

$$(X^T X)^{-1} = \frac{1}{n} \begin{pmatrix} n+4 & -(n+2) \\ -(n+2) & n+1 \end{pmatrix}$$

Giving least squares estimates of β as

$$\hat{\beta}_{ls} = (X^T X)^{-1} X^T y = \frac{1}{n} \begin{pmatrix} n+4 & -(n+2) \\ -(n+2) & n+1 \end{pmatrix} \begin{pmatrix} \sum_1^{n+1} y_i \\ \sum_1^{n+1} y_i + y_{n+1} \end{pmatrix}$$

Simplifying

$$\hat{\beta}_{ls} = \begin{pmatrix} \frac{2\sum y_i - (n+2) y_{n+1}}{n} \\ \frac{(n+1) y_{n+1} - \sum y_i}{n} \end{pmatrix}$$

3. The design matrix can be written as

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & 1 & \dots & 0 \\ 1 & 0 & 0 & 0 & \ddots & 0 \\ 1 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}_{n \times (n+1)} = \begin{pmatrix} \mathbf{1}_n & I_{n \times n} \end{pmatrix}$$

It is easy to notice that the vector $(1, -1, \dots, -1)$ is orthogonal to all rows of X (check that the inner products are 0). That is,

$$X \begin{pmatrix} 1 \\ -1 \\ \vdots \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Further, notice that the submatrix constructed by removing the first column from X is just the idenity matrix of size n, thus all except the first column of X are linearly independent (the first column is the sum of all other columns). This implies, $rank(X) = n$.

Combining these two observations, we conclude that $(1, -1, \dots, -1)$ is the only non-zero vector orthogonal to $\mathcal{R}(X)$. Another way to write this down is

$$\mathcal{R}(X) = \{x \in \mathbb{R}^{n+1} | x^T (1, -1, \dots, -1) = 0\}$$

This let's us easily verify whether any particular linear combination of β's in estimable or not.

(a) $(1, 0, 1, 0, \dots)^T (1, -1, \dots, -1) = 0$ so $\beta_0 + \beta_2$ is estimable

(b) $(0, 1, 0, 0, \dots)^T (1, -1, \dots, -1) \neq 0$ so β_1 is not estimable

(c) $(0, -1, 1, 0, \dots)^T (1, -1, \dots, -1) = 0$ so $\beta_1 - \beta_2$ is estimable

(d) $(0, 1, 1, 1, -3, 0 \dots)^T (1, -1, \dots, -1) = 0$ so $\beta_1 + \beta_2 + \beta_3 - 3\beta_4$ is estimable

For the second part, we try three different solutions. Firstly, notice that the normal equations can be parsed as follows

$$\begin{aligned} n\beta_0 + \beta_1 + \dots \beta_n &= \sum y_i \\ \beta_0 + \beta_1 &= y_1 \\ \beta_0 + \beta_2 &= y_2 \\ &\vdots \\ \beta_0 + \beta_n &= y_n \end{aligned}$$

Solution 1: By inspection we notice that

$$\hat{\beta} = \begin{pmatrix} 0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

satisfies the normal equations. Further, using this $\hat{\beta}$, we find that $\hat{y}_i = y_i$, so the residual sum of squares $= \sum_i (\hat{y}_i - y_i)^2 = 0$. Since, this is the minimum possible value for residual sum of squares, $\hat{\beta}$ is indeed a least squares solution. So, using $\widehat{c^T\beta} = c^T\hat{\beta}$, we get

(a) Least squares estimate of $\beta_0 + \beta_2$ is y_2

(b) β_1 is not estimable as shown before

(c) Least squares estimate of $\beta_1 - \beta_2$ is $y_1 - y_2$

(d) Least squares estimate of $\beta_1 + \beta_2 + \beta_3 - 3\beta_4$ is $y_1 + y_2 + y_3 - 3y_4$

Solution 2: A closely related solution is to notice that from the normal equations (specifically from the third equation),

$$\beta_0 + \beta_2 = y_2$$

Similarly, subtracting the second from the third,

$$\beta_1 - \beta_2 = y_1 - y_2$$

Adding the second, third and fourth and subtracting the fifth,

$$\beta_1 + \beta_2 + \beta_3 - 3\beta_4 = y_1 + y_2 + y_3 - 3y_4$$

Solution 3: Here we attempt the full singular value decomposition of X^TX. We have already established that all the singular vectors must be orthogonal to the vector $(1, -1, -1, \cdots - 1)^T$. Now we pull another singular vector out of the hat,

$$X^TX \begin{pmatrix} n \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} n & \mathbf{1}^T \\ \mathbf{1} & I \end{pmatrix} \begin{pmatrix} n \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} n^2 + n \\ (n+1)\mathbf{1} \end{pmatrix} = (n+1) \begin{pmatrix} n \\ \mathbf{1} \end{pmatrix}$$

This gives us $\sigma^2 = n + 1$ and

$$v_1 = \frac{1}{n^2 + n} \begin{pmatrix} n \\ \mathbf{1} \end{pmatrix}$$

Any other v's, that is $v_2, \ldots, v_n$ have to be orthogonal to v_1 in addition to being orthogonal to $(1, -1, \ldots, -1)$. Let's write a candidate v as $(x_1, \ldots, x_n)$. Then from the above mentioned orthogonality constraints we know,

$$\begin{aligned} -nx_1 &= x_2 + \cdots + x_{n+1} \\ x_1 &= x_2 + \cdots + x_{n+1} \end{aligned}$$

Or,

$$\begin{aligned} x_1 &= 0 \\ x_2 + \cdots + x_{n+1} &= 0 \end{aligned}$$

Now, post multiplying X^TX with such a vector,

$$X^TXv = \begin{pmatrix} n & \mathbf{1}^T \\ \mathbf{1} & I \end{pmatrix} \begin{pmatrix} 0 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_2 + \cdots + x_n \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = 1 \times \begin{pmatrix} 0 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Take a moment to understand what this means. This means that all the other singular values, $\sigma_2^2 = \cdots = \sigma_n^2 = 1$ and any set of $n-1$ orthogonal vectors which are orthogonal to v_1 and $(1, -1, \ldots, -1)$ are good enough to be $v_2, \ldots, v_n$. We don't need to fix values for these $v_2, \ldots, v_n$.

Finally,

$$X^T X = (n+1) v_1^T v_1 + \sum_2^n v_i^T v_i$$

Or,

$$(X^T X)^- = \frac{1}{n+1} v_1^T v_1 + \sum_2^n v_i^T v_i = X^T X + \left(\frac{1}{n+1} - (n+1)\right) v_1^T v_1$$

Finally giving us,

$$\hat{\beta}_{ls} = (X^T X)^- X^T y = (X^T X)^{-1} \begin{pmatrix} \sum y_i \\ y_1 \\ \vdots \\ y_n \end{pmatrix} =$$

After some algebra from this expression we read off the same estimates.

4. Here, it is immediate to notice that the linear model is not identifiable. Constructing a design matrix $X_{252\times 5}$ with columns **1**, `AGE`, `WEIGHT`, `HEIGHT` and `AGE + 10*WEIGHT + 3*HEIGHT`, $\text{rank}(X) = 4$. In such cases, one can construct infinitely many equally valid choices of $\hat{\beta}$ each differing by multiples of vectors in the null space of X (orthogonal to row space of X).

   (a) The first estimate can be obtained by running `lm` on (`BODYFAT`,X), which gives an expected `NA` in the last entry which can be interpreted as 0:

   $$\hat{\beta}(1) = \texttt{17.7673848 0.1697902 0.1981519 -0.5943339 NA}$$

   The second estimate is by singular value decomposition, $\hat{\beta} = (X^T X)^- X^T y$,

   $$\hat{\beta}(2) = \texttt{17.7673848 0.1664721 0.1649710 -0.6042881 0.0033180}$$

   The third can be constructed by adding a multiple of v_5 to $\hat{\beta}(2)$, in this case we choose the multiple to be 1,

   $$\hat{\beta}(3) = \texttt{17.767385 0.071556 -0.784187 -0.889035 0.098234}$$

   (b) β_1 is not estimable. It is clear that since X is rank-deficient, the linear model can be rewritten as

   $$\texttt{BODYFAT} \approx \beta_0 + (\beta_1 + \beta_4)\texttt{AGE} + (\beta_2 + 10\beta_4)\texttt{WEIGHT} + (\beta_3 + 3\beta_4)\texttt{HEIGHT}$$

   where the natural parameters to estimate are $\beta_0, \beta_1 + \beta_4, \beta_2 + 10\beta_4, \beta_3 + 3\beta_4$. Since β_1 is not a linear combination of them it can not be estimated.
   This can be further rigorised by observing that v_5 is the only direction orthogonal to $\mathcal{R}(X)$, (similar to problem 3). Now one can verify that $(0,1,0,0,0)^T v_5 = \texttt{-0.0949158} \neq 0$ so $(0,1,0,0,0)^T \beta = \beta_1$ is not estimable.

   (c) The least squares estimate of $\beta_0, \beta_1 + \beta_4, \beta_2 + 10\beta_4, \beta_3 + 3\beta_4$ are `17.7673848 0.1697902 0.1981519 -0.5943339`

   (d) The estimates can be read off from the output of the R code provided in the question

5. Notice that, the columns of X are orthogonal to each other. In fact,

$$X^T X = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix}$$

Leading to least squares estimates

$$\hat{\beta}_{ls} = \frac{1}{8}\begin{pmatrix} \sum_i y_i \\ (y_1+y_3+y_5+y_7)-(y_2+y_4+y_6+y_8) \\ (y_1+y_2+y_5+y_6)-(y_3+y_4+y_7+y_8) \\ (y_1+y_4+y_5+y_8)-(y_2+y_3+y_6+y_7) \end{pmatrix}$$

6. In general, under the model $y = X\beta + e$, where e are iid mean 0 variance σ^2, the least squares estimate has mean β and covariance $\sigma^2(X^TX)^{-1}$. Here, $(X^TX)^{-1} = \frac{1}{8}I_4$ implying that every β_j can be estimated with variance $\sigma^2/8$.

7. For any non-negative matrix Σ we have the following general result,

$$(\Sigma^{-1})_{ii} \geq \frac{1}{\Sigma_{ii}}$$

In our case, we have,

$$\mathrm{Var}(\hat{\beta}_i) = \sigma^2((X^TX)^{-1})_{ii} \geq \frac{\sigma^2}{\sum_j x_{ji}^2} \geq \frac{\sigma^2}{8}$$

since $x_{ji} \in \{0, -1, 1\}$. In conclusion it is not possible to have a design matrix that leads to a variance strictly smaller than $\sigma^2/8$ for any $\hat{\beta}_i$