Homework Two

Statistics 151a (Linear Models)

Due by 11:59 pm on September 18, 2018

September 6, 2018

1. Consider the linear model $Y = X\beta + e$ with $\beta = (\beta_0, \beta_1, \ldots, \beta_p)^T$. Suppose I can find real numbers $a_1, \ldots, a_n$ such that

$$\mathbb{E}\left(a_1 Y_1 + \ldots a_n Y_n\right) = \beta_1.$$

   Show then that β_1 is estimable. (**0.8 points**).

2. Consider the data: $Y_i = \beta_0 + \beta_1 + e_i$ where $e_1, \ldots, e_n$ are uncorrelated errors with mean zero and variance σ^2.

   a) Write this model in the form $Y = X\beta + e$ with $\beta = (\beta_0, \beta_1)^T$. Specify the matrix X. (**0.3 point**)

   b) Write down the normal equations. Find a solution to them. Is the solution unique? (**0.5 points**).

   c) What is the least squares estimate of $\beta_0 + \beta_1$? (**0.5 points**).

   d) Is β_1 estimable? (**0.5 points**)

   e) Consider now another observation $Y_{n+1} = \beta_0 + 2\beta_1 + e_{n+1}$ where $e_1, \ldots, e_{n+1}$ are uncorrelated errors with mean zero and variance σ^2. Write this model in the form $Y = X\beta + e$ and calculate the least squares estimate of β. (**1 points**).

3. Consider the model: $y_i = \beta_0 + \beta_i + e_i$ for $i = 1, \ldots, n$ where $e_1, \ldots, e_n$ are mean zero and uncorrelated errors. For each of the following parameter functions, specify whether they are estimable or not. If estimable, provide their least squares estimate. If not, explain why.

   a) $\beta_0 + \beta_2$ (**0.5 points**)

   b) β_1 (**0.5 points**)

c) $\beta_1 - \beta_2$ (**0.5 points**)

d) $\beta_1 + \beta_2 + \beta_3 - 3\beta_4$ (**0.5 points**).

4. In the Bodyfat dataset (available in bcourses/ Files/Lab 1), consider the linear model:

$$\text{BODYFAT} = \beta_0 + \beta_1\text{AGE} + \beta_2\text{WEIGHT} + \beta_3\text{HEIGHT} + \beta_4\left(\text{AGE} + 10{*}\text{WEIGHT} + 3{*}\text{HEIGHT}\right) + e$$

   a) Using R, provide three different least squares estimates of $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T$ (**1 points**).

   b) Is β_1 estimable? Why or why not? (**0.5 points**).

   c) Find the least squares estimates (using R) of β_0, $(\beta_1 + \beta_4), (\beta_2 + 10\beta_4)$ and $\beta_3 + 3\beta_4$. (**0.5 points**).

   d) Can the estimates above be read off from the output to the following command in R? (**1 point**).

```
summary(lm(BODYFAT ~ AGE + WEIGHT + HEIGHT + I(AGE + 10*WEIGHT +
3*HEIGHT), data = bodyfat))
```

5. Suppose there are 4 objects whose individual weights $\beta_1, \ldots, \beta_4$ need to be estimated. We have a weighing balance which gives measurements with error having mean zero and variance σ^2. One approach is to weigh each object a number of times and take the average measurement value as the estimate of its weight. Such a procedure needs a total of 32 weighings (8 for each of the 4 objects) to estimate the weight of each object with precision (variance) $\sigma^2/8$.

   Another method is to weigh the objects in combinations. Each operation consists in placing some of the objects in one pan of the balance and the others in the other pan. One then places some weights in the two pans to achieve equilibrium. This results in an observational equation of the type

$$y = x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + x_4\beta_4 + e$$

   where x_i is 0, 1 or -1 according as the ith object is not used, placed in the left pan or in the right pan of the balance and y is the weight required for equilibrium. After n measurements, one can get data that can be represented in an $n \times 1$ vector Y and an $n \times 4$ matrix X.

a) Suppose $n = 8$ and

$$X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

What are the least squares estimates of $\beta_1, \beta_2, \beta_3$ and β_4? (**0.7 points**).

b) For X as above, what is the Covariance matrix of $\hat{\beta}$? Show that this scheme gives a way of taking 8 weighings and estimating all the weights with individual precision $\sigma^2/8$. This should be contrasted with the naive weighing scheme described previously that takes 32 weighings to get estimates with precision $\sigma^2/8$. (**0.7 points**).