Statistics 151 (Linear Modelling -Theory and Applications)
Homework One

Due on 09 February, 2015

30 January, 2015

1. Consider the linear model $Y = X\beta + e$ where $\mathbb{E}e = 0$ and $Cov(e) = \sigma^2 I_n$. 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.

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.

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

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

   (d) Is β_1 estimable?

   (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.

3. Consider a simple one-way Analysis of Variance 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) $\beta_0 + \beta_2$

   (b) β_1

   (c) $\beta_1 - \beta_2$

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

4. In the Bodyfat dataset, consider the linear model:

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

   (a) Is β_1 estimable?

   (b) Find the least squares estimates (using R) of β_0, $(\beta_1 + \beta_4), (\beta_2 + 10\beta_4)$ and β_3.

   (c) Can the estimates above be read off from the output to the following command in R?

```
summary(lm(BODYFAT ~ AGE + WEIGHT + HEIGHT + I(AGE + 10*WEIGHT), data
= bodyfata.dataset))
```

5. Consider simple linear regression where there is one response variable y and an explanatory variable x and there are n subjects with values $y_1, \ldots, y_n$ and $x_1, \ldots, x_n$.

   (a) Write down (no need to calculate) the least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ of β_0 and β_1 in the model $y_i = \beta_0 + \beta_1 x_i + e_i$.

   (b) Write down (again no need to calculate) the estimates $\hat{\alpha}_0$ and $\hat{\alpha}_1$ of α_0 and α_1 in the model $x_i = \alpha_0 + \alpha_1 y_i + e_i$.

   (c) Intuition might suggest that $\hat{\alpha}_1 = 1/\hat{\beta}_1$. Is this true?

   (d) Consider the BODYFAT dataset with $y =$ BODYFAT and $x =$ THIGH. Plot the data and the two lines $y = \hat{\beta}_0 + \hat{\beta}_1 x$ and $x = \hat{\alpha}_0 + \hat{\alpha}_1 y$.

6. Consider the Anscombe dataset available in R which can be accessed (and plotted) via

```
library(datasets)
a <- anscombe
par(mfrow=c(2,2))
plot(a$x1,a$y1, main=paste("Dataset One"))
plot(a$x2,a$y2, main=paste("Dataset Two"))
plot(a$x3,a$y3, main=paste("Dataset Three"))
plot(a$x4,a$y4, main=paste("Dataset Four"))
```

   (a) For each of these four datasets, fit a linear model for the response variable on the explanatory variable (including the intercept term). Plot these four datasets (in the same graphics window as above) along with the fitted regression lines. Does the linear model make sense for these datasets?

   (b) In each of the four datasets, predict the response variable when the explanatory variable equals 10. Do these predictions make sense?

7. 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 is the least squares estimate of $\beta_1, \beta_2, \beta_3$ and β_4?

(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 $\sigma^2/8$.

(c) **(Bonus)** Does there exist a scheme of designing the 8×4 matrix X (each of whose elements is one among 0, 1 and -1) so that the variance of any of the 4 weight estimates is strictly smaller than $\sigma^2/8$? Why or why not?