

Stat 151A - Homework 1

Michael Knopf

February 9, 2015

1. Consider the linear model $Y = X\beta + e$ where $\mathbb{E}e = 0$ and $\text{Cov}(e) = \sigma^2 I_n$. Suppose I can find real numbers a_1, \dots, a_n such that

$$\mathbb{E}[a_1 Y_1 + \dots + a_n Y_n] = \beta_1.$$

Show then that β_1 is estimable.

Proof. Let $a = (a_1, \dots, a_n)$. Then $\beta_1 = \mathbb{E}[a^T Y] = a^T \mathbb{E}[X\beta + e] = a^T (X\beta + 0) = a^T X\beta = (X^T a)^T \beta$. Since $X^T a \in \text{Col}(X^T)$, β_1 is estimable.

□

2. Consider the data: $Y_i = \beta_0 + \beta_1 + e_i$ where e_1, \dots, 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 .

$$Y = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

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

Proof. The normal equations are given by

$$\begin{aligned} X^T X \hat{\beta} &= X^T Y \\ \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} &= \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \\ \begin{pmatrix} n & n \\ n & n \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} &= \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n Y_i \end{pmatrix} \\ \begin{pmatrix} \hat{\beta}_0 + \hat{\beta}_1 \\ \hat{\beta}_0 + \hat{\beta}_1 \end{pmatrix} &= \begin{pmatrix} \bar{Y} \\ \bar{Y} \end{pmatrix}. \end{aligned}$$

Therefore, the solution set for the normal equations is $\{(\hat{\beta}_0, \hat{\beta}_1) : \hat{\beta}_0 + \hat{\beta}_1 = \bar{Y}\}$. One solution is $\hat{\beta} = (\bar{Y}, 0)$. However, this solution is not unique. □

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

The least squares estimate of $\beta_0 + \beta_1$ is \bar{Y} .

(d) Is β_1 estimable?

No. The unique expression of β_1 in the form $\lambda^T \beta$ is $(0 \ 1) \beta$, but $\begin{pmatrix} 0 \\ 1 \end{pmatrix} \notin \text{Col}(X^T) = \text{span} \left(\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} \right)$

(e) Consider now another observation $Y_{n+1} = \beta_0 + 2\beta_1 + e_{n+1}$ where e_1, \dots, 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 .

Proof.

$$Y = \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} e_1 \\ \vdots \\ e_n \\ e_{n+1} \end{pmatrix}$$

The normal equations are given by

$$\begin{aligned} X^T X \hat{\beta} &= X^T Y \\ \begin{pmatrix} 1 & \cdots & 1 & 1 \\ 1 & \cdots & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} &= \begin{pmatrix} 1 & \cdots & 1 & 1 \\ 1 & \cdots & 1 & 2 \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \\ Y_{n+1} \end{pmatrix} \\ \begin{pmatrix} n+1 & n+2 \\ n+2 & n+4 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} &= \begin{pmatrix} \sum_{i=1}^{n+1} Y_i \\ \sum_{i=1}^{n+1} Y_i + Y_{n+1} \end{pmatrix} \end{aligned}$$

Note that $\det(X^T X) = (n+1)(n+4) - (n+2)^2 = n$. We may assume that $n \neq 0$, thus $X^T X$ is invertible. Multiplying both sides by $(X^T X)^{-1}$ gives

$$\begin{aligned} \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} &= \frac{1}{n} \begin{pmatrix} n+4 & -n-2 \\ -n-2 & n+1 \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{n+1} Y_i \\ \sum_{i=1}^{n+1} Y_i + Y_{n+1} \end{pmatrix} \\ &= \frac{1}{n} \begin{pmatrix} 2 \sum_{i=1}^{n+1} Y_i - (n+2)Y_{n+1} \\ -\sum_{i=1}^{n+1} Y_i + (n+1)Y_{n+1} \end{pmatrix} = \frac{n+1}{n} \begin{pmatrix} 2\bar{Y} - \frac{n+2}{n+1}Y_{n+1} \\ -\bar{Y} + Y_{n+1} \end{pmatrix} \end{aligned}$$

which is 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 $i = 1, \dots, n$ where e_1, \dots, 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) $\beta_0 + \beta_2$
Yes

(b) β_1
No

(c) $\beta_1 - \beta_2$
Yes

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

Proof. The design matrix and its transpose are given below.

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \quad X^T = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

Let v_1, \dots, v_n be the columns of X^T , so that its column space is $\text{span}(\{v_1, \dots, v_n\})$. A parameter is estimable if and only if we can express it as $\lambda^T \beta$ for some $\lambda \in \text{Col}(X^T)$. So we express each given parameter in this form and check whether the corresponding λ is in the column space of X^T .

$$\begin{aligned} \beta_0 + \beta_2 &= v_2^T \beta \\ \beta_1 &= (0 \ 1 \ 0 \ \cdots \ 0) \beta \\ \beta_1 - \beta_2 &= (v_1 - v_2)^T \beta \\ \beta_1 + \beta_2 + \beta_3 - 3\beta_4 &= (v_1 + v_2 + v_3 - 3v_4)^T \beta \end{aligned}$$

The only parameter for which λ which is not in $\text{Col}(X^T)$ is β_1 .

To find the least squares estimates of these parameters, note that

$$X^T X \beta = X^T Y$$

$$\begin{pmatrix} n & 1 & 1 & \cdots & 1 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_n \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

$$\begin{pmatrix} n\hat{\beta}_0 + \hat{\beta}_1 + \cdots + \hat{\beta}_n \\ \hat{\beta}_0 + \hat{\beta}_1 \\ \hat{\beta}_0 + \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_0 + \hat{\beta}_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}.$$

The least squares estimates of the estimable parameters are

$$\begin{aligned} \hat{\beta}_0 + \hat{\beta}_2 &= Y_2 \\ \hat{\beta}_1 - \hat{\beta}_2 &= (\hat{\beta}_0 + \hat{\beta}_1) - (\hat{\beta}_0 + \hat{\beta}_2) = Y_1 - Y_2 \\ \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 - 3\hat{\beta}_4 &= (\hat{\beta}_0 + \hat{\beta}_1) + (\hat{\beta}_0 + \hat{\beta}_2) + (\hat{\beta}_0 + \hat{\beta}_3) - 3(\hat{\beta}_0 + \hat{\beta}_4) \\ &= Y_1 + Y_2 + Y_3 - 3Y_4 \end{aligned}$$

□

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

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

(a) Is β_1 estimable?

No

(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 + HEIGHT + I(AGE + 10*WEIGHT), data = bodyfat.dataset))
```

Yes

Proof. (a) β_1 is not estimable. Since $\beta_1 = (0 \ 1 \ 0 \ 0 \ 0)\beta$, we would need $(0 \ 1 \ 0 \ 0 \ 0)^T \in \text{Col}(X^T)$. So assume

$$\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \sum_{i=1}^{252} a_i \begin{pmatrix} 1 \\ x_{i1} \\ x_{i2} \\ x_{i3} \\ x_{i1} + 10x_{2i} \end{pmatrix}.$$

The resulting system of equations requires that

$$\sum_{i=1}^{252} a_i x_{i1} = 1 \quad \sum_{i=1}^{252} a_i x_{i2} = 0 \quad \sum_{i=1}^{252} a_i x_{i1} + 10 \sum_{i=1}^{252} a_i x_{i2} = 0$$

all be solved simultaneously by some a_1, \dots, a_{252} . However, substituting the second equation into the third gives $\sum_{i=1}^{252} a_i x_{i1} = 0$, which contradicts the first.

(b) We can edit the formula to leave out the last term $I(\text{AGE} + 10 \cdot \text{WEIGHT})$. The estimates for the coefficients given for (INTERCEPT) , AGE , WEIGHT , and HEIGHT are β_0 , $(\beta_1 + \beta_4)$, $(\beta_2 + 10\beta_4)$, and β_3 , respectively:

```
body = read.csv("BodyFat.csv")
#ft = lm(BODYFAT ~ AGE + WEIGHT + HEIGHT + I(AGE + 10*HEIGHT), data = body)
ft = lm(BODYFAT ~ AGE + WEIGHT + HEIGHT, data = body)
summary(ft)
```

Output:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	17.72142	6.92955	2.557	0.0111 *
AGE	0.15583	0.02739	5.690	3.57e-08 ***
WEIGHT	0.18373	0.01216	15.107	< 2e-16 ***
HEIGHT	-0.55099	0.09904	-5.563	6.85e-08 ***

(c) If we keep the last term in the formula, R actually gives these same coefficients for each of these variables, meaning that it is actually estimating β_0 , $(\beta_1 + \beta_4)$, $(\beta_2 + 10\beta_4)$, and β_3 for both formulas:

```
body = read.csv("BodyFat.csv")
ft = lm(BODYFAT ~ AGE + WEIGHT + HEIGHT + I(AGE + 10*HEIGHT), data = body)
#ft = lm(BODYFAT ~ AGE + WEIGHT + HEIGHT, data = body)
summary(ft)
```

Output:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	17.72142	6.92955	2.557	0.0111 *
AGE	0.15583	0.02739	5.690	3.57e-08 ***
WEIGHT	0.18373	0.01216	15.107	< 2e-16 ***
HEIGHT	-0.55099	0.09904	-5.563	6.85e-08 ***
I(AGE + 10 * HEIGHT)	NA	NA	NA	NA

□

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, \dots, y_n and x_1, \dots, 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 $y_i = \beta_0 + \beta_1 x_i + e_i$.

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = r \frac{s_y}{s_x}$$

(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 $x_i = \alpha_0 + \alpha_1 y_i + e_i$.

$$\hat{\alpha}_0 = \bar{x} - \hat{\alpha}_1 \bar{t} \qquad \hat{\alpha}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} = r \frac{s_x}{s_y}$$

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

Proof. No. $1/\hat{\beta}_1 = \frac{1}{r} \frac{s_x}{s_y} = r \frac{s_x}{s_y} = \hat{\alpha}_1$ if and only if $r = \pm 1$. However, it is true in general that $\hat{\alpha}_1 = r^2/\hat{\beta}_1$.

□

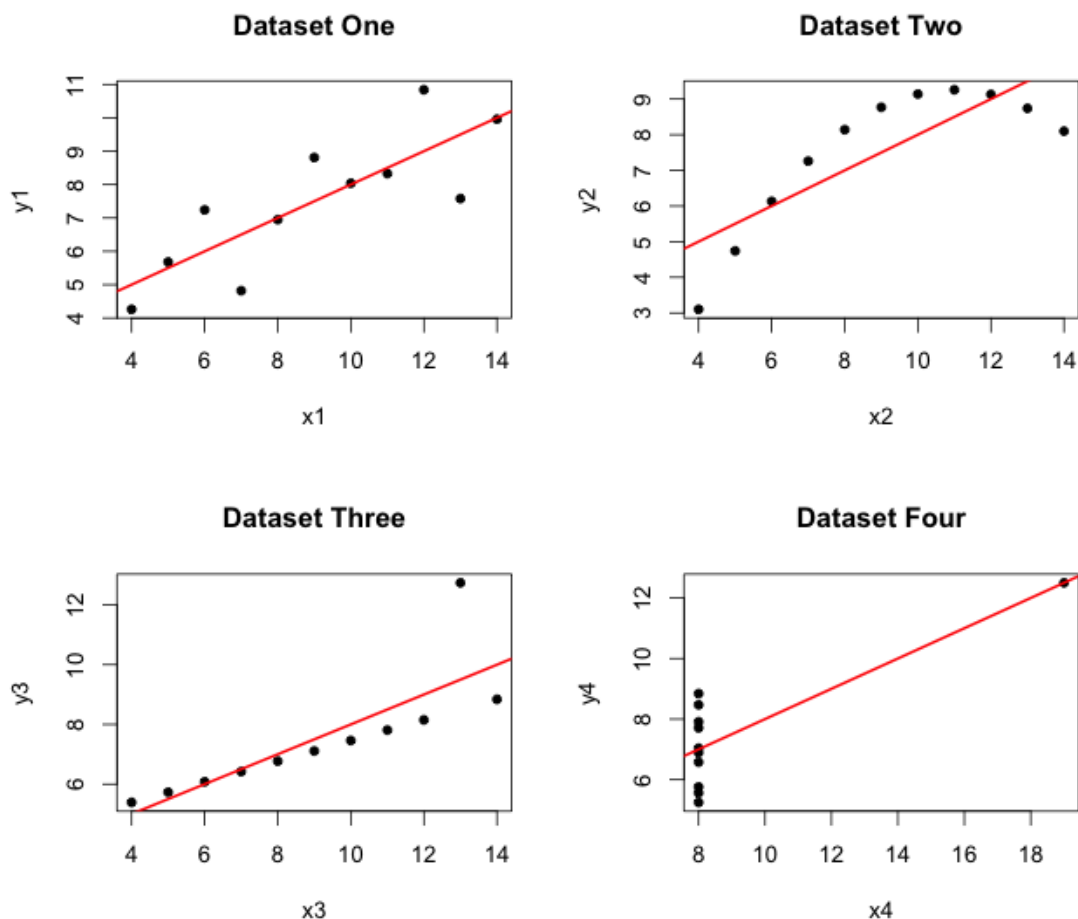
(d) Consider the BODYFAT dataset with $y = \text{BODYFAT}$ and $x = \text{THIGH}$. Plot the data and the two lines $y = \hat{\beta}_0 + \hat{\beta}_1 x$ and $x = \hat{\alpha}_0 + \hat{\alpha}_1 y$.



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?

The respective predicted responses for $x = 10$ are 8.001, 8.00909, 7.999727, and 8.000818. For dataset 1, this prediction makes sense. The dataset appears to have a linear relationship. Dataset two is clearly not linear (it looks quadratic), and dataset three contains an outlier that probably should not have been included in the model. Dataset four is a horrible set to make a linear model on, and the prediction is basically useless.

7. Suppose there are 4 objects whose individual weights β_1, \dots, β_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 of 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 i th 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 ?

Proof.

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix}$$

$$X^T Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{pmatrix} = \begin{pmatrix} y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8 \\ y_1 - y_2 + y_3 - y_4 + y_5 - y_6 + y_7 - y_8 \\ y_1 + y_2 - y_3 - y_4 + y_5 + y_6 - y_7 - y_8 \\ y_1 - y_2 - y_3 + y_4 + y_5 - y_6 - y_7 + y_8 \end{pmatrix}$$

So the estimates of the parameters are given by

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = I_4 \hat{\beta} = \frac{1}{8} X^T X = \frac{1}{8} X^T Y = \frac{1}{8} \begin{pmatrix} y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8 \\ y_1 - y_2 + y_3 - y_4 + y_5 - y_6 + y_7 - y_8 \\ y_1 + y_2 - y_3 - y_4 + y_5 + y_6 - y_7 - y_8 \\ y_1 - y_2 - y_3 + y_4 + y_5 - y_6 - y_7 + y_8 \end{pmatrix}.$$

□

(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 $\sigma^2/8$.

Proof. Since $X^T X$ is invertible, we have

$$\text{Cov}(\hat{\beta}) = (X^T X)^{-1} \sigma^2 = \frac{\sigma^2}{8} I_4.$$

So the variance of each $\hat{\beta}_i$ is $\frac{\sigma^2}{8}$. Since we accomplish the same precision with one fourth as many observations, this is obviously the better experiment. □

(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?

Proof. No. Suppose we wanted to minimize the variance of one of the estimates, regardless of how it affects the variance of the others. Our best strategy would be to observe the weight of this object 8 times and ignore the others, since observing the others does not help us estimate the weight of this one.

In this experiment, there would be only one parameter, β_1 , and the design matrix would be

$$X = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Thus $X^T X = (8)$, so $\text{Cov}(\hat{\beta}) = \left(\frac{\sigma^2}{8} \right) = \text{Var}(\hat{\beta}_1)$. Diverting our attention to observing the other weights will at best increase this variance.

Equivalently, we could use the 8×4 design matrix

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

β_1 is estimable and the least squares estimate would be \bar{Y} , which has a variance of $\frac{\sigma^2}{8}$. □