STAT 350 Lecture 2

Reading: Chapter 5 sections 1–4.

Matrix form of a linear model

Stack Y_i, μ_i and ϵ_i into vectors:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Define

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \quad X = \begin{bmatrix} x_{1,1} & \cdots & x_{1,p} \\ x_{2,1} & \cdots & x_{2,p} \\ & \vdots & \\ x_{n,1} & \cdots & x_{n,p} \end{bmatrix}_{n\times p}$$

Note

$$X\beta = \begin{bmatrix} x_{1,1}\beta_1 + \cdots + x_{1,p}\beta_p \\ \vdots \\ x_{n,1}\beta_1 + \cdots + x_{n,p}\beta_p \end{bmatrix} = \mu$$

so

$$\mu = X\beta$$

Finally

$$Y = X\beta + \epsilon$$

is our original set of n model equations written in vector matrix form.

Assumptions so far:

$$\begin{aligned} \mathrm{E}(\epsilon_i) &= 0 \\ Y &= \mu + \epsilon \\ \mu &= X\beta \end{aligned}$$

Still to come: independence, homoscedasticity, normality.

Examples: please take the point that this is a very large class of models.

1. One sample problem:
   - $Y_1, \ldots, Y_n$ measured under "identical" conditions.
   - So $\mu_1, \ldots, \mu_n = \beta_1$, say.
   - $X = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_{n\times 1}$

- $\beta = [\beta_1]_{1\times 1}$ (so $p = 1$).
- $Y = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \beta + \epsilon.$

2. Two sample problem: $n = r + s$

$$\mu_1 = \cdots = \mu_r = \beta_1 \qquad \mu_{r+1} = \cdots = \mu_{r+s} = \beta_2$$

For $i \le r$

$$Y_i = \beta_1 + \epsilon_i \qquad \mathrm{E}(Y_i) = \beta_1$$

For $r < i \le r + s$

$$Y_i = \beta_2 + \epsilon_i \qquad \mathrm{E}(Y_i) = \beta_2$$

In matrix form

$$Y = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \epsilon$$

Sometimes it is convenient to write:

$$X^T = \left[\begin{array}{c|c} \overbrace{\begin{matrix} 1 & \cdots & 1 \\ 0 & \cdots & 0 \end{matrix}}^{r \text{ cols}} & \overbrace{\begin{matrix} 0 & \cdots & 0 \\ 1 & \cdots & 1 \end{matrix}}^{s \text{ cols}} \end{array} \right]$$

which is a **partitioned** matrix where I have described the **transpose** of X.

3. Simple linear regression:

$$Y_i = TL$$

$$D_i = \text{Dose}$$

The model

$$Y_i = \beta_1 + \beta_2 D_i + \epsilon_i$$

gives

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \quad X^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ D_1 & D_2 & \cdots & D_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & D_1 \\ \vdots & \vdots \\ 1 & D_n \end{bmatrix}$$

4. Polynomial models: "polynomial regression". In we had the quadratic model:

$$Y_i = \beta_1 + D_i\beta_2 + D_i^2\beta_3 + \epsilon_i$$

for which

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \qquad X^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ D_1 & D_2 & \cdots & D_n \\ D_1^2 & D_2^2 & \cdots & D_n^2 \end{bmatrix}$$

In general we might fit a polynomial of degree $p-1$ to get

$$Y_i = \beta_1 + D_i\beta_2 + \cdots + D_i^{p-1}\beta_p + \epsilon_i$$

for which

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \qquad X^T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ D_1 & D_2 & \cdots & D_n \\ \vdots & \vdots & \cdots & \vdots \\ D_1^{p-1} & D_2^{p-1} & \cdots & D_n^{p-1} \end{bmatrix}$$

5. Analysis of Covariance: fitting two straight lines

Consider the TL data again but now suppose that samples 1 to r were "bleached" (left in the sun for several hours before analysis) and samples $r+1$ to s were "unbleached". We combine the 2 sample problem with the straight line problem:

$$\begin{aligned} \mu_i &= \beta_1 + \beta_2 D_i \qquad & i &= 1, \ldots, r \\ \mu_i &= \beta_3 + \beta_4 D_i \qquad & i &= r+1, \ldots, r+s \end{aligned}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

$$X^T = \left[\begin{array}{ccc|ccc} 1 & \cdots & 1 & 0 & \cdots & 0 \\ D_1 & \cdots & D_r & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 \\ 0 & \cdots & 0 & D_1 & \cdots & D_r \end{array}\right]$$

Special case: "No interaction" of Bleach and Dose: the effect of dose is the same for bleached and unbleached samples. That is:

$$\beta_2 = \beta_4$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

$$X^T = \left[\begin{array}{ccc|ccc} 1 & \cdots & 1 & 0 & \cdots & 0 \\ D_1 & \cdots & D_r & D_1 & \cdots & D_r \\ 0 & \cdots & 0 & 1 & \cdots & 1 \end{array}\right]$$

Note: we usually re-order the parameters in this case to get

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_3 \\ \beta_2 \end{bmatrix}$$

$$X^T = \left[\begin{array}{ccc|ccc} 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 \\ D_1 & \cdots & D_r & D_1 & \cdots & D_r \end{array}\right]$$

6. Weighing designs: (a simple example mostly for illustration) Idea: weigh two objects with (true) weights β_1 and β_2.

$$\begin{aligned} Y_1 &= \text{measured weight of Object 1} \\ Y_2 &= \text{measured weight of Object 2} \\ Y_3 &= \text{measured weight of Object 1 and 2 together} \end{aligned}$$

   Now we have

$$\mu_1 = \beta_1 \qquad \mu_2 = \beta_2 \qquad \mu_3 = \beta_1 + \beta_2$$

   and get

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 1\beta_1 + 0\beta_2 \\ 0\beta_1 + 1\beta_2 \\ 1\beta_1 + 1\beta_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

   so that

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

   Notice that $\mathrm{E}(Y_i) =$ true weight is a physically meaningful and **important** assumption. This sort of assumption may well be **wrong**.

7. One way layout (ANOVA). Example has data Y_{ij} being the blood coagulation time for rat number j fed diet number i for $i = 1, 2, 3, 4$. There were 4 rats for diet 1, 6 for diets 2 and 3 and 8 rats fed diet 4. We use μ_{ij} as notation for $\mathrm{E}(Y_{ij})$. The idea is that all the rats fed diet 1 have the same mean coagulation time β_1 so $\mu_{11} = \mu_{12} = \mu_{13} = \mu_{14} = \beta_1$. (In fact it is pretty common notation to use μ_1 for β_1 but this will conflict, for the time being, with my notation for the mean of the first Y.) If we stack up the Ys we get

$$Y = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ \vdots \\ Y_{26} \\ \vdots \\ Y_{48} \end{bmatrix} \qquad \mu = \begin{bmatrix} \beta_1 \\ \beta_1 \\ \beta_1 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_2 \\ \vdots \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

Again we have $\mu = X\beta$.

Jargon: X is called a "design matrix".

The one way layout as a linear model

The sum of squares decomposition in one example

The data consist of blood coagulation times for 24 animals fed one of 4 different diets. Here are the data with the 4 diets being the 4 columns.

$$\begin{bmatrix} 62 & 63 & 68 & 56 \\ 60 & 67 & 66 & 62 \\ 63 & 71 & 71 & 60 \\ 59 & 64 & 67 & 61 \\ & 65 & 68 & 63 \\ & 66 & 68 & 64 \\ & & & 63 \\ & & & 59 \end{bmatrix}$$

The usual ANOVA model equation is

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

which we can write in matrix form by stacking up the observations into a column.

$$\begin{bmatrix} 62 \\ 60 \\ 63 \\ 59 \\ 63 \\ 67 \\ 71 \\ 64 \\ 65 \\ 66 \\ 68 \\ 66 \\ 71 \\ 67 \\ 68 \\ 68 \\ 56 \\ 62 \\ 60 \\ 61 \\ 63 \\ 64 \\ 63 \\ 59 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \\ \epsilon_{14} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{23} \\ \epsilon_{24} \\ \epsilon_{25} \\ \epsilon_{26} \\ \epsilon_{31} \\ \epsilon_{32} \\ \epsilon_{33} \\ \epsilon_{34} \\ \epsilon_{35} \\ \epsilon_{36} \\ \epsilon_{41} \\ \epsilon_{42} \\ \epsilon_{43} \\ \epsilon_{44} \\ \epsilon_{45} \\ \epsilon_{46} \\ \epsilon_{47} \\ \epsilon_{48} \end{bmatrix}$$

Let X denote the 24×4 design matrix in this formula. Usually we reparametrize the model in the form

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

which would lead to a design matrix which looked like X above with an extra column on the left all of whose entries are equal to 1. The parameter vector β would now be

$$\beta^T = [\mu \quad \alpha_1 \quad \alpha_2 \quad \alpha_3 \quad \alpha_4 \quad]$$

It will turn out that trying to use this parametrization the different parameters are not *identifiable*, that is, they cannot be separately estimated, because making μ bigger by a certain amount and each α_i smaller by the same amount leaves the data unchanged. We usually solve this problem by *defining*

$$\mu = (n_1\mu_1 + \cdots + n_k\mu_k)/(n_1 + \cdots + n_k)$$

and

$$\alpha_i = \mu_i - \mu\,.$$

Now it is automatic that $\sum \alpha_i = 0$ so we usually eliminate α_4 by replacing it in the model equation by the quantity

$$-(\alpha_1 + \alpha_2 + \alpha_3).$$

This leads to

$$\beta^T = [\mu \quad \alpha_1 \quad \alpha_2 \quad \alpha_3 \quad]$$

and

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{bmatrix}$$