MAST30025: Linear Statistical Models

Weeks 12 and 13 Lab

1. The following data were collected, to compare two treatments. The treatments were randomly assigned to test subjects.

treatment 1		treatment 2	
subject	response	subject	response
10	7.5	11	9.5
9	9.6	6	9.7
5	8.4	2	10.8
12	10.6	8	11.9
7	9.9	4	10.0
1	10.6	3	12.9

   (a) Estimate the difference between treatment effects, and test if it is significantly different from 0.

   (b) Now suppose that it is discovered that the response can be affected by the season, and that the data was collected over a period of six months, in the order given by the table. That is, a month was spent collecting each row of the table.

   We re-express the experiment by blocking: each month (row of the table) is considered one block, and we model the data as an additive two-factor model (the factors being the treatment and the block). Using this model, repeat your analysis. Is the estimate different? Is the p-value different?

2. In lectures, we showed that for the (randomised) complete block design

$$y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij},$$

   a solution to the reduced normal equations for $\boldsymbol{\tau} = (\tau_1, \dots, \tau_k)^T$ is given by

$$(\bar{y}_{\cdot 1} - \bar{y}_{\cdot\cdot}, \dots, \bar{y}_{\cdot k} - \bar{y}_{\cdot\cdot})^T.$$

   Here we suppose that we have b blocks and k treatments.

   Consider now the completely randomised design, with k treatments and b replications of each treatment

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}.$$

   Treating μ as a nuisance parameter, obtain the reduced normal equations for $\boldsymbol{\tau}$, then show that they admit the solution

$$(\bar{y}_1 - \bar{y}_{\cdot}, \dots, \bar{y}_k - \bar{y}_{\cdot})^T.$$

3. Suppose we have a (randomised) complete block design, $y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij}$, with b blocks and k treatments.

Let $\mathbf{c}^T$ be a treatment contrast, so that $\mathbf{c}^T\boldsymbol{\tau}$ is estimable, in which case

$$\text{Var } \mathbf{c}^T\boldsymbol{\tau} = \frac{\sigma^2}{b}\sum_{j=1}^{k} c_j^2.$$

(a) Give an approximate $100(1-\alpha)\%$ CI for $\mathbf{c}^T\boldsymbol{\tau}$, using the percentage point from a normal rather than the correct percentage point from a t distribution. (This is reasonable if the degrees of freedom are large.)

(b) Now suppose that you know σ^2 (perhaps you have an estimate from a pilot study), and that you think a plausible alternative to $\mathbf{c}^T\boldsymbol{\tau} = 0$ is given by some $\mathbf{c}^T\boldsymbol{\tau}^* \neq 0$. How large should b be to give a power of $100(1-\alpha)\%$ against this alternative (roughly)?

4. Consider the following data:

Response	Block	Treatment
1.245	1	1
1.804	1	2
2.468	2	1
6.664	2	3
5.573	3	1
-0.560	3	4
7.880	4	2
10.469	4	3
0.457	5	2
-3.621	5	4
-4.291	6	3
-9.384	6	4

(a) Show that this data comes from a balanced incomplete block design, and give t, b, k, r and λ.

(b) Give the design matrix X^A for a model with block and treatment effects (and an overall mean).

(c) Using this model, estimate $\tau_1 - \tau_2$, the difference between the first two treatment effects, and its variance. Write the variance estimate as $s^2\mathbf{c}^T(X^{AT}X^A)^c\mathbf{c}$ for a suitable $\mathbf{c}$.

(d) Give the design matrix X^B for a model with just treatment effects (and an overall mean).

(e) Using this model, estimate $\tau_1 - \tau_2$, the difference between the first two treatment effects, and its variance. Write the variance estimate as $s^2\mathbf{c}^T(X^{BT}X^B)^c\mathbf{c}$ for a suitable $\mathbf{c}$.

(f) Show that when going from model A (BIBD) to model B (CRD) the term $\mathbf{c}^T(X^TX)^c\mathbf{c}$ decreases, but s^2 increases markedly. What does this indicate?

(g) Is your estimate for $\tau_1 - \tau_2$ the same or different for the two models? Why?

5. Consider the BIBD model, with t treatments and b blocks of size k. Let λ be the number of times each pair appears, and write the design as

$$\mathbf{y} = X_1\boldsymbol{\alpha} + X_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}.$$

Show that for this model, contrasts in $\boldsymbol{\tau}$ are estimable.

If $\mathbf{c}^T\boldsymbol{\tau}$ is a contrast, show that an unbiased estimate is $(k/\lambda t)\mathbf{c}^T\mathbf{q}$, where

$$\mathbf{q} = \mathbf{t} - X_2^T X_1 \mathbf{b},$$

and $\mathbf{t}$ are the treatment totals and $\mathbf{b}$ the block totals.

6. An experimenter is tasked with designing an experiment to compare three treatment levels. There is a known confounding factor, so a blocked design is appropriate. Consider the following two designs, each using four blocks of size three:

Design 1	block 1:	1	2	3
	block 2:	2	1	3
	block 3:	1	3	2
	block 4:	3	2	1
Design 2	block 1:	1	1	2
	block 2:	2	2	3
	block 3:	3	3	1
	block 4:	1	2	3

(a) Which design is a complete block design?

(b) Write down the design matrix for each design. Hence show that $\tau_2 - \tau_1$ is estimable in each case.

(c) For each design, in terms of the unknown error variance σ^2, what is the variance of the estimator for $\tau_2 - \tau_1$, the difference between the first two treatment effects?

Based on this, which design is better?

7. In some situations, it is sensible to think of block effects as random. For example, experiments performed on a single day might be considered as a single block, subject to some effect for conditions on that day.

Consider the following model for an experiment with fixed treatment effects $\boldsymbol{\tau}$ and random block effects $\boldsymbol{\beta}$ (independent of the error $\boldsymbol{\varepsilon}$):

$$\mathbf{y} = X_1\boldsymbol{\beta} + X_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}, \quad \mathbb{E}\boldsymbol{\varepsilon} = \mathbf{0}, \quad \text{Var}\,\boldsymbol{\varepsilon} = \sigma^2 I, \quad \mathbb{E}\boldsymbol{\beta} = \mu\mathbf{1}, \quad \text{Var}\,\boldsymbol{\beta} = \sigma_\beta^2 I.$$

(a) Find $\mathbb{E}\mathbf{y}$ and $V = \text{Var}\,\mathbf{y}$.

(b) Give a solution to the generalised least squares problem:

$$\min_{\mathbf{t}} (\mathbf{y} - X_2\mathbf{t})^T V^{-1} (\mathbf{y} - X_2\mathbf{t}).$$

(c) A problem with the generalised least squares above is that μ may not be zero, so that if we write $\mathbf{y} = X_2\boldsymbol{\tau} + \boldsymbol{\varepsilon}'$, then $\boldsymbol{\varepsilon}' = \boldsymbol{\varepsilon} + X_1\boldsymbol{\beta}$ does not have a zero mean.

To get around this, first suppose that each block is of size k, so

$$X_1 = \begin{bmatrix} \mathbf{1}_k & \mathbf{0}_k & \cdots & \mathbf{0}_k \\ \mathbf{0}_k & \mathbf{1}_k & \cdots & \mathbf{0}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_k & \mathbf{0}_k & \cdots & \mathbf{1}_k \end{bmatrix},$$

then suppose that U is such that $X_1^T U = 0$.

Put $\mathbf{y}_1 = U^T\mathbf{y}$ and $\mathbf{y}_2 = X_1^T\mathbf{y}$, then show that we can write them as linear models whose errors have mean zero.

(d) Show that $\mathrm{Cov}(\mathbf{y}_1, \mathbf{y}_2) = \mathbb{E}(\mathbf{y}_1 - \mathbb{E}\mathbf{y}_1)(\mathbf{y}_2 - \mathbb{E}\mathbf{y}_2)^T = 0$.