Lecture 25 Friday March 24

Exam 2 Wed March 29 Lectures 11-25; Hws 4, 5; Quizzes 2, 3; Project 1 - recognize R code & output

PLAN

Two-factor model: Improved explanation of constraints p. 122

Exercise w/ data notation

Least-squares estimates

Corn-Yield Example

Two-way ANOVA, factor effects model: Explanation of constraints

We have two factors: Factor A at $\underline{a}$ levels
Factor B at $\underline{b}$ levels.

The factors are crossed, so there are ab treatments, and the first model we consider ~~has~~ allows ab unique means, $\{\mu_{ij}\}_{\substack{i=1,\ldots,a \\ j=1,\ldots,b}}$

Our factor-effects model equation, though, has $(\underbrace{1}_{\mu_{\cdot\cdot}} + \underbrace{a}_{\alpha_i\text{'s}} + \underbrace{b}_{\beta_j\text{'s}} + \underbrace{ab}_{(\alpha\beta)_{ij}\text{'s}})$ parameters,

so at first glance this model is <u>not</u> <u>the</u> <u>same</u> as the cell-means model.

The constraints on the factor-effects parameters resolve the paradox.

The constraints can be proved from the definitions of $\mu_{\cdot\cdot}$, α_i, etc.

p. 122

Show: ① $\sum_{i=1} \alpha_i = 0$, ② $\sum_{j=1} (\alpha\beta)_{ij} = 0$

Proof

① $\underset{\text{def. of } \alpha_i}{\sum_{i=1}^{a} \alpha_i =} \sum_{i=1}^{a} (\mu_{i\cdot} - \mu_{\cdot\cdot}) \underset{\text{property of summation}}{=} \sum_{i=1}^{a} \mu_{i\cdot} - \sum_{i=1}^{a} \mu_{\cdot\cdot} \underset{\substack{\text{because avg. of } \mu_{i\cdot}\text{'s is } \mu_{\cdot\cdot} \\ \text{AND because } \sum_{i=1}^{a} c = ac.}}{=} a\mu_{\cdot\cdot} - a\mu_{\cdot\cdot} = 0$

② Won't show this (same idea as for ①)

Exercise for Lecture 24:

Show $\bar{Y}_{\cdot j \cdot}$ = average of cell means in j^{th} column

Least-squares estimates for the two-way ANOVA model

Factor effects model equation:

$$\mu_{ij} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

$\overline{Y}_{ij.}$

Least-squares estimates of μ_{ij} are the cell means $\overline{Y}_{ij.}$, fitted values are $\hat{Y}_{ijk} = \overline{Y}_{ij.}$, and residuals are $e_{ijk} = Y_{ijk} - \hat{Y}_{ijk} = Y_{ijk} - \overline{Y}_{ij.}$.

Estimates of the model equation parameters:

$$\widehat{\mu}_{..} = \overline{Y}_{...}$$

$$\widehat{\alpha}_i = \overline{Y}_{i..} - \overline{Y}_{...}$$

$$\widehat{\beta}_j = \overline{Y}_{.j.} - \overline{Y}_{...}$$

$$\widehat{(\alpha\beta)}_{ij} = \overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...}$$

note $\hat{\mu}_{ij} - (\hat{\mu}_{..} + \hat{\alpha}_i + \hat{\beta}_j)$

I. II.

Explain why the cell means are the least-squares estimates, and illustrate why the factor-effects estimates are these.

I. View the model as a one-way model with $r = ab$ treatments and n observations per treatment. We earlier derived the least-squares estimator to be the treatment means (pp. 67-68). QED

II. Show $\hat{\mu}_{\cdot\cdot} = \bar{Y}_{\cdot\cdot\cdot}$

Proof. Now $\mu_{\cdot\cdot} = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \mu_{ij}$ so by "plugging in" $\hat{\mu}_{ij}$'s for μ_{ij}, we get

$$\hat{\mu}_{\cdot\cdot} = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \bar{Y}_{ij\cdot}$$

$$= \frac{1}{ab} \sum_{i} \sum_{j} \sum_{k} \frac{1}{n} Y_{ijk} \quad \text{by def. of } \bar{Y}_{ij\cdot}$$

$$= \frac{1}{abn} \sum \sum \sum Y_{ijk} = \frac{1}{abn} Y_{\cdot\cdot\cdot} = \bar{Y}_{\cdot\cdot\cdot}$$

Recall the Corn Yield data (p. 108)

Example: Corn Yield $n = 5$ plots per “cell” (A “cell” is a particular combination of the levels of the two factors, e.g. “low, low” means “low fertilizer” and “low manure”). The summary statistics are given in the two tables below.

Cell Means

	Fertilizer	
Manure	Low	High
Low	11.3	13.9
High	14.0	15.1

Cell SDs

	Fertilizer	
Manure	Low	High
Low	1.9	1.7
High	1.8	1.3

Ex. Corn Yield

$\bar{Y}_{11.} = 11.3$, $\bar{Y}_{12.} = 13.9$, $\bar{Y}_{21.} = 14.0$, $\bar{Y}_{22.} = 15.1$

$\bar{Y}_{1..} = \dfrac{11.3 + 13.9}{2} = 12.6$, $\bar{Y}_{2..} = \dfrac{14 + 15.1}{2} = 14.55$

$\bar{Y}_{.1.} = \dfrac{11.3 + 14.0}{2} = 12.65$, $\bar{Y}_{.2.} = \dfrac{13.9 + 15.1}{2} = 14.50$

$\bar{Y}_{...} = 13.5\cancel{}75 = 13.575$

Get least-squares estimates next.

This will lead to Sums of squares

— several of them ...

(to be continued)

$\hat{\mu}_{..}$ $\cancel{\mu_{..}} = \bar{Y}_{...} = 13.575$

$\hat{\alpha}_1 = \underbrace{\bar{Y}_{1..} - \bar{Y}_{...}}_{\hat{\mu}_{1.} - \hat{\mu}_{..}} = 12.6 - 13.575 = -.975$, $\hat{\alpha}_2 = \hat{\mu}_{2.} - \hat{\mu}_{..} = 14.55 - 13.575 = .975$

$\hat{\beta}_1 = \hat{\mu}_{.1} - \hat{\mu}_{..} = 12.65 - 13.575 = -.970$, $\hat{\beta}_2 = .970$

$(\widehat{\alpha\beta}_{11}) = \hat{\mu}_{11} - \hat{\mu}_{1.} - \hat{\mu}_{.1} + \hat{\mu}_{..} = 11.3 - 12.6 - 12.65 + 13.575 = -.375$, $(\widehat{\alpha\beta})_{12} = .375$

$(\widehat{\alpha\beta}_{21}) = .375$, $(\widehat{\alpha\beta})_{22} = -.375$

<u>EX.</u> (con.) The point of the factor-effects model is to get sums of squares for each type of parameter

---

Let's do the calculations for the Corn Yield data

$$SSA = nb\sum_{i=1}^{a}\left(\bar{Y}_{i..} - \bar{Y}_{...}\right)^2 = nb\sum_{i=1}^{a}\left(\hat{\alpha}_i\right)^2 = 10(-.975)^2 + 10(.975)^2 = 19.01$$

main effect of A

$df(A) = a-1 = 2-1 = 1$

$$SSB = na\sum_{j=1}^{b}\left(\bar{Y}_{.j.} - \bar{Y}_{...}\right)^2 = na\sum_{j=1}^{b}\hat{\beta}_j^2 = 10(-.970)^2 + 10(.970)^2 = 18.818$$

main effect of B

$df(B) = b-1 = 1$

$$SSAB = n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\widehat{\alpha\beta}\right)_{ij}^2 = 5\left[(.375)^2 \times 4\right] = 2.8125$$

$df(AB) = (a-1)(b-1) = (1)(1) = 1$

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b} s_{ij}^2\,(n-1) = 4\left[1.9^2 + 1.7^2 + 1.8^2 + 1.3^2\right] = 45.72$$

w/ $df = ab(n-1) = 16$

$= abn - ab$

$= n_T - r$

one-way ANOVA notation