Stat 565
1-23-18
①

Another possible power-related goal:

Find the sample size necessary to detect a certain increase in standard deviation.

Model: $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, but now assume that the τ_i terms are random variables, independent of the ε_{ij}'s.

Under H_0, $y_{ij} = \mu + \varepsilon_{ij}$

$$V(y_{ij}) = V(\varepsilon_{ij}) = \sigma^2$$

②

Under H_1, $y_{ij} = \mu + \tau_i + \varepsilon_{ij}$

$$V(y_{ij}) = V(\tau_i) + V(\varepsilon_{ij})$$

$$V(y_{ij}) = \frac{1}{a}\sum \tau_i^2 + \sigma^2$$

Set $$\sqrt{\frac{\frac{1}{a}\sum \tau_i^2 + \sigma^2}{\sigma^2}} = 1 + p$$

$$\frac{1}{a\sigma^2}\sum \tau_i^2 + 1 = (1+p)^2$$

$$\varphi = \sqrt{\frac{n\sum \tau_i^2}{a\sigma^2}} = \sqrt{n[(1+p)^2 - 1]}$$

Now follow Thursday's pattern

③

Randomized Complete Block Design

RCBD

Example: Test 4 fertilizers

There are 3 different fields, each with different growing conditions (sunlight, soil, water)

1 factor: fertilizer $a = 4$

1 blocking factor: field $b = 3$

④

y_{ij} = measurement for fertilizer i, block j

Model: $y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$

$i = 1, \ldots, a \quad j = 1, \ldots, b$

$\sum_{i=1}^{a} \tau_i = 0, \quad \sum_{j=1}^{b} \beta_j = 0, \quad E[\varepsilon_{ij}] \equiv 0$

$V[\varepsilon_{ij}] \equiv \sigma^2$

Assume the ε_{ij}'s are normal & independent

⑤

Use least squares to estimate the parameters

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij} - \mu - \tau_i - \beta_j\right)^2$$

Result: $\hat{\mu} = \bar{y}_{..}$, $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$, $\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$

Now $\hat{y}_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..})$

$$= \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}$$

⑥

<u>Decompose SS_{TOT}</u>

$$SS_{TOT} = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij} - \bar{y}_{..}\right)^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij} - \hat{y}_{ij} + \hat{y}_{ij} - \bar{y}_{..}\right)^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\left(\underline{\underline{y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}}} + \underline{\bar{y}_{i.}} + \underline{\underline{\bar{y}_{.j}}} - \underline{\bar{y}_{..}} - \underline{\underline{\bar{y}_{..}}}\right)^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\left[(\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})\right]^2$$

(7)

$$= b\sum_{i=1}^{a}(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2 + a\sum_{j=1}^{b}(\bar{y}_{\cdot j}-\bar{y}_{\cdot\cdot})^2$$

$$+ \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij}-\bar{y}_{i\cdot}-\bar{y}_{\cdot j}+\bar{y}_{\cdot\cdot})^2$$

+ 3 crossproducts that are 0

$$= SS_{TRT} + SS_{BLK} + SS_E$$

(8)

ANOVA table

Source	SS	df	MS	F
TRT	SS_{TRT}	$a-1$	$\frac{SS}{df}$	MS_{TRT}/MS_E
BLK	SS_{BLK}	$b-1$		—
ERR	SS_E	$(a-1)(b-1)$		—
TOT	SS_T	$N-1$	—	—

$$df_E = N-1-(a-1)-(b-1)$$
$$= ab-1-a+1-b+1 = ab-a-b+1$$
$$= (a-1)(b-1)$$

Matrix form of the RCBD (9)

$$Y = X\beta + \varepsilon \qquad (y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij})$$

$$\begin{bmatrix} y_{11}\\ \vdots \\ y_{1b}\\ \hline y_{21}\\ \vdots\\ y_{2b}\\ \hline \vdots \\ \vdots \\ \hline y_{a1}\\ \vdots\\ y_{ab}\end{bmatrix}_{N\times 1} = \left[\begin{array}{c|cccc|cccc}
1 & 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & & \vdots & 0 & 1 & \ddots & \vdots \\
 & & & & & & & & 1 \\
1 & 1 & 0 & \cdots & 0 & -1 & \cdots & & -1 \\ \hline
1 & 0 & 1 & 0 \cdots & 0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & 0 & 1 & \ddots & \vdots \\
 & & & & & & & & 1 \\
1 & 0 & 1 & 0 \cdots & 0 & -1 & \cdots & & -1 \\ \hline
\vdots & & & \vdots & & & \vdots & & \\ \hline
1 & -1 & \cdots & & -1 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & & & & 0 & 1 & \ddots & \vdots \\
 & & & & & & & & 1 \\
1 & -1 & \cdots & & -1 & -1 & \cdots & & -1
\end{array}\right] \begin{bmatrix}\mu\\ \tau_1\\ \vdots\\ \tau_{a-1}\\ \beta_1\\ \vdots\\ \beta_{b-1}\end{bmatrix} + \varepsilon_{N\times 1}$$

(10)

What happens if there is missing data?

2 solutions

① Exact method

Enter the data as a regression problem.
One or more rows will be missing.

Compute SS_E & df_E for this full model.

Now, run a reduced model: $y_{ij} = \mu + \beta_j + \varepsilon_{ij}$

(i.e., take out all of the columns of X corresponding to the τ_i's.)

(11)

Compute SS_E & df_E for the reduced model.

$$F = \frac{\left[\dfrac{SS_{E(red)} - SS_{E(full)}}{df_{E(red)} - df_{E(full)}}\right]}{MSE_{full}}$$

"Addition sum of squares F test"

(12)

(2) Approximate method (Imputation)

Suppose we have exactly 1 missing value.
Treat it as if it were another parameter to be estimated.

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot}\right)^2$$

$$= \sum\sum y_{ij}^2 + b\sum \bar{y}_{i\cdot}^2 + a\sum \bar{y}_{\cdot j}^2 + N\bar{y}_{\cdot\cdot}^2$$

$$- 2\sum_i\sum_j y_{ij}\bar{y}_{i\cdot} - 2\sum_j\sum_i y_{ij}\bar{y}_{\cdot j} + 2\sum\sum y_{ij}\bar{y}_{\cdot\cdot}$$

$$+ 2 \sum_i \sum_j \bar{y}_{i\cdot} \bar{y}_{\cdot j} - 2 \sum_i \sum_j \bar{y}_{i\cdot} \bar{y}_{\cdot\cdot}$$

$$- 2 \sum\sum \bar{y}_{\cdot j} \bar{y}_{\cdot\cdot}$$

$$= \sum\sum y_{ij}^2 + \underline{b \sum \bar{y}_{i\cdot}^2} + \underline{a \sum \bar{y}_{\cdot j}^2} + N \bar{y}_{\cdot\cdot}^2$$

$$- \underline{2b \sum \bar{y}_{i\cdot}^2} - \underline{2a \sum} \bar{y}_{\cdot j}^2 + 2N \bar{y}_{\cdot\cdot}^2$$

$$+ 2N \bar{y}_{\cdot\cdot}^2 - 2N \bar{y}_{\cdot\cdot}^2 - 2N \bar{y}_{\cdot\cdot}^2$$

$$= \sum\sum y_{ij}^2 - b \sum \bar{y}_{i}^2 - a \sum \bar{y}_{\cdot j}^2 + N \bar{y}_{\cdot\cdot}^2$$