Ex. 1 Beetles,
1-factor ANOVA

Response Y : size of beetle
Factor α : location, 3 levels (A, B, C)

Assume :
At location (level) j:
$Y_{ij} \sim N(\mu_j, \sigma^2)$, $i = 1, \dots, 10$
← equal variance

independent observations

[boxplot]

$H_0: \mu_1 = \mu_2 = \mu_3$, H_1: at least one μ_j differs

"Natural model" : $\hat{Y}_{ij} = \bar{Y}_j$ (average at level j)

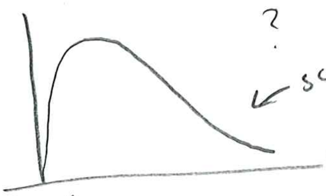
Partitioning of variation
← set $\bar{Y}_j = \hat{Y}_{ij}$, we recognize $SST = SSR + SSE$

$$SST = SSA^{\leftarrow \alpha} + SSE$$

$$\sum_{j=1}^{m}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y})^2 = \underbrace{\sum_{j=1}^{m}\sum_{i=1}^{n_j}(\bar{Y}_j - \bar{Y})^2}_{\text{between-groups variability}} + \underbrace{\sum_{j=1}^{m}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_j)^2}_{\text{within-group variability}} \quad (\sigma^2)$$

[Discuss]

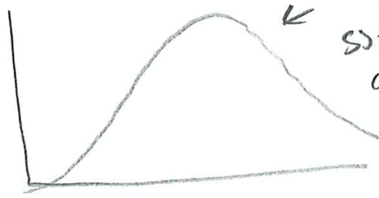
Sketch , SSA under H_0



? ← some higher observations due to chance (σ^2)

↑ most close to zero

H_1, sketch



← higher values of SSA due to actual different means.

$$F = \frac{SSA / m-1}{SSE / n-m} \overset{H_0}{\sim} F_{(m-1, n-m)}$$

← test for difference between m group means.

↑ where did this come from? (follows from our previous knowledge of linear model)

Aim: Present 1-factor ANOVA

as $Y = X\beta + \varepsilon$

Ex.2 Y: response, α = factor, 3 levels, 2 observation per level

Model: $Y_{ij} = \mu_j + \varepsilon_{ij}$ $\quad \varepsilon_{ij} \sim N(0, \sigma^2)$ (independent)

$H_0: \mu_1 = \mu_2 = \mu_3$ $\quad H_1$: at least one μ_j differs.

$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$ $\quad \varepsilon_{ij} \sim N(0, \sigma^2)$

μ: grand, global mean

α_j: effect of level j

$\alpha_j = \mu - \mu_j$

restriction $\sum_{j=1}^{m} \alpha_j = 0$ (otherwise μ is not a grand mean)

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$ $\quad H_1$: at least one α_j differs

Data:

y_{11}, y_{12} } level 1 $\quad \mu_1 = \mu + \alpha_1$

y_{21}, y_{22} } level 2 $\quad \mu_2 = \mu + \alpha_2$

y_{31}, y_{32} } level 3 $\quad \mu_3 = \mu + \alpha_3$

$Y \nearrow$

$\to$ Need $\boxed{X\beta}$

"naive guess"

Try :

$$\beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

$$X\beta = \begin{bmatrix} \mu+\alpha_1 \\ \mu+\alpha_1 \\ \mu+\alpha_2 \\ \mu+\alpha_2 \\ \mu+\alpha_3 \\ \mu+\alpha_3 \end{bmatrix}$$

mean of each group (and as the E(Y)).

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

<u>NB</u> X is not full rank.
4 columns, rank$(X) = 3$ $\rightarrow$ Not possible to estimate β.

Solution A: Dummy-variable coding (repetition)

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \qquad \beta = \begin{bmatrix} \mu_1 \\ \mu_2-\mu_1 \\ \mu_3-\mu_1 \end{bmatrix} = \begin{bmatrix} \mu+\alpha_1 \\ \alpha_2-\alpha_1 \\ \alpha_3-\alpha_1 \end{bmatrix} \Bigg\} \; (*)$$

$$X\beta = \begin{bmatrix} \mu+\alpha_1 \\ \mu+\alpha_1 \\ \mu+\alpha_1+(\alpha_2-\alpha_1) \\ \vdots \\ \mu+\alpha_1+(\alpha_3-\alpha_1) \end{bmatrix} = \begin{bmatrix} \mu+\alpha_1 \\ \mu+\alpha_1 \\ \mu+\alpha_2 \\ \mu+\alpha_2 \\ \mu+\alpha_3 \\ \mu+\alpha_3 \end{bmatrix} \leftarrow \text{what we wanted}$$

test $H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$

(*) $\rightarrow$ $H_0: \beta_1 = \beta_2 = 0$ (significance of regression) $\rightarrow$

From R:

$F = 3.069$
$P = 0.06584$

Solution B : Effect coding

Recall restriction $\alpha_1 + \alpha_2 + \alpha_3 = 0$

$\rightarrow \alpha_3 = -\alpha_1 - \alpha_2$

need estimates for these, and μ.

$$\beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{array}{l} \leftarrow \text{global mean} \\ \leftarrow \beta_1 \\ \leftarrow \beta_2 \end{array}$$

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{array}{l} \left.\vphantom{\begin{matrix}1\\1\end{matrix}}\right\} \text{level 1} \\ \left.\vphantom{\begin{matrix}1\\1\end{matrix}}\right\} \text{level 2} \\ \left.\vphantom{\begin{matrix}1\\1\end{matrix}}\right\} \text{level 3} \end{array}$$

$$X\beta = \begin{array}{l} \mu + \alpha_1 \\ \mu + \alpha_1 \\ \mu + \alpha_2 \\ \mu + \alpha_2 \\ \mu + (-\alpha_1 - \alpha_2) \\ \mu + \underbrace{(-\alpha_1 - \alpha_2)}_{\alpha_3} \end{array}$$

(as before)

$H_0 : \alpha_1 = \alpha_1 = \alpha_3 = 0$

$\hookrightarrow H_0 : \beta_1 = \beta_2 = 0$ $\leftarrow$ test of significance of regression

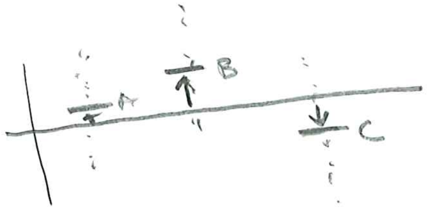
From R : $F = 3.069$, $p = 0.06584$ } same as before.

$\hat{\alpha}_1 = 0.067$

$\hat{\alpha}_2 = 1.367$

$\hat{\alpha}_3 = -\hat{\alpha}_1 - \hat{\alpha}_2 = -1.434$

$\hat{\sigma} = 2.203$ $\leftarrow$ much larger than effects



$\hat{\mu} = 23.9$

Ex. 2 Machine operator, 2-factor ANOVA (no interaction)

Response Y: time to assemble product (sec)

Factor α: machine, $m_\alpha = 4$ levels

Factor γ: operator, $m_\gamma = 6$ levels

Experiment: Each operator handles all machines (once) in random order.

Data
$$\left.\begin{array}{l} y_{11} \\ y_{12} \\ \vdots \\ y_{14} \end{array}\right\} \text{operator } 1$$
$$\left.\begin{array}{l} y_{21} \\ \vdots \\ y_{24} \end{array}\right\} \text{operator } 2 \qquad \text{etc.}$$

model:

$$Y_{ijk} = \underbrace{\mu}_{\text{global mean}} + \underbrace{\alpha_j}_{\text{effect of machine } j} + \underbrace{\gamma_k}_{\text{effect of operator } k} + \varepsilon_{ijk}, \qquad \sum_{j=1}^{m_\alpha} \alpha_j = 0, \quad \sum_{k=1}^{m_\gamma} \gamma_k = 0$$

[Discuss model assumptions]

Expected time, random machine and operator $= \mu$

'Effect of machine j' always α_j regardless of operator

'Effect of operator k' always γ_k regardless of machine

$Y = X\beta + \varepsilon$ $\qquad X$ effect coding (se slide)

$$\beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \end{bmatrix}$$

Find estimates:

- Effect of machine 1 $\quad -0.8208$
- — " — 4 $\quad (1-(-0.8208)-(-0.7375)-0.4458 = 1.11$
- Effect of operator 4 $\quad 0.3292$
- Estimated time machine 1, operator 2

$42.1208 - 0.8208 - 1.5958$

$= 39.7042$ seconds.

Tests : • significance of machine $H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$

• significance of operator $H_0: \gamma_1 = \cdots = \gamma_5 = 0$

→ testing subvector

! anova (model fit) → automatic in R

machine $F = 3.3388$ $p = 0.047904$

operator $F = 5.2944$ $p = 0.005328$

Bonferroni? $\frac{0.05}{2} = 0.025$, operator significant.

? Interaction between machine and operator?

2-factor ANOVA with interaction

$$Y_{ijk} = \mu + \alpha_j + \gamma_k + (\alpha\gamma)_{jk} + \varepsilon_{ijk}$$

$$\sum_j \alpha_j = 0 \qquad \sum_j (\alpha\gamma)_{jk} = 0$$

$$\sum_k \gamma_k = 0 \qquad \sum_k (\alpha\gamma)_{jk} = 0$$

Interpretation:

Operator k, handling machine j,

expected time $\mu + \alpha_j + \gamma_k + (\alpha\gamma)_{jk}$

$$Y = X\beta + \varepsilon$$

$$\beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_3 \\ \gamma_1 \\ \vdots \\ \gamma_5 \\ (\alpha\gamma)_{11} \\ (\alpha\gamma)_{12} \\ (\alpha\gamma)_{13} \\ \vdots \\ (\alpha\gamma)_{51} \\ (\alpha\gamma)_{52} \\ (\alpha\gamma)_{53} \end{pmatrix}$$

operator 1, machines 1-3

operator 5, machines 1-5.

← length of β = $1 + 3 + 5 + 3 \cdot 5 = 24$

we only have 24 observat