Random Effects

- One-way random effects of ANOVA
- Two-way mixed & random effects of ANOVA

Recall: Two-way ANOVA model

(Y_{ijk}) $\quad 1 \le i \le r, \quad 1 \le j \le m, \quad 1 \le k \le n_{ij}$

first grouping variable — m groups in the second — n_{ij} samples in $(i,j)^{th}$ cell

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

$\varepsilon_{ijk} \sim N(0, \sigma^2)$

Constraints: $\cdot \sum_{i=1}^{r} \alpha_i = 0 \quad \cdot \sum_{j=1}^{m} \beta_j = 0 \quad \cdot \sum_{i=1}^{\ell} \gamma_{ij} = 0 \ \forall\ 1 \le j \le m \quad \cdot \sum_{j=1}^{n} \gamma_{ij} = 0 \ \forall\ 1 \le i \le r$

Random vs fixed effects:

- repeated measures of more than one (identical) measurement is taken on the same individual

- in such cases the "group" effect α_i is best thought of as random ($\because$ sampling a subset of the entire population)

  enable us to view the levels we observe in that group to be samples from a population

  Examples: Collecting data from different medical centres

  "centre" might be thought of as random.

Example 2: How much calcium is there in 500 ml of buttermilk?

(vary across brands)

- Pick 4 brands of buttermilk - record amount of calcium in 500ml packet in 8 bottles

"individuals" — "repeated measurement"

- Quantities of interest
- Grand mean Calcium Content
- variability across brands

One-way random effect model

- take n- identical measurements from r subjects

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \qquad 1 \le i \le r \qquad 1 \le j \le n$$

$\varepsilon_{ij} \overset{d}{=} N(0, \sigma^2)$ (independent) $\quad 1 \le i \le r, \ 1 \le j \le n$

$\alpha_i \overset{d}{=} N(0, \sigma_\mu^2)$

Q: Is $\mu = 0$? , Is $\sigma_\mu^2 = 0$?

Implications :

- Y_{ij} are No longer independent

$$\mathrm{Cov}(Y_{ij}, Y_{k\ell}) = \sigma_\mu^2 \delta_{ik} + \sigma^2 \delta_{j\ell} \qquad \delta_{ij} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{o.w} \end{cases}$$

- make MLE more complicated
- OLS only parameter to estimate was σ^2

One way random ANOVA table :

Source	Sum of Squares	df	(Exercise) $E(\cdot)$
Treatment	$SS_{treatment} = \sum_{i=1}^{r} n(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	$r-1$	$\sigma^2 + n\sigma_\mu^2$
Error	$SS_{Error} = \sum_{i=1}^{r} \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_{i\cdot})^2$	$(n-1)r$	σ^2

$H_0: \sigma_\mu^2 = 0$

$$\frac{MS_{Treatment}}{MS_{Error}} = \frac{SS_{treatment}/r-1}{SS_{error}/(n-1)r} \sim F_{r-1,(n-1)r}$$

Inference on μ :

$$E[\bar{Y}_{\cdot\cdot}] = \mu \quad , \quad \mathrm{Var}(\bar{Y}_{\cdot\cdot}) = \frac{n\sigma_\mu^2 + \sigma^2}{rn} \quad (\text{Ex.})$$

$$\Rightarrow \frac{\bar{Y}_{\cdot\cdot} - \mu}{\sqrt{\frac{SS_{treatment}}{(r-1)rn}}} \sim t_{r-1}$$

Estimate σ^2_μ

From the ANOVA table

$$\sigma^2_\mu = \left(E\left[\frac{SS_{treatment}}{r-1} - \frac{SS_{Error}}{(n-1)r} \right] \right) \frac{1}{n}$$

(Estimate) $$\hat{\sigma}^2_\mu = \left(\frac{SS_{treatment}}{r-1} - \frac{SS_{Error}}{(n-1)r} \right) \frac{1}{n}$$ (Caution: $\hat{\sigma}^2_\mu < 0$)

Example: Manufacturing company (improve productivity)

- m employees - n days
- r machines
- employees and machines as random (interactions)

Two way random effect model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \qquad \begin{array}{l} 1 \le i \le r \\ 1 \le j \le m \\ 1 \le k \le n \end{array}$$

$$\left.\begin{array}{l} \alpha_i \overset{d}{=} N(0, \sigma_\alpha^2) \\ \beta_j \overset{d}{=} N(0, \sigma_\beta^2) \\ \gamma_{ij} \overset{d}{=} N(0, \sigma_{\alpha\beta}^2) \\ \varepsilon_{ijk} \overset{d}{=} N(0, \sigma^2) \end{array}\right\} \quad 1 \le j \le m \quad 1 \le i \le r \quad 1 \le k \le n \quad \Rightarrow$$

$Cov(Y_{ijk}, Y_{stu}) = \dots$ (Ex) $\neq 0$

Two way random Anova table

(Ex.)

Source	Sum of squares	df	Expected
A	$SS_A := nm\sum_{i=1}^{r}(\overline{Y}_{i..} - \overline{Y}_{...})^2$	$r-1$	$\sigma^2 + nm\,\sigma_\alpha^2 + n\,\sigma_{\alpha\beta}^2$
B	$SS_B := nr\sum_{j=1}^{m}(\overline{Y}_{.j.} - \overline{Y}_{...})^2$	$m-1$	$\sigma^2 + nr\,\sigma_\beta^2 + n\,\sigma_{\alpha\beta}^2$
Interaction	$SS_I := n\sum_{i=1}^{r}\sum_{j=1}^{m}(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...})^2$	$(m-1)(r-1)$	$\sigma^2 + n\,\sigma_{\alpha\beta}^2$
Error	$SS_{Error} = \sum_{i=1}^{r}\sum_{j=1}^{m}\sum_{k=1}^{n}(Y_{ijk} - \overline{Y}_{ij.})^2$	$(n-1)rm$	σ^2

To test $H_0: \sigma_\alpha^2 = 0$ (like before, use SS_A & SS_I)

$H_0: \sigma_{\alpha\beta}^2 = 0$ (use SS_{AB} & SS_{Error})

Mixed Effect

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \qquad 1 \le i \le r, \quad 1 \le j \le m, \quad 1 \le k \le n$$

$$\left.\begin{array}{l} \alpha_i \overset{d}{=} N(0, \sigma_\alpha^2) \\ \beta_j \ \text{constant} \\ \gamma_{ij} \overset{d}{=} N\left(0, \frac{(m-1)\sigma_{\alpha\beta}^2}{m}\right) \\ \varepsilon_{ijk} \overset{d}{=} N(0, \sigma^2) \end{array}\right\} \quad \begin{array}{l} 1 \le j \le m \quad 1 \le i \le r \\ 1 \le k \le n \end{array} \quad \Rightarrow$$

Constraints:

- $\sum_{j=1}^{m} \beta_j = 0$; $\sum_{j=1}^{m} \sum_{i=1}^{r} \gamma_{ij} = 0$; $Cov(\gamma_{ij}, \gamma_{kl}) = -\frac{\sigma_{\alpha\beta}^2}{m}$

$\Rightarrow$ • $Cov(Y_{ijk}, Y_{stn}) = \dots$ (Ex) $\neq 0$

Anova table

Source	Sum of squares	df	Expected value
A	$SS_A := nm \sum_{i=1}^{r} (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$r-1$	$\sigma^2 + nm\sigma_\alpha^2$
B	$SS_B := nr \sum_{j=1}^{m} (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$m-1$	$\sigma^2 + nr \frac{\sum_{j=1}^{m} \beta_j^2}{m-1} + n\sigma_{\alpha\beta}^2$
Interaction	$SS_I := n \sum_{i=1}^{r} \sum_{j=1}^{m} (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$(m-1)(r-1)$	$\sigma^2 + n\sigma_{\alpha\beta}^2$
Error	$SS_{Error} = \sum_{i=1}^{r} \sum_{j=1}^{m} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2$	$(n-1)rm$	σ^2

Testing

$H_0: \sigma_\alpha^2 = 0$; use SS_A & SS_{Error}

$H_0: \beta_1 = \dots = \beta_m$; use SS_B & SS_I

$H_0: \sigma_{\alpha\beta} = 0$; SS_I & SS_{Error}.

Estimate

$$\hat{\sigma}_B = nr\,(\underbrace{MS_B - MS_I}_{\text{l.c. of } \chi^2})$$

(C.I: difficulty)