$$\underline{Y_{ijk}} - \underbrace{\overline{Y}_{...}}_{\hat{\mu}} = \underbrace{(Y_{ijk} - \overline{Y}_{ij\cdot})}_{\hat{\epsilon}_{ijk}} + \underbrace{(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{...})}_{\hat{\alpha}_i} + \underbrace{(\overline{Y}_{\cdot j\cdot} - \overline{Y}_{...})}_{\hat{\beta}_j} + \underbrace{(\overline{Y}_{ij\cdot} - \overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot j\cdot} + \overline{Y}_{...})}_{\hat{\delta}_{ij}}$$

Note.

$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$

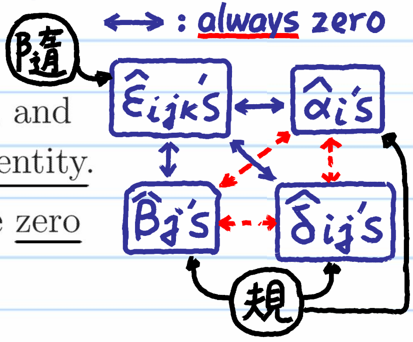
$= \overline{\mu} + \alpha_i + \beta_j + \delta_{ij} + \varepsilon_{ijk}$

$= \hat{\overline{\mu}} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\delta}_{ij} + \hat{\varepsilon}_{ijk}$

$= \hat{\mu}_{ij} + \hat{\varepsilon}_{ijk}$

$\leftarrow\text{-}\text{-}\rightarrow$: equals zero ∵ balanced

$\longleftrightarrow$: always zero



- Taking square on both sides, expanding the terms, and summing them over all (i, j, k)'s, we get the SS identity.
- (exercise) The sums of the cross-product terms are zero because of the balanced condition.

  ↳ brings in "orthogonality"

Question 10.

What is the information of data provided by each of these SS's?

Theorem 16 (Expectations of SS_A, SS_B, SS_{AB}, and SS_E)

Consider the model ($\triangledown$) in LNp.38. Suppose (i) μ_{ij} $(= \overline{\mu} + \alpha_i + \beta_j + \delta_{ij})$ is the mean of F_{ij}, (ii) all F_{ij}'s have same variance σ^2, and (iii) all $K_{ij} = K$.

Then,

(1) $E(SS_A) = J\boxed{K}(\sum_i \alpha_i^2) + (I-1)\sigma^2$ ← ≥ 0

(2) $E(SS_B) = I\boxed{K}(\sum_j \beta_j^2) + (J-1)\sigma^2$ ← ≥ 0

(3) $E(SS_{AB}) = \boxed{K}(\sum_i \sum_j \delta_{ij}^2) + (I-1)(J-1)\sigma^2$ ← ≥ 0

What happen

(1) if $H_0^{(A)}$ is true?

(2) if $H_0^{(B)}$ is true?

(3) if $H_0^{(AB)}$ is true?

When K↑, $E(SS_{A/B/AB})$↑ ⇒ easier to detect small difference

dimension / d.f.

Why do they have the 2nd term?

(4) $E(SS_E) = I\,J\,(K-1)\,\sigma^2$ ⟵ e ⟶ $S_p^2 = SS_E/[IJ(K-1)]$ is unbiased

cf.

$\overline{Y}_{i\cdot\cdot} = \hat{\overline{\mu}} + \hat{\alpha}_i$ → **Proof** of (1). By the Thm 2 in LNp.10, with the role of Z_k being played by $\overline{Y}_{i\cdot\cdot}$ and that of $\overline{Z}$ being played by $\overline{Y}_{...}$, we have

independent (table in LNp.48) $\frac{1}{I}\sum_i \overline{Y}_{i\cdot\cdot} = \overline{Y}_{...}$

∵ unbiased → ● $E(\overline{Y}_{i\cdot\cdot}) = \overline{\mu} + \alpha_i$ for $i = 1, \ldots, I$, and $E(\overline{Y}_{...}) = \overline{\mu}$,

- $Var(\overline{Y}_{i\cdot\cdot}) = \sigma^2/(JK)$ for $i = 1, \ldots, I$, $\quad Var(\overline{Y}_{...}) = \frac{\sigma^2}{N} = \frac{\sigma^2}{IJK} = \frac{1}{I}\frac{\sigma^2}{JK}$
- $E(SS_A) = E\left(JK[\sum_i (\overline{Y}_{i\cdot\cdot} - \overline{Y}_{...})^2]\right) = J\,K\,[\sum_i E(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{...})^2]$

$$= \cancel{JK}\sum_i\left(\alpha_i^2 + \frac{I-1}{I} \times \frac{\sigma^2}{\cancel{JK}}\right) = J\,K\,(\textstyle\sum_i \alpha_i^2) + (I-1)\sigma^2.$$

$\overline{Y}_{\cdot j\cdot} = \hat{\overline{\mu}} + \hat{\beta}_j$ → **Proof** of (2). It can be proved by the Thm 2 in LNp.10, with the role of Z_k being played by $\overline{Y}_{\cdot j\cdot}$ and that of $\overline{Z}$ being played by $\overline{Y}_{...}$ (exercise).

independent (table in LNp.48) $\frac{1}{J}\sum_j \overline{Y}_{\cdot j\cdot} = \overline{Y}_{...}$

i, j fixed → **Proof** of (4). It can be proved by the Thm 2 in LNp.10, with the role of Z_k being played by Y_{ijk} and that of $\overline{Z}$ being played by $\overline{Y}_{ij\cdot}$ (exercise).

independent $\frac{1}{K}\sum_k Y_{ijk} = \overline{Y}_{ij\cdot}$

Note. Y_{ijk} & $\overline{Y}_{ij\cdot}$ have same mean μ_{ij}

Proof of (3). By applying the Thm 2 in LNp.10 to SS_{TOT}, with the role of Z_k being played by Y_{ijk} and that of $\overline{Z}$ being played by $\overline{Y}_{...}$, we get

independent $\frac{1}{N}\sum_i\sum_j\sum_k Y_{ijk} = \overline{Y}_{...}$

- $E(Y_{ijk}) = \overline{\mu} + \alpha_i + \beta_j + \delta_{ij}$ for any (i, j, k), and $E(\overline{Y}_{...}) = \overline{\mu}$,

$\overline{Y}_{ij\cdot} - \hat{\alpha}_i - \hat{\beta}_j = \hat{\overline{u}} + \hat{\delta}_{ij}$

Why not use

$Z_k \longleftrightarrow \overline{Y}_{ij\cdot} - (\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot\cdot\cdot}) - (\overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdot\cdot\cdot})$

$\overline{Z} \longleftrightarrow \overline{Y}_{\cdot\cdot\cdot}$

not indep.

- $Var(Y_{ijk}) = \sigma^2$ for any (i,j,k),
- $E(SS_{TOT}) = E\left[\sum_i \sum_j \sum_k (Y_{ijk} - \overline{Y}_{\cdot\cdot\cdot})^2\right] = \sum_i \sum_j \sum_k E(Y_{ijk} - \overline{Y}_{\cdot\cdot\cdot})^2$

$$= \sum_i \sum_j \sum_k \left[(\alpha_i + \beta_j + \delta_{ij})^2 + (N-1)\sigma^2/N\right]$$

$$= (N-1)\sigma^2 + \sum_i \sum_j \sum_k (\alpha_i + \beta_j + \delta_{ij})^2$$

$$= (N-1)\sigma^2 + JK\left(\sum_i \alpha_i^2\right) + IK\left(\sum_j \beta_j^2\right) + K\left(\sum_i \sum_j \delta_{ij}^2\right).$$

The last equality holds due to the linear constraints on the parameters. For example, the cross-product terms involving α_i's and β_j's is

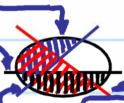
$\sum_i \alpha_i = 0$
$\sum_j \beta_j = 0$
$\sum_i \delta_{ij} = \sum_j \delta_{ij} = 0$

$$\sum_i \sum_j \sum_k \alpha_i \beta_j = K\left(\sum_i \alpha_i\right)\left(\sum_j \beta_j\right) = 0.$$

The desired expression for $^*E(SS_{AB})$ now follows, since

$$E(SS_{TOT}) = E(SS_E) + E(SS_A) + E(SS_B) + {}^*E(SS_{AB}).$$

$\Omega\ (u_{ij} = \overline{u} + \alpha_i + \beta_j + \delta_{ij})$



$u_{ij} = \overline{u} + \alpha_i + \beta_j \leftarrow H_0^{(AB)}$

$H_0^{(A)}$: $u_{ij} = \overline{u} + \beta_j + \delta_{ij}$

$H_0^{(B)}$: $u_{ij} = \overline{u} + \alpha_i + \delta_{ij}$

Theorem 17 (Distribution of SS_A, SS_B, SS_{AB}, and SS_E, (IJ)-sample normal model)

Consider the model ($\square$) in LNp.45. ← assume normality

(1) Under $\Omega = H_0^{(A)} \cup H_A^{(A)} = H_0^{(B)} \cup H_A^{(B)} = H_0^{(AB)} \cup H_A^{(AB)}$,

$$SS_E \sim \sigma^2 \chi^2_{IJ(K-1)} \quad \Leftrightarrow \quad SS_E/\sigma^2 \sim \chi^2_{IJ(K-1)}.$$

$N - IJ$

dimension = degrees of freedom

(2) Under $H_0^{(A)}: \alpha_1 = \cdots = \alpha_I = 0$,

$$SS_A \sim \sigma^2 \chi^2_{I-1} \quad \Leftrightarrow \quad SS_A/\sigma^2 \sim \chi^2_{I-1}.$$

---

(3) Under $H_0^{(B)}: \beta_1 = \cdots = \beta_J = 0$,

$$SS_B \sim \sigma^2 \chi^2_{J-1} \quad \Leftrightarrow \quad SS_B/\sigma^2 \sim \chi^2_{J-1}.$$

dimension = degrees of freedom

(4) Under $H_0^{(AB)}$: all δ_{ij}'s are 0,

$$SS_{AB} \sim \sigma^2 \chi^2_{(I-1)(J-1)} \quad \Leftrightarrow \quad SS_{AB}/\sigma^2 \sim \chi^2_{(I-1)(J-1)}.$$

(5) Under Ω, the SS_E, SS_A, SS_B and SS_{AB} are independent. ← $\because$ balanced

proof of Thm 5 (LNp.15-16)

cf.

Proof of (1). Let $s_{ij}^2 = \frac{1}{K-1}\sum_{k=1}^{K}(Y_{ijk} - \overline{Y}_{ij\cdot})^2$ be the sample variance in the (i,j)th sample, $i = 1, \ldots, I$, $j = 1, \ldots, J$. Because

$\sigma^2 \leftarrow e$

$[IJ(K-1)]S_p^2$

- $Y_{ij1}, \ldots, Y_{ijK} \sim$ i.i.d. $N(\mu_{ij}, \sigma^2) \Rightarrow (K-1)s_{ij}^2/\sigma^2 \sim \chi^2_{K-1}$ ← check LNp.15 & Review 1 in LN, CH11, p.6
- $SS_E/\sigma^2 = \sum_i \sum_j (K-1)s_{ij}^2/\sigma^2$,
- s_{ij}^2's are independent because all the Y_{ijk}'s are independent,

we have $SS_E/\sigma^2 \sim \chi^2_{IJ(K-1)}$ since the sum of m chi-square random variables, each with n_i degrees of freedom, follows a chi-square distribution with $n_1 + \cdots + n_m$ degrees of freedom. check the table in LNp.48

Proof of (2). Consider the *independent* random sample $\{\overline{Y}_{1\cdot\cdot}, \ldots, \overline{Y}_{I\cdot\cdot}\}$. Notice that $\overline{Y}_{\cdot\cdot\cdot}$ and $\frac{1}{I-1}\sum_{i=1}^{I}(\overline{Y}_{i\cdot\cdot} - \overline{Y}_{\cdot\cdot\cdot})^2$ are the sample mean and the sample variance of $\{\overline{Y}_{1\cdot\cdot}, \ldots, \overline{Y}_{I\cdot\cdot}\}$, respectively. Under $H_0^{(A)}: \alpha_1 = \cdots = \alpha_I = 0$, In Ω, $Y_{ijk} \sim N(\bar{\mu}+\alpha_i+\beta_j+\delta_{ij}, \sigma^2)$

- for $i = 1, \ldots, I$,

$\alpha_i = 0$ (in $H_0^{(A)}$)

$$E(\overline{Y}_{i\cdot\cdot}) = \frac{1}{JK}\sum_j\sum_k(\bar{\mu}+\beta_j+\delta_{ij}) = \frac{1}{J}(J\bar{\mu}+\underbrace{\textstyle\sum_j\beta_j}_{=0}+\underbrace{\textstyle\sum_j\delta_{ij}}_{=0}) = \bar{\mu}$$

$= \bar{\mu}+\alpha_i$ in $H_A^{(A)}$

and $Var(\overline{Y}_{i\cdot\cdot}) = \sigma^2/(JK) \quad\Rightarrow\quad \overline{Y}_{1\cdot\cdot}, \ldots, \overline{Y}_{I\cdot\cdot} \sim$ i.i.d. $N(\bar{\mu}, \sigma^2/(JK))$, ← like 1-sample data

- thus, we have

check LNp.50

$$SS_A = JK\left\{(I-1)\left[\underbrace{\frac{1}{I-1}\sum_{i=1}^{I}(\overbrace{\overline{Y}_{i\cdot\cdot}-\overline{Y}_{\cdot\cdot\cdot}}^{\hat{\alpha}_i})^2}_{\text{sample variance}}\right]\right\} \overset{e}{\sim} \cancel{JK}\times\frac{\sigma^2}{\cancel{JK}}\chi^2_{I-1}.$$

Proof of (3). Similar to (2) by considering $\{\overline{Y}_{\cdot 1\cdot}, \ldots, \overline{Y}_{\cdot J\cdot}\}$. ← check the table in LNp.48

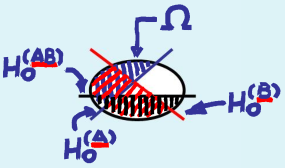
Proofs of (4) and (5). Skipped. They can be proved under the framework of linear model with the "orthogonality" property brought by balanced data.

Theorem 18 (ANOVA for two-way layout) ←cf→ Thm6 (1-way ANOVA) in LNp.16~18

Consider the model (□) in LNp.45, and the null hypotheses (LNp.49):

assume normality

$$H_0^{(A)}, \quad H_0^{(B)}, \quad H_0^{(AB)}$$



- Define mean square (MS):

$$MS_E = SS_E/[IJ(K-1)], \qquad MS_A = SS_A/(I-1),$$
$$MS_B = SS_B/(J-1), \qquad MS_{AB} = SS_{AB}/[(I-1)(J-1)].$$

- Test $H_0^{(A)}: \alpha_1 = \cdots = \alpha_I = 0$ vs. $H_A^{(A)}$: at least one of α_i's is not 0
  - test statistic $F_A = \dfrac{MS_A}{MS_E} = \dfrac{SS_A/(I-1)}{SS_E/[IJ(K-1)]}$. A justification is:

Thm16 (LNp.51)

$*$ under $H_0^{(A)}$, $E(MS_A) = E(MS_E) = \sigma^2$, the F_A should be close to 1,

$*$ under $H_A^{(A)}$, $E(MS_A) = \frac{JK}{I-1}\underbrace{(\sum_i\alpha_i^2)}_{\geq 0} + \sigma^2 > \sigma^2 = E(MS_E)$,

the F_A should be large (i.e., data with larger F_A ⇒ more extreme).

Thm17 (LNp.54-55)

  - null distribution: Under $H_0^{(A)}$,

$$\boxed{\text{independent}} \quad F_A = \frac{(\overbrace{SS_A/\sigma^2}^{\sim\chi^2_{I-1}\text{ (under }H_0^{(A)})})/(I-1)}{(\underbrace{SS_E/\sigma^2}_{\chi^2_{IJ(K-1)}\text{ (under }\Omega)\sim})/[IJ(K-1)]} \sim F_{I-1,\,IJ(K-1)}$$

  - rejection region at level α: reject $H_0^{(A)}$ if $F_A > F_{I-1,\,IJ(K-1)}(\alpha)$ ← $(1-\alpha)$-quantile (cf. α_i's)

- Test $H_0^{(B)}: \beta_1 = \cdots = \beta_J = 0$ vs. $H_A^{(B)}$: at least one of β_j's is not 0
  - test statistic $F_B = \dfrac{MS_B}{MS_E} = \dfrac{SS_B/(J-1)}{SS_E/[IJ(K-1)]}$. Same justification as F_A.
  - null distribution: Under $H_0^{(B)}$,

$$\boxed{\text{independent}} \quad F_B = \frac{(\overbrace{SS_B/\sigma^2}^{\sim\chi^2_{J-1}\text{ (under }H_0^{(B)})})/(I-1)}{(\underbrace{SS_E/\sigma^2}_{\chi^2_{IJ(K-1)}\text{ (under }\Omega)\sim})/[IJ(K-1)]} \sim F_{J-1,\,IJ(K-1)}$$

- rejection region at level α: reject $H_0^{(B)}$ if $F_B > F_{J-1,\,IJ(K-1)}(\alpha)$

- Test $H_0^{(AB)}$: all δ_{ij}'s are 0 vs. $H_A^{(AB)}$: at least one of δ_{ij}'s is not 0
  - test statistic $F_{AB} = \dfrac{MS_{AB}}{MS_E} = \dfrac{SS_{AB}/[(I-1)(J-1)]}{SS_E/[IJ(K-1)]}$. Same justification as that for F_A.
  - null distribution: Under $H_0^{(AB)}$, $\rightarrow \sim \chi^2_{(I-1)(J-1)}$ (under $H_0^{(AB)}$)

$$\boxed{\text{independent}}\quad F_{AB} = \frac{(SS_{AB}/\sigma^2)/[(I-1)(J-1)]}{(SS_E/\sigma^2)/[IJ(K-1)]} \sim F_{(I-1)(J-1),\,IJ(K-1)}$$

$\chi^2_{IJ(K-1)}$ (under Ω) $\sim \leftarrow$

  - rejection region at level α: reject $H_0^{(AB)}$ if $F_{AB} > F_{(I-1)(J-1),\,IJ(K-1)}(\alpha)$

- Two-way ANOVA table ↱ If treat the IJ samples as 1-way layout, $SS_{within} = SS_E$, $SS_{between} = SS_A + SS_B + SS_{AB}$

	Source	SS	df		MS ($=SS/df$)		F
規律	I A	SS_A	$I-1$	$\Sigma_i \alpha_i = 0$	$SS_A/(I-1)$	$\div$	MS_A/MS_E
	II B	SS_B	$J-1$	$\Sigma_j \beta_j = 0$	$SS_B/(J-1)$	$\div$	MS_B/MS_E
	III AB	SS_{AB}	$(I-1)(J-1)$	$\Sigma_i \delta_{ij}=0$ $\Sigma_j \delta_{ij}=0$	$SS_{AB}/[(I-1)(J-1)]$	$\div$	MS_{AB}/MS_E
隨機	Error	SS_E	$IJ(K-1)$	$\hat{\mu}_{ij}$'s	$SS_E/[IJ(K-1)]$ $= S_p^2 \xrightarrow{e} \sigma^2$		use same standard
	Total	SS_{TOT}	$IJK-1$	$\hat{\mu}$			

$\oplus$ $\oplus$

Example 7 (Iron retention)

- An experiment was performed to determine whether two forms of iron, Fe^{2+} and Fe^{3+}, are retained differently.

crossing

	B		
A	0.3	1.2	10.2
Fe^{2+}	18 obs.	18 obs.	18 obs.
Fe^{3+}	18 obs.	18 obs.	18 obs.

↰ 6 cells. balanced data

- Design:
  - 108 mice were randomly divided into 6 groups of 18 each
  - 3 groups were given Fe^{2+} in 3 different concentrations, 10.2, 1.2, and 0.3 millimolar
    high → median ← low
  - the other 3 groups were given Fe^{3+} at the same 3 concentrations
  - percentage of iron retained was calculated for each mouse → Y_{ijk}
- factors, levels, and replicates

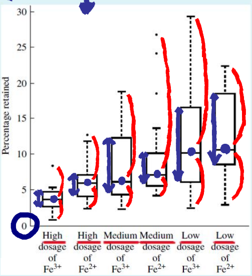
2-way layout

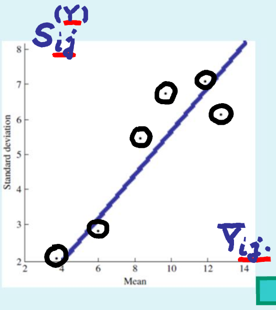
  - factor A: iron, with 2 levels: Fe^{2+} and Fe^{3+} $\Rightarrow I = 2$
  - factor B: dosage (concentration), with 3 levels: 10.2, 1.2, 0.3 $\Rightarrow J = 3$
  - for each of the 6 level combinations, there are 18 replicates $\Rightarrow K = 18$
- Some plots of Y_{ijk}'s:
  - **Figure 12.4, 12.6** (textbook): boxplots of Y_{ijk}'s and a plot of $s_{ij}^{(Y)}$ vs. $\overline{Y}_{ij\cdot}$

standard deviation of the (i,j)th sample

6-sample data in 1-way layout

cf.





- boxplots show the data is quite skewed to the right
- Figure 12.6 shows the error variance increases with the mean
- They show Y_{ijk}'s are not normally distributed with equal variance

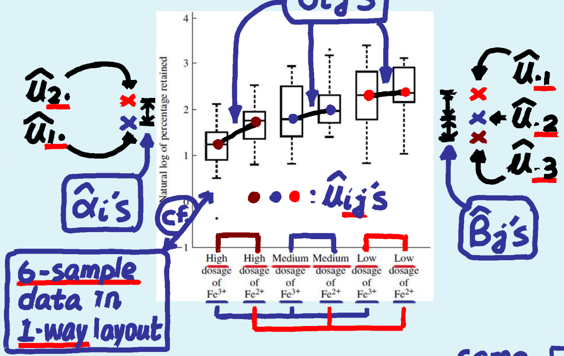
  $\Rightarrow$ **Q**: What can be the remedy for these problems?

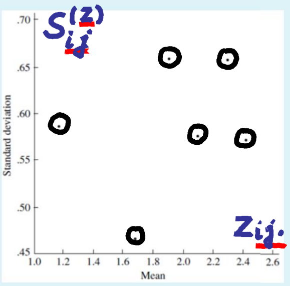
Y_{ijk}'s do not fit the ANOVA model (□) in LNp.45

- Consider the transformation

$$Z_{ijk} = \log(Y_{ijk}):$$

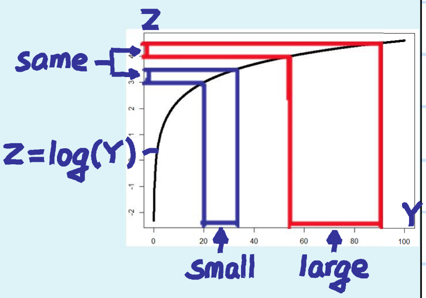
- **Figure 12.5, 12.7** (textbook): boxplots of Z_{ijk}'s and a plot of $s_{ij}^{(Z)}$ vs. $\overline{Z}_{ij\cdot}$





- boxplots show
  - distribution of Z_{ijk} is more symmetrical,
  - interquartile ranges are less variable,
- Figure 12.7 shows the log transformation is successful in stabilizing the variance

  constant variance



- **Q**: Why can it work?
- Statistical modeling: Assume Z_{ijk}'s follow the model (□) in LNp.45.