Ch 7: ANALYSIS OF VARIANCE AND COVARIANCE

7.1 Fuzzy definition

ANOVA:	All independent variables categorical
	(sex, race, treatment)
ANCOVA:	Plus some continuous
	(height, age, blood pressure)

The one-way model

1. (Ex) The 3 sample problem:

μ_1	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}
μ_2	y_{21}	y_{22}	y_{23}		
μ_3	y_{31}	y_{32}	y_{33}	y_{34}	

   μ_i = mean of population i

2. (One-way ANOVA)

$$y_{ij} = \mu_i + e_{ij} \qquad i = 1, \cdots, r \; ; \; j = 1, \cdots, n_i$$

   y_{ij} = response

   i = independent variable

   μ_i = regression function

   In the example μ_i is the population mean.

3. (General linear model)

$$\begin{aligned} y &= \mu + e \;\;,\;\; \mu \in \mathcal{M} \\ y &= (y_{ij}) = \text{ 2-way table} \\ \mu &= (\mu_{ij}) = \text{ 2-way table} \\ \mathcal{M} &= \{(\mu_{ij}) : \mu_{ij} = \mu_i \text{ some } \mu_i\} \end{aligned}$$

In the example y and μ have the form

$$\begin{pmatrix} * & * & * & * & * \\ * & * & * & & \\ * & * & * & * & \end{pmatrix}$$

and

$$\mu = \begin{pmatrix} \mu_1 & \mu_1 & \mu_1 & \mu_1 & \mu_1 \\ \mu_2 & \mu_2 & \mu_2 & & \\ \mu_3 & \mu_3 & \mu_3 & \mu_3 & \end{pmatrix} \longleftrightarrow \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}$$

4. (Th) dim $\mathcal{M} = r =$ number of rows

5. (The basic hypothesis)

$$\mu_1 = \cdots = \mu_r$$

The corresponding restricted model is

$$y_{ij} = \nu + e_{ij}$$

6. (Note) dim $\mathcal{H} = 1$

Pf:

$$\mu = \begin{pmatrix} \nu & \nu & \nu & \nu & \nu \\ \nu & \nu & \nu & & \\ \nu & \nu & \nu & \nu & \end{pmatrix} \longleftrightarrow \nu$$

7. (Least squares estimates) From Ch 6 Prob 14

$$\begin{aligned}\hat{\mu}_i &= \bar{y}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} \\ \hat{\nu} &= \bar{y}_{..} = \frac{1}{n}\sum\sum y_{ij} \quad , \quad n = \sum n_i\end{aligned}$$

8. (F-statistic) Ch 6 Prob 15

$$\begin{aligned}\hat{y}_{ij} &= \bar{y}_{i.} \\ \hat{\hat{y}} &= \bar{y}_{..} \\ F &= \frac{n-p}{p-q}\frac{\|\hat{y}-\hat{\hat{y}}\|^2}{\|y-\hat{y}\|^2} = \frac{n-r}{r-1}\frac{\sum\sum(\bar{y}_{i.}-\bar{y}_{..})^2}{\sum\sum(y_{ij}-\bar{y}_{i.})^2}\end{aligned}$$

9. (ANOVA table)

Source	SS	df	MS*
between	$\sum\sum(\bar{y}_{i.}-\bar{y}_{..})^2$	$r-1$	MS_b
within	$\sum\sum(y_{ij}-\bar{y}_{i.})^2$	$n-r$	MS_w

* MS = SS/df

10. (Th) If the y_{ij} are independent $N(\mu_i, \sigma^2)$ and

$$\mu_1 = \cdots = \mu_r$$

then

$$F = \mathrm{MS}_b/\mathrm{MS}_w \sim F(r-1, n-r)$$

Pf: Th 6.15

11. (Confidence intervals) If the y_{ij} are independent $N(\mu_i, \sigma^2)$ and $a \neq b$

$$\frac{\bar{y}_{a.} - \bar{y}_{b.} - (\mu_a - \mu_b)}{(\text{MS}_w(\frac{1}{n_a} + \frac{1}{n_b}))^{1/2}} \sim t(n-r)$$

Pf:

$$\begin{aligned} \text{var}(\bar{y}_{a.} - \bar{y}_{b.}) &= \sigma^2/n_a + \sigma^2/n_b \\ \widehat{\text{var}}(\bar{y}_{a.} - \bar{y}_{b.}) &= \hat{\sigma}^2(\frac{1}{n_a} + \frac{1}{n_b}) = \text{MS}_w(\frac{1}{n_a} + \frac{1}{n_b}) \end{aligned}$$

Use Th 6.14.

12. (Alternate form of the model)

$$y_{ij} = \nu + \alpha_i + e_{ij} \quad , \quad \mu_i = \nu + \alpha_i$$

For identification one usually assumes

$$\sum \alpha_i = 0$$

but we don't because SAS doesn't.

13. (Ex 7.1, p198) Coin data.

$$y_{ij} = \nu + \alpha_i + e_{ij}$$

Dummy variables

1. (Representation for μ_i)

$$\mu_i = \nu + \alpha_i = \nu + \alpha_1 d_i^{(1)} + \cdots + \alpha_r d^{(r)}$$

where $d_i^{(k)} = \delta_{ik} =$ Kronecker δ.

2. (Note) This is a regression model with carriers

$$\mathbf{1}, d^{(1)}, \cdots, d^{(r)}$$

3. (Ex) When i has 3 values SAS uses

	ν	α_1	α_2	α_3	
i	$\mathbf{1}$	$d^{(1)}$	$d^{(2)}$	$d^{(3)}$	μ_i
1	1	1	0	0	$\nu + \alpha_1$
2	1	0	1	0	$\nu + \alpha_2$
3	1	0	0	1	$\nu + \alpha_3$

4. (Note) Other systems use other dummy variables.

The simple two-way additive model

1. (Ex)

coin type (rows) × fountains (columns)

y_{11}	y_{12}	y_{13}
y_{21}	y_{22}	y_{23}

2. (Model)

$$\begin{aligned} y_{ij} &= \nu + \alpha_i + \beta_j + e_{ij} \qquad i = 1, \cdots, r \;;\; j = 1, \cdots, c \\ \mu_{ij} &= \nu + \alpha_i + \beta_j \\ &= \text{fnc}(i) + \text{fnc}(j) = \text{ additive model} \end{aligned}$$

3. (Basic property)

$$\begin{aligned} \mu_{aj} - \mu_{bj} &= \alpha_a - \alpha_b \quad \text{independent of } j \\ \mu_{ia} - \mu_{ib} &= \beta_a - \beta_b \quad \text{independent of } i \end{aligned}$$

4. (Least squares fit)

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

Pf: Page 104, $y - \hat{y} \perp \mathcal{M}$.

5. (Th) $\dim \mathcal{M} = r + c - 1$

Pf: For the additive model

$$(\mu_{ij}) = \begin{matrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{matrix} \longleftrightarrow \begin{matrix} * & * & * & * \\ * & & & \\ * & & & \end{matrix}$$

Thus

$$\dim \mathcal{M} = r + c - 1$$

6. (The basic row hypothesis)

$$\alpha_1 = \cdots = \alpha_r$$

The corresponding restricted model is

$$y_{ij} = \nu + \beta_j + e_{ij}$$

The least squares fit:

$$\hat{\hat{y}}_{ij} = \bar{y}_{.j}$$

7. (Th) $\quad \dim \mathcal{H}_{row} = c$

8. (The F-statistic)

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

$$\begin{aligned}
\|\hat{y} - \hat{\hat{y}}\|^2 &= \sum\sum(\bar{y}_{i.} - \bar{y}_{..})^2 = \mathrm{SS}_{row} \\
\|y - \hat{y}\|^2 &= \sum\sum(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 = \mathrm{SS}_{res} \\
\mathrm{df}_{row} &= p - q = (r + c - 1) - c = r - 1 \\
\mathrm{df}_{res} &= n - p = rc - r - c + 1 = (r-1)(c-1)
\end{aligned}$$

$$F_{row} = \mathrm{MS}_{row}/\mathrm{MS}_{res}$$

9. (ANOVA table)

Source	SS	df	MS
row	$\sum\sum(\bar{y}_{i.} - \bar{y}_{..})^2$	$r-1$	MS_{row}
col	$\sum\sum(\bar{y}_{.j} - \bar{y}_{..})^2$	$c-1$	MS_{col}
res	$\sum\sum(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	$(r-1)(c-1)$	MS_{res}

10. (Th) If the e_{ij} are independent $N(0, \sigma^2)$, then

(a) If $\alpha_1 = \cdots = \alpha_r$

$$F_{row} = \mathrm{MS}_{row}/\mathrm{MS}_{res} \sim F(r-1, (r-1)(c-1))$$

(b) If $\beta_1 = \cdots = \beta_c$

$$F_{col} = MS_{col}/MS_{res} \sim F(c-1, (r-1)(c-1))$$

Pf: Theorem 6.15

11. (Example) Two coin types and three fountains.

$$y = \nu + \alpha_i + \beta_j + e$$

Estimability

1. (Def) In the linear regression model

$$y = X\beta + e$$

a linear combination $c'\beta$ is called estimable iff for some ℓ

$$c'\beta = \ell'\mu$$

for all β and μ such that $\mu = X\beta$.

2. (Th) Let $\hat{y}$ be the least squares fit for the model in Def 1.

If $c'\beta$ is estimable

(a) $c'\hat{\beta} = \ell'\hat{y}$ ($\hat{\beta}$ need not be unique)

If also $E\,e = 0$

(b) $c'\hat{\beta}$ is an unbiased estimate of $c'\beta$

If also $\operatorname{cov} e = \sigma^2 I$

(c) $c'\hat{\beta}$ is a BLUE of $c'\beta$

Pf:

(a) $\hat{y} = X\hat{\beta} \Longrightarrow c'\hat{\beta} = \ell'\hat{y}$

(b) $E\,\ell'\hat{y} = \ell'\mu = c'\beta$

(c) The Gauss-Markov Th

3. (Def) $c'\hat{\beta} = \ell'\hat{y}$ is called the least squares estimate of $c'\beta$.

4. (Note) The rest of this section applies to the simple two-way additive model.

5. (Ex) In the simple two-way additive model

$$\mu_{ij} = \nu + \alpha_i + \beta_j$$

   (a) none of the parameters ν, α_i, β_j are estimable.

   (b) all differences $\alpha_a - \alpha_b$ are estimable.

   Pf: (a) Let $\nu = 2$, all $\alpha_i = -1$, and all $\beta_j = -1$. Then $\mu_{ij} = 0$ for all i and j. If ν, α_i, and β_j were estimable they would also have to be zero because they would then be linear combinations of the μ_{ij}. Since ν, α_i, and β_j are not zero, they are not estimable.

   (b) Since

$$\alpha_a - \alpha_b = \mu_{a1} - \mu_{b1}.$$

   $\alpha_a - \alpha_b$ is a linear combination of the μ_{ij} and hence estimable.

6. (Ex) The LS estimate of $\alpha_a - \alpha_b$ is

$$\bar{y}_{a.} - \bar{y}_{b.}$$

Pf:

$$\begin{aligned}
\alpha_a - \alpha_b &= \mu_{a1} - \mu_{b1} \\
(\alpha_a - \alpha_b)\hat{} &= \hat{y}_{a1} - \hat{y}_{b1} \\
&= (\bar{y}_{a.} + \bar{y}_{.1} - \bar{y}_{..}) - (\bar{y}_{b.} + \bar{y}_{.1} - \bar{y}_{..}) \\
&= \bar{y}_{a.} - \bar{y}_{b.}
\end{aligned}$$

7. (Note) If the e_{ij} are independent $N(0, \sigma^2)$,

$$\frac{\bar{y}_{a.} - \bar{y}_{b.} - (\alpha_a - \alpha_b)}{(2\mathrm{MS}_{res})^{1/2}} \sim t(\mathrm{df}_{res})$$

Pf:

$$\begin{aligned}
\mathrm{var}(\bar{y}_{a.} - \bar{y}_{b.}) &= \sigma^2/c + \sigma^2/c \\
\widehat{\mathrm{var}}(\bar{y}_{a.} - \bar{y}_{b.}) &= 2\hat{\sigma}^2/c = 2\mathrm{MS}_{res}/c
\end{aligned}$$

Apply Th 6.11.

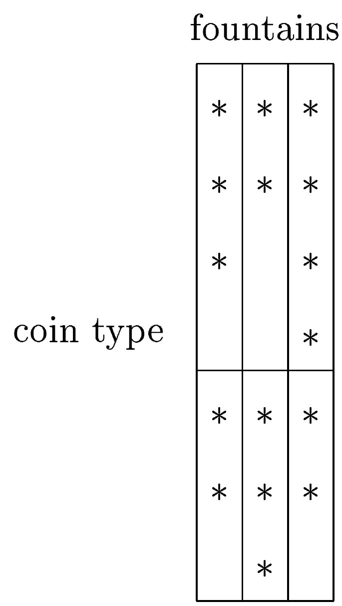
8. (Ex 7.2, p 206) Coin data

$$y = \nu + \alpha_i + \beta_j + e$$

$$
\begin{aligned}
(\alpha_2 - \alpha_1)\hat{} &= \bar{y}_{2.} - \bar{y}_{1.} = 171.6 - 142.3 = 29.3 = \text{ output} \\
(2\text{MS}_{res})^{1/2} &= (2 \cdot 1037/3)^{1/2} = 26.29 = \text{ output}
\end{aligned}
$$

The general two-way additive model

1. (Ex)



2. (Model)

$$
\begin{aligned}
y_{ijk} &= \nu + \alpha_i + \beta_j + e_{ijk} \qquad k = 1, \cdots, n_{ij} \\
\tau_{ij} &= \nu + \alpha_i + \beta_j = \text{ reg fnc } = \text{ cell response} \\
\mu_{ijk} &= \tau_{ij} = \text{ theoretical response} \quad , \quad k = 1, \cdots, n_{ij}
\end{aligned}
$$

3. (Def) The model is:

   (a) **balanced** if all n_{ij} are equal

   (b) **incomplete** if some $n_{ij} = 0$

4. (Note) In the unbalanced case there is no "summation formula" ANOVA table like Table 7.4. Moreover,

$$(\alpha_a - \alpha_b)\hat{} \neq \bar{y}_{a..} - \bar{y}_{b..}$$

5. (ANOVA table)

Source	SS	df	MS
row	SS_{row}	df_{row}	MS_{row}
col	SS_{col}	df_{col}	MS_{col}
res	SS_{res}	df_{res}	MS_{res}

6. (Sums of squares)

(a) $\text{SS}_{res} = \|y - \hat{y}\|^2$

where $\hat{y}$ is the fit for the model.

$$y_{ijk} = \nu + \alpha_i + \beta_j + e_{ijk}$$

(b) $\text{SS}_{row} = \|\hat{y} - \hat{\hat{y}}\|^2$

where $\hat{\hat{y}}$ is the fit for the restricted model

$$y_{ijk} = \nu + \beta_j + e_{ijk}$$

(c) $\text{SS}_{col} = \|\hat{y} - \hat{\hat{y}}\|^2$

where $\hat{\hat{y}}$ is the fit for

$$y_{ijk} = \nu + \alpha_i + e_{ijk}$$

7. (Note) SS_{row} and SS_{col} are the difference sums of squares for the hypotheses:

$$\alpha_1 = \cdots = \alpha_r$$

and

$$\beta_1 = \cdots = \beta_b$$

8. (Degrees of freedom)

$$\text{df} = \text{rank}\,\text{SS}$$

9. (Th) If the e_{ijk} are independent $N(0, \sigma^2)$

$$\begin{aligned} \alpha_1 = \cdots = \alpha_r &\implies \frac{\text{MS}_{row}}{\text{MS}_{res}} \sim F(\text{df}_{row}, \text{df}_{res}) \\ \beta_1 = \cdots = \beta_c &\implies \frac{\text{MS}_{col}}{\text{MS}_{res}} \sim F(\text{df}_{col}, \text{df}_{res}) \end{aligned}$$

Pf: The fundamental F-theorem, Th 6.16.

10. (Note) If every cell response τ_{ij} is estimable

$$\begin{aligned} \text{df}_{res} &= n - r - c + 1 \\ \text{df}_{row} &= r - 1 \\ \text{df}_{col} &= c - 1 \end{aligned}$$

Pf: See p210 (From the simple additive model result)

11. (Ex) Consider

*	*	*
*	*	

all τ_{ij} are estimable

Here

$$\text{df}_{res} = n - r - c + 1 = 5 - 2 - 3 + 1 = 1$$

Consider

*	

only τ_{11} is estimable

Here

$$\text{df}_{res} = 0 \neq 1 - 2 - 2 + 1 = n - r - c + 1$$

12. (Note) For complete models all τ_{ij} are estimable.

Pf: $\tau_{ij} = \mu_{ij1}$

13. (Note) If all τ_{ij} are estimable, all differences

$$\alpha_a - \alpha_b \text{ and } \beta_a - \beta_b$$

are estimable

Pf: $\alpha_a - \alpha_b = \tau_{a1} - \tau_{b1} = \ell' \mu$

14. (Confidence intervals and tests) Let θ denote all of the model parameters. If $\delta = c'\theta$ is estimable, then $\delta = \ell'\mu$ for some ℓ and $\hat{\delta} = \ell'\hat{y}$ is the least squares estimate of δ.

Under normality
$$\frac{\hat{\delta} - \delta}{\widehat{\text{std}}\,\hat{\delta}} \sim t(\text{df}_{res})$$

Pf: Equation (6.15).

15. (Ex 7.3, p211) Coin data

$$\begin{aligned} y_{ijk} &= \nu + \alpha_i + \beta_j + e_{ijk} \\ (n_{ij}) &= \begin{pmatrix} 6 & 8 & 5 \\ 6 & 7 & 5 \end{pmatrix} \end{aligned}$$

SAS specification is the same as for the simple additive model. The data set, however, is larger.

Since this model is complete

$$\begin{aligned} \mathrm{df}_{res} &= n - r - c + 1 = 37 - 2 - 3 + 1 = 33 \\ \mathrm{df}_{row} &= r - 1 = 1 \\ \mathrm{df}_{col} &= c - 1 = 2 \end{aligned}$$

The General two-way model

1. (Model)

$$y_{ijk} = \nu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \qquad k = 1, \cdots, n_{ij}$$

   Cell responses:

$$\tau_{ij} = \nu + \alpha_i + \beta_j + \gamma_{ij}$$

   No constraints. The τ_{ij} are arbitrary.

2. (Additivity hypothesis)

$$\gamma_{ij} = 0 \quad \text{all} \quad i, j$$

   This is called the **additivity** or no **interaction** hypothesis.

3. (Interaction plot)

parallel $\Longrightarrow$ no interaction

4. (ANOVA table)

Source	SS	df	MS
row	SS_{row}	df_{row}	MS_{row}
col	SS_{col}	df_{col}	MS_{col}
int	SS_{int}	df_{int}	MS_{int}
res	SS_{res}	df_{res}	MS_{res}

5. (Sums of squares)

SS	hypothesis
row	$\bar{\tau}_{1.} = \cdots = \bar{\tau}_{r.}$
col	$\bar{\tau}_{.1} = \cdots = \bar{\tau}_{.c}$
int	$\gamma_{ij} = 0$ all i, j

6. (Degrees of freedom)

$$\text{df} = \text{rank}\,\text{SS}$$

7. (Th) If the e_{ijk} are independent $N(0, \sigma^2)$

$$\begin{aligned} \text{row hyp} &\implies \frac{\text{MS}_{row}}{\text{MS}_{res}} \sim F(\text{df}_{row}, \text{df}_{res}) \\ \text{col hyp} &\implies \frac{\text{MS}_{col}}{\text{MS}_{res}} \sim F(\text{df}_{col}, \text{df}_{res}) \\ \text{int hyp} &\implies \frac{\text{MS}_{int}}{\text{MS}_{res}} \sim F(\text{df}_{int}, \text{df}_{res}) \end{aligned}$$

8. (Note) All τ_{ij} are estimable if and only if the general two-way model is complete.

Pf: (if)

$$\text{complete} \implies \tau_{ij} = \mu_{ij1} \implies \text{all } \tau_{ij} \text{ est}$$

(only if)

$$\text{incomplete} \implies n_{ab} = 0 \text{ some } a \text{ and } b$$

Set all parameters zero except $\gamma_{ab} = 1$. Then

$$\mu_{ijk} = \nu + \alpha_i + \beta_j + \gamma_{ij} = 0 \quad \text{all} \quad i, j, k$$

Since $\tau_{ab} = \gamma_{ab} = 1$, τ_{ab} is not estimable.

9. (Th) For a complete model

$$\begin{aligned} \text{df}_{row} &= r - 1 \\ \text{df}_{col} &= c - 1 \\ \text{df}_{int} &= (r-1)(c-1) \\ \text{df}_{res} &= n - rc \end{aligned}$$

Pf: Text

10. (Ex 7.4) Complete general two-way model. Coin data.

$$y_{ijk} = \nu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

11. (Interaction plot)

$$\begin{aligned} \hat{\tau}_{ij} &= \hat{y}_{ijk} \\ \text{plot} \qquad & \hat{y} * j = i \end{aligned}$$

12. (Incomplete general two-way model)

* * *	* *	* * *
* *	* * *	empty

Not all τ_{ij} are estimable and you may not get what you hoped.

PROC GLM gives:

$$\text{SS}_{res} \quad , \quad \text{SS}_{int}$$

as defined above, but does not give:

$$\text{SS}_{row} \quad , \quad \text{SS}_{col}$$

Instead it uses:

$$\begin{aligned} \text{row hyp:} \quad & \tau_{11} + \tau_{12} = \tau_{21} + \tau_{22} \\ \text{col hyp:} \quad & \tau_{11} + \tau_{21} = \tau_{12} + \tau_{22} \\ & \tau_{11} + \tau_{12} = 2\tau_{13} \end{aligned}$$

Problems arise with ESTIMATE and LSMEANS statements also.

Higher order models

1. (Additive 3-way model)

$$y_{ijkl} = \nu + \alpha_i + \beta_j + \gamma_k + e_{ijkl}$$

PROC GLM;

CLASS I J K;

MODEL Y = I J K;

The Type III sums of squares for I tests the hyp:

$$\alpha_1 = \cdots = \alpha_a$$

ESTIMATE 'A2-A3' I -1 1 estimates:

$$\alpha_1 - \alpha_2 \quad \text{(if it is estimable)}$$

2. (General 3-way model)

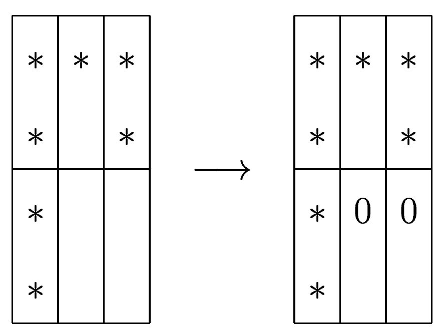
$$y_{ijkl} = \nu+\alpha_i+\beta_j+\gamma_k+(\alpha\beta)_{ij}+(\alpha\gamma)_{ik}+(\beta\gamma)_{jk}+(\alpha\beta\gamma)_{ijk}+e_{ijkl}$$

If all cell responses τ_{ijk} are estimable, the Type III sums of squares are difference sums of squares for the hypotheses:

Type III SS	hypothesis
I	$\bar{\tau}_{i..}$ is constant
⋮	⋮
I*J	$\tau_{ij.}$ is additive
⋮	⋮
IJK	$(\alpha\beta\gamma)_{ijk} = 0$

3. (Other models) Statements similar to those for the general 3-way model hold in general when all cell responses are estimable.

4. (Testing all cells estimable) See p223.



MODEL df $\longrightarrow$ MODEL df

no change $\Longleftrightarrow$ all cells estimable

Analysis of covariance

1. (Model)

$$y_{ij} = \alpha_i + \beta x_{ij} + e_{ij} \qquad i = 1, \cdots, r \ \ ; \ \ i = 1, \cdots, n_i$$

   Let

$$f(i, x) = \alpha_i + \beta x = \text{ reg fun}$$

   Note this is additive in i and x

2. (Picture)

   Note equal slope.

3. (Note) The regression line for each group passes through the group mean.

$$\bar{y}_{i.} = \hat{\alpha}_i + \hat{\beta}\bar{x}_{i.}$$

Pf: For any sequence α_i

$$\begin{aligned}
\sum_i \sum_j (y_{ij} - \hat{\alpha}_i - \hat{\beta} x_{ij})\alpha_i &= 0 \\
\sum_j (y_{ij} - \hat{\alpha}_i - \hat{\beta} x_{ij}) &= 0 \qquad \text{all } i \\
\bar{y}_{i.} - \hat{\alpha}_i - \hat{\beta}\bar{x}_{i.} &= 0
\end{aligned}$$

4. (Def) The adjusted group mean is

$$\tau_i = f(i, \bar{x}_{..})$$

5. (Note) If β is estimable it's least squares estimate is

$$\hat{\tau}_i = \bar{y}_{i.} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})$$

The last term looks like an adjustment.

Pf:

$$\begin{aligned}
\hat{\tau}_i &= \hat{\alpha}_i + \hat{\beta}\bar{x}_{..} \\
&= \bar{y}_{i.} - \hat{\beta}\bar{x}_{i.} + \hat{\beta}\bar{x}_{..} \\
&= \bar{y}_{i.} - \hat{\beta}(\bar{x}_{i.} - \bar{x}..)
\end{aligned}$$

6. (Standard tests)

$$\begin{aligned}
&\text{between:} \quad \alpha_1 = \cdots = \alpha_r \\
&\text{slope:} \qquad \beta = 0
\end{aligned}$$

7. (ANCOVA table)

source	SS	df	MS
between	SS_{btw}	df_{btw}	MS_{btw}
slope	SS_{reg}	df_{reg}	MS_{reg}
res	SS_{res}	df_{res}	MS_{res}

8. (Th) If β is estimable

$$\begin{aligned} df_{btw} &= r - 1 \\ df_{reg} &= 1 \\ df_{res} &= n - r - 1 \end{aligned}$$

Pf: Easy

9. (Ex 7.7, p226) Coins.

$$\begin{aligned} i &= \text{ coin type} \qquad x = \text{ fountain depth} \\ y_{ij} &= \nu + \alpha_i + \beta x_{ij} + e_{ij} \end{aligned}$$

- MODEL Y = I X   gives the ANCOVA table
- ESTIMATE 'B' X 1   estimates β = slope
- ESTIMATE 'A2-A1' I -1 1   estimates $\alpha_2 - \alpha_1 = \tau_2 - \tau_1$

- LSMEANS I estimates $\tau_i = f(i, \bar{x})$ = adj group mean

10. (Equality of slope test) Test $\beta_1 = \cdots = \beta_r$ in the model

$$\begin{aligned} y_{ij} &= \alpha_i + \beta_i x_{ij} + e_{ij} \\ F &= \text{MS}_{eql}/MS_{res} = \text{ GLM F-statistic} \end{aligned}$$

11. (Th) If all β_i are estimable

$$\begin{aligned} \text{df}_{eql} &= r - 1 \\ \text{df}_{res} &= n - 2r \end{aligned}$$

12. (Equality of slope test SAS) Model

$$\begin{aligned} y &= \nu + \alpha_i + \beta x + \gamma_i x + e \\ \text{Y} &= \text{I X I*X} \end{aligned}$$

Test: $\gamma_j = 0$ which is the I*X term.

13. (Other generalizations)

Two covariates	$y = \alpha_i + \beta_1 x_1 + \beta_2 x_2 + e$
Two-way additive	$y = \nu + \alpha_i + \beta_j + \gamma x + e$

Transformation and diagnostics

1. (Note) Balance $\Longrightarrow$ equal leverage

   $\Longrightarrow$ no leverage or influence indices.

   Residuals, however, are important.

2. (Note) A model is called saturated when the cell responses are arbitrary.

   Saturated model $\Longrightarrow$ unbiased and linear

   $\Longrightarrow$ transformations may be used for normalization.

   One-way model:

$$\begin{aligned} y_{ij} &= \alpha_i + e_{ij} \\ \log\, y_{ij} &= \log\, \alpha_i + e_{ij} \end{aligned}$$

   Both are linear and unbiased. Here we have transformed both sides.

3. (Transforming to an additive model)

$$\begin{aligned} y &\approx \alpha_i \beta_j \\ \log\, y &\approx \log\, \alpha_i + \log\, \beta_j \end{aligned}$$

4. (Ex 7.8, p231) Survival times, Table 7.14

$$s = \nu + \alpha_i + \beta_j + e$$

$$i = \text{ poison } , \quad j = \text{ treatment}$$

Fig 7.12

Fig 7.13

$$y = \frac{1}{s} = \text{ rate of dyeing}$$

$$y = \nu + \alpha_i + \beta_j + e$$

Fig 7.14

$$F_{int} = 1.21 \quad \text{not sig (Prob 19)}$$