Regression Model Approach

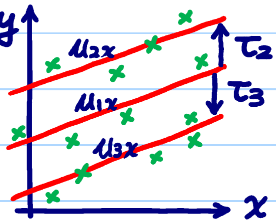
add a constraint on $\tau_1, \cdots, \tau_k$, say, $\tau_1 = 0$

reference line

potato

canna

corn



$\mu_{1,x} = E(y_{1,x}) = \eta + \quad + \gamma x$

$\mu_{2,x} = E(y_{2,x}) = \eta + \tau_2 + \gamma x$

$\mu_{3,x} = E(y_{3,x}) = \eta + \tau_3 + \gamma x$

η, γ: intercept & slope of $\mu_{1,x}$

$\tau_2 = \mu_{2,x} - \mu_{1,x}$

$\tau_3 = \mu_{3,x} - \mu_{1,x}$

- Model :

baseline (reference)

functional form

$$
\begin{array}{lllll}
y_{1j} = \eta + \quad + \gamma \times x_{1j} + \varepsilon_{1j}, & j = 1, \cdots, 13, & i = 1 & \text{(canna)} & \\
y_{2j} = \eta + \tau_2 + \gamma \times x_{2j} + \varepsilon_{2j}, & j = 1, \cdots, 19, & i = 2 & \text{(corn)} & (6) \\
y_{3j} = \eta + \tau_3 + \gamma \times x_{3j} + \varepsilon_{3j}, & j = 1, \cdots, 17, & i = 3 & \text{(potato)} &
\end{array}
$$

$\tau_1 = 0$ → treatment codings

Matrix form: $Y = X\beta + \varepsilon$

model matrix (MM $\overset{cf.}{\longleftrightarrow}$ DM)

where

- τ_1 is set to zero (baseline constraint),
- η = intercept,
- γ = regression coefficient for thickness,
- τ_2 = canna vs. corn, and
- τ_3 = canna vs. potato.

$$
\begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ y_{31} \\ \vdots \\ y_{3n_3} \end{bmatrix}
=
\begin{bmatrix} 1 & 0 & 0 & x_{11} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & x_{1n_1} \\ 1 & 1 & 0 & x_{21} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & x_{2n_2} \\ 1 & 0 & 1 & x_{31} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 1 & x_{3n_3} \end{bmatrix}
\begin{bmatrix} \eta \\ \tau_2 \\ \tau_3 \\ \gamma \end{bmatrix}
+ \underset{\sim}{\varepsilon}
$$

(Write the model matrix for (6)).

(exercise) What if sum codings are used?

- Run regression analysis in the usual way.

Regression Analysis of Starch Experiment

Q: how to perform sequential ANOVA for this case?

It shows the necessarity to to add the covariate, which plays a role similar to block factor in the analysis, but usually not orthogonal to treatment factor.

Table 21: Tests, Starch Experiment (cf)

drop-one test

	Effect	Estimate	Standard Error	t	p-value
η	intercept	158.261	179.775	0.88	0.38
γ	thickness	62.501	17.060	3.66	0.00
τ_2	canna vs. corn	−83.666	86.095	−0.97	0.34
τ_3	canna vs. potato	70.360	67.781	1.04	0.30
	corn vs. potato	154.026	107.762	1.43	0.16

Ω: model (6) in LNp.41

ω: model (6) in LNp.41 with τ_2 setting to be 0

both models contain the term "γx"

H_0: $\tau_2 = \tau_3$ ← ω: merge τ_2, τ_3 columns in X (MM)

- In the table, corn vs. potato $= \hat{\tau}_3 - \hat{\tau}_2 = 70.360 - (-83.666) = 154.026$.

$\mu_{3,x} - \mu_{2,x} = (\mu_{3,x} - \mu_{1,x}) - (\mu_{2,x} - \mu_{1,x}) = \tau_3 - \tau_2$

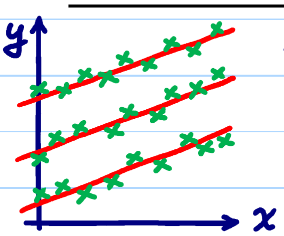
- No pair of film types has any significant difference after adjusting for thickness effect. (So, how should the choice be made between the three film types?) Most of the variation is explained by the covariate thickness.

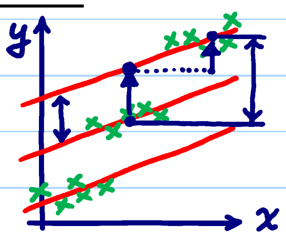
"no significant difference" can also be useful information for making decision.

Q: What if we fit the model: $y \sim \beta_0 + \text{starch} + \varepsilon$?

$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$

(B) $H_1X_2\beta_2$ $\quad$ $(I - H_1)X_2\beta_2$ (W)





Multiple Comparisons

① obtain $t_{ij} = \frac{\hat{\mu}_{jx} - \hat{\mu}_{ix}}{s.e.(\hat{\mu}_{jx} - \hat{\mu}_{ix})}$
② determine critical value (or p-value)

- $Var(\hat{\tau}_3)$ and $Var(\hat{\tau}_2)$ can be obtained from regression output Table 21 (LNp.4-42).
- From (1.33) of textbook (p.22),
  $Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$. → $\sigma^2 \begin{bmatrix} \ddots & & & & \\ & (X^TX)^{-1}_{ii} & \cdots & (X^TX)^{-1}_{ij} & \cdots \\ & & \ddots & & \\ & & & (X^TX)^{-1}_{jj} & \\ & & & & \ddots \end{bmatrix}$ ① ③ ②
  t_{12}, t_{13}
- Using this, for H_0: $\tau_2 = \tau_3$ (i.e., $\tau_3 - \tau_2 = 0$), the variance of $\hat{\tau}_3 - \hat{\tau}_2$ can be found as

$$t_{23} = \frac{(\hat{\tau}_3 - \hat{\tau}_2) - 0}{s.e.(\hat{\tau}_3 - \hat{\tau}_2)}$$

$$Var(\hat{\tau}_3 - \hat{\tau}_2) = \underbrace{Var(\hat{\tau}_3)}_{②} + \underbrace{Var(\hat{\tau}_2)}_{①} - 2\,\underbrace{Cov(\hat{\tau}_3, \hat{\tau}_2)}_{③}.$$

- The degrees of freedom for the t-statistic is same as that of the residuals. The p-values for the three tests are given in Table 21 (LNp.4-42).
  for specific (i, j)
- For simultaneous testing, use adjusted p-values (LNp.4-24).
  for all (i, j)'s

ANCOVA Table

Analysis of covariance

Apply sequential ANOVA to the model with covariate — treated as block factor and usually not orthogonal to treatment factor

anova($y \sim \beta_0 +$ thickness$_e$ + starch)

Note. covariate must appear before treatment factor

(M1) $y \sim \beta_0$
(M2) $y \sim \beta_0 +$ thickness$_e$
(M3) $y \sim \beta_0 +$ thickness$_e$ + starch

Table 22: ANCOVA Table, Starch Experiment

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F
thickness	1	2553357	2553357	94.19
starch	2	56725	28362	1.05
residual	45	1219940	27110	

Ω: (M2), ω: (M1) — thickness
Ω: (M3), ω: (M2) — starch
not orthogonal
df. of residuals under (M3)
cf. the t-value in Table 21 (LNp.42)
not significant

(exercise) Q. What if we use anova($y \sim \beta_0 +$ starch + thickness$_e$)? (starch: qualitative, thickness: quantitative)

Note. It is possible that we are not interested in "starch" (↔ block), but are interested in the coefficient of "thickness" (↔ treatment)

❖ **Reading**: textbook, 3.10

Example of Split-plot Design: Wood Experiment

- Experiment objective : to study the water resistant property of wood.
  response y
- Two treatment factors:
  - A — wood pretreatments: a_1, a_2;
    qualitative, 2 levels
  - B — types of stain: b_1, b_2, b_3, b_4.
    qualitative, 4 levels
- Completely randomized design: randomly apply the 8 level combinations of A and B to 8 wood panels, such as in Table 23.
- Problem: inconvenient to apply the pretreatments to a small wood panel.

★ Exp'tal unit : a small wood panel

homogeneous → EU$_1$ • • • EU8

completely randomized → $(a_1,b_1)\cdots(a_2,b_4)$

(check 2-way layout in LNp.4-15)

★ Design matrix

pretreatment	stain	y
a_1	b_1	□⋯□
a_1	b_2	□⋯□
a_1	b_3	⋮
a_1	b_4	
a_2	b_1	
a_2	b_2	
a_2	b_3	□⋯□
a_2	b_4	□⋯□

Data matrix

of ε's = # of EUs

★ conceptual model : $y \sim B_0 + A + B + A\times B + \varepsilon$

1 parameter; 3 parameters; 3 parameters

EUs → x → y, ε

Q: What does ε represent?

★ $\mathrm{cov}(\underset{\sim}{y}) = \sigma^2 I$

Q: How many times pretreatments (a_1 or a_2) are applied?
Ans: 8 times → 8 EUs

Table 23: Completely Randomized Version of the Wood Experiment

Run (EUs)	1	2	3	4	5	6	7	8
Pretreatment (A)	a_1	a_2	a_2	a_1	a_2	a_1	a_1	a_2
Stain (B)	b_2	b_4	b_1	b_1	b_3	b_4	b_3	b_2

← # of distinct level combinations of A and B

Split-plot Design

- Alternative Design: split-plot design in Table 24.

Table 24: Split-Plot Version of the Wood Experiment

First panel				Second panel			
Pretreated with a_1				Pretreated with a_2			
b_3	b_2	b_4	b_1	b_2	b_1	b_4	b_3

Design matrix and data matrix look the same as those in 2-way layout

★ Design matrix

pretreatment	stain	y
a_1	b_1	□
a_1	b_2	□
a_1	b_3	⋮
a_1	b_4	⋮
a_2	b_1	⋮
a_2	b_2	
a_2	b_3	□
a_2	b_4	□

Data matrix

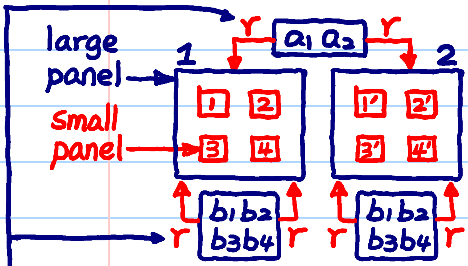
SPD   CRD

規律 X B
structures in y
隨機 ε

Q: How many times pretreatments (a_1 or a_2) are applied?
Ans: 2 times

- Justification: Easier to apply pretreatment to *large* wood panels.

★ Exp'tal units (Q: what are the exp'tal units):



different restricted randomization schemes for A & B.

- EUs for A and B are different.
- EUs for A : large panels
- EUs for B : small panels
- large panels are split into small panels → "small panel" is nested in "large panel"

LNp.4-15

EUs	y
1	□
2	□
3	•
4	•
5	•
6	
7	□
8	□

large panel	small panel	y
1	1	□
1	2	□
1	3	•
1	4	•
2	1'	•
2	2'	
2	3'	□
2	4'	□

Split-plot Design (Cont'd)

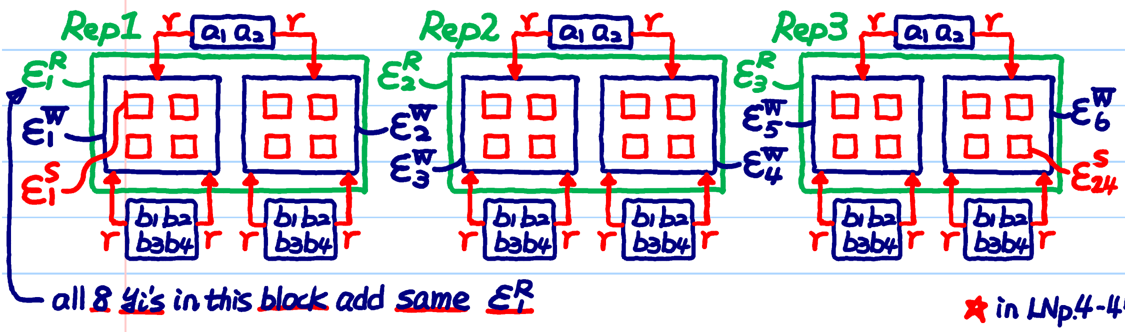
Why is it called "plot"?

- Split-plot design (and the name) has its origin in agriculture.
- Some factors need to be applied to large plots, called *whole plots*. In the example, the two big wood panels to which pretreatement a_1 and a_2 are applied are whole plots.
- Split each whole plot into smaller plots, called *subplots*. In the example, the four small wood panels within the large panels are subplots.

Exp'tal unit $\leftrightarrow$ ε (check LNp.4-45)

- Wood Experiment: A: whole-plot factor; B: subplot factor
  - $\varepsilon_1^R, \varepsilon_2^R, \varepsilon_3^R \leftarrow$ 3 replications (treated as 3 blocks with random effects), 1 block factor with 3 levels
  - $\varepsilon_1^W, \cdots, \varepsilon_6^W \leftarrow$ 6 whole plots (two large panels for a_1 and a_2 per replication),
  - $\varepsilon_1^S, \cdots, \varepsilon_{24}^S \leftarrow$ 24 subplots (four small panels for b_1, b_2, b_3, and b_4 per large panel).

similar to the 1-way random effects model discussed in LNp 3-31~38: $\tau_i + \varepsilon_{ij}$



all 8 y_i's in this block add same ε_1^R

★ Conceptual model:

$$y \sim B_0 + A + B + A\times B + \varepsilon^R + \varepsilon^W + \varepsilon^S$$

$E(y)$: $B_0 + A + B + A\times B$; $cov(y)$: $\varepsilon^R + \varepsilon^W + \varepsilon^S$

★ in LNp.4-45 (cf.)

Data from the Wood Experiment

3 parameters about error variance in Σ: $\sigma_R^2, \sigma_W^2, \sigma_S^2$

$\mathcal{R}_1 \subset \mathcal{R}_2 \subset \mathcal{R}_3$, dim= 3 6 24

σ_W^2 ($\mathcal{R}_2$), σ_S^2 ($\mathcal{R}_3$), W_A, $W_{A\times B}$, W_B, σ_R^2 ($\mathcal{R}_1$)

Whole plot	Sub-plot	Pretreatment (A)	Stain type (B)	Replication (Rep)	Resistance (Y)
4	1 ε_1^S	a_2 -1	b_2 0 1 0	1	53.5
4	2 ε_2^S	a_2	b_4 -1 -1 -1	1	32.5
4	3	a_2	b_1 1 0 0	1	46.6
4	4	a_2 -1	b_3 0 0 1	1	35.4
1	5	a_1 1	b_3	1	40.8
1	6	a_1	b_1	1	43.0
1	7	a_1	b_2	1	51.8
1	8	a_1 1	b_4	1	45.5
5		a_2 -1	b_4	2	44.6
5		a_2	b_1	2	52.2
5		a_2	b_3	2	45.9
5		a_2 -1	b_2	2	48.3
2		a_1 1	b_2	2	60.9
2		a_1	b_4	2	55.3
2		a_1	b_3	2	51.1
2		a_1 1	b_1	2	57.4
6		a_2 -1	b_1	3	32.1
6		a_2	b_4	3	30.1
6		a_2	b_2	3	34.4
6		a_2 -1	b_3	3	32.2
3		a_1 1	b_1 1 0 0	3	52.8
3		a_1	b_3 0 0 1	3	51.7
3	23 ε_{23}^S	a_1	b_4 -1 -1 -1	3	55.3
3	24 ε_{24}^S	a_1 1	b_2 0 1 0	3	59.2

Whole plots: $\varepsilon_4^W, \varepsilon_1^W, \varepsilon_5^W, \varepsilon_2^W, \varepsilon_6^W, \varepsilon_3^W$; Replications: $\varepsilon_1^R, \varepsilon_2^R, \varepsilon_3^R$

sum codings

assume all ε's are independent

$cov(y_1, y_2) = cov(\varepsilon_1^R + \varepsilon_4^W + \varepsilon_2^S, \varepsilon_1^R + \varepsilon_4^W + \varepsilon_3^S) = Var(\varepsilon_1^R + \varepsilon_4^W) = \sigma_R^2 + \sigma_W^2$

$cov(y_1, y_2) = cov(\varepsilon_1^R + \varepsilon_4^W + \varepsilon_4^S, \varepsilon_1^R + \varepsilon_1^W + \varepsilon_8^S) = Var(\varepsilon_1^R) = \sigma_R^2$

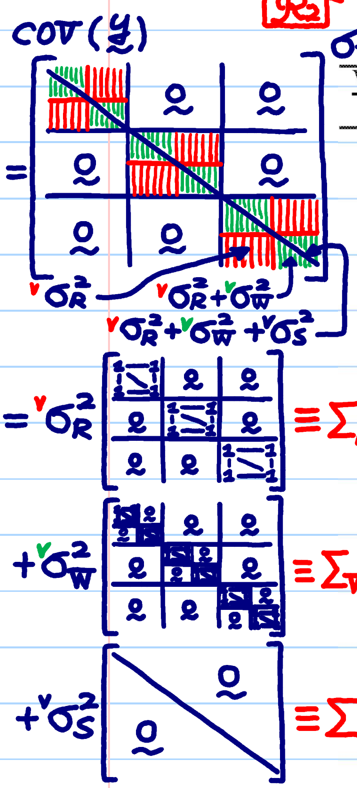
$Var(y) = \sigma_R^2 + \sigma_W^2 + \sigma_S^2$

$cov(y_1, y_2) = cov(\varepsilon_2^R + \varepsilon_2^W + \varepsilon_{16}^S, \varepsilon_3^R + \varepsilon_6^W + \varepsilon_{17}^S) = 0$

$W_A \subset \mathcal{R}_2 \cap \mathcal{R}_1^\perp = S_2$

$W_B \subset \mathcal{R}_3 \cap \mathcal{R}_2^\perp = S_3$

$W_{A\times B} \subset \mathcal{R}_3 \cap \mathcal{R}_2^\perp = S_3$



$cov(y) = \sigma_R^2 (\equiv \Sigma_R) + \sigma_W^2 (\equiv \Sigma_W) + \sigma_S^2 (\equiv \Sigma_S) \equiv \Sigma \; (= \Sigma_R + \Sigma_W + \Sigma_S)$

$Y \sim N(X\beta, \Sigma)$ cf. REM in LNp.3-32; not $\sigma^2 I$