A Brief Note on Strip-Plot Design

- Major distinction between strip-plot design and split-plot design

check LNp.46 →	split-plot design	strip-plot design
(EU) plot structure	nesting (SP nested in WP)	crossing ← random

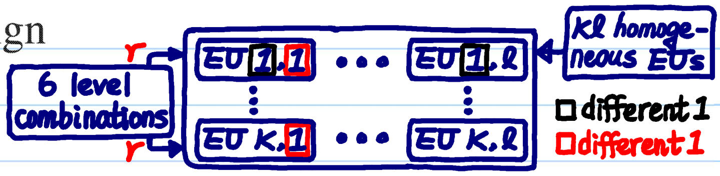
cf. (LNp.29) fixed

- An example. laundry experiment: Recall: row-column structure (LNp.27)

A: 2 washing machines a_1, a_2; B: 3 dryers b_1, b_2, b_3; EUs: cloth samples

⊖ Completely randomized design

★ do washing $k\ell$ times
★ do drying $k\ell$ times



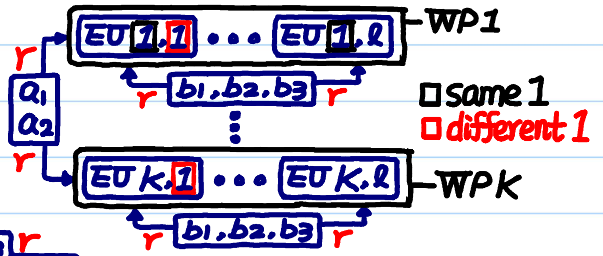
of y's $= k\ell$

same model matrix

⊖ Split-plot deisgn

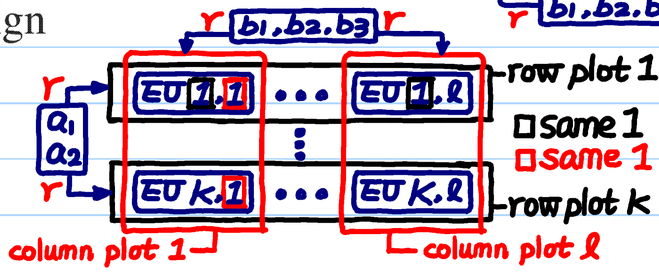
- A: WP factor, B: SP factor

★ do washing k times
★ do drying $k\ell$ times



⊖ Strip-plot design

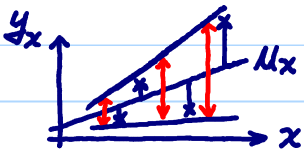
- A: row plot factor
  B: column plot factor
  ★ do washing k times
  ★ do drying ℓ times



❖ **Further Reading**: Cheng (2014), *Theory of Factorial Design*, Chapter 1, Sec. 7.11

Transformation of Response ← (LM, LNp.8-19~21)

- Transform y before fitting a regression model.



Theory: Suppose in the model ← linear structure $X\beta$ & constant variance

$$y_{\mathbf{x}} = \mu_{\mathbf{x}} + \varepsilon_{\mathbf{x}}, \quad \sigma_{\mathbf{x}} = [Var(y_{\mathbf{x}})]^{1/2}, \quad \sigma_{\mathbf{x}} \propto \mu_{\mathbf{x}}^{\alpha}.$$

$= [Var(\varepsilon_{\mathbf{x}})]^{1/2}$ (★) $\overset{\text{cf.}}{\longleftrightarrow}$ constant variance assumption in LM.

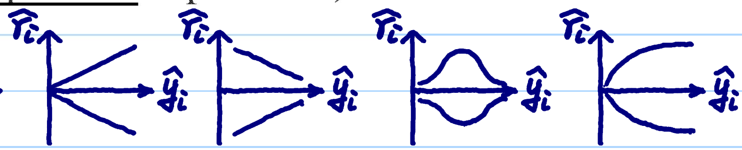
This can be detected by plotting residuals

model: $f(y_{\mathbf{x}}) = X\beta + \varepsilon$

$r_{ij} = y_{ij} - \bar{y}_{i\cdot}$ against $\bar{y}_{i\cdot}$ (for replicated experiment) or

$r_i = y_i - \hat{y}_i$ against $\hat{y}_i$ (for unreplicated experiment).

(What pattern to look for?)



residual plot
$\hat{\varepsilon}_i \leftrightarrow \hat{y}_i$
$\sigma_{\mathbf{x}} \leftrightarrow \mu_{\mathbf{x}}$

Note. The plot shows the relationship btwn σ and μ (not σ^2 and μ)

- Error transmission formula:

transformation $\quad z_{\mathbf{x}} = f(y_{\mathbf{x}}) \approx f(\mu_{\mathbf{x}}) + f'(\mu_{\mathbf{x}})(y_{\mathbf{x}} - \mu_{\mathbf{x}}).$

by (★)

a constant over $\mathbf{x}$ $\rightarrow$ $c = Var(z_{\mathbf{x}}) \approx [f'(\mu_{\mathbf{x}})]^2 Var(y_{\mathbf{x}}) = [f'(\mu_{\mathbf{x}})]^2 \sigma_{\mathbf{x}}^2 \propto [f'(\mu_{\mathbf{x}})]^2 \mu_{\mathbf{x}}^{2\alpha}$

$$\Rightarrow f'(\mu) \propto \mu^{-\alpha}$$

$$f(\mu) = \int f'(\mu)\,d\mu \propto \int \mu^{-\alpha} d\mu = \begin{cases} \mu^{1-\alpha}, & \text{if } \alpha \neq 1 \\ \ln(\mu), & \text{if } \alpha = 1 \end{cases} \xrightarrow{\lambda = 1-\alpha} \begin{cases} \mu^{\lambda}, & \lambda \neq 0 \\ \ln(\mu), & \lambda = 0 \end{cases}$$

Power (Box-Cox) Transformation

model (∇): $z = f_\lambda(y) = X\beta + \epsilon$, i.e., $f_\lambda(y) \sim N(X\beta, \sigma^2 I)$

can use MLE to estimate λ, β, σ^2

$$z_{\mathbf{x}} = f_\lambda(y_{\mathbf{x}}) = \begin{cases} \frac{y_{\mathbf{x}}^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \ln y_{\mathbf{x}}, & \lambda = 0, \end{cases} \quad (8)$$

λ: treated as a parameter

$\lim_{\lambda \to 0} f_\lambda(y) = \ln(y)$

$$f'_\lambda(\mu_{\mathbf{x}}) = \mu_{\mathbf{x}}^{\lambda-1},$$

by (*) in LNp.4-68

$$\sqrt{Var(z_{\mathbf{x}})} \approx |f'_\lambda(\mu_{\mathbf{x}})|\sigma_{\mathbf{x}} = \mu_{\mathbf{x}}^{\lambda-1}\sigma_{\mathbf{x}} \propto \mu_{\mathbf{x}}^{\lambda-1}\mu_{\mathbf{x}}^{\alpha} = \mu_{\mathbf{x}}^{\lambda+\alpha-1}. \propto C$$

$0 \Rightarrow \lambda = 1-\alpha$

- Choosing $\lambda = 1-\alpha$ would make $Var(z_{\mathbf{x}})$ *nearly constant* over $\mathbf{x}$.
- Since α is unknown, λ can be chosen by some statistical criterion (e.g., maximum likelihood). A simpler method is to try a few selected values of λ (see Table 28 in LNp.4-70). In each transform, analyze the $z_{\mathbf{x}}$ data and choose the transformation (i.e., λ value) such that

use model (∇)

(a) it gives a parsimonious model, ← "few" very significant effects

(b) no unusual pattern in the residual plots,

(c) good interpretation of the transformation.

Example of (c): $y_{\mathbf{x}}$ = survival time, $y_{\mathbf{x}}^{-1}$ = rate of dying in the example of Box-Cox(1964).

單位時間/次 次/單位時間

Variance Stabilizing Transformations

Their relationship may be identified from: (i) residual plot of $\hat{\epsilon}_i$ vs. $\hat{y}_i$ → α (=1−λ) (ii) MLE (or confidence interval) of λ

Table 28: Variance Stabilizing Transformations

$\sigma_{\mathbf{x}} \propto \mu_{\mathbf{x}}^{\alpha}$	α	$\lambda\ (=1-\alpha)$	Transformation
$\sigma_{\mathbf{x}} \propto \mu_{\mathbf{x}}^{3}$	3	−2	reciprocal squared
$\sigma_{\mathbf{x}} \propto \mu_{\mathbf{x}}^{2}$	2	−1	reciprocal
$\sigma_{\mathbf{x}} \propto \mu_{\mathbf{x}}^{3/2}$	3/2	−1/2	reciprocal square root
$\sigma_{\mathbf{x}} \propto \mu_{\mathbf{x}}$	1	0	log
$\sigma_{\mathbf{x}} \propto \mu_{\mathbf{x}}^{1/2}$	1/2	1/2	square root
$\sigma_{\mathbf{x}} \propto$ constant	0	1	original scale
$\sigma_{\mathbf{x}} \propto \mu_{\mathbf{x}}^{-1/2}$	−1/2	3/2	3/2 power
$\sigma_{\mathbf{x}} \propto \mu_{\mathbf{x}}^{-1}$	−1	2	square

original scale ← no transformation

transformations with good interpretation

Analysis of Drill Experiment

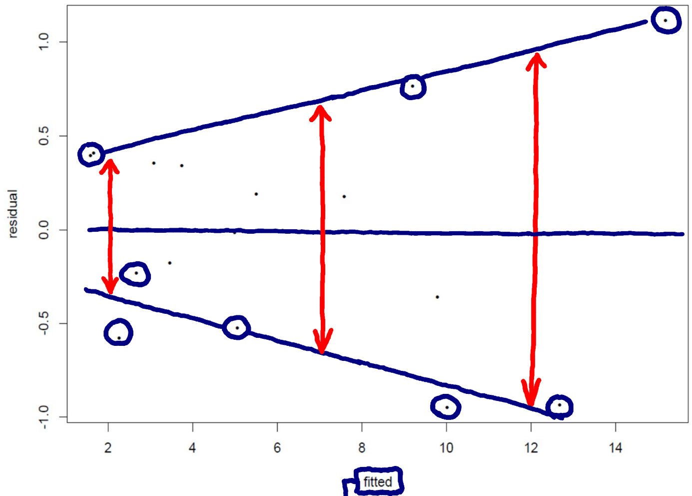
- Data in Table 3.40 of textbook (p.135).
  - four factors A, B, C and D, each at two levels
  - use a 2^4 design ← full factorial design
  - fit a model with 4 main effects and 6 two-factor interactions (2fi's)

  $\leftarrow (4)$ $\qquad \leftarrow \binom{4}{2}$

d.f. left for residuals $= 2^4 - 1 - 4 - 6 = 5$

The $\hat{r}$-vs-$\hat{y}$ plot shows an increasing pattern.

Show the pattern btwn $\sigma_x \leftrightarrow \mu_x$



$\sigma_x \propto \mu_x^{1}$

$\Rightarrow \alpha = 1$

$\Rightarrow \lambda = 1 - \alpha = 0$

$\Rightarrow$ suggest log-transformation of y_x

Figure 2: r_i vs. $\hat{y}_i$, Drill Experiment

$\leftarrow \hat{\mu}_x$

Scaled lambda plot

- For each of the eight transformations λ values in Table 28 (LNp.4-70), a model of main effects and 2fi's is fitted to the transformed $z_x = f_\lambda(y_x)$. The t-statistic values for the 10 effects are displayed in Figure 3 (LNp.4-73).

- **Comments on the plot.**
  - For the log transformation ($\lambda = 0$), the largest t statistics (C, B, and D) stand out.
  - The next best is $\lambda = -1/2$, but not as good as log transformation (Why ? It has a 2fi BC, but the log transform removes the term BC.)
  - On the original scale ($\lambda = 1$), the four main effects do not separate apart.

- **Conclusion :** Use log transformation.

Q. Why not draw $\hat{\beta}_\lambda$? cf. Q: Why use t-statistic?

① Eliminate unit. (Note. z_x has different units for different λ's)
② For different λ's, can use same critical value to declare significance

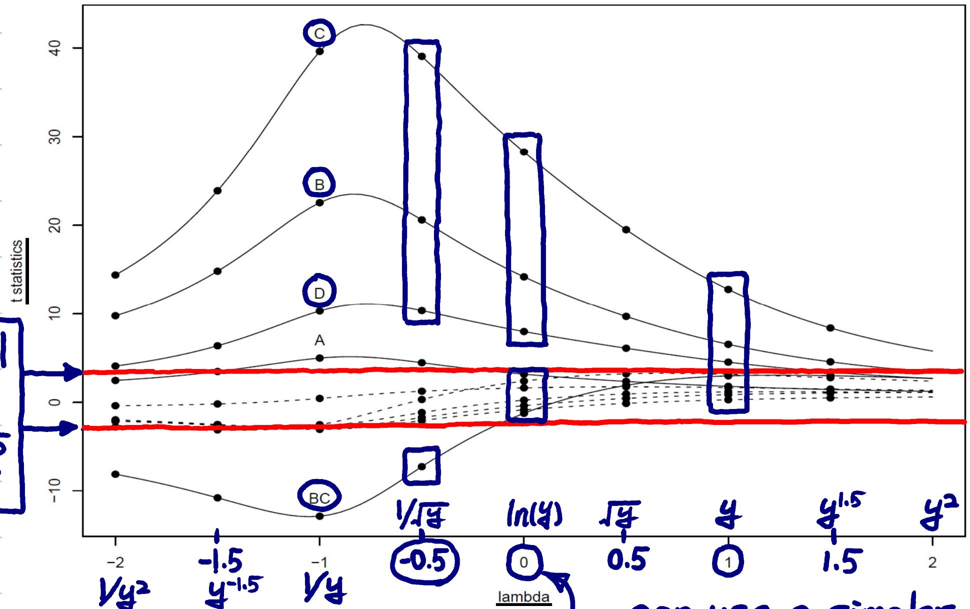
Scaled lambda plot : Drill Experiment

(exercise) use the Box-Cox transformation method to obtain the MLE and confidence interval of λ.

Note 1. Observe how effect significance changes with λ.

Note 2. The plot does not show how good the fitting is (say, $R^2 = ?$)

critical value $t_{5,0.025} = 2.57$



can use a simpler model to explain the response

Figure 3: Scaled λ Plot (lambda denotes the power λ in the transformation (8), LNp.4-69)

❖ **Reading**: textbook, 3.11