# ANOVA Decomposition (Cont'd)

Table 26: Wood Experiment (← LNp 5-48): Summarized data for whole plot analysis

$S_2$ ; $S_2 = \mathcal{R}_2 \ominus \mathcal{R}_1$

$\sum_{j=1}^{J} y_{kij}$: (LNp.53) they are proportional to the projection of $\underset{\sim}{y}$ onto the 6-dimensional space $\mathcal{R}_2$

| | Rep 1 | Rep 2 | Rep 3 | Total |
|---|---|---|---|---|
| $a_1$ | 181.1 | 224.7 | 219.0 | 624.8 |
| $a_2$ | 168.0 | 191.0 | 128.8 | 487.8 |
| Total | 349.1 | 415.7 | 347.8 | 1112.6 |

$P_{\mathcal{R}_2}\underset{\sim}{y} = (\bar{y}^W_1, \ldots, \bar{y}^W_1, \ldots, \bar{y}^W_6, \ldots, \bar{y}^W_6)^T$ $\}\, k_W$ times ; $P_{\mathcal{R}_1}\underset{\sim}{y} = (\bar{y}^R_1, \ldots, \bar{y}^R_1, \bar{y}^R_2, \ldots, \bar{y}^R_2, \bar{y}^R_3, \ldots, \bar{y}^R_3)^T$ $\}\, k_R$ times

$W_1 \oplus W_0$ ; $W_0$

$\|P_{\mathcal{R}_2}\underset{\sim}{y}\|^2 = \sum_{\ell=1}^{n_W} k_W \bar{y}_\ell^{W\,2} = \left[\sum_{\ell=1}^{n_W} (\bar{y}^W_\ell \times k_W)^2\right]/k_W$

from data in LNp.48 ; $T_1 \oplus W_0 = S_1$

only 6 obs. ⇒ total d.f. $= 6 = n_W$ ↳ like a 2-way layout

(1) $W_1$ ⟶
$$SS_A = (624.8^2 + 487.8^2)/\underset{3\times4}{12} - 1112.6^2/\underset{6\times4}{24} = 782.04,$$

check LNp 58

$\oplus$

(2) $T_1$ ⟶
$$SS_{\text{Rep}} = (349.1^2 + 415.7^2 + 347.8^2)/\underset{2\times4}{8} - 1112.6^2/\underset{6\times4}{24} = 376.99,$$

$\oplus$

(2) $T_2$ ⟶
$$SS_{\text{whole}} = SS_{\text{Rep}\times A} = 398.37,$$

$= \|P_{\mathcal{R}_2}\underset{\sim}{y}\|^2 - 1112.6^2/24 - 782.04 - 376.99$

$MS_{whole} = \frac{SS_{whole}}{\dim(T_2)} \xrightarrow{\text{estimate}} k_W\sigma_W^2 + \sigma_S^2$

$S_2 = \mathcal{R}_2 \ominus W_0$
(6) (1)

$$\underset{T_3}{SS_{\text{sub}}} = 927.88 - \underset{T_2}{SS_{\text{whole}}} - \underset{T_1}{SS_{\text{Rep}}} = 152.52.$$

$T_1 \oplus T_2 \oplus T_3 = V_k^\perp$

can use $\bar{y}^W_1, \ldots, \bar{y}^W_{n_W}$ to calculate these SS's

check Table 25 in LNp 57, $\|P_{V_k^\perp}\underset{\sim}{y}\|^2$



# Expected Mean Squares in ANOVA

By (L4)(L5) in LNp.55

$e_1 = \sigma_S^2 + k_W\sigma_W^2 + k_R\sigma_R^2$

$W_0 \oplus T_1 = S_1 \supset T_1$ ; $\dim(S_1) = n_R$

| Source | Effect | df | E(Mean Squares) |
|---|---|---|---|
| Replicate (or block) | $\tau_k$ (random effects) | $n-1$ ($= n_R - \dim(W_0)$) | $\sigma_S^2 + J\sigma_W^2 + IJ\sigma_\tau^2$ |
| $A$ | $\alpha_i$ (fixed effects) | $I-1$ ($= \dim(W_1)$) | $\sigma_S^2 + J\sigma_W^2 + \frac{nJ\sum_{i=1}^I \alpha_i^2}{I-1}$ |
| Whole plot error (Rep×A) | $\varepsilon^W_{ki}$ | $(I-1)(n-1)$ ($= (n_W - n_R) - \dim(W_1)$) | $\sigma_S^2 + J\sigma_W^2$ |
| $B$ | $\beta_j$ (fixed effects) | $J-1$ ($= \dim(W_2)$) | $\sigma_S^2 + \frac{nI\sum_{j=1}^J \beta_j^2}{J-1}$ |
| $A \times B$ | $(\alpha\beta)_{ij}$ | $(I-1)(J-1)$ ($= \dim(W_3)$) | $\sigma_S^2 + \frac{n\sum_{i=1}^I\sum_{j=1}^J (\alpha\beta)_{ij}^2}{(I-1)(J-1)}$ |
| Subplot error ((Rep×B)⊕(Rep×A×B)) | $\varepsilon^S_{kij}$ | $I(J-1)(n-1)$ ($= (N - n_W) - \dim(W_2) - \dim(W_3)$) | $\sigma_S^2$ |

$S_2$: $W_1$ → $A$, $T_2$ → Whole plot error ; $\dim(S_1 \oplus S_2) = n_W$ ; $n = n_R$, $nI = n_W$

$S_3$: $W_2$ → $B$, $W_3$ → $A \times B$, $T_3$ → Subplot error

$E[\|P_{W_1}\underset{\sim}{\varepsilon}\|^2/\dim(W_1)]$ ; $e_2 = \sigma_S^2 + k_W\sigma_W^2$ ; $\|P_{W_1}X\underset{\sim}{\beta}\|^2/\dim(W_1)$

By (L4) in LNp. 55 & $\sum_i \alpha_i = 0$

what if $\alpha_1 = \cdots = \alpha_I = 0$?

$\|P_{W_2}X\underset{\sim}{\beta}\|^2/\dim(W_2)$ ; what if $\beta_1 = \cdots = \beta_J = 0$?

$\|P_{W_3}X\underset{\sim}{\beta}\|^2/\dim(W_3)$ ; what if $\alpha\beta_{ij}$'s $= 0$?

$e_3$

- Proofs are similar to but more tedious than in one-way random effects model (LNp.3-33∼36).

# Hypothesis Testing

By (L3) in LNp.54
(L5) in LNp.55

$$F_A = \frac{MS_A}{MS_{\text{whole}}} \Rightarrow H_0^1 : \alpha_1 = \cdots = \alpha_I,$$

$MS_A \leftarrow W_1 \subset S_2$

$MS_{\text{whole}} \leftarrow T_2 \subset S_2$

$\hat{e}_2 = \widehat{\sigma_S^2 + k_W \sigma_W^2}$

under $H_0^1$, $P_{W_1} X\beta = 0$
$\Rightarrow SS_A \sim e_2 \chi^2_{\dim(W_1)}$

under $H_A^1 \Rightarrow SS_A \sim e_2 \times$ noncentral $\chi^2$

under $H_0^1 \cup H_A^1$,
$SS_{\text{whole}} \sim e_2 \chi^2_{\dim(T_2)}$

$$F_B = \frac{MS_B}{MS_{\text{sub}}} \Rightarrow H_0^2 : \beta_1 = \cdots = \beta_J,$$

$MS_B \leftarrow W_2 \subset S_3$

$MS_{\text{sub}} \leftarrow T_3 \subset S_3$

$\hat{e}_3 = \hat{\sigma}_S^2$

apply similar argument as for $H_0^1$

Q: How to estimate $\sigma_S^2$, $\sigma_W^2$, $\sigma_R^2$?

$$F_{AB} = \frac{MS_{A\times B}}{MS_{\text{sub}}} \Rightarrow H_0^3 : (\alpha\beta)_{ij} = \text{constant},$$

$MS_{A\times B} \leftarrow W_3 \subset S_3$

$MS_{\text{sub}} \leftarrow T_3 \subset S_3$

apply similar argument as for $H_0^1$

$$i = 1, \ldots, I,\ j = 1, \ldots, J.$$

$\hat{e}_1 = \widehat{\sigma_S^2 + k_W \sigma_W^2 + k_R \sigma_R^2}$

$$F_{Rep} = \frac{MS_{\text{Rep}}}{MS_{\text{whole}}} \Rightarrow H_0^4 : \sigma_\tau = 0.$$

$MS_{\text{Rep}} \leftarrow T_1 \subset S_1$

$S_2 \supset T_2 \rightarrow MS_{\text{whole}}$

$\sigma_\tau \rightarrow \sigma_R$

What happens if $\tau_k$'s are treated as fixed effects?
$\mathbb{R}^N \left\{ \begin{array}{l} = W_0 \oplus W_\tau \oplus W_1 \oplus \cdots \oplus W_{ij} \oplus T_2 \oplus \cdots \\ = \underbrace{\quad}_{T_1} \quad \underbrace{\quad}_{\mathcal{R}_2} \quad \oplus S_3 \end{array} \right.$
(check LNp.53)

The ANOVA for one-way REM in LNp.3-33~34 $\overset{\text{cf.}}{\longleftrightarrow}$

$F_{\dim(T_1),\dim(T_2)} \xleftarrow[\text{null}]{\text{under}} \dfrac{MS_{T_1}}{MS_{T_2}}$

$F_{\dim(T_2),\dim(T_3)} \xleftarrow[\text{null}]{\text{under}} \dfrac{MS_{T_2}}{MS_{T_3}}$

$MS_{T_1} \xrightarrow{\text{est.}} e_1$
$\text{IV}$ when $\sigma_R^2 = 0$
$MS_{T_2} \xrightarrow{\text{est.}} e_2$
$\text{IV}$ when $\sigma_W^2 = 0$
$MS_{T_3} \xrightarrow{\text{est.}} e_3$



# Correct ANOVA Analysis

Table25(LNp.57) $\overset{\text{cf.}}{\longleftrightarrow}$ Table 27: Correct ANOVA Table, Wood Experiment

| | Source | Degrees of Freedom | Sum of Squares | | Mean Squares | F | p-value |
|---|---|---|---|---|---|---|---|
| $S_1$ $T_1 \rightarrow$ | Replicate | $3-1=2$ ($\dim(S_1)$) | 376.99 | $k_R\sigma_R^2 + k_W\sigma_W^2 + \sigma_S^2$ | 188.50 | 0.95 | 0.513 |
| $W_1 \rightarrow$ | A | $2-1=1$ | $\boxed{782.04}$ | $8$; Note. $188.5 \le 199.19$ | $\boxed{782.04}$ | 3.93 | 0.186 ✗ |
| $S_2$ $T_2 \rightarrow$ | Whole plot error | $3-1=2$ ($\dim(S_2)$) | 398.37 | $k_W\sigma_W^2 + \sigma_S^2$; $k_W = 4$ | 199.19 | | |
| | | | $927.88 = \oplus$ | Note. usually, $\sigma_W^2 > \sigma_S^2$ | | | |
| $W_2 \rightarrow$ | B | $4-1=3$ | $\boxed{266.00}$ | | $\boxed{88.67}$ | 6.98 | 0.006 ✓ |
| $W_3 \rightarrow$ | $A\times B$ | $(2-1)(4-1)=3$ | $\boxed{62.79}$ | | 20.93 | 1.65 | 0.230 ✗ |
| $S_3$ $T_3 \rightarrow$ | Subplot error | $18-3-3=12$ ($\dim(S_3)$) | 152.52 | $\hat{\sigma}_S^2$ | 12.71 | | |
| $\mathcal{R}_3 \ominus W_0 \rightarrow$ | Total | 23 | 2038.72 | | | | |

Q: Why does A become insignificant?

Q: Why does B become significant?

3 ($n_R$) blocks, size 2

6 ($n_W$) whole plots, size 4

24 ($N$) sub-plots

$\mathcal{R}_2$ (dim=$n_W$) ← check Table 26 (LNp.61)

$\mathcal{R}_3$ (dim=$N$)

It is like treating $\tau_k+\alpha_i+\epsilon^W_{ki}$ (LNp.58) as a fixed block effect

| Source | d.f. | SS | MS | $F$ |
|---|---|---|---|---|
| Replicate | 2 | 376.99 | 188.50 | 0.95 |
| $A$ | 1 | 782.04 | 782.04 | 3.93 |
| Whole plot error | 2 | 398.37 | 199.19 | |
| $B$ | 3 | 266.00 | 88.67 | 6.98 |
| $A\times B$ | 3 | 62.79 | 20.93 | 1.65 |
| Subplot error | 12 | 152.52 | 12.71 | |
| Total | 23 | 2038.72 | | |

| Source | d.f. | SS | MS | $F$ |
|---|---|---|---|---|
| 6 blocks | 5 | 1557.4 | 311.48 | |
| $B$ | 3 | 266.00 | 88.67 | 6.98 |
| $A\times B$ | 3 | 62.79 | 20.93 | 1.65 |
| Subplot error | 12 | 152.52 | 12.71 | |
| Total | 23 | 2038.72 | | |

$\dim(S_2)=n_W-n_R$

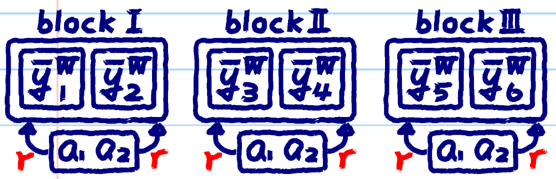
Randomized Block Design (check graphs in LNp.55)

$\dim(S_3)=N-n_W$

Whole plots : EU structure (3 blocks, size 2)

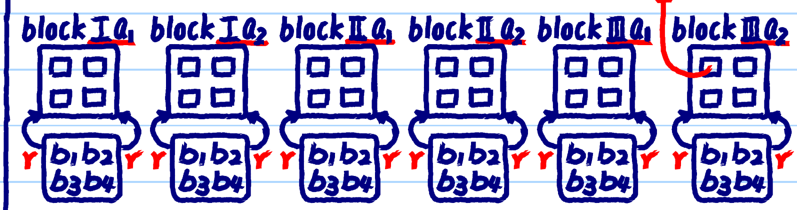
Sub-plots : EU structure (6 blocks, size 4)

$y_{ijk}$




★ total d.f. = 6 = # of WP EUs $n_W$

★ # of blocks = 3 = # of replicates $n_R$




★ total d.f. = 24 = # of SP EUs $N$

★ # of blocks = 6 = # of WP EUs $n_W$



# Analysis Results

only $A$ is significant

- Only $B$ is significant. ←cf.→ ANOVA results from 2-way layout (LNp.57) ↳ $cov(\underset{\sim}{y})=\sigma^2 I$
- Explanation for discrepancy:

$$\underbrace{MS_{\text{whole}} = 199.19}_{k_W\sigma_W^2+\sigma_S^2\ (T_2)\ \leftrightarrow\ A} \gg \underbrace{MS_{\text{Residual}} = 57.99}_{\hat{\sigma}^2 \text{ in 2-way layout}} \gg \underbrace{MS_{\text{sub}} = 12.71}_{\hat{\sigma}_S^2\ (T_3)\ \to\ B,\ A\times B}.$$

↳ $V_K^\perp = T_1\oplus T_2\oplus T_3$

check Table 25 (LNp.57) Table 27 (LNp.64)

- To test $H_0^4:\sigma_\tau=0$, use

null dist. $F_{2,2}$ ⟶ $$\frac{MS_{\text{Rep}}\ (T_1)}{MS_{\text{whole}}\ (T_2)}=\frac{188.5}{199.19}=0.95. < 1 \Rightarrow \hat{\sigma}_R^2=0$$

↳check LNp.3-37

$\Rightarrow$ no significant difference between three replications.

- When does testing $H_0^4$ make sense? ← treat block effects as random effects and are interested in the population of all blocks. The $n_R$ replicates should be a representative sample of the population.

❖ **Reading**: textbook, 3.9

# 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 / fixed |

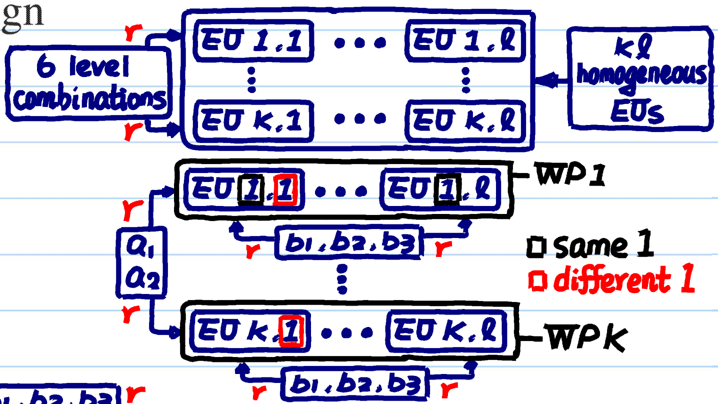
cf.

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

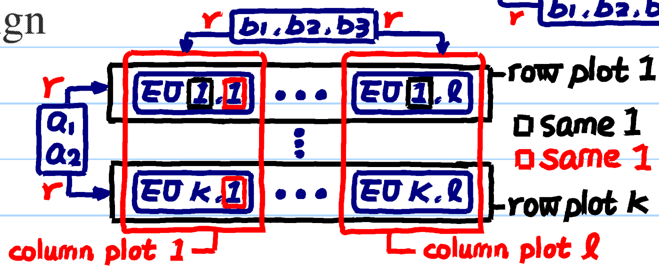
  $A$: 2 washing machines $a_1, a_2$; $B$: 3 dryers $b_1, b_2, b_3$; $EU$s: cloth samples

  – Completely randomized design
    - ★ do washing $k\ell$ times
    - ★ do drying $k\ell$ times

  – Split-plot deisgn
    - • A: WP factor, B: SP factor
    - ★ do washing $k$ times
    - ★ do drying $k\ell$ times



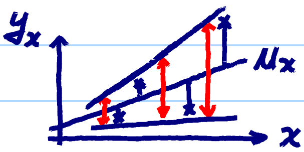

  – Strip-plot design
    - • A: row plot factor
    - B: column plot factor
    - ★ do washing $k$ times
    - ★ do drying $\ell$ times




❖ **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

  $$y_{\mathbf{x}} = \mu_{\mathbf{x}} + \varepsilon_{\mathbf{x}}, \quad \sigma_{\mathbf{x}} = [Var(y_{\mathbf{x}})]^{1/2} = [Var(\varepsilon_{\mathbf{x}})]^{1/2}, \quad \sigma_{\mathbf{x}} \propto \mu_{\mathbf{x}}^{\alpha}. \xleftrightarrow{\text{cf.}} \text{constant variance assumption in LM}$$

  (✱)

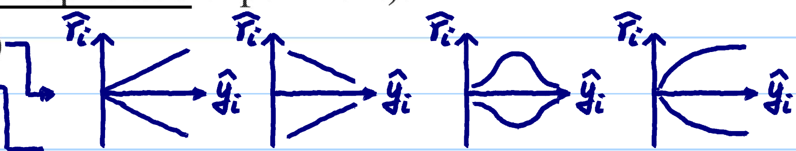
  This can be detected by plotting residuals

  $r_{ij} = y_{ij} - \bar{y}_{i\cdot}$ 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
$\hat{\varepsilon}_i \leftrightarrow \hat{y}_i$
$\sigma_x \leftrightarrow \mu_x$




Note. The plot shows the relationship btwn $\sigma$ and $\mu$ (not $\sigma^2$ and $\mu$)

- Error transmission formula:

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

by (✱)

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

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

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

# Power (Box-Cox) Transformation

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

can use MLE to estimate $\lambda, \beta, \sigma^2$

$$z_x = f_\lambda(y_x) = \begin{cases} \dfrac{y_x^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \ln y_x, & \lambda = 0, \end{cases} \quad (8)$$

← $\lambda$: treated as a parameter

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

$$f'_\lambda(\mu_x) = \mu_x^{\lambda-1},$$

$$\sqrt{Var(z_x)} \approx |f'_\lambda(\mu_x)|\sigma_x = \mu_x^{\lambda-1}\sigma_x \propto \mu_x^{\lambda-1}\mu_x^\alpha = \mu_x^{\lambda+\alpha-1}. \propto C$$

by (*) in LNp.4-68

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

- Choosing $\lambda = 1-\alpha$ would make $Var(z_x)$ *nearly constant* over **x**.
- Since $\alpha$ is unknown, $\lambda$ can be chosen by some statistical criterion (e.g., maximum likelihood). A simpler method is to try a few selected values of $\lambda$ (see Table 28 in LNp.4-70). In each transform, analyze the $z_x$ data and choose the transformation (i.e., $\lambda$ 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): $y_x$ = survival time, $y_x^{-1}$ = 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 $\hat{\varepsilon}_i$ vs. $\hat{y}_i$ → $\alpha$ (= 1−$\lambda$) (ii) MLE (or confidence interval) of $\lambda$

Table 28: Variance Stabilizing Transformations

| $\sigma_x \propto \mu_x^\alpha$ | $\alpha$ | $\lambda\ (=1-\alpha)$ | Transformation |
|---|---|---|---|
| $\sigma_x \propto \mu_x^3$ | 3 | −2 | reciprocal squared |
| $\sigma_x \propto \mu_x^2$ | 2 | −1 | reciprocal |
| $\sigma_x \propto \mu_x^{3/2}$ | 3/2 | −1/2 | reciprocal square root |
| $\sigma_x \propto \mu_x$ | 1 | 0 | log |
| $\sigma_x \propto \mu_x^{1/2}$ | 1/2 | 1/2 | square root |
| $\sigma_x \propto$ constant | 0 | 1 | original scale ← no transformation |
| $\sigma_x \propto \mu_x^{-1/2}$ | −1/2 | 3/2 | 3/2 power |
| $\sigma_x \propto \mu_x^{-1}$ | −1 | 2 | square |

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)

$\binom{4}{1}$ $\binom{4}{2}$

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 $\sigma_x \leftrightarrow \mu_x$

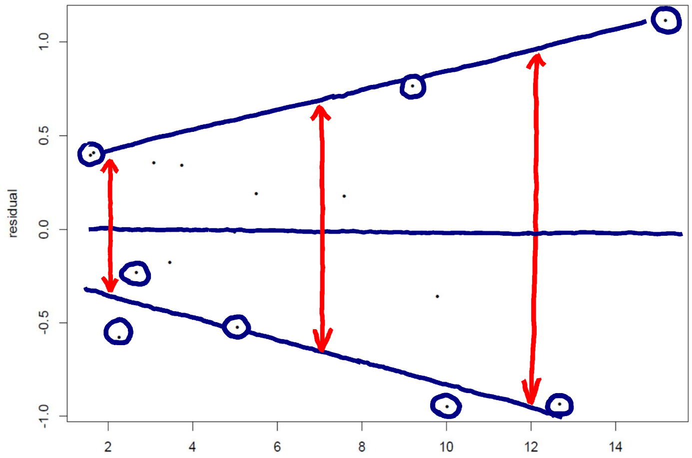



Figure 2: $r_i$ vs. $\hat{y}_i$, Drill Experiment

$\hat{\mu}_x$

$\sigma_x \propto \mu_x^1$
$\Rightarrow \alpha = 1$
$\Rightarrow \lambda = 1 - \alpha = 0$
$\Rightarrow$ suggest log-transformation of $y_x$



# Scaled lambda plot

- For each of the eight transformations $\lambda$ 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 $\lambda$'s)
② For different $\lambda$'s, can use same critical value to declare significance

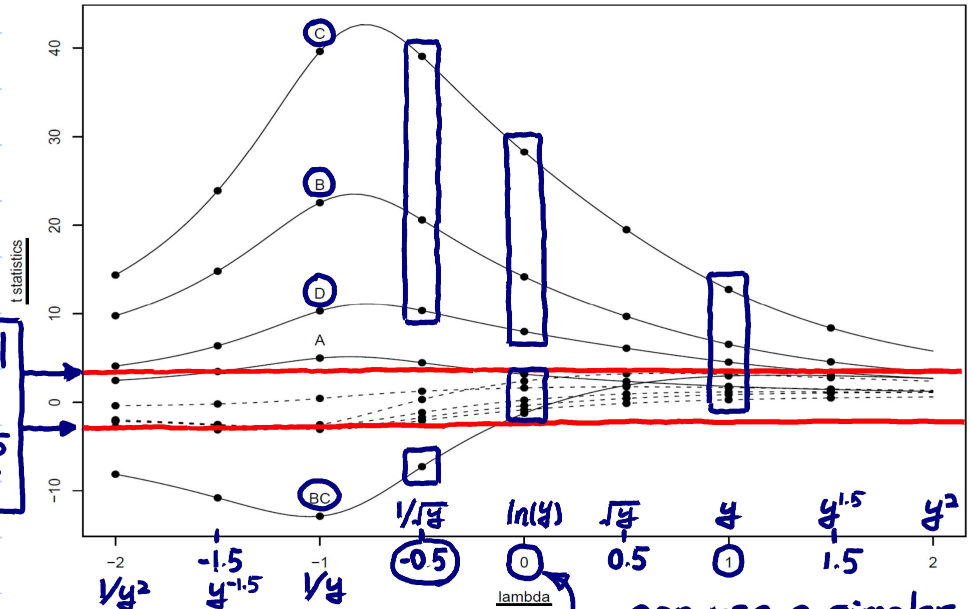
# Scaled lambda plot : Drill Experiment

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

Note 1. Observe how effect significance changes with $\lambda$.

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 $\lambda$ Plot (lambda denotes the power $\lambda$ in the transformation (8), LNp.4-69)

❖ **Reading**: textbook, 3.11