Unit 3: Experiments with More Than One Factor

Sources : Chapter 3.

- Paired comparison design (Section 3.1).
- Randomized block design (Section 3.2).
- Two-way and multi-way layout with fixed effects (Sections 3.3 and 3.5).
- Latin and Graeco-Latin square design (Sections 3.6 and 3.7).
- Balanced incomplete block design (Section 3.8).
- Split-plot design (Section 3.9).
- Analysis of covariance (ANCOVA) (Section 3.10).
- Transformation of response (Section 3.11).

Sewage Experiment

- **Objective :** To compare two methods MSI and SIB for determining chlorine content in sewage effluents; y = residual chlorine reading.

Table 1: Residual Chlorine Readings, Sewage Experiment

Sample	Method MSI	Method SIB	d_i
1	0.39	0.36	-0.03
2	0.84	1.35	0.51
3	1.76	2.56	0.80
4	3.35	3.92	0.57
5	4.69	5.35	0.66
6	7.70	8.33	0.63
7	10.52	10.70	0.18
8	10.92	10.91	-0.01

- **Experimental Design :** Eight samples were collected at different doses and contact times. Two methods were applied to each of the eight samples. It is a *paired comparison* design because the pair of treatments are applied to the same samples (or units).

Paired Comparison Design vs. Unpaired Design

- *Paired Comparison Design* : Two treatments are randomly assigned to each $block$ of two units. Can eliminate block-to-block variation and is effective if such variation is large.
  Examples : pairs of twins, eyes, kidneys, left and right feet.
  (Subject-to-subject variation much larger than within-subject variation).

- *Unpaired Design* : Each treatment is applied to a $separate$ set of units, or called the *two-sample* problem. Useful if pairing is unnecessary; also it has more degrees of freedom for error estimation (see page 5).

Paired t tests

- **Paired t test :** Let y_{i1}, y_{i2} be the responses of treatments 1 and 2 for unit $i, i = 1, \ldots N$. Let $d_i = y_{i2} - y_{i1}$, $\bar{d}$ and s_d^2 the sample mean and variance of d_i.

$$t_{paired} = \sqrt{N}\bar{d}/s_d$$

The two treatments are declared significantly different at level α if

$$|t_{paired}| > t_{N-1,\alpha/2}. \tag{1}$$

Unpaired t tests

- **Unpaired t test :** The unpaired t test is appropriate if we randomly choose N of the $2N$ units to receive one treatment and assign the remaining N units to the second treatment. Let $\bar{y_i}$ and s_i^2 be the sample mean and sample variance for the ith treatment, i = 1 and 2. Define

$$t_{unpaired} = (\bar{y_2} - \bar{y_1})/\sqrt{(s_2^2/N) + (s_1^2/N)}.$$

The two treatments are declared significantly different at level α if

$$|t_{unpaired}| > t_{2N-2,\alpha/2}. \tag{2}$$

Note that the degrees of freedom in (1) and (2) are $N - 1$ and $2N - 2$ respectively. The unpaired t test has more df's but make sure that the unit-to-unit variation is under control (if this method is to be used).

Analysis Results : t tests

$$
\begin{aligned}
t_{paired} &= \frac{0.4138}{0.321/\sqrt{8}} = \frac{0.4138}{0.1135} = 3.645, \\
t_{unpaired} &= \frac{5.435 - 5.0212}{\sqrt{(17.811 + 17.012)/8}} = \frac{0.4138}{2.0863} = 0.198.
\end{aligned}
$$

The p values are

$$
\begin{aligned}
Prob(|t_7| > 3.645) &= 0.008, \\
Prob(|t_{14}| > 0.198) &= 0.848.
\end{aligned}
$$

- Unpaired t test fails to declare significant difference because its denominator 2.0863 is too large. Why ? Because the denominator contains the sample-to-sample variation component.

Analysis Results : ANOVA and F tests

- Wrong to analyze by ignoring pairing. A better explanation is given by ANOVA.
- F statistic in ANOVA for paired design equals t^2_{paired}; similarly, F statistic in ANOVA for unpaired design equals $t^2_{unpaired}$. Data can be analyzed in two equivalent ways.
- In the correct analysis (Table 2), the total variation is decomposed into three components; the largest one is the sample-to-sample variation (its $MS = 34.77$). In the unpaired analysis (Table 3), this component is mistakenly included in the residual SS, thus making the F test powerless.

ANOVA Tables

Table 2: ANOVA Table, Sewage Experiment

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F
sample	7	243.4042	34.77203	674.82
method	1	0.6848	0.68476	13.29
residual	7	0.3607	0.05153	

Table 3: ANOVA Table Ignoring Pairing, Sewage Experiment

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F
method	1	0.6848	0.68476	0.04
residual	14	243.7649	17.41178	

Randomized Block Design : Girder Experiment

- Recall the principles of blocking and randomization in Unit 1. In a randomized block design (RBD), k treatments are randomly assigned to each block (of k units); there are in total b blocks. Total sample size $N = bk$.
- Paired comparison design is a special case with $k = 2$. (Why ?)
- **Objective :** To compare four methods for predicting the shear strength for steel plate girders ($k = 4, b = 9$).

Table 4: Strength Data, Girder Experiment

(Block)	Method			
Girder	Aarau	Karlsruhe	Lehigh	Cardiff
S1/1	0.772	1.186	1.061	1.025
S2/1	0.744	1.151	0.992	0.905
S3/1	0.767	1.322	1.063	0.930
S4/1	0.745	1.339	1.062	0.899
S5/1	0.725	1.200	1.065	0.871
S1/2	0.844	1.402	1.178	1.004
S2/2	0.831	1.365	1.037	0.853
S3/2	0.867	1.537	1.086	0.858
S4/2	0.859	1.559	1.052	0.805

Model and Estimation

Model for RBD :

$$y_{ij} = \eta + \alpha_i + \tau_j + \epsilon_{ij},\ i = 1, \ldots, b;\ j = 1, \ldots, k,$$

where y_{ij} = observation of the jth treatment in the ith block, α_i = ith block effect, τ_j = jth treatment effect, ϵ_{ij} = errors, independent $N(0, \sigma^2)$.

$$\begin{aligned} y_{ij} &= \hat{\eta} + \hat{\alpha}_i + \hat{\tau}_j + r_{ij} \\ &= \bar{y}_{\cdot\cdot} + (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) + (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}) + (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot}), \end{aligned}$$

where

$$\hat{\eta} = \bar{y}_{\cdot\cdot},\ \hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot},\ \hat{\tau}_j = \bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot},\ r_{ij} = y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot},$$
$$\bar{y}_{i\cdot} = k^{-1} \textstyle\sum_{j=1}^{k} y_{ij},\ \bar{y}_{\cdot j} = b^{-1} \sum_{i=1}^{b} y_{ij},\ \bar{y}_{\cdot\cdot} = (bk)^{-1} \sum_{i=1}^{b} \sum_{j=1}^{k} y_{ij}.$$

ANOVA

Subtracting $\bar{y}_{..}$, squaring both sides and summing over i and j yields

$$\begin{aligned}\sum_{i=1}^{b}\sum_{j=1}^{k}(y_{ij}-\bar{y}_{..})^2 &= \sum_{i=1}^{b}k(\bar{y}_{i.}-\bar{y}_{..})^2+\sum_{j=1}^{k}b(\bar{y}_{.j}-\bar{y}_{..})^2 \\ &\quad +\sum_{i=1}^{b}\sum_{j=1}^{k}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2 \\ &= SS_b+SS_t+SS_r.\end{aligned}$$

Table 5: ANOVA Table for Randomized Block Design

Source	Degrees of Freedom	Sum of Squares
block	$b-1$	$\sum_{i=1}^{b} k(\bar{y}_{i.}-\bar{y}_{..})^2$
treatment	$k-1$	$\sum_{j=1}^{k} b(\bar{y}_{.j}-\bar{y}_{..})^2$
residual	$(b-1)(k-1)$	$\sum_{i=1}^{b}\sum_{j=1}^{k}(y_{ij}-\bar{y}_{i.}-\bar{y}_{.j}+\bar{y}_{..})^2$
total	$bk-1$	$\sum_{i=1}^{b}\sum_{j=1}^{k}(y_{ij}-\bar{y}_{..})^2$

Testing and Multiple Comparisons

- $H_0 : \tau_1 = \cdots = \tau_k$, can be tested by using the F statistic

$$F = \frac{SS_t/(k-1)}{SS_r/(b-1)(k-1)}, \tag{3}$$

  The F test rejects H_0 at level α if $F > F_{k-1,(b-1)(k-1),\alpha}$.

- *If H_0 is rejected, multiple comparisons of the τ_j should be performed.*
  t statistics for making multiple comparisons :

$$t_{ij} = \frac{\bar{y}_{\cdot j} - \bar{y}_{\cdot i}}{\hat{\sigma}\sqrt{1/b + 1/b}}, \tag{4}$$

  where $\hat{\sigma}^2$ is the mean square error in the ANOVA table.

- At level α, the Tukey multiple comparison method identifies "treatments i and j as different" if

$$|t_{ij}| > \frac{1}{\sqrt{2}} q_{k,(b-1)(k-1),\alpha}.$$

Simultaneous Confidence Intervals

By solving

$$\frac{|(\bar{y}_{\cdot j} - \bar{y}_{\cdot i}) - (\tau_j - \tau_i)|}{\hat{\sigma}\sqrt{2/b}} \leq \frac{1}{\sqrt{2}} q_{k,(b-1)(k-1),\alpha}$$

for $\tau_j - \tau_i$, the simultaneous confidence intervals for $\tau_j - \tau_i$ are

$$\bar{y}_{\cdot j} - \bar{y}_{\cdot i} \pm q_{k,(b-1)(k-1),\alpha} \frac{\hat{\sigma}}{\sqrt{b}}$$

for all i and j pairs.

Analysis of Girder Experiment : F test

Table 6: ANOVA Table, Girder Experiment

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F
girder	8	0.089	0.011	1.62
method	3	1.514	0.505	73.03
residual	24	0.166	0.007	

- The F statistic in (3) has the value

$$\frac{1.514/3}{0.166/24} = 73.03.$$

Therefore, the p value for testing the difference between methods is $Prob(F_{3,24} > 73.03)$=0.00. The small p value suggests that the methods are different.

Analysis of Girder Experiment : Multiple Comparisons

Table 7: Multiple Comparison t Statistics, Girder Experiment

A vs. K	A vs. L	A vs. C	K vs. L	K vs. C	L vs. C
13.91	6.92	2.82	-6.99	-11.09	-4.10

- The means for the four methods, A for Aarau, K for Karlsruhe, L for Lehigh and C for Cardiff are 0.7949, 1.3401, 1.0662 and 0.9056.

- The multiple comparison t statistics based on (4) are displayed in Table 7. For example, the A vs. K t statistic is $t_{12} = \frac{1.3401-0.7949}{\sqrt{0.007}\sqrt{2/9}} = 13.91$. With $\alpha = 0.05$, $t_{24,0.05/(6\times 2)} = 2.875$ for the Bonferroni method. Since $k = 4$ and $\binom{k}{2} = 6$, $\frac{1}{\sqrt{2}} q_{4,24,0.05} = \frac{3.90}{1.414} = 2.758$ for the Tukey method. Again, Tukey method is more powerful. (Why ?)

Two-way layout

- This is similar to RBD. The only difference is that here we have two treatment factors instead of one treatment factor and one block factor. Interested in assessing interaction effect between the two treatments. In blocking, block$\times$treatment interaction is assumed negligible.
- **Bolt experiment :** The goals was to test if there is any difference between two test media (bolt, mandrel) and among three plating methods (C&W, HT, P&O). Response y is the torque of the locknut.

Table 8: Torque Data, Bolt Experiment

	C&W	HT	P&O
Bolt	20, 16, 17, 18, 15, 16, 19, 14, 15, 24	26, 40, 28, 38, 38, 30, 26, 38, 45, 38	25, 40, 30, 17, 16, 45, 49, 33, 30, 20
Mandrel	24, 18, 17, 17, 15, 23, 14, 18, 12, 11	32, 22, 30, 35, 32, 28, 27, 28, 30, 30	10, 13, 17, 16, 15, 14, 11, 14, 15, 16

Model and Estimation

- Model :

$$y_{ijl} = \eta + \alpha_i + \beta_j + \omega_{ij} + \epsilon_{ijl},\ i = 1, \ldots, I;\ j = 1, \ldots, J;\ l = 1, \ldots, n \quad (5)$$

where y_{ijl} = observation for the lth replicate of the ith level of factor A and the jth level of factor B, α_i = ith main effect for A, β_j = jth main effect for B, ω_{ij} = (i, j)th interaction effect between A and B and ϵ_{ijl} = errors, independent $N(0, \sigma^2)$.

- Estimation :

$$\begin{aligned} y_{ijl} &= \hat{\eta} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\omega}_{ij} + r_{ijl} \\ &= \bar{y}_{\cdots} + (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}) + (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}) \\ &\quad + (y_{ijl} - \bar{y}_{ij\cdot}), \end{aligned}$$

where

$$\hat{\eta} = \bar{y}_{\cdots},\ \hat{\alpha}_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots},\ \hat{\beta}_j = \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots},$$

$$\hat{\omega}_{ij} = \bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots},\ r_{ijl} = y_{ijl} - \bar{y}_{ij\cdot},$$

ANOVA

Table 9: ANOVA Table for Two-Way Layout

Source	Degrees of Freedom	Sum of Squares
A	$I-1$	$nJ\sum_{i=1}^{I}(\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot\cdot\cdot})^2$
B	$J-1$	$nI\sum_{j=1}^{J}(\bar{y}_{\cdot j\cdot}-\bar{y}_{\cdot\cdot\cdot})^2$
$A\times B$	$(I-1)(J-1)$	$n\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{ij\cdot}-\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot j\cdot}+\bar{y}_{\cdot\cdot\cdot})^2$
residual	$IJ(n-1)$	$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{l=1}^{n}(y_{ijl}-\bar{y}_{ij\cdot})^2$
total	$IJn-1$	$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{l=1}^{n}(y_{ijl}-\bar{y}_{\cdot\cdot\cdot})^2$

Analysis of Bolt Experiment

Table 10: ANOVA Table, Bolt Experiment

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F
test	1	821.400	821.400	22.46
plating	2	2290.633	1145.317	31.31
test $\times$ plating	2	665.100	332.550	9.09
residual	54	1975.200	36.578	

- **Conclusions :** Both factors and their interactions are significant. Multiple comparisons of C&W, HT and P&O can be performed by using Tukey method with $k = 3$ and 54 error df's.
- Another method is considered in the following pages.

Two Qualitative Factors: a Regression Modeling Approach

- Motivation: need to find a model that allows the comparison and estimation between levels of the qualitative factors. The parameters α_i and β_j in model (5) are not estimable without putting constraints. For qualitative factors, use the **baseline constraint** $\alpha_1 = \beta_1 = 0$ and $w_{1j} = w_{i1} = 0, i = 1, 2, j = 1, 2, 3$ for the bolt experiment.

- It can be shown that

$$
\begin{aligned}
E(y_{11}) = \eta,\ E(y_{12}) = \eta + \beta_2,\ E(y_{13}) = \eta + \beta_3, \\
E(y_{21}) = \eta + \alpha_2,\ E(y_{22}) = \eta + \alpha_2 + \beta_2 + \omega_{22}, \\
E(y_{23}) = \eta + \alpha_2 + \beta_3 + \omega_{23}.
\end{aligned}
$$

Regression Model (continued)

In the regression model $y = X\beta + \epsilon$,

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Regression Model (continued)

Interpretation of parameters

$$
\begin{aligned}
\eta &= E(y_{11}), \\
\alpha_2 &= E(y_{21}) - E(y_{11}), \\
\beta_2 &= E(y_{12}) - E(y_{11}), \\
\beta_3 &= E(y_{13}) - E(y_{11}), \\
\omega_{22} &= (E(y_{22}) - E(y_{21})) - (E(y_{12}) - E(y_{11})), \\
\omega_{23} &= (E(y_{23}) - E(y_{21})) - (E(y_{13}) - E(y_{11})).
\end{aligned}
$$

Regression Analysis Results

Table 11: Tests, Bolt Experiment

Effect	Estimate	Standard Error	t	p value
η	17.4000	1.9125	9.10	0.000
α_2	-0.5000	2.7047	-0.18	0.854
β_2	17.3000	2.7047	6.40	0.000
β_3	13.1000	2.7047	4.84	0.000
ω_{22}	-4.8000	3.8251	-1.25	0.215
ω_{23}	-15.9000	3.8251	-4.16	0.000

- Significant effects: $\hat{\beta}_2$ (C & W and H & T are different), $\hat{\beta}_3$ (C & W and P & O are different), $\hat{\omega}_{23}$ (difference between C&W and P&O varies from bolt to mandrel); $\hat{\alpha}_2$ not significant suggests no difference between bolt and mandrel.

Adjusted p Values

The p values in Table 11 are for each individual effect. Since five effects (excluding η) are considered $simultaneously$, we should, strictly speaking, adjust the p values when making a $joint$ statement about the five effects. In the spirit of the Bonferroni method (again justified by the Bonferroni's inequality in (2.15) of the book), we multiply the individual p value by the number of tests to obtain **adjusted p value**. For $\hat{\omega}_{23}$, the adjusted p value is $5 \times 0.0001 = 0.0005$, still very significant. The adjusted p values, for $\hat{\beta}_2$ and $\hat{\beta}_3$ are smaller.

Box-Whisker Plot : Bolt Experiment

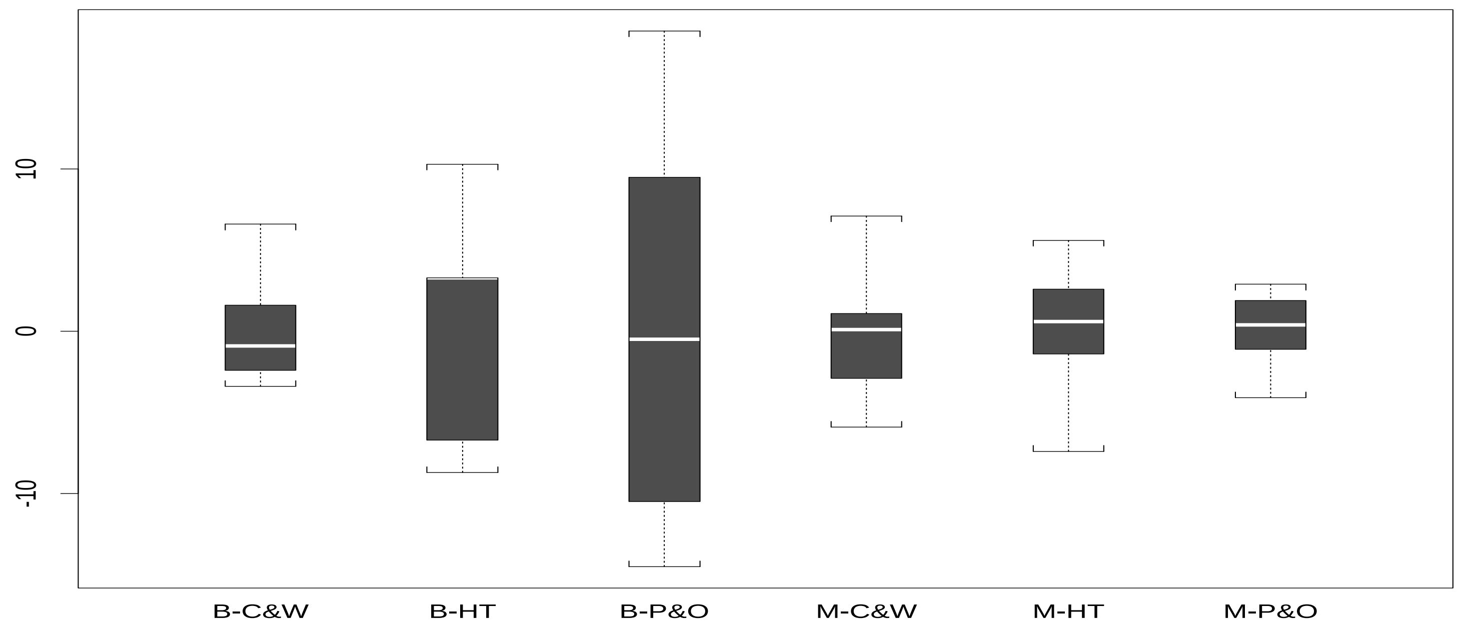

Figure 1: Box-Whisker Plots of Residuals, Bolt Experiment

The plot suggests that the constant variance assumption in (5) does not hold and that the variance of y for bolt is larger than that for mandrel. These are aspects that cannot be discovered by regression analysis alone.

Multiple-Way Layout

Table 12: ANOVA Table for Three-Way Layout

Source	df	Sum of Squares
A	$I-1$	$\sum_{i=1}^{I} nJK(\hat{\alpha}_i)^2$
B	$J-1$	$\sum_{j=1}^{J} nIK(\hat{\beta}_j)^2$
C	$K-1$	$\sum_{k=1}^{K} nIJ(\hat{\delta}_k)^2$
$A \times B$	$(I-1)(J-1)$	$\sum_{i=1}^{I} \sum_{j=1}^{J} nK(\widehat{(\alpha\beta)}_{ij})^2$
$A \times C$	$(I-1)(K-1)$	$\sum_{i=1}^{I} \sum_{k=1}^{K} nJ(\widehat{(\alpha\delta)}_{ik})^2$
$B \times C$	$(J-1)(K-1)$	$\sum_{j=1}^{J} \sum_{k=1}^{K} nI(\widehat{(\beta\delta)}_{jk})^2$
$A \times B \times C$	$(I-1)(J-1)(K-1)$	$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} n(\hat{\gamma}_{ijk})^2$
residual	$IJK(n-1)$	$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{n} (y_{ijkl} - \bar{y}_{ijk\cdot})^2$
total	$IJKn-1$	$\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{l=1}^{n} (y_{ijkl} - \bar{y}_{\cdots})^2$

- $\hat{\alpha}_i$, $\hat{\beta}_j$, $\widehat{\alpha\beta}_{ij}$, etc given in (3.35) of the book.
- Estimation, F test, residual analysis are similar to those for two-way layout.

Latin Square Design : Wear Experiment

Wear Experiment : Testing the abrasion resistance of rubber-covered fabric, $y =$ loss in weight over a period of time.

One treatment factor : Material type A, B, C, D.

Two blocking factors : (1) four positions on the tester,
(2) four applications (four different times for setting up the tester)

Latin square design of order k : Each of the k Latin letters (i.e., treatments) appears once in each row and once in each column.

It is an extension of RBD to accommodate two blocking factors. Randomization applied to assignments to rows, columns, treatments. (Collection of Latin Square Tables given in Appendix 3A of WH).

Wear Experiment : Design and Data

Table 13: Latin Square Design (columns correspond to positions, rows correspond to applications and Latin letters correspond to materials), Wear Experiment

	Position			
Application	1	2	3	4
1	C	D	B	A
2	A	B	D	C
3	D	C	A	B
4	B	A	C	D

Table 14: Weight Loss Data, Wear Experiment

	Position			
Application	1	2	3	4
1	235	236	218	268
2	251	241	227	229
3	234	273	274	226
4	195	270	230	225

Model for Latin Square Design

Model:

$$y_{ijl} = \eta + \alpha_i + \beta_j + \tau_l + \epsilon_{ijl},$$

$l =$ Latin letter in the (i, j) cell of the Latin Square,
$\alpha_i = i$th row effect,
$\beta_j = j$th column effect,
$\tau_l = l$th treatment (i.e., Latin letter) effect,
ϵ_{ijl} are independent $N(0, \sigma^2)$.

There are only k^2 values in the triplet (i, j, l) dictated by the particular LS; this set is denoted by S.

$$\begin{aligned} y_{ijl} &= \hat{\eta} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\tau}_l + r_{ijl} \\ &= \bar{y}_{\cdots} + (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}) + (\bar{y}_{\cdot\cdot l} - \bar{y}_{\cdots}) \\ &\quad + (y_{ijl} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot l} + 2\bar{y}_{\cdots}), \end{aligned}$$

ANOVA decomposition: similar formula (see (3.40) of WH)

ANOVA for Latin Square Design

Table 15: ANOVA Table for Latin Square Design

Source	Degrees of Freedom	Sum of Squares
row	$k-1$	$k\sum_{i=1}^{k}\left(\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot\cdot\cdot}\right)^2$
column	$k-1$	$k\sum_{j=1}^{k}\left(\bar{y}_{\cdot j\cdot}-\bar{y}_{\cdot\cdot\cdot}\right)^2$
treatment	$k-1$	$k\sum_{l=1}^{k}\left(\bar{y}_{\cdot\cdot l}-\bar{y}_{\cdot\cdot\cdot}\right)^2$
residual	$(k-1)(k-2)$	$\sum_{(i,j,l)\in S}\left(y_{ijl}-\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot j\cdot}-\bar{y}_{\cdot\cdot l}+2\bar{y}_{\cdot\cdot\cdot}\right)^2$
total	k^2-1	$\sum_{(i,j,l)\in S}\left(y_{ijl}-\bar{y}_{\cdot\cdot\cdot}\right)^2$

Table 16: ANOVA Table, Wear Experiment

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F
application	3	986.5	328.833	5.37
position	3	1468.5	489.500	7.99
material	3	4621.5	1540.500	25.15
residual	6	367.5	61.250	

F Test and Multiple Comparisons

- $H_0 : \tau_1 = \cdots = \tau_k$, can be tested by using the F statistic

$$F = \frac{SS_t/(k-1)}{SS_r/(k-1)(k-2)},$$

  The F test rejects H_0 at level α if $F > F_{k-1,(k-1)(k-2),\alpha}$.

- If H_0 is rejected, multiple comparisons of the τ_j should be performed. t statistics for making multiple comparisons :

$$t_{ij} = \frac{\bar{y}_{\cdot\cdot j} - \bar{y}_{\cdot\cdot i}}{\hat{\sigma}\sqrt{1/k + 1/k}},$$

  where $\hat{\sigma}^2$ is the mean square error in the ANOVA table.

- At level α, the Tukey multiple comparison method identifies "treatments i and j as different" if

$$|t_{ij}| > \frac{1}{\sqrt{2}} q_{k,(k-1)(k-2),\alpha}.$$

Analysis Results

- The p values for application and position are 0.039(=$Prob(F_{3,6} > 5.37)$) and 0.016(=$Prob(F_{3,6} > 7.99)$), respectively. This indicates that blocking is important.

- The treatment factor (material) has the most significance as indicated by a p value of 0.0008 (= $Prob(F_{3,6} > 25.15)$).

- With k=4 and $(k-1)(k-2)$=6, the critical value for the Tukey multiple comparison method is

$$\frac{1}{\sqrt{2}} q_{4,6,0.05} = \frac{4.90}{\sqrt{2}} = 3.46$$

  at the 0.05 level.

- By comparing the multiple comparisons t statistics given in Table 17 with 3.46, material A and B, A and C, A and D and B and C are identified as different at 0.05 level.

Multiple Comparisons Tables

Table 17: Multiple Comparison t Statistics, Wear Experiment

A vs. B	A vs. C	A vs. D	B vs. C	B vs. D	C vs. D
-8.27	-4.34	-6.37	3.93	1.90	-2.03

Table 18: ANOVA Table (Ignoring Blocking), Wear Experiment

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F
material	3	4621.5	1540.500	6.55
residual	12	2822.5	235.21	

Effectiveness of blocking:

With blocking, $Pr(F_{3,6} > 25.15)$ = 0.0008,

Without blocking, $Pr(F_{3,12} > 6.55)$ = 0.007.

Therefore blocking can make a difference in decision making if treatment effects are smaller.

Graeco-Latin Square Design

- Two Latin squares are $orthogonal$ if each pair of letters appears once in the two squares, when superimposed. The super-imposed square is called a *Graeco-Latin square*.

$$\boxed{\begin{array}{ccc} A\alpha & B\beta & C\gamma \\ B\gamma & C\alpha & A\beta \\ C\beta & A\gamma & B\alpha \end{array}}$$

- Useful for studying four factors (1 treatment, 3 blocking factors; or 2 treatment, 2 blocking factors etc.) allowing one more factor to be studied than in LS.

Model and ANOVA in Graeco-Latin Square Design

Model:

$$y_{ijlm} = \eta + \alpha_i + \beta_j + \tau_l + \zeta_m + \epsilon_{ijlm},$$

(Similar interpretation as in LS, and ζ_m is the mth effect of Greek letters). F test and Tukey's multiple comparisons similar formulae.

Table 19: ANOVA Table for Graeco-Latin Square Design

Source	Degrees of Freedom	Sum of Squares
row	$k-1$	$k\sum_{i=1}^{k}(\bar{y}_{i\cdots} - \bar{y}_{\cdots\cdot})^2$
column	$k-1$	$k\sum_{j=1}^{k}(\bar{y}_{\cdot j\cdot\cdot} - \bar{y}_{\cdots})^2$
Latin letter	$k-1$	$k\sum_{l=1}^{k}(\bar{y}_{\cdot\cdot l\cdot} - \bar{y}_{\cdots\cdot})^2$
Greek letter	$k-1$	$k\sum_{m=1}^{k}(\bar{y}_{\cdots m} - \bar{y}_{\cdots\cdot})^2$
residual	$(k-3)(k-1)$	by subtraction
total	k^2-1	$\sum_{(i,j,l,m)\in S}(y_{ijlm} - \bar{y}_{\cdots\cdot})^2$

Incomplete Blocking

- Blocking is $incomplete$ if the number of treatments t is greater than the block size k. This happens if the nature of blocking makes it difficult to form blocks of large size.
- Example : wine or ice cream tasting, block size limited by taste buds.
- On the other hand, RBD has $complete$ blocking.
- Example: Tire wear experiment. Compare four components A,B,C,D in terms of wear. Because of manufacturing limitations, each tire can be divided into only three sections with each section being made of one compound.

Table 20: Wear Data, Tire Experiment

	Compound			
Tire	A	B	C	D
1	238	238	279	
2	196	213		308
3	254		334	367
4		312	421	412

Balanced Incomplete Block Design (BIBD)

- A BIBD has t treatments, and b blocks of size k, $t > k$, each treatment replicated r times, such that each pair of treatments appear in the $same$ number (denoted by λ) of blocks.
  In the wear experiment, $t = 4$, $k = 3$, $b = 4$, $r = 3$ and $\lambda = 2$.
  Two basic relations:

$$\begin{aligned} bk &= rt, \\ r(k-1) &= \lambda(t-1). \end{aligned}$$

  (Proof of (i) and (ii).)

- For given k, t and b, a BIBD may or may not exist. When it does not, either adjust the values of k, t, b to get a BIBD, or if not possible, find a partially balanced incomplete block design (PBIBD) (which is not covered in the book). Tables of BIBD or PBIBD in books like Cochran and Cox (1957).

Example of Split-plot Design: Wood Experiment

- Experiment objective : to study the water resistant property of wood.
- Two factors: A—wood pretreatment, 2-level; B—type of stain, 4-level.
- Completely randomized design: randomly apply the 8 combinations of A and B to 8 wood panels, such as in Table 21.
- Problem: inconvenient to apply the pretreatment to a small wood panel.

Table 21: Completely Randomized Version of the Wood Experiment

Run	1	2	3	4	5	6	7	8
Pretreatment (A)	$A1$	$A2$	$A2$	$A1$	$A2$	$A1$	$A1$	$A2$
Stain (B)	$B2$	$B4$	$B1$	$B1$	$B3$	$B4$	$B3$	$B2$

Split-plot Design

- Alternative Design: split-plot design in Table 22.

Table 22: Split-Plot Version of the Wood Experiment

First panel Pretreated with $A1$				Second panel Pretreated with $A2$			
$B3$	$B2$	$B4$	$B1$	$B2$	$B1$	$B4$	$B3$

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

Split-plot Design (Cont'd)

- 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 A1 and A2 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.
- Wood Experiment: 3 replications, 6 whole plots (two large panels for A1 and A2 per replication).

Data from the Wood Experiment

Whole Plot	Pretreatment type (A)	Stain type (B)	Replication (Rep)	Resistance (Y)
4	2	2	1	53.5
4	2	4	1	32.5
4	2	1	1	46.6
4	2	3	1	35.4
5	2	4	2	44.6
5	2	1	2	52.2
5	2	3	2	45.9
5	2	2	2	48.3
1	1	3	1	40.8
1	1	1	1	43.0
1	1	2	1	51.8
1	1	4	1	45.5
2	1	2	2	60.9
2	1	4	2	55.3
2	1	3	2	51.1
2	1	1	2	57.4
6	2	1	3	32.1
6	2	4	3	30.1
6	2	2	3	34.4
6	2	3	3	32.2
3	1	1	3	52.8
3	1	3	3	51.7
3	1	4	3	55.3
3	1	2	3	59.2

Incorrect Model and Analysis

- Two-way layout model (factors A and B with n replicates):

$$y_{ijk} = \eta + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \boldsymbol{\epsilon}_{ijk},$$
$$i = 1, \ldots, I; \quad j = 1, \ldots, J; \quad k = 1, \ldots, n,$$

  where $I = 2$, $J = 4$, $n = 3$.

- ANOVA (table on next page) shows that only factor A is significant; neither B nor $A \times B$ is significant.

- The model is wrong: A and B use *different* randomization schemes. The error component should be separated into two parts–the whole plot error and the subplot error. To test the significance of various effects, we need to compare their respective mean squares with two different error components.

Incorrect ANOVA Table

Table 23: Incorrect ANOVA Table, Wood Experiment

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F
A	1	782.04	782.04	13.49
B	3	266.00	88.67	1.53
$A \times B$	3	62.79	20.93	0.36
Residual	16	927.88	57.99	
Total	23	2038.72		

- Only A is significant.

Correct Model

$$y_{ijk} = \eta + \tau_k + \alpha_i + (\tau\alpha)_{ki} + \beta_j + (\alpha\beta)_{ij} + (\tau\beta)_{kj} + (\tau\alpha\beta)_{kij} + \epsilon'_{ijk},$$
$$i = 1, \ldots, I; \quad j = 1, \ldots, J; \quad k = 1, \ldots, n, \tag{6}$$

- y_{ijk}: observation for the kth replicate of the $i$the level of factor A and the jth level of factor B, ϵ'_{ijk} are independent error terms.
- Treat τ_k as a **random** effect (because there are potentially many other possible replications).

Terms representing the whole plot		Terms representing the subplot	
τ_k	effect of kth replicate	β_j	jth main effect of B
α_i	ith main effect for A	$(\alpha\beta)_{ij}$	(i, j)th interaction $A \times B$
$(\tau\alpha)_{ki}$	(k, i)th interaction replicate$\times A$	$(\tau\beta)_{kj}$	(k, j)th interaction replicate$\times B$
		$(\tau\alpha\beta)_{kij}$	(k, i, j)th interaction replicate $\times A \times B$

Model for Split-plot Design

- Model (6) can be viewed as a three-way layout with α_i and β_j as *fixed* effects, τ_k as *random* effects, $\tau_k \sim N(0, \sigma_\tau^2)$.
- whole plot error: $(\tau\alpha)_{ki} \sim N(0, \sigma_{\tau\alpha}^2)$, for testing α effects.
- subplot error: $\epsilon_{kij} = (\tau\beta)_{kj} + (\tau\alpha\beta)_{kij} + \epsilon'_{ijk} \sim N(0, \sigma_\epsilon^2)$, for testing β and $\alpha\beta$ effects.
- Model (6) can be rewritten as

$$y_{ijk} = \eta + \tau_k + \alpha_i + (\tau\alpha)_{ki} + \beta_j + (\alpha\beta)_{ij} + \epsilon_{kij}. \tag{7}$$

- Subplot error is usually *smaller* than whole plot error because subplots are more homogeneous than whole plots.

ANOVA Decomposition

- Use the zero-sum constraints: $\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = \sum_{i=1}^{I} \sum_{j=1}^{J} (\alpha\beta)_{ij} = 0$, break up the total sum of squares as a three-way layout with factors A, B, and Rep:

$$SST = SS_{\text{Rep}} + SS_A + SS_B + SS_{\text{Rep}\times A} + SS_{A\times B} + SS_{\text{Rep}\times B} + SS_{\text{Rep}\times A\times B}.$$

- Define the sum of squares for the whole plot error SS_{whole} and the sum of squares for the subplot error SS_{sub} as:

$$\begin{aligned} SS_{\text{whole}} &= SS_{\text{Rep}\times A}, \\ SS_{\text{sub}} &= SS_{\text{Rep}\times B} + SS_{\text{Rep}\times A\times B}. \end{aligned}$$

ANOVA decomposition for the split-plot model:

$$SST = SS_{\text{Rep}} + SS_A + SS_{\text{whole}} + SS_B + SS_{A\times B} + SS_{\text{sub}}.$$

ANOVA Decomposition (Cont'd)

Table 24: Wood Experiment : Summarized data for whole plot analysis

	Rep 1	Rep 2	Rep 3	Total
$A1$	181.1	224.7	219.0	624.8
$A2$	168.0	191.0	128.8	487.8
Total	349.1	415.7	347.8	1112.6

$$
\begin{aligned}
SS_A &= (624.8^2 + 487.8^2)/12 - 1112.6^2/24 = 782.04, \\
SS_{\text{Rep}} &= (349.1^2 + 415.7^2 + 347.8^2)/8 - 1112.6^2/24 = 376.99, \\
SS_{\text{whole}} &= SS_{\text{Rep}\times A} = 398.37, \\
SS_{\text{sub}} &= 927.88 - SS_{\text{whole}} - SS_{\text{Rep}} = 152.52.
\end{aligned}
$$

Expected Mean Squares in ANOVA

Source	Effect	df	E(Mean Squares)
Replicate	τ_k	$n-1$	$\sigma_\epsilon^2 + J\sigma_{\tau\alpha}^2 + IJ\sigma_\tau^2$
A	α_i	$I-1$	$\sigma_\epsilon^2 + J\sigma_{\tau\alpha}^2 + \frac{nJ\sum_{i=1}^I \alpha_i^2}{I-1}$
Whole plot error	$(\tau\alpha)_{ki}$	$(I-1)(n-1)$	$\sigma_\epsilon^2 + J\sigma_{\tau\alpha}^2$
B	β_j	$J-1$	$\sigma_\epsilon^2 + \frac{nI\sum_{j=1}^J \beta_j^2}{IJ-1}$
$A \times B$	$(\alpha\beta)_{ij}$	$(I-1)(J-1)$	$\sigma_\epsilon^2 + \frac{n\sum_{i=1}^I \sum_{j=1}^J (\alpha\beta)_{ij}^2}{(I-1)(J-1)}$
Subplot error	ϵ_{kij}	$I(J-1)(n-1)$	σ_ϵ^2

Proofs are similar to but more tedious than in one-way random effects model.

Hypothesis Testing

$$F_A = \frac{MS_A}{MS_{\text{whole}}} \quad \Rightarrow \quad H_{01} : \alpha_1 = \ldots = \alpha_I,$$

$$F_B = \frac{MS_B}{MS_{\text{sub}}} \quad \Rightarrow \quad H_{02} : \beta_1 = \ldots = \beta_J,$$

$$F_{AB} = \frac{MS_{A\times B}}{MS_{\text{sub}}} \quad \Rightarrow \quad H_{03} : (\alpha\beta)_{ij} = \text{constant},\ i = 1, \ldots, I,\ j = 1, \ldots, J.$$

$$F_{Rep} = \frac{MS_{\text{Rep}}}{MS_{\text{whole}}} \quad \Rightarrow \quad H_{04} : \sigma_\tau = 0.$$

Correct ANOVA Analysis

Table 25: Correct ANOVA Table, Wood Experiment

Source	Degrees of Freedom	Sum of Squares	Mean Squares	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		

Analysis Results

- Only B is significant.
- Explanation:
  $MS_{\text{whole}} = 199.19 \gg MS_{\text{Residual}} = 57.99 \gg MS_{\text{sub}} = 12.71.$
- To test $H_{04} : \sigma_\tau = 0$, use

$$\frac{MS_{\text{Rep}}}{MS_{\text{whole}}} = \frac{188.5}{199.19} = 0.95.$$

  $\Rightarrow$ no significant difference between three replications.
- When does testing H_{04} make sense?

Analysis of Covariance: Starch Experiment

- Data in Table 3.34 of WH. Goal: To compare the three treatments (canna, corn, potato) for making starch film, y = break strength of film, covariate x = film thickness. Known that x affects y (thicker films are stronger); thickness cannot be controlled but are measured after films are made. **Question**: How to perform treatment comparisons by incorporating the effect of the covariate x?

- Model:

$$y_{ij} = \eta + \tau_i + \gamma x_{ij} + \epsilon_{ij}, \; i = 1, \ldots, k, \; j = 1, \ldots, n_i,$$

τ_i = ith treatment effect
x_{ij} = covariate value,
γ = regression coefficient for the x_{ij}
ϵ_{ij} independent $N(0, \sigma^2)$.

Special cases:

1. When γx_{ij}=0 (i.e., x_{ij} not available or no x effect), one-way layout.
2. When τ_i=0 (no treatment effect), simple linear regression.

Regression Model Approach

Model :

$$\begin{aligned} y_{1j} &= \eta + \gamma x_{1j} + \epsilon_{1j}, & j = 1, \cdots, 13,\ i = 1 \quad &\text{(canna)} \\ y_{2j} &= \eta + \tau_2 + \gamma x_{2j} + \epsilon_{2j}, & j = 1, \cdots, 19,\ i = 2 \quad &\text{(corn)} \\ y_{3j} &= \eta + \tau_3 + \gamma x_{3j} + \epsilon_{3j}, & j = 1, \cdots, 17,\ i = 3 \quad &\text{(potato)} \end{aligned} \tag{8}$$

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.

(Write the model matrix for (8)).

Run regression analysis in the usual way.

Regression Analysis of Starch Experiment

Table 26: Tests, Starch Experiment

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

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

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.

Multiple Comparisons

$Var(\hat{\tau}_3)$ and $Var(\hat{\tau}_2)$ can be obtained from regression output. From (1.33) of WH,

$$Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}.$$

Using this, $t_{(\hat{\tau}_3 - \hat{\tau}_2)}$ can be found as
$Var(\hat{\tau}_3 - \hat{\tau}_2) = Var(\hat{\tau}_3) + Var(\hat{\tau}_2) - 2Cov(\hat{\tau}_3, \hat{\tau}_2)$. 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 26. For simultaneous testing, use adjusted p values.

ANCOVA Table

Table 27: 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	

Transformation of Response

- Transform y before fitting a regression model.
  Theory: Suppose in the model $y = \mu + \epsilon$, $\sigma_y = [Var(y)]^{1/2}$, $\sigma_y \propto \mu^\alpha$. 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 ?)

- Error transmission formula:

$$z = f(y) \approx f(\mu) + f'(\mu)(y - \mu).$$

$$\sigma_z^2 = Var(z) \approx (f'(\mu))^2 Var(y) = (f'(\mu))^2 \sigma_y^2.$$

Power (Box-Cox) Transformation

$$
\begin{aligned}
z = f(y) &= \begin{cases} \dfrac{y^{\lambda}-1}{\lambda}, & \lambda \neq 0, \\ \ln y, & \lambda = 0, \end{cases} \qquad (9) \\
f'(\mu) &= \mu^{\lambda-1}, \\
\sigma_z \approx |f'(\mu)|\sigma_y &= \mu^{\lambda-1}\sigma_y \propto \mu^{\lambda-1}\mu^{\alpha} = \mu^{\lambda+\alpha-1}.
\end{aligned}
$$

- Choosing $\lambda = 1 - \alpha$ would make σ_z *nearly constant*.
- 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 each transform, analyze the z data and choose the transformation (i.e., λ value) such that
  (a) it gives a parsimonious model,
  (b) no unusual pattern in the residual plots,
  (c) good interpretation of the transformation.

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

Variance Stabilizing Transformations

Table 28: Variance Stabilizing Transformations

$\sigma_y \propto \mu^\alpha$	α	$\lambda = 1 - \alpha$	Transformation
$\sigma_y \propto \mu^3$	3	-2	reciprocal squared
$\sigma_y \propto \mu^2$	2	-1	reciprocal
$\sigma_y \propto \mu^{3/2}$	3/2	$-1/2$	reciprocal square root
$\sigma_y \propto \mu$	1	0	log
$\sigma_y \propto \mu^{1/2}$	1/2	1/2	square root
$\sigma_y \propto \text{constant}$	0	1	original scale
$\sigma_y \propto \mu^{-1/2}$	$-1/2$	3/2	3/2 power
$\sigma_y \propto \mu^{-1}$	-1	2	square

Analysis of Drill Experiment

- Data in Table 3.40 of WH. Four factors A,B,C and D, each at two levels, using a 2^4 design. Fit a model with 4 main effects and 6 two-factor interactions (2fi's). The $\hat{r}$-vs-$\hat{y}$ plot shows an increasing pattern.

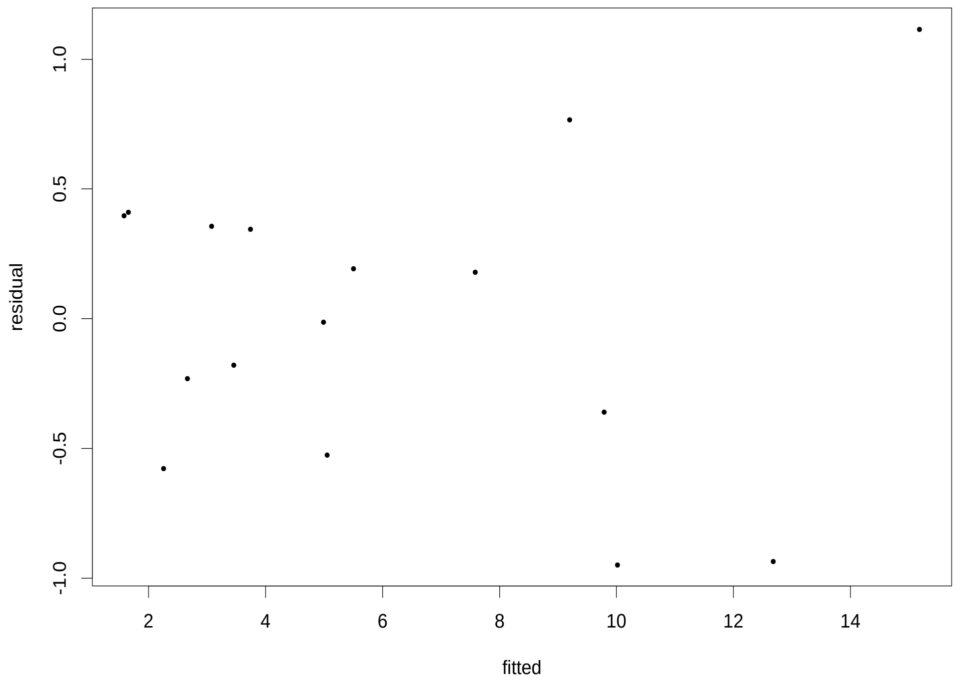

Figure 2: r_i vs. $\hat{y}_i$, Drill Experiment

Scaled lambda plot

- For each of the eight transformations λ values in Table 28, a model of main effects and 2fi's is fitted to the transformed $z = f(y)$. The t statistic values for the 10 effects are displayed.

- **Comments on 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 (Why ? It has an interaction BC). The log transform removes the interaction term BC.

  On the original scale ($\lambda = 1$), the four main effects do not separate apart.

- **Conclusion :** Use log transformation.

Scaled lambda plot : Drill Experiment

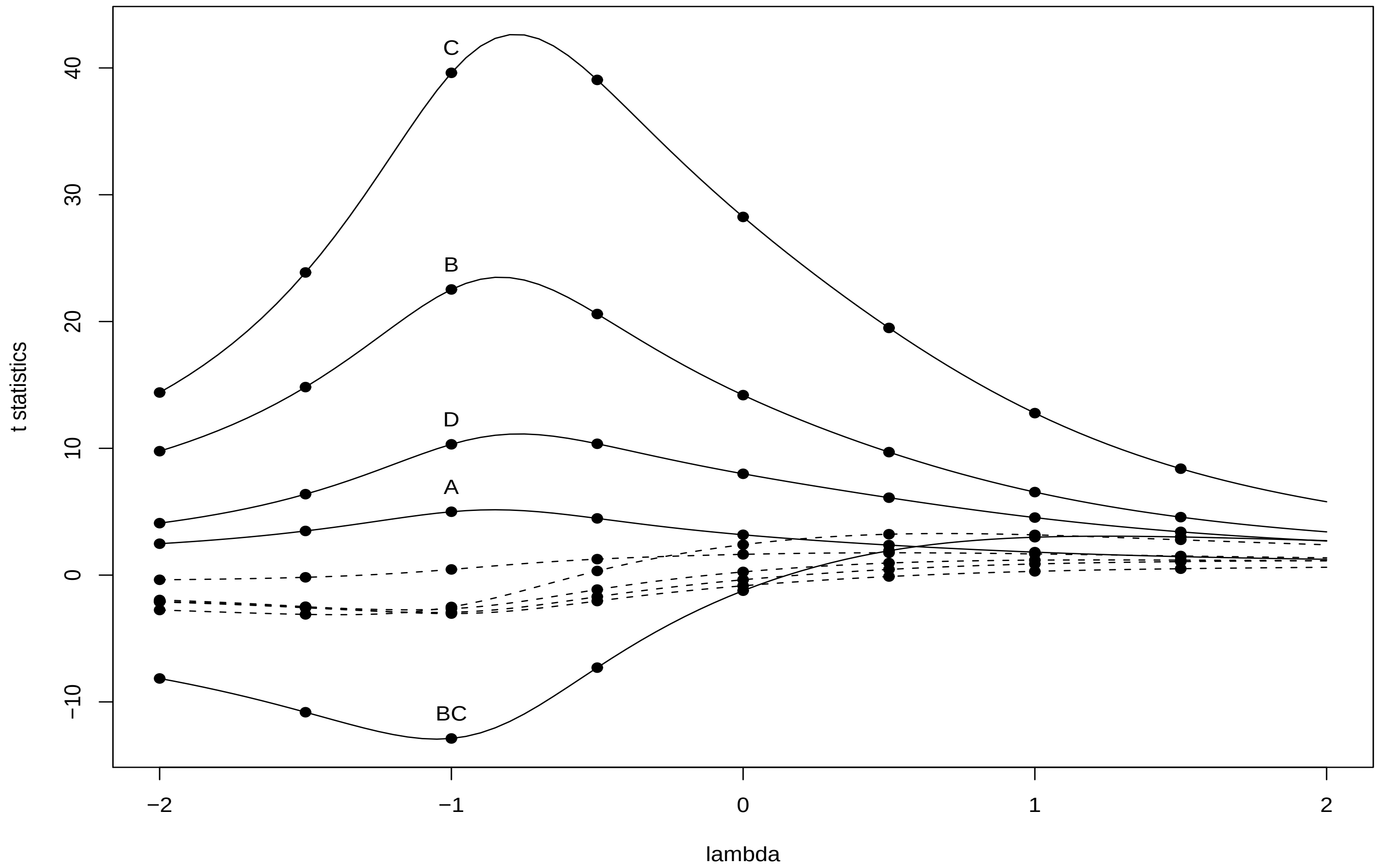

Figure 3: Scaled λ Plot (lambda denotes the power λ in the transformation (9))

Comments on Board