Analysis Results : t tests

$d_1, \cdots, d_N$ (within block comparison)

identical cf.

$$t_{paired} = \frac{0.4138}{0.321/\sqrt{8}} = \frac{\overset{=\bar{d}}{0.4138}}{\underset{s.e.(\bar{d})}{0.1135}} = 3.645,$$

t_{paired}: 1-sample problem

$$t_{unpaired} = \frac{5.435-5.0212}{\sqrt{(17.811+17.012)/8}} = \frac{\overset{=\bar{y}_2-\bar{y}_1}{0.4138}}{\underset{=s.e.(\bar{y}_2-\bar{y}_1)}{2.0863}} = 0.198.$$

cf.

$t_{unpaired}$: one-way layout (2-sample problem)

2.0863: much larger (why?)

The p values are

waste d.f. to estimate block effects

Q: What if the sample-to-sample variation is small?

$$Prob(|t_7| > 3.645) = 0.008, \leftarrow \text{significant}$$

$$Prob(|t_{14}| > 0.198) = 0.848. \leftarrow \text{insignificant}$$

cf. 7: $N-1$; 14: $2N-2$

cf.

Note. $t_{7,\alpha/2} > t_{14,\alpha/2}$

- 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.

between block variation

if large ⇒ significant block effects

Analysis Results : (sequential) ANOVA and F-tests

Recall 1. equivalence btwn t-test & F-test

Recall 2. sequential ANOVA ← cf. → drop-one ANOVA

- Wrong to analyze by ignoring pairing.

  A better explanation is given by ANOVA.

Under sum coding for factors

model matrix

$$X = \Big[\ \begin{matrix}1\\ \vdots\\ 1\end{matrix}\ \Big|\ \underbrace{\cdots}_{x_b}\ \Big|\ \underbrace{\cdots}_{x_t}\ \Big]$$

$W_0 = span\{\mathbb{1}\}$
$W_b = span\{x_b\}$
$W_t = span\{x_t\}$

$\{\mathbb{1}\} \perp \{x_b\} \perp \{x_t\}$ ← ∵ balance

⇒ sequential = drop-one

⇒ sequential ANOVA does not depend on order of effect entry

Note. ANOVA does not depend on coding

- Data can be analyzed in two equivalent ways.
  - F statistic in ANOVA for paired design equals t^2_{paired}; (2 factors)
  - F statistic in ANOVA for unpaired design equals $t^2_{unpaired}$. (1 factor)

$\Omega: y \sim \beta_0 + sample + method + \varepsilon$
$\omega: y \sim \beta_0 + sample + \varepsilon$

$\Omega: y \sim \beta_0 + method + \varepsilon^*$
$\omega: y \sim \beta_0 + \varepsilon^*$

- In the correct analysis (Table 2, LNp.4-7), the total variation is decomposed into three components; the largest one is the sample-to-sample variation ($MS = 34.77$). In the unpaired analysis (Table 3, LNp. 4-7), this component is mistakenly included in the residual SS, thus making the F test powerless.

ANOVA Tables

In sequential ANOVA, block factor should always be put before treatment factor.

$\|P_{\mathcal{W}_b}(\underset{\sim}{y})\|^2$

$\|P_{\mathcal{W}_t}(\underset{\sim}{y})\|^2$

$\|P_{\Omega^\perp}(\underset{\sim}{y})\|^2$

[規律$_b$ / 規律$_t$] 隨機

anova(y ~ Bo+sample +method)

Table 2: ANOVA Table, Sewage Experiment

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F
sample	7 = N-1	243.4042	34.77203	674.82
method	1 = 2-1	0.6848	0.68476	13.29
residual	7 = N-1	0.3607	0.05153 $\hat{\sigma}^2_p$	
total	15 = 2N-1 ← 16 obs	☐		

Ω_b: y~Bo+sample
ω_b: y~Bo

Ω_t: y~Bo+sample +method
ω_t: y~Bo+sample

identical (∵ orthogonality)

cf.

Q: What if sample-to-sample variation is small?

anova(y~ Bo+method)

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F
method	1	0.6848	0.68476	0.04
residual	14	243.7649	17.41178 $\hat{\sigma}^2_{up}$	
total	15	☐		

much larger

Ω: y~Bo+method
ω: y~Bo

Table 3: ANOVA Table Ignoring Pairing, Sewage Experiment

❖ **Reading**: textbook, 3.1

Randomized (Complete) Block Design: Girder Experiment

combine the concepts in
- one-way layout
- paired designs

- **Objective :** To compare four methods for predicting the shear strength for steel plate girders ($k = 4, b = 9$).

✱ response: strength

✱ treatment factor: method (qualitative)
4 levels - A, K, L, C

✱ block factor
Girder (qualitative)
9 levels - S1/1, ⋯, S4/2
block size = 4

✱ Exp'tal units: a section of a steel plate girder

S1/1 (block 1)
EU11 ••• EU14
r A, K, L, C r
⋮
S4/2 (block 9)
EU91 ••• EU94
r A, K, L, C r

restricted randomization

within block comparison

Table 4: Strength Data, Girder Experiment

(Block)	Method			
Girder	Aarau	Karlsruhe	Lehigh	Cardiff
block 1 → 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
block 9 → S4/2	0.859	1.559	1.052	0.805

✱ Design matrix — Data matrix

Girder	Method	
S1/1	A	0.772
S1/1	K	•
S1/1	L	•
S1/1	C	•
⋮	⋮	
S4/2	A	
S4/2	K	
S4/2	L	
S4/2	C	0.805

same as in paired design

same as in one-way layout

single-replicate design

ANOVA & multiple comparison

- 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 $N = bk$.
  → or $k \times \ell$ units, ℓ: # of replicates — or $N = bk\ell$
- Paired comparison design is a special case of RBD with $k = 2$. (Why?)

Model and Estimation

- Model for RBD :

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

where

y_{ij} = observation of the jth treatment in the ith block,

α_i = ith block effect, ← fixed effects

τ_j = jth treatment effect,

ε_{ij} = errors, independent $N(0,\sigma^2)$.

The following results also hold for the case of kl units in a block. $y_{ij} \to y_{ijm}$, $\varepsilon_{ij} \to \varepsilon_{ijm}$, $m = 1, \dots, l$.

before exp't → conceptual model: $y \sim B_0 + \text{Girder} + \text{Method} + \varepsilon$

8 parameters ↑ ↑ 3 parameters

(Note. main-effect-only model, no interactions) → $\hat{\sigma}$

over-parameterised model: can do ANOVA (∵ ω_i, Ω_i still well-defined), but cannot do estimation (need to add some constraints on α_i's & τ_j's)

What if α_i's treated as random effects? Check split-plot design (LNp.45~66)

Under sum coding

$$\begin{bmatrix} \alpha_b = -(\alpha_1 + \cdots + \alpha_{b-1}), \\ \tau_k = -(\tau_1 + \cdots + \tau_{k-1}) \end{bmatrix}$$

model matrix

$$X = \Big[\ \underset{}{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\ \Big|\ \underbrace{\ \cdots\ }_{X_b}\ \Big|\ \underbrace{\ \cdots\ }_{X_t}\ \Big]$$

$W_0 = \text{span}\{\mathbb{1}\}$, $W_b = \text{span}\{X_b\}$, $W_t = \text{span}\{X_t\}$

$\{\mathbb{1}\} \perp \{X_b\} \perp \{X_t\}$ ← ∵ balance

$\dim(W_0) = 1$, $\dim(W_b) = b-1$, $\dim(W_t) = k-1$

$\Omega = W_0 \oplus W_b \oplus W_t$

$\Rightarrow \dim(\Omega^\perp) = (b-1)(k-1)$

- Estimation

use the same method as given in LNp3-4~5 to show $P_{W_0 \oplus W_b}(y) = \bar{y}_{i\cdot}$, $P_{W_0 \oplus W_t}(y) = \bar{y}_{\cdot j}$

$$\begin{aligned} y_{ij} &= \hat{\eta} + \hat{\alpha}_i + \hat{\tau}_j + r_{ij} \\ \underset{\sim}{y} &= \underbrace{\bar{y}_{\cdot\cdot}}_{P_{W_0}(y)} + \underbrace{(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})}_{P_{W_b}(y)} + \underbrace{(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})}_{P_{W_t}(y)} \\ &\quad \underbrace{+ (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})}_{P_{\Omega^\perp}(y)}, \end{aligned}$$

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

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

$\bar{y}_{i\cdot} = \frac{1}{k}\sum_{j=1}^k y_{ij}$, $\bar{y}_{\cdot j} = \frac{1}{b}\sum_{i=1}^b y_{ij}$, $\bar{y}_{\cdot\cdot} = \frac{1}{bk}\sum_{i=1}^b \sum_{j=1}^k y_{ij}$

$$\begin{bmatrix} \hat{\eta} \\ \hat{\alpha}_i \\ \hat{\tau}_j \end{bmatrix} = \hat{\underset{\sim}{\beta}} = (X^TX)^{-1}X^TY = \begin{bmatrix} \bar{y}_{\cdot\cdot} \\ (X_b^TX_b)^{-1}X_b^TY \\ (X_t^TX_t)^{-1}X_t^TY \end{bmatrix}$$

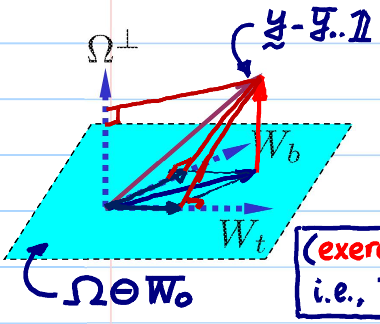
$$\hat{Y} = X\hat{\beta} = X(X^TX)^{-1}X^TY = \underbrace{\bar{y}_{\cdot\cdot}\mathbb{1}}_{P_{W_0}(y)} + \underbrace{X_b(X_b^TX_b)^{-1}X_b^TY}_{P_{W_b}(y)} + \underbrace{X_t(X_t^TX_t)^{-1}X_tY}_{P_{W_t}(y)}$$

ANOVA

$$\|y\|^2 = \|P_{W_0}(y)\|^2 + \|P_{W_b}(y)\|^2 + \|P_{W_t}(y)\|^2 + \|P_{\Omega^\perp}(y)\|^2$$

- **ANOVA decomposition**. Subtracting $\bar{y}_{\cdot\cdot}$, squaring both sides and summing over i and j yields

↑ remove $P_{W_0}(y)$



$$\begin{aligned} \underbrace{\sum_{i=1}^b \sum_{j=1}^k (y_{ij} - \bar{y}_{\cdot\cdot})^2}_{\text{total variation in } y,\ \|P_{\mathbb{R}^N \ominus W_0}(y)\|^2} &= \underbrace{\sum_{i=1}^b k(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}_{\|P_{W_b}(y)\|^2} + \underbrace{\sum_{j=1}^k b(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2}_{\|P_{W_t}(y)\|^2} \\ &\quad + \underbrace{\sum_{i=1}^b \sum_{j=1}^k (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2}_{\|P_{\Omega^\perp}(y)\|^2} \\ &= SS_b + SS_t + SS_r. \end{aligned}$$

$\mathbb{R}^N \ominus W_0 = W_b \oplus W_t \oplus \Omega^\perp$

sources of variation

(exercise) Note. ∵ orthogonality, i.e., $W_0 \perp W_b \perp W_t \perp \Omega^\perp$

Table 5: ANOVA Table for Randomized Block Design

→ usually is the 1st effects to enter the model

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

(exercise) check (P4) in LNp. 2-36 & LNp. 3-6

Note. In this case, ∵ orthogonality, anova($y \sim B_0$ + block + treatment) & anova($y \sim B_0$ + treatment + block) have same output

same as the df's for interactions ← → can do $k \times b$-run RBD because of assuming main-effect model in conceptual modeling

Testing and Multiple Comparisons

ANOVA & multiple comparison for RBD use the same principles as in one-way layout, except that the variation caused by block factor must be removed from error variation.

how? adding block effects into model

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

$$F = \frac{\overbrace{SS_t}^{RSS_\omega - RSS_\Omega}/\overbrace{(k-1)}^{df_\omega - df_\Omega}}{\underbrace{SS_r}_{RSS_\Omega}/\underbrace{(b-1)(k-1)}_{df_\Omega}}, \qquad (3)$$

$\Omega: y \sim B_0 + \text{Girder} + \text{Method} + \varepsilon$ $(y = \eta + \alpha_i + \tau_j + \varepsilon)$

$\omega: y \sim B_0 + \text{Girder} + \varepsilon$ $(y = \eta + \alpha_i + \varepsilon)$

Recall. LNp.3-9~13

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.*

– The t-statistics for making multiple comparisons :

$\mathrm{Var}(\bar y_{\cdot j} - \bar y_{\cdot i}) = \sigma^2/b + \sigma^2/b$

$H_0^{ij}: \tau_i = \tau_j$ $(\tau_j - \tau_i = 0)$

– for specific (i,j) → use t-dist. for null

– for all (i,j)'s

$$t_{ij} = \frac{\overbrace{\bar y_{\cdot j} - \bar y_{\cdot i}}^{\hat\tau_j - \hat\tau_i} - 0}{\underset{\text{block effects removed}}{\hat\sigma}\sqrt{1/b + 1/b}}, \qquad (4)$$

where $\hat\sigma^2$ is the mean square error in the ANOVA table. $\hat\sigma^2 = \frac{SS_r}{(b-1)(k-1)}$

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

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

← studentized range dist.

cf. → Alternative: Bonferroni Method

Simultaneous Confidence Intervals

Recall. duality between confidence interval and test.

model: $y = \eta + \text{block} + \text{treatment} + \varepsilon$

τ_j's

$H_0^{ij}: \tau_j - \tau_i = C$ for all (i,j)

- By solving

$$\frac{\left|(\bar y_{\cdot j} - \bar y_{\cdot i}) - \overbrace{(\tau_j - \tau_i)}^{C}\right|}{\hat\sigma\sqrt{2/b}} \le \frac{1}{\sqrt{2}} q_{k,(b-1)(k-1),\alpha}$$

← # of τ_j's

×

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

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

$\hat\sigma$ ← block effects removed

for all i and j pairs.

Analysis of Girder Experiment : F-test

Table 6: ANOVA Table, Girder Experiment

anova($y \sim \beta_0 + \text{girder} + \text{method}$)

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F
girder	$8 = 9-1\ (b-1)$	$0.089 \leftarrow \|P_{\omega_b}(y)\|^2$	0.011	1.62
method	$3 = 4-1\ (k-1)$	$1.514 \leftarrow \|P_{\omega_t}(y)\|^2$	0.505	73.03
residual	$24 = (b-1)(k-1)$	$0.166 \leftarrow \|P_{\Omega^\perp}(y)\|^2$	0.007 $\leftarrow \hat\sigma^2$	
total	$35 \leftarrow 36$ obs	$\square \leftarrow \|P_{\mathbb{R}^N \ominus \omega_0}(y)\|^2 = \|y\|^2 - \|P_{\omega_0}(y)\|^2$		

girder: not of interest

method: of main interest

$\Omega_b: y \sim \beta_0 + \text{girder} + \varepsilon$

$\omega_b: y \sim \beta_0 + \varepsilon$

1.62: not significant ⇒ homogeneous btwn blocks ⇒ (maybe) not necessary to consider it as a block factor in future exp't

- The F-statistic in (3) (LNp.4-11) has the value

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

$\Omega_t: y \sim \beta_0 + \text{girder} + \text{method} + \varepsilon$

$\omega_t: y \sim \beta_0 + \text{girder} + \varepsilon$

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

Answer to "how different" problem

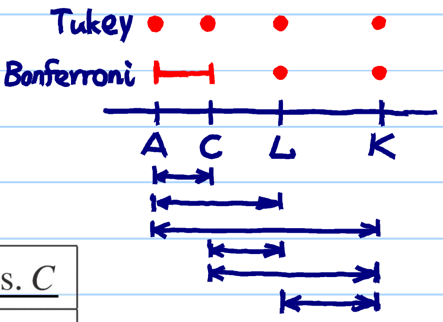


Table 7: Multiple Comparison t Statistics, Girder Experiment

	A vs. K	A vs. L	A vs. C	K vs. L	K vs. C	L vs. C
t_{ij}	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.

  $\bar y_{\cdot i}$'s

- The multiple comparison t-statistics based on (4) (LNp.4-11) are displayed in Table 7. For example, the A vs. K t-statistic is

$$t_{12} = \frac{1.3401 - 0.7949}{\sqrt{\underset{\hat\sigma^2}{0.007}}\sqrt{\underset{2(1/b)}{2/9}}} = 13.91.$$

With $\alpha = 0.05$, $t_{24,0.05/(6\times2)} = 2.875$ for the Bonferroni method. Since $k = 4$ and $\binom{k}{2} = 6$, $\frac{1}{\sqrt2} q_{4,24,0.05} = \frac{3.90}{1.414} = 2.758$ for the Tukey method. Again, Tukey method is more powerful. (Why ?)

6: # of tests

$k = 4$: # of $\bar y_{\cdot i}$'s

❖ **Reading**: textbook, 3.2

Two-way layout

nesting e.g. plant < A → operator − 1, 2, 3; B → operator − 1′, 2′, 3′

cf. ↕

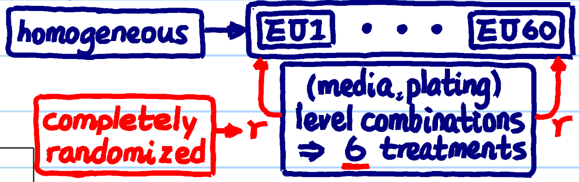
↳ treatment factors (crossing)

- **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.

★ response : torque

★ treatment factors :
① media (qualitative) ② plating (qualitative)
2 levels - B, M   3 levels - C, H, P

★ Exp'tal units : a locknut, 60 EUs



homogeneous → EU1 • • • EU60

completely randomized → r

(media, plating) level combinations ⇒ 6 treatments

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

★ Each treatment repeats 10 times
⇒ 10 replicates (Why need replicate?)

★ Design matrix   Data matrix

$$\left[\begin{array}{cc|c} \text{media} & \text{plating} & y \\ B & C & \square \cdots \square \\ B & H & \vdots \\ B & P & \vdots \\ M & C & \vdots \\ M & H & \vdots \\ M & P & \square \cdots \square \end{array}\right]$$

total dfs : 60
− 6 for β
− 54 for σ^2

- This is similar to RBD. The only difference is that we have two treatment factors instead of one treatment factor and one block factor. Also interested in assessing interaction effects between the two treatment factors. In blocking, block×treatment interaction is assumed negligible.

different EU structure
different randomization scheme

Conceptual model : ← before exp't

$y \sim \beta_0 + \underline{\text{media}} + \underline{\text{plating}}$

1 parameter   2 parameters

$+ \underline{\text{media} \times \text{plating}} + \varepsilon$

$1 \times 2 = 2$ parameters