Experimental Design
Note 4
Randomized complete block design (RCBD)

Keunbaik Lee

Sungkyunkwan University

Nuisance Factor

Control 해야 하지만 관심사항이 아닐 때

Nuisance Factor (may be present in experiment)

- Has effect on response but its effect is not of interest
- if unknown → Protecting experiment through randomization
- If known (measurable) but uncontrollable → Analysis of Covariance (Chapter 15 Section 3)
- If known and controllable → Blocking

Example: Penicillin Experiment I

- In this experiment, four penicillin manufacturing processes (A, B, C, and D) where being investigated. Yield was the response. It was known that an important raw material, corn steep liquor, was quite variable. The experiment and its results were given below:

block 안에서 Randomization

	blend 1	blend 2	blend 3	blend 4	blend 5
A	89_1	84_4	81_2	87_1	79_3
B	88_3	77_2	87_1	92_3	81_4
C	97_2	92_3	87_4	89_2	80_1
D	94_4	79_1	85_3	84_4	88_2

factor

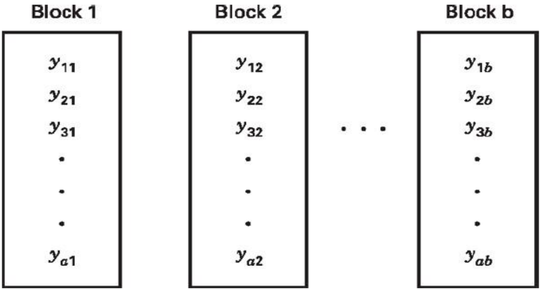
blocks

- Blend is a nuisance factor, treated as a block factor.

Example: Penicillin Experiment II

- (Complete) Blocking: all the treatments are applied within each block, and they are compared within blocks.
- Advantage: Eliminate blend-to-blend (between-block) variation from experimental error variance when comparing treatments.
- Cost: degree of freedom.

Randomized Complete Block Design (RCBD) I



- *b* blocks each consisting of (partitioned into) *a* experimental units.

Randomized Complete Block Design (RCBD) II

- a treatments are randomly assigned to the experimental units within each block.
- Typically after the runs in one block have been conducted, then move to another block.
- Typical blocking factors: day, batch of raw material etc.
- Results in restriction on randomization because randomization is only within blocks.
- Data within a block are related to each other, When $a = 2$, randomized complete block design becomes paired two sample case.

Statistical Model: two-way ANOVA I

- b blocks and a treatments
- Statistical model is

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, \text{ for } i = 1, 2, \cdots, a; \; j = 1, 2, \cdots, b$$

  where μ is the grand mean, τ_i is the ith treatment effect, β_j is the $j$the block effect, and $\epsilon_{ij} \sim^{iid} N(0, \sigma^2)$.
- The model is additive because within a fixed block, the block effect is fixed; for a fixed treatment, the treatment effect is fixed across blocks. In other words, blocks and treatments do not interact.
- parameter constraints: $\sum_{i=1}^{a} \tau_i = 0$; $\sum_{j=1}^{b} \beta_j = 0$

Estimates for Parameters

- Rewrite observation y_{ij} as:

$$y_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})$$

- Compared with the model

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}.$$

- The estimates for the parameters are

$$\begin{aligned} \hat{\mu} &= \bar{y}_{..}, \\ \hat{\tau}_i &= \bar{y}_{i.} - \bar{y}_{..}, \\ \hat{\beta}_j &= \bar{y}_{.j} - \bar{y}_{..}, \\ \hat{\epsilon}_{ij} &= y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..} \end{aligned}$$

Sum of Squares (SS) I

- Can partition $SS_T = \sum_i \sum_j (y_{ij} - \bar{y}_{\cdot\cdot})^2$ into

$$b\sum_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 + a\sum_j(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2 + \sum_i\sum_j(y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2,$$

$$SS_{Treatment} = b\sum_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = b\sum_i \hat{\tau}_i^2, \quad df = a-1,$$

$$SS_{Block} = a\sum_j(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2 = a\sum_j \hat{\beta}_j^2, \quad df = b-1,$$

$$SSE = \sum_i\sum_j(y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2 = \sum_i\sum_j \hat{\epsilon}_{ij}^2, \; df = (a-1)(b-1).$$

Hence,

$$SST = SS_{Treatment} + SS_{Block} + SSE$$

- The Mean Squares are $MS_{Treatment} = SS_{Treatment}/(a-1)$, $MS_{Block} = SS_{Block}/(b-1)$, and $MSE = SSE/(a-1)(b-1)$.

Testing Basic Hypotheses I

- $H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$ vs H_1 : at least one is not
- Can show:

$$\frac{E(MS_{treatment})}{E(MS_E)} = \frac{\sigma^2 + \frac{b}{a-1}\sum_{i=1}^{a}\tau_i^2}{\sigma^2}$$

$$E(MSE) = \sigma^2$$

$$E(MS_{Treatment}) = \sigma^2 + b\sum_i \tau_i^2/(a-1)$$

$$E(MS_{Block}) = \sigma^2 + a\sum_j \beta_j^2/(b-1)$$

- Use F-test to test H_0 :

$$F_0 = \frac{MS_{Treatment}}{MSE} = \frac{SS_{Treatment}/(a-1)}{SSE/((a-1)(b-1))}$$

Testing Basic Hypotheses II

- Caution testing block effects
  - Usually not of interest.
  - Randomization is restricted: Differing opinions on F-test for testing blocking effect.
  - Can use ratio MS_{Block}/MSE to check if blocking successful.
  - Block effects can be random effects (considered fixed effects in this chapter).

Analysis of Variance (ANOVA) Table I

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Blocks	SS_{Block}	$b-1$	MS_{Block}	
Treatment	$SS_{Treatment}$	$a-1$	$MS_{Treatment}$	$MS_{Teatment}/MSE$
Error	SSE	$(a-1)(b-1)$	MSE	
Total	SS_T	$ab-1$		

SS_{model} $a+b-2$ MS_{model}

$$F_0 = \frac{MS_{model}}{MS_E} \overset{H_0}{\sim} F_{a+b-2,(a-1)(b-1)}$$

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_a = \beta_1 = \beta_2 = \cdots = \beta_b = 0$$

$$SS_T = \sum_i \sum_j y_{ij}^2 - y_{..}^2/N,$$

$$SS_{Treatment} = \frac{1}{b}\sum_i y_{i.}^2 - y_{..}^2/N,$$

$$SS_{Block} = \frac{1}{a}\sum_j y_{.j}^2 - y_{..}^2/N,$$

$$SSE = SS_T - SS_{Treatment} - SS_{Block}.$$

Decision rule: If $F_0 > F_{\alpha,a-1,(a-1)(b-1)}$, then reject H_0.

Another example I

An experiment was designed to study the performance of our different detergents in cleaning clothes. The following "cleanness" readings (higher=cleaner) were obtained specially designed equipment for three different types of common stains. Is there a difference between the detergents?

	Stain 1	Stain 2	Stain 3
Detergent 1	45	43	51
Detergent 2	47	46	52
Detergent 3	48	50	55
Detergent 4	42	39	49

Another example II

$$\sum_i \sum_j y_{ij} = 565 \text{ and } \sum_i \sum_j y_{ij}^2 = 26867$$

$y_{1\cdot} = 139,\; y_{2\cdot} = 145,\; y_{3\cdot} = 153,$ and $y_{4\cdot} = 128;$

$y_{\cdot 1} = 182,\; y_{\cdot 2} = 176,$ and $y_{\cdot 3} = 207$

$SS_T = 26867 - 565^2/12 = 265$

$SS_{Trt} = (139^2 + 145^2 + 153^2 + 128^2)/3 - 565^2/12 = 111$

$SS_{Block} = (182^2 + 176^2 + 207^2)/4 - 565^2/12 = 135$

$SSE = 265 - 111 - 135 = 19$

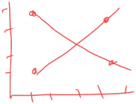
$F_0 = (111/3)/(19/6) = 11.6,$ P-value < 0.01

Checking Assumptions (Diagnostics) I

- Assumptions
  - Model is additive (no interaction between treatment effects and block effects) (additivity assumption).
  - Errors are independent and normally distributed.
  - Constant variance.
- Checking normality:
  - Histogram, QQ plot of residuals, Shapiro-Wilk Test.
- Checking constant variance
  - Residual Plot: Residuals vs $\hat{y}_{ij}$
  - Residuals vs blocks
  - Residuals vs treatments
- Additivity
  - Residual plot: residuals vs $\hat{y}_{ij}$

Checking Assumptions (Diagnostics) II

  - If residual plot shows curvilinear pattern, interaction between treatment and block likely exists.
  - Interaction: block effects can be different for different treatments.



- Formal test: Tukey's One-degree Freedom Test of Non-additivity.
- If interaction exists, usually try transformation to eliminate interaction.

See block-design.SAS.

Model adequacy checking: additivity

- We used a linear statistical model for RCBD:

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$$

- In other word, it is additive.
- The linear model is very useful, but in some situations it may be inadequate.
  - i.e., there may be an interaction between the treatment and $(\tau\beta)_{ij}$ block.

위 식에 τ_i와 β_i 간의 Interaction term 을 추가하면 ϵ_{ij}에 할당할 Df 가 없어진다. 그래서 이러한 문제를 해결하고자 $y_{ij} = \mu + \tau_i + \beta_j + \gamma\tau_i\beta_j + \epsilon_{ij}$ 를 Full model 이라 가정하고 $y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$ 를 reduced model 이라 가정한 뒤,

$$F_0 = \frac{(SS_E(R) - SS_E(F))}{MS_E(F)} \overset{H_0}{\sim} F_{1,\,df(F)}$$

Tukey's Test for Non-additivity I

- Additivity assumption (or no interaction assumption) is crucial for block designs or experiments.
- To check the interaction between block and treatment fully needs $(a-1)(b-1)$ degree of freedom. It is not affordable when without replicates.
- Instead consider a special type of interaction. Assume following model

Full model : $$y_{ij} = \mu + \tau_i + \beta_j + \gamma\tau_i\beta_j + \epsilon_{ij}$$, $SS_E(F)$, $SS_{trt}(F)$

Reduced model : $y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$, $SS_E(R)$, $SS_{trt}(R)$

$\Rightarrow F = \dfrac{SS_E(R) - SS_E(F)}{MS_E(F)} \sim F_{1, df(SS_E(F))}$

Tukey's Test for Non-additivity II

- $H_0 : \gamma = 0$ vs $H_1 : \gamma \neq 0$
  Sum of Squares caused by possible interaction:

  $$SS_N = \frac{\left[\sum_i \sum_j y_{ij} y_{i\cdot} y_{\cdot j} - y_{\cdot\cdot}(SS_{Trt} + SS_{Block} + y_{\cdot\cdot}^2/ab)\right]^2}{abSS_{Trt}SS_{Block}}, \quad df = 1.$$

  Remaining error SS: $SSE' = SSE - SS_N$,
  $df = (a-1)(b-1) - 1/$
  Test Statistic:

  $$F_0 = \frac{SS_N/1}{SSE'/[(a-1)(b-1)-1]} \sim F_{1,(a-1)(b-1)-1}$$

- Decision rule: Reject H_0 if $F_0 > F_{\alpha,1,(a-1)(b-1)-1}$.

Tukey's Test for Non-additivity III

- A Convenient Procedure to Calculate SS_N, SSE' and F_0
  - Fit additive model $y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$
  - fitted values
  - Obtain $\hat{y}_{ij}$ and $q_{ij} = \hat{y}_{ij}^2$ $= (\hat{\mu} + \hat{\tau}_i + \hat{\beta}_j)^2$
  - Fit the model $y_{ij} = \mu + \tau_i + \beta_j + q_{ij} + \epsilon_{ij}$
    Use the test for q_{ij} in the ANOVA table with type III SS and ignore the test for the treatment and block factors.

See block-design-additivity.SAS.

Interaction term 으로 $\tau_i \cdot \beta_j$ 와 같은 형태가 아닌 q_{ij} 의 형태로 넣는 이유는 어차피

해당 실험의 목적이 additive 인가 non-additive, 즉 quadratic 인가를 가려내기 위함이기 때문

Type I, III sum of squres

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$$

$y_i = \beta_0 + \varepsilon_i$

$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$

$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$

$SS_E(R) - SS_E(F)$ sequentially

- Type I (sequential) Sums of Squares
  - The Type I Sums of Squres for β_1 are the Sums of Squares obtained from fitting β_1 over and above the mean;
  - The Type I Sums of Squres for β_2 are the Sums of Squares obtained from fitting β_2 after β_1.
  - etc

Full model은 $y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$ 로 고정되고,
Reduced model은 parameters 중 하나씩 뺀 식들이다.

- Type III (marginal) Sums of Squares
  - The Sums of Squares obtained by fitting each effect after all the other terms in the model.
  - The Type III (marginal) Sums of Squares do not depend upon the order in which effects are specified in the model.

More info: http://afni.nimh.nih.gov/sscc/gangc/SS.html

Post- ANOVA Treatments Comparison I

- Multiple Comparisons/Contrasts
  - procedures (methods) are similar to those for Completely Randomized Design (CRD)
    n is replaced by b in all formulas
    Degree of freedom for error is $(b-1)(a-1)$
- Example: Comparison of Detergents See block-design-additivity.SAS.
  - Tukey's Method ($\alpha = .05$)
    $q_\alpha(a, df) = q_\alpha(4, 6) = 4.896$.
    $CD = \frac{q_\alpha(4,6)}{\sqrt{2}}\sqrt{MSE\left(\frac{1}{b}+\frac{1}{b}\right)} = 4.896\sqrt{\frac{19}{6\times 3}} = 5.001$

Post- ANOVA Treatments Comparison II

Comparison of Treatment Means

Treatments			
4	1	2	3
42.67	46.33	48.33	51.00
A	A		
	B	B	B

Random Blocks and/or Treatments I

Assuming that the RCBD model is appropriate, it the blocks are random and the treatments are fixed, then

$$
\begin{aligned}
& y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, \\
& \beta_j \overset{iid}{\sim} N(0, \sigma^2_\beta), \quad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2),
\end{aligned}
$$

for $i = 1, \ldots, a$; $j = 1, \ldots, b$. b_i and ϵ_{ij} are independent. Then we have

$$
\begin{aligned}
E(y_{ij}) &= \mu + \tau_i,, \quad var(y_{ij}) = \sigma^2_\beta + \sigma^2, \\
cov(y_{ij}, y_{i'j'}) &= 0, \quad j \neq j', \\
cov(y_{ij}, y_{i'j}) &= \sigma^2_\beta, \quad i \neq i',
\end{aligned}
$$

Thus, the variance ofthe observations is constant, the covariance between any two observations in different blocks is zero, but the covariance between two observations from the same block is σ_β^2. The expected mean squares from the usual ANOVA partitioning of the total sum of squares are

$$E(MS_{Treatment}) = \sigma^2 + \frac{b\sum_{i=1}^{a}\tau_i^2}{a-1},$$
$$E(MS_{Block}) = \sigma^2 + a\sigma_\beta^2,$$
$$E(MSE) = \sigma^2.$$

Random Blocks and/or Treatments III

The appropriate statistic for testing the null hypothesis of no treatment effects (all $\tau_i = 0$) is

$$F_0 = \frac{MS_{Treatment}}{MSE}$$

which is exactly the same test statistic we used in the case where the blocks were fixed. Based on the expected mean squares, we can obtain an ANOVA-type estimator of the variance component for blocks as

$$\hat{\sigma}^2_\beta = \frac{MS_{Block} - MSE}{a}.$$

Missing Values I

- When missing
  - Orthogonality lost
  - Design unbalanced
- Procedures
  - Exact (Regression) approach: Use Type III SS's (general regression significant test)
  - Approximate approach: Estimate missing value
    Choose value to minimize *SSE*
    Take derivative and set equal to zero

Missing Values II

$$
\begin{aligned}
SS_{\mathsf{E}} &= \sum\sum y_{ij}^2 - y_{..}^2/ab - \frac{1}{b}\sum y_{i.}^2 + y_{..}^2/ab - \frac{1}{a}\sum y_{.j}^2 + y_{..}^2/ab \\
&= x^2 - \frac{1}{b}(y'_{i.} + x)^2 - \frac{1}{a}(y'_{.j} + x)^2 + \frac{1}{ab}(y'_{..} + x)^2 + R
\end{aligned}
$$

$$
x = \frac{ay'_{i.} + by'_{.j} - y'_{..}}{(a-1)(b-1)}
$$

- Example: detergent comparison example
  - Suppose $y_{4,2} = 37$ is missing

Missing Values III

- Estimate approach

$$y'_{4.} = 91, \quad y'_{..} = 528, \quad y'_{.2} = 139$$

  Estimate is

$$x = \frac{4(91) + 3(139) - 528}{6} = 42.17$$

  Do analysis but adjust error degrees of freedom
- Estimate: $\hat{\sigma}^2 = 1.097$ (must divide by 5 not 6)
- See missing_block.SAS.
- $F_0 = \frac{71.95/3}{5.49/5} = 21.84$, p-value$= 0.0027$
- Remark: Approximate approach produces a biased mean square for treatment.
  Exact analysis approach is preferred.