Lecture 3

Completely Randomized Designs: Model, Estimation, Inference

STAT 8050 Design and Analysis of Experiments
January 16, 2020

Whitney Huang
Clemson University

Statistical Model

Let Y_{ij} be the random variable that represents the response for the j^{th} experimental unit to treatment i. Also, let $\mu_i = \mathrm{E}(Y_{ij})$ be the mean response for the i^{th} treatment. We have

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i,$$

where ϵ_{ij} is the random variable representing error associated with Y_{ij} with $\mathrm{E}(\epsilon_{ij}) = 0$. This is called a means model.

Statistical Model

Let Y_{ij} be the random variable that represents the response for the j^{th} experimental unit to treatment i. Also, let $\mu_i = \mathrm{E}(Y_{ij})$ be the mean response for the i^{th} treatment. We have

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i,$$

where ϵ_{ij} is the random variable representing error associated with Y_{ij} with $\mathrm{E}(\epsilon_{ij}) = 0$. This is called a means model.

Alternatively, we could let $\mu_i = \mu + \alpha_i$, which leads to

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i.$$

This is called an effects model

Distributional Assumption on Error

In both the means model and the effects model. We further assume

$$\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2),$$

and ϵ_{ij}'s are independent to each other.

This yields

$$Y_{ij} \sim \mathrm{N}(\mu + \alpha_i, \sigma^2) \qquad \text{Effects Model}$$

$$Y_{ij} \sim \mathrm{N}(\mu_i, \sigma^2) \qquad \text{Means Model}$$

Distributional Assumption on Error

In both the means model and the effects model. We further assume

$$\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2),$$

and ϵ_{ij}'s are independent to each other.

This yields

$$Y_{ij} \sim \mathrm{N}(\mu + \alpha_i, \sigma^2) \qquad \text{Effects Model}$$
$$Y_{ij} \sim \mathrm{N}(\mu_i, \sigma^2) \qquad \text{Means Model}$$

Note: We make the common variance assumption here

Effects Model Properties

The model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i,$$

- is overparameterized

Effects Model Properties

The model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i,$$

- is overparameterized

Effects Model Properties

The model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i,$$

- is overparameterized
- is nonidentifiable

Effects Model Properties

The model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i,$$

- is overparameterized
- is nonidentifiable

Effects Model Properties

The model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i,$$

- is overparameterized
- is nonidentifiable

Example

Suppose $g = 2$, then we have to esimtate μ, α_1, and α_2.

$$\mu = 10, \alpha_1 = -1, \alpha_2 = 1,$$
$$\text{and } \mu = 11, \alpha_1 = -2, \alpha_2 = 0.$$

Effects Model Properties

CLEMSON UNIVERSITY

The model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, g, \quad j = 1, \cdots, n_i,$$

- is overparameterized
- is nonidentifiable

Example

Suppose $g = 2$, then we have to esimtate μ, α_1, and α_2.

$$\mu = 10, \alpha_1 = -1, \alpha_2 = 1,$$
$$\text{and } \mu = 11, \alpha_1 = -2, \alpha_2 = 0.$$

$\Rightarrow$ each yield $Y_{1j} \sim \mathrm{N}(9, \sigma^2)$ and $Y_{2j} \sim N(11, \sigma^2)$

Dot Notation

- $\bar{Y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$ – i^{th} treatment mean

Dot Notation

- $\bar{Y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$ – i^{th} treatment mean

Dot Notation

- $\bar{Y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$ – i^{th} treatment mean
- $Y_{i\cdot} = \sum_{j=1}^{n_i} Y_{ij}$ – Total for i^{th} treatment

Dot Notation

- $\bar{Y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$ – i^{th} treatment mean
- $Y_{i\cdot} = \sum_{j=1}^{n_i} Y_{ij}$ – Total for i^{th} treatment

Dot Notation

- $\bar{Y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$ – i^{th} treatment mean
- $Y_{i\cdot} = \sum_{j=1}^{n_i} Y_{ij}$ – Total for i^{th} treatment
- $Y_{\cdot\cdot} = \sum_{i=1}^{g}\sum_{j=1}^{n_i} Y_{ij} = \sum_{i=1}^{j} Y_{i\cdot}$ – Total of all observations

Dot Notation

- $\bar{Y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$ – i^{th} treatment mean
- $Y_{i\cdot} = \sum_{j=1}^{n_i} Y_{ij}$ – Total for i^{th} treatment
- $Y_{\cdot\cdot} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} Y_{ij} = \sum_{i=1}^{j} Y_{i\cdot}$ – Total of all observations

Dot Notation

- $\bar{Y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$ – i^{th} treatment mean
- $Y_{i\cdot} = \sum_{j=1}^{n_i} Y_{ij}$ – Total for i^{th} treatment
- $Y_{\cdot\cdot} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} Y_{ij} = \sum_{i=1}^{j} Y_{i\cdot}$ – Total of all observations
- $\bar{Y}_{\cdot\cdot} = \frac{1}{N} \sum_{i=1}^{g} \sum_{j=1}^{n_i} Y_{ij}$ – Grand mean of all observations where $N = \sum_{i=1}^{g} n_i$

Least Squares Estimation

To estimate $\mu, \alpha_1, \cdots, \alpha_g$, we find the values for these parameters that minimize

$$\sum_{i=1}^{g} \sum_{j=1}^{n_i} e_{ij}^2 = \sum_{i=1}^{g} \sum_{j=1}^{n_i} \left(Y_{ij} - (\mu + \alpha_i) \right)^2 .$$

Least Squares Estimation

To estimate $\mu, \alpha_1, \cdots, \alpha_g$, we find the values for these parameters that minimize

$$\sum_{i=1}^{g} \sum_{j=1}^{n_i} e_{ij}^2 = \sum_{i=1}^{g} \sum_{j=1}^{n_i} \left(Y_{ij} - (\mu + \alpha_i)\right)^2 .$$

To obtain the estimates, we have a system of $g+1$ equations with $g+1$ unknowns. Unfortunately, we only have g treatment means that can be used to solve this system of equations $\Rightarrow$ no unique solution exists for $\hat{\mu}, \hat{\alpha}_1, \cdots, \hat{\alpha}_g$

Least Squares Estimation

To estimate $\mu, \alpha_1, \cdots, \alpha_g$, we find the values for these parameters that minimize

$$\sum_{i=1}^{g}\sum_{j=1}^{n_i} e_{ij}^2 = \sum_{i=1}^{g}\sum_{j=1}^{n_i} \left(Y_{ij} - (\mu + \alpha_i)\right)^2 .$$

To obtain the estimates, we have a system of $g + 1$ equations with $g + 1$ unknowns. Unfortunately, we only have g treatment means that can be used to solve this system of equations $\Rightarrow$ no unique solution exists for $\hat{\mu}, \hat{\alpha}_1, \cdots, \hat{\alpha}_g$

Typically constraints are used to obtain solutions and hence estimators.

Note: Different software uses different constraints

Constraints

Constraint	$\hat{\mu}$	$\hat{\alpha}_i$	$\hat{\mu} + \hat{\alpha}_i$	$\hat{\alpha}_i - \hat{\alpha}_{i'}$
$\hat{\alpha}_g = 0$				
$\hat{\mu} = 0$				
$\sum_{i=1}^{g} n_i\hat{\alpha}_i = 0$				

Constraints

Constraint	$\hat{\mu}$	$\hat{\alpha}_i$	$\hat{\mu} + \hat{\alpha}_i$	$\hat{\alpha}_i - \hat{\alpha}_{i'}$
$\hat{\alpha}_g = 0$				
$\hat{\mu} = 0$				
$\sum_{i=1}^{g} n_i \hat{\alpha}_i = 0$				

$\hat{\mu}$ and $\hat{\alpha}_i$ depends upon the constraint used. $\hat{\mu} + \hat{\alpha}_i$ and $\hat{\alpha}_i - \hat{\alpha}_{i'}$ are invariant to the constraint used.

Constraints

Constraint	$\hat{\mu}$	$\hat{\alpha}_i$	$\hat{\mu} + \hat{\alpha}_i$	$\hat{\alpha}_i - \hat{\alpha}_{i'}$
$\hat{\alpha}_g = 0$				
$\hat{\mu} = 0$				
$\sum_{i=1}^{g} n_i\hat{\alpha}_i = 0$				

$\hat{\mu}$ and $\hat{\alpha}_i$ depends upon the constraint used. $\hat{\mu} + \hat{\alpha}_i$ and $\hat{\alpha}_i - \hat{\alpha}_{i'}$ are invariant to the constraint used.

Note: If we use the **means model**, $\hat{\mu}_i = \bar{Y}_{i\cdot}$, and we do not have these issues here, but we will have other issues later on.

Analysis of Variance (ANOVA)

The total variation is represented by total sum of squares SS_T:

$$\text{SS}_T = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2$$

This quantity can be decomposed to variation between treatments (SS_{TRT}) and variation within treatment (SS_E):

$$\text{SS}_T = \sum_{i=1}^{j} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{..})^2 = \underbrace{\sum_{i=1}^{g} n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2}_{\text{SS}_{TRT}} + \underbrace{\sum_{i=1}^{g} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2}_{\text{SS}_E}$$

Equivalent Computational Formulae

$$\text{SS}_T = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{\cdot\cdot}\right)^2 = \sum_{i=1}^{g}\sum_{j=1}^{n_i} Y_{ij}^2 - \frac{Y_{\cdot\cdot}^2}{N}$$

Equivalent Computational Formulae

$$\text{SS}_T = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{\cdot\cdot}\right)^2 = \sum_{i=1}^{g}\sum_{j=1}^{n_i} Y_{ij}^2 - \frac{Y_{\cdot\cdot}^2}{N}$$

Equivalent Computational Formulae

$$\mathrm{SS}_T = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{\cdot\cdot}\right)^2 = \sum_{i=1}^{g}\sum_{j=1}^{n_i} Y_{ij}^2 - \frac{Y_{\cdot\cdot}^2}{N}$$

$$\mathrm{SS}_{TRT} = \sum_{i=1}^{g} n_i\left(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}\right)^2 = \sum_{i=1}^{g}\sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^{g}\frac{Y_{i\cdot}^2}{n_i}$$

CLEMSON UNIVERSITY

Equivalent Computational Formulae

$$\mathsf{SS}_T = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{\cdot\cdot}\right)^2 = \sum_{i=1}^{g}\sum_{j=1}^{n_i} Y_{ij}^2 - \frac{Y_{\cdot\cdot}^2}{N}$$

$$\mathsf{SS}_{TRT} = \sum_{i=1}^{g} n_i\left(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}\right)^2 = \sum_{i=1}^{g}\sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^{g}\frac{Y_{i\cdot}^2}{n_i}$$

Equivalent Computational Formulae

$$\text{SS}_T = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{\cdot\cdot}\right)^2 = \sum_{i=1}^{g}\sum_{j=1}^{n_i} Y_{ij}^2 - \frac{Y_{\cdot\cdot}^2}{N}$$

$$\text{SS}_{TRT} = \sum_{i=1}^{g} n_i \left(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}\right)^2 = \sum_{i=1}^{g}\sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^{g} \frac{Y_{i\cdot}^2}{n_i}$$

$$\text{SS}_E = \sum_{i=1}^{g}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2 = \sum_{i=1}^{g} \frac{Y_{i\cdot}^2}{n_i} - \frac{Y_{\cdot\cdot}^2}{N}$$

Mean Squares

Dividing mean squares by their associated degrees of freedom yield "variance-like" quantities called mean squares.

Mean Squares

Dividing mean squares by their associated degrees of freedom yield "variance-like" quantities called mean squares.

- We have $N - 1$ total degrees of freedom (Why?)

Mean Squares

Dividing mean squares by their associated degrees of freedom yield “variance-like” quantities called mean squares.

- We have $N - 1$ total degrees of freedom (Why?)

Mean Squares

Dividing mean squares by their associated degrees of freedom yield "variance-like" quantities called mean squares.

- We have $N-1$ total degrees of freedom (Why?)
- We have ___ treatment degrees of freedom ⇒

Mean Squares

Dividing mean squares by their associated degrees of freedom yield “variance-like” quantities called mean squares.

- We have $N - 1$ total degrees of freedom (Why?)
- We have ___ treatment degrees of freedom ⇒

Mean Squares

Dividing mean squares by their associated degrees of freedom yield “variance-like” quantities called mean squares.

- We have $N-1$ total degrees of freedom (Why?)
- We have ___ treatment degrees of freedom ⇒

$$\text{MS}_{TRT} = \frac{\text{SS}_{TRT}}{g-1}$$

- We have ___ error degrees of freedom ⇒

Mean Squares

Dividing mean squares by their associated degrees of freedom yield “variance-like” quantities called mean squares.

- We have $N - 1$ total degrees of freedom (Why?)

- We have ___ treatment degrees of freedom ⇒

$$\text{MS}_{TRT} = \frac{\text{SS}_{TRT}}{g-1}$$

- We have ___ error degrees of freedom ⇒

Mean Squares

Dividing mean squares by their associated degrees of freedom yield "variance-like" quantities called mean squares.

- We have $N - 1$ total degrees of freedom (Why?)
- We have ___ treatment degrees of freedom $\Rightarrow$

$$\text{MS}_{TRT} = \frac{\text{SS}_{TRT}}{g - 1}$$

- We have ___ error degrees of freedom $\Rightarrow$

$$\text{MS}_{E} = \frac{\text{SS}_{E}}{N - g}$$

Mean Squares Cont'd

Note that

$$\text{MS}_E = \frac{1}{N-g} \underbrace{\sum_{i=1}^{g} (n_i - 1) s_i^2}_{\text{SS}_E}$$

provides an **unbiased** estimator of σ^2 **regardless of whether the treatment population means differ or not.**

Mean Squares Cont'd

Note that

$$\text{MS}_E = \frac{1}{N-g} \underbrace{\sum_{i=1}^{g} (n_i - 1) s_i^2}_{\text{SS}_E}$$

provides an **unbiased** estimator of σ^2 **regardless of whether the treatment population means differ or not.**

Also, it can be shown that

$$\text{MS}_{TRT} = \frac{1}{g-1} \underbrace{\sum_{i=1}^{g} n_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2}_{\text{SS}_{TRT}}$$

is an **unbiased** estimator of σ^2 **if all treatment population means are equal.**

Mean Squares Cont'd

If

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_g = 0$$

is true, then MS_{TRT} and MS_E will be "similar". Otherwise, they will be different. We can show that

$$\mathrm{E}(\mathrm{MS}_{TRT}) = \sigma^2 + \sum_{i=1}^{g} n_i \alpha_i^2/(g-1) \geq \sigma^2 = \mathrm{E}(\mathrm{MS}_E)$$

$\Rightarrow$ if H_0 is false, MS_{TRT} will tend to be larger than MS_E.

ANOVA Table

Source	df	SS	MS	EMS
Treatment	$g-1$	SS_{TRT}	$\text{MS}_{TRT} = \frac{\text{SS}_{TRT}}{g-1}$	$\sigma^2 + \frac{\sum_{i=1}^{g} n_i \alpha_i^2}{g-1}$
Error	$N-g$	SS_E	$\text{MS}_E = \frac{\text{SS}_E}{N-g}$	σ^2
Total	$N-1$	SS_T		

Testing for treatment effects

$$H_0 : \alpha_i = 0 \quad \text{for all } i$$
$$H_a : \alpha_i \neq 0 \quad \text{for some } i$$

Test statistics: $F = \frac{\text{MS}_{TRT}}{\text{MS}_E}$. Under H_0, the test statitic follows an F-distribution with $g-1$ and $N-g$ degrees of freedom

F-Test

Reject H_0 if

$$F_{obs} > F_{g-1,N-g;\alpha}$$

for an α-level test, $F_{g-1,N-g;\alpha}$ is the $100 \times (1-\alpha)\%$ percentile of a central F-distribution with $g-1$ and $N-g$ degrees of freedom.

F-Test

Reject H_0 if

$$F_{obs} > F_{g-1,N-g;\alpha}$$

for an α-level test, $F_{g-1,N-g;\alpha}$ is the $100 \times (1-\alpha)\%$ percentile of a central F-distribution with $g-1$ and $N-g$ degrees of freedom.

P-value

The P-value of the F-test is the probability of obtaining F at least as extreme as F_{obs}, that is, $\mathrm{P}(F > F_{obs})$.

F-Test

Reject H_0 if

$$F_{obs} > F_{g-1,N-g;\alpha}$$

for an α-level test, $F_{g-1,N-g;\alpha}$ is the $100 \times (1-\alpha)\%$ percentile of a central F-distribution with $g-1$ and $N-g$ degrees of freedom.

P-value
The P-value of the F-test is the probability of obtaining F at least as extreme as F_{obs}, that is, $\mathrm{P}(F > F_{obs})$.

We reject H_0 if P-value $< \alpha$.

F Distribution and the F-Test

Consider the observed F test statistic: $F_{obs} = \frac{\text{MS}_{TRT}}{\text{MS}_{E}}$

F Distribution and the F-Test

Consider the observed F test statistic: $F_{obs} = \frac{\text{MS}_{TRT}}{\text{MS}_E}$

- Should be “near” 1 if the treatment means are equal

F Distribution and the F-Test

Consider the observed F test statistic: $F_{obs} = \frac{\text{MS}_{TRT}}{\text{MS}_E}$

- Should be "near" 1 if the treatment means are equal

F Distribution and the F-Test

Consider the observed F test statistic: $F_{obs} = \frac{\text{MS}_{TRT}}{\text{MS}_E}$

- Should be “near” 1 if the treatment means are equal
- Should be “larger than” 1 if treatment means are not equal

F Distribution and the F-Test

Consider the observed F test statistic: $F_{obs} = \frac{\text{MS}_{TRT}}{\text{MS}_E}$

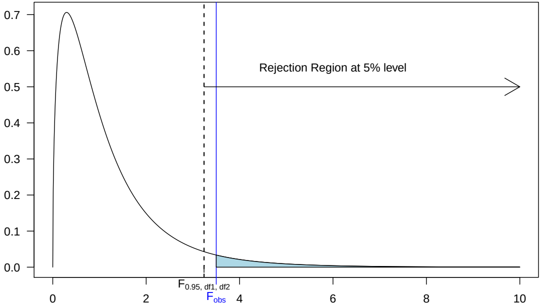
- Should be "near" 1 if the treatment means are equal
- Should be "larger than" 1 if treatment means are not equal

F Distribution and the F-Test

Consider the observed F test statistic: $F_{obs} = \frac{\text{MS}_{TRT}}{\text{MS}_E}$

- Should be "near" 1 if the treatment means are equal
- Should be "larger than" 1 if treatment means are not equal

$\Rightarrow$ We use the null distribution $F \sim F_{df_1=g-1, df_2=N-g}$ to quantify if F_{obs} is large enough to reject H_0



CLEMSON UNIVERSITY

CLEMSON UNIVERSITY

Example

An experiment was conducted to determine if experience has an effect on the time it takes for mice to run a maze. Four treatment groups, consisting of mice having been trained on the maze one, two, three and four times were run through the maze and their times recorded. Three mice were originally assigned to each group, but it was discovered that some lab assistants, in an attempt to win a bet, gave one mouse a stimulant and another mouse a sedative. These mice were removed from the analysis.

Training runs	1	2	3	4
Times	11, 9	7,8,9	6,5,7	5,3
$y_{i\cdot}$	20	24	18	8
n_i	2	3	3	2
s_i^2				

Example Cont'd

Training runs	1	2	3	4
Times	11, 9	7,8,9	6,5,7	5,3
$y_{i\cdot}$	20	24	18	8
n_i	2	3	3	2
s_i^2				

- Write down the model.

Example Cont'd

Training runs	1	2	3	4
Times	11, 9	7,8,9	6,5,7	5,3
$y_{i\cdot}$	20	24	18	8
n_i	2	3	3	2
s_i^2				

- Write down the model.

Example Cont'd

Training runs	1	2	3	4
Times	11, 9	7,8,9	6,5,7	5,3
$y_{i\cdot}$	20	24	18	8
n_i	2	3	3	2
s_i^2				

- Write down the model.
- Fill out the ANOVA table and test whether the time to run the maze is affected by training. Use a significant level of .05.