Lecture 12

Multisample inference: Analysis of Variance

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Public Health

Learning Objectives

1. Describe Analysis of Variance (ANOVA)
2. Explain the Rationale of ANOVA
3. Compare Experimental Designs
4. Test the Equality of 2 or More Means
   a. Completely Randomized Design
   b. Randomized Block Design
   c. Factorial Design
5. Use of Computer Program

wilopo@yahoo.com

Analysis of Variance

A analysis of variance **is a technique that partitions the total sum of squares of deviations of the observations about their mean into portions associated with independent variables in the experiment and a portion associated with error**

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Public Health

Analysis of Variance

The ANOVA table was previously discussed in the context of regression models with quantitative independent variables, in this chapter the focus will be on nominal independent variables (factors)

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Public Health

Analysis of Variance

A ***factor*** **refers to a categorical quantity under examination in an experiment as a possible cause of variation in the response variable.**

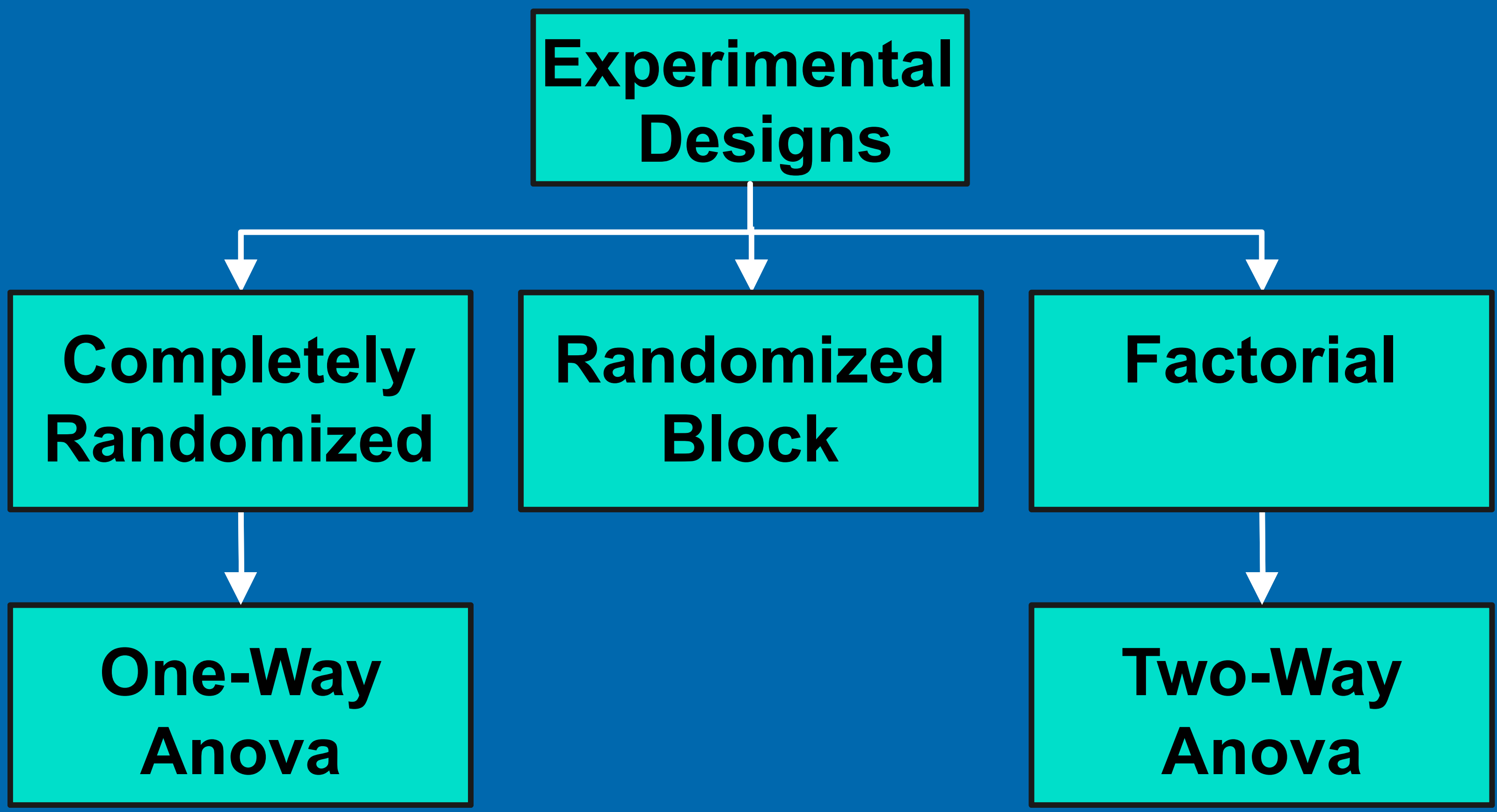
sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Public Health

Analysis of Variance

Levels **refer to the categories, measurements, or strata of a factor of interest in the experiment.**

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Public Health

Types of Experimental Designs



Completely Randomized Design

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Public Health

Completely Randomized Design

1. Experimental Units (Subjects) Are Assigned Randomly to Treatments
   - Subjects are Assumed Homogeneous

2. One Factor or Independent Variable
   - 2 or More Treatment Levels or groups
3. Analyzed by One-Way ANOVA

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

One-Way ANOVA F-Test

1. **Tests the Equality of 2 or More (*p*) Population Means**

2. **Variables**
   - **One Nominal Independent Variable**
   - **One Continuous Dependent Variable**

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

One-Way ANOVA F-Test Assumptions

1. **Randomness & Independence of Errors**
2. **Normality**
   - Populations (for each condition) are Normally Distributed
3. **Homogeneity of Variance**
   - Populations (for each condition) have Equal Variances

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

One-Way ANOVA F-Test Hypotheses

H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_p$

- All Population Means are Equal
- No Treatment Effect

H_a: Not All μ_j Are Equal

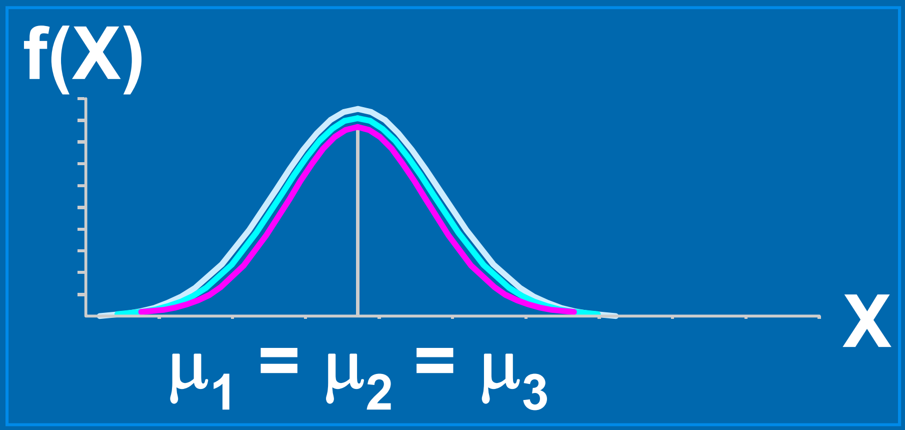
- At Least 1 Pop. Mean is Different
- Treatment Effect
- NOT $\mu_1 \neq \mu_2 \neq \ldots \neq \mu_p$

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

One-Way ANOVA F-Test Hypotheses

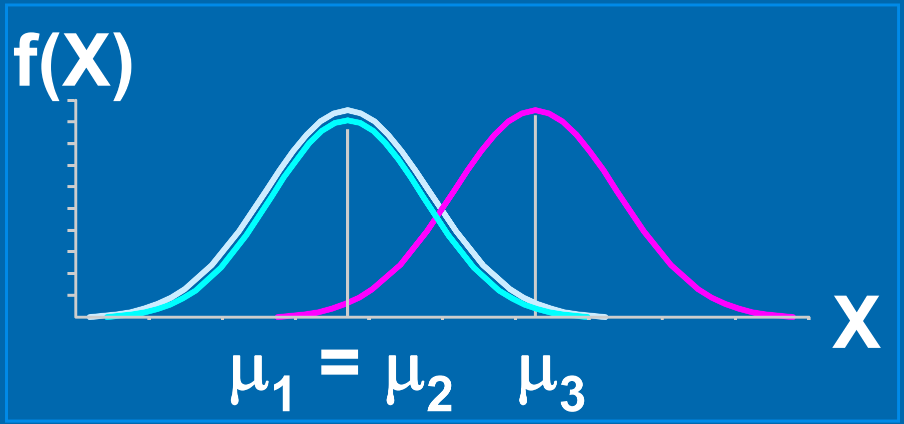
H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_p$

- All Population Means are Equal
- No Treatment Effect



H_a: Not All μ_j Are Equal

- At Least 1 Pop. Mean is Different
- Treatment Effect
- NOT $\mu_1 = \mu_2 = \ldots = \mu_p$
- Or $\mu_i \neq \mu_j$ for some *i*, *j*.



One-Way ANOVA
Basic Idea

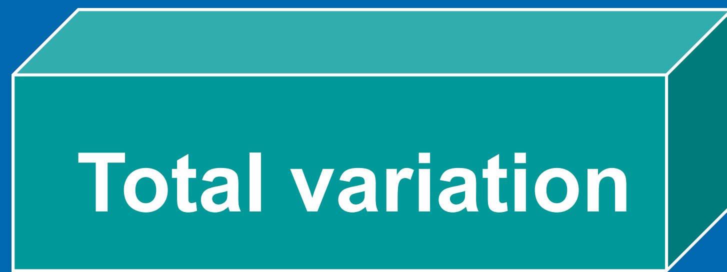
1. **Compares 2 Types of Variation to Test Equality of Means**
2. **If Treatment Variation Is Significantly Greater Than Random Variation then Means Are Not Equal**
3. **Variation Measures Are Obtained by 'Partitioning' Total Variation**

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

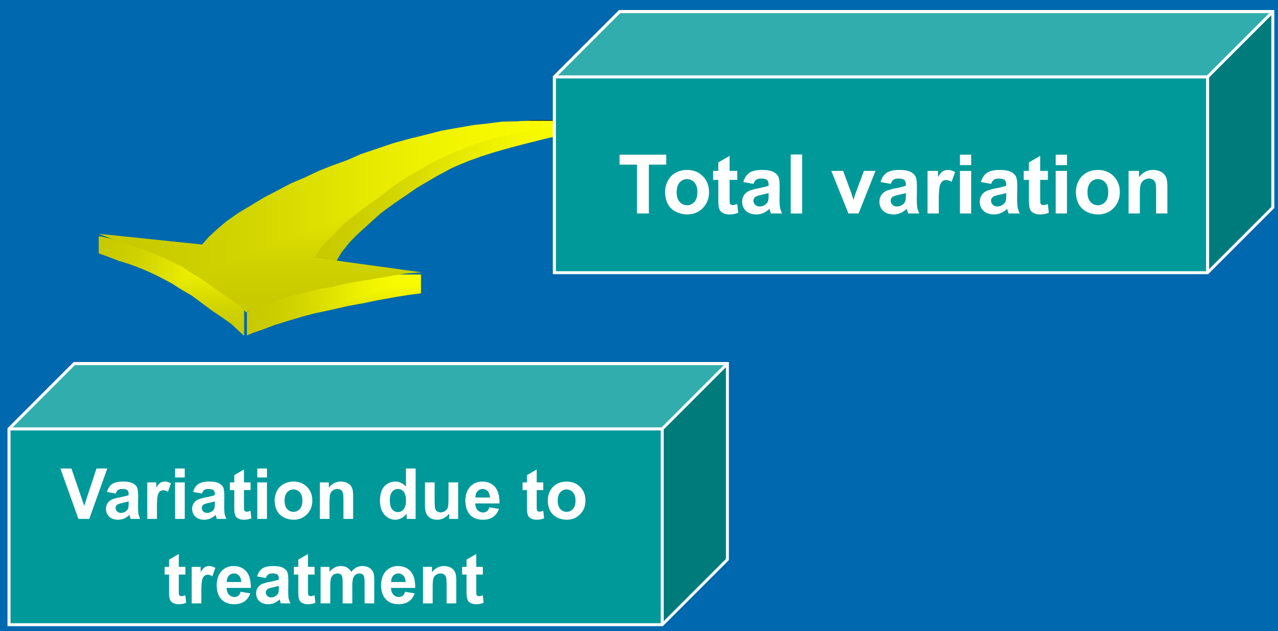
One-Way ANOVA
Partitions Total Variation

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

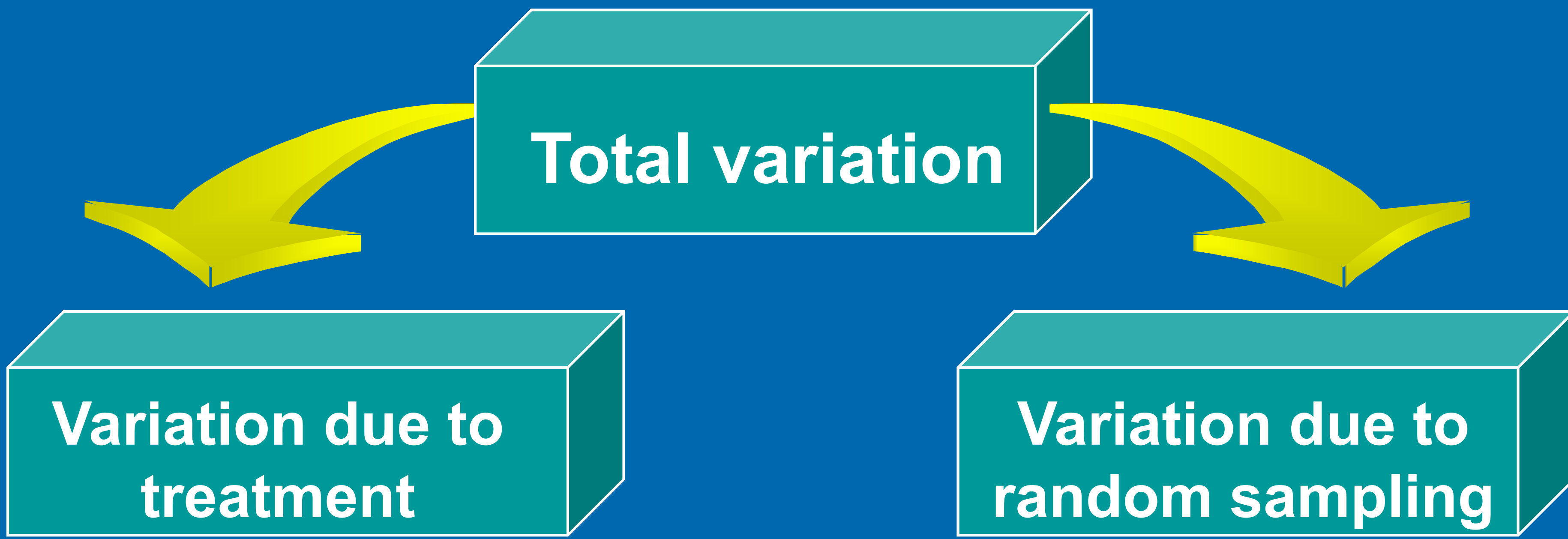
One-Way ANOVA Partitions Total Variation



One-Way ANOVA Partitions Total Variation

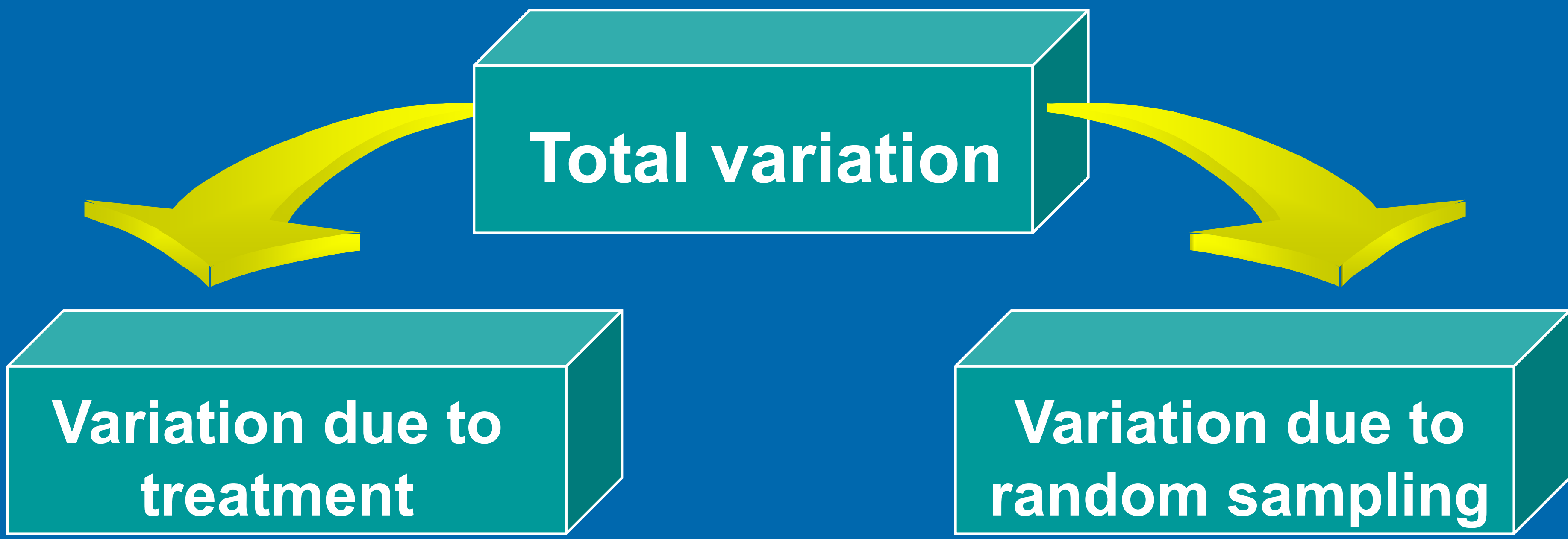


One-Way ANOVA
Partitions Total Variation



wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

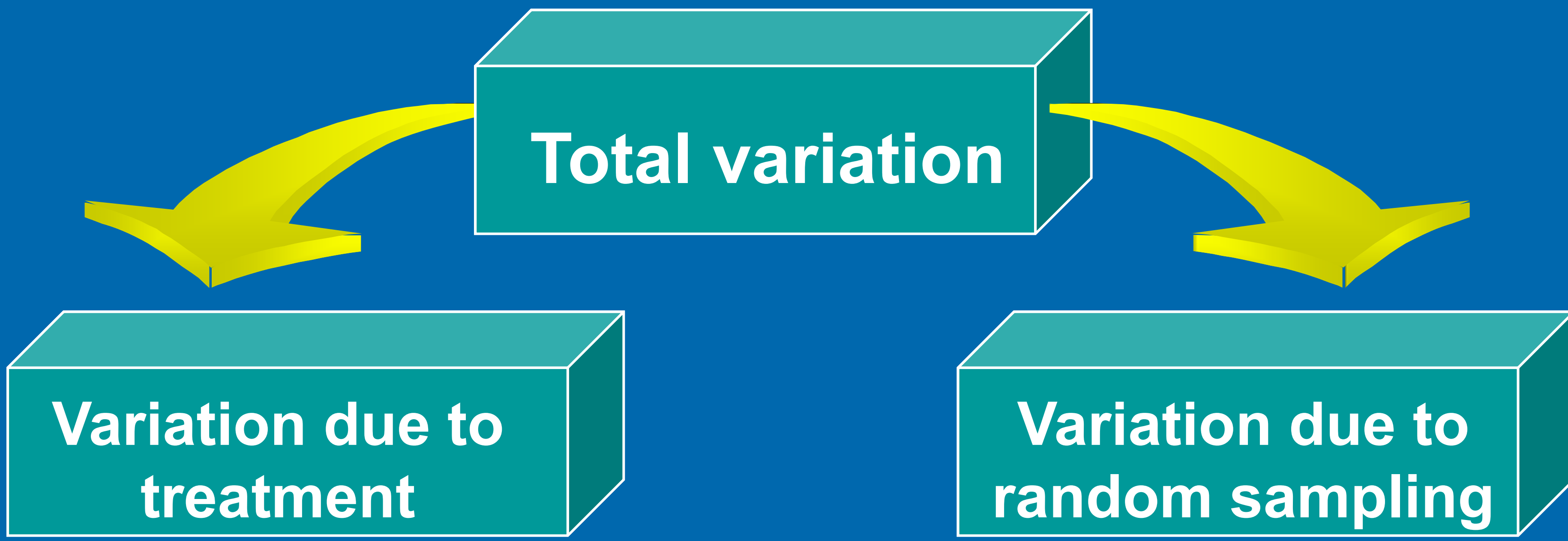
One-Way ANOVA Partitions Total Variation



Sum of Squares Among
Sum of Squares Between
Sum of Squares Treatment
Among Groups Variation

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

One-Way ANOVA Partitions Total Variation



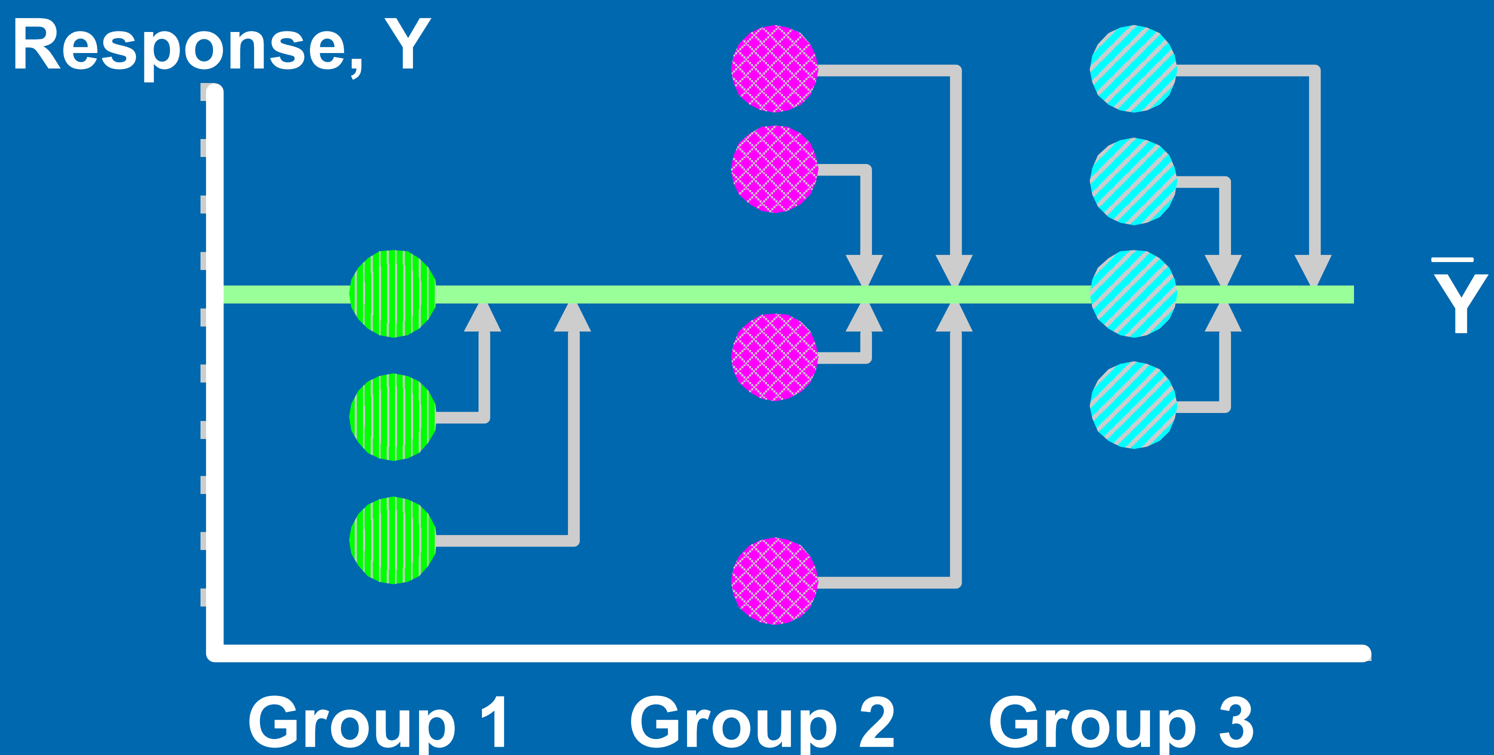
Sum of Squares Among
Sum of Squares Between
Sum of Squares Treatment (SST)
Among Groups Variation

Sum of Squares Within
Sum of Squares Error (SSE)
Within Groups Variation

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Total Variation

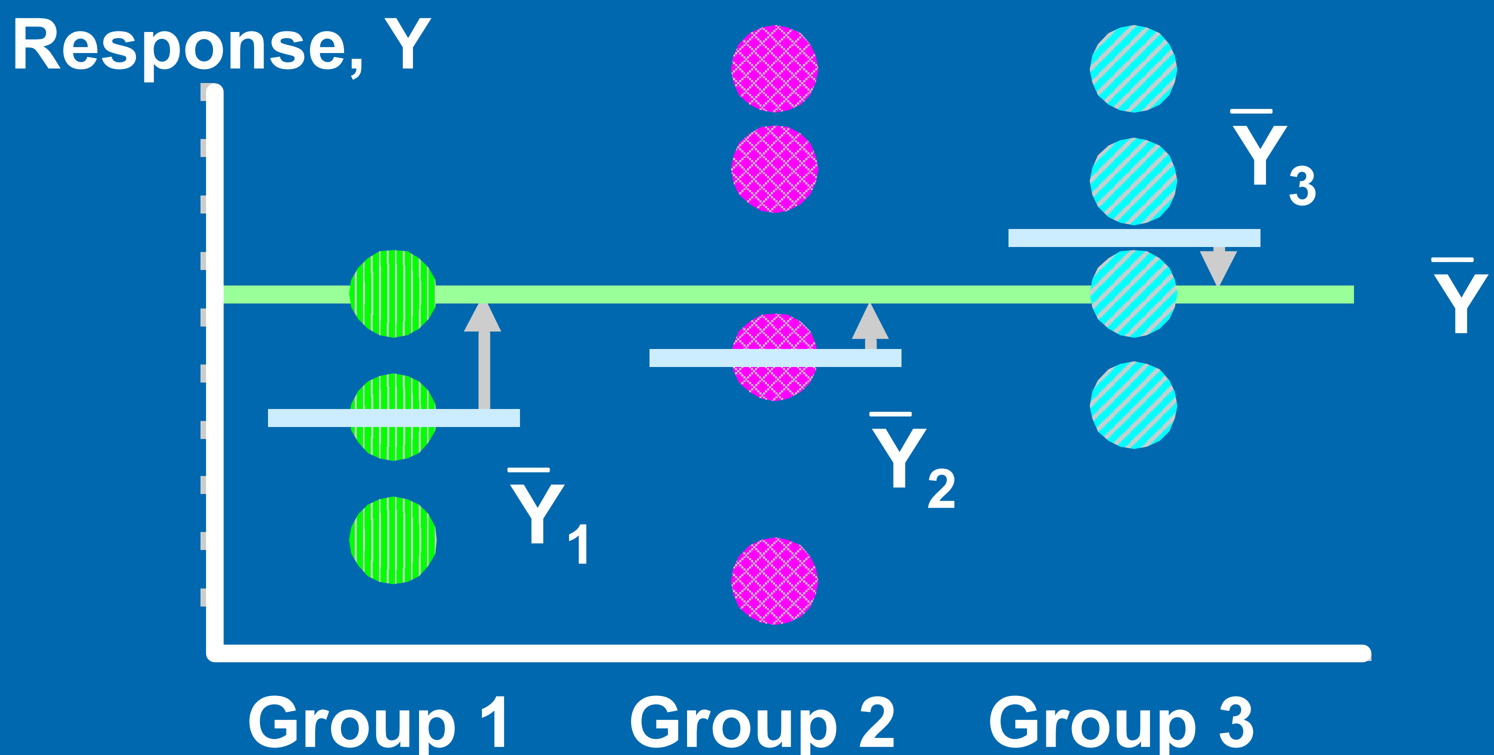
$$SS(Total) = (Y_{11} - \overline{Y})^2 + (Y_{21} - \overline{Y})^2 + \ldots + (Y_{ij} - \overline{Y})^2$$



wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Treatment Variation

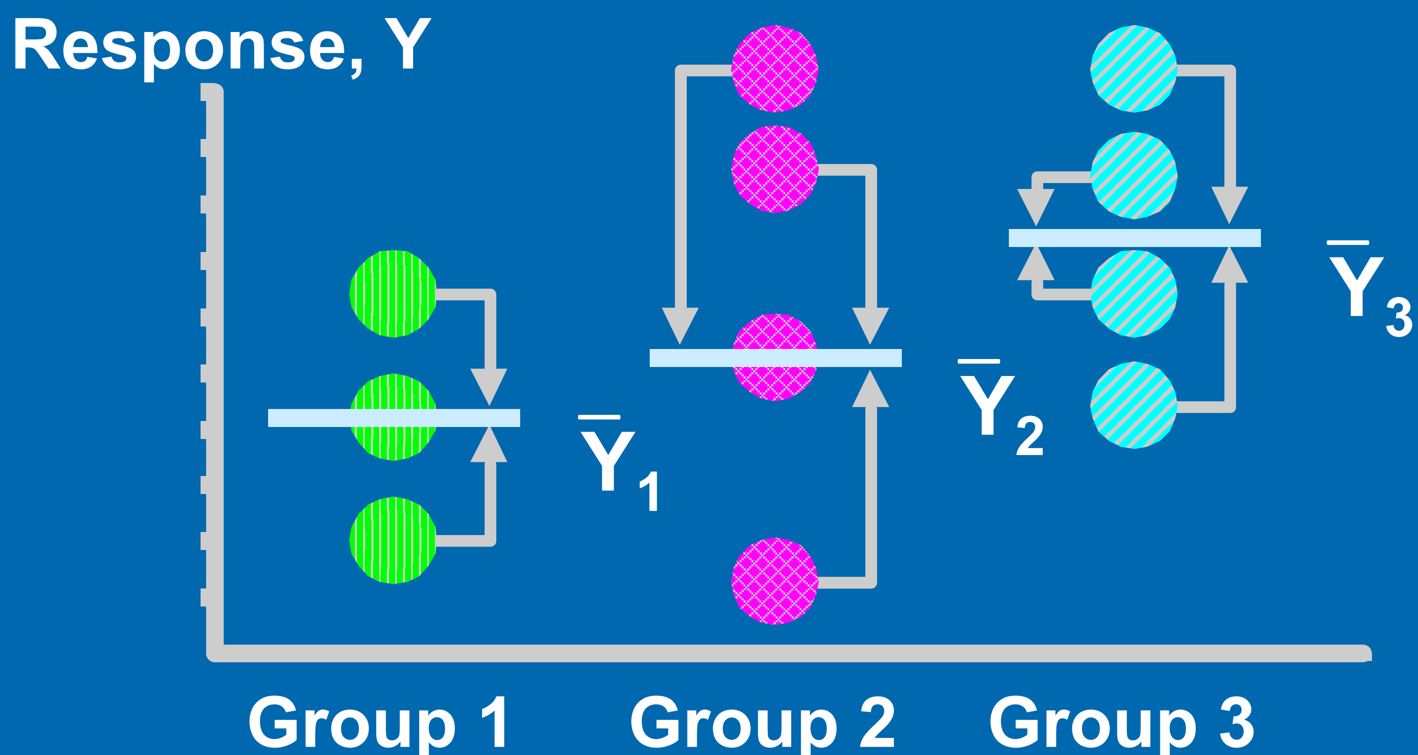
$$SST = n_1\left(\bar{Y}_1 - \bar{Y}\right)^2 + n_2\left(\bar{Y}_2 - \bar{Y}\right)^2 + \ldots + n_p\left(\bar{Y}_p - \bar{Y}\right)^2$$



wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Random (Error) Variation

$$SSE = \left(Y_{11} - \bar{Y}_1\right)^2 + \left(Y_{21} - \bar{Y}_1\right)^2 + \ldots + \left(Y_{pj} - \bar{Y}_p\right)^2$$



wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

One-Way ANOVA F-Test
Test Statistic

- **1. Test Statistic**
  - ***F = MST / MSE*** $= \dfrac{STT/(p-1)}{SSE/(n-p)}$
    - *MST* Is Mean Square for Treatment
    - *MSE* Is Mean Square for Error
- **2. Degrees of Freedom**
  - $\nu_1 = p - 1$
  - $\nu_2 = n - p$
    - *p* = # Populations, Groups, or Levels
    - *n* = Total Sample Size

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

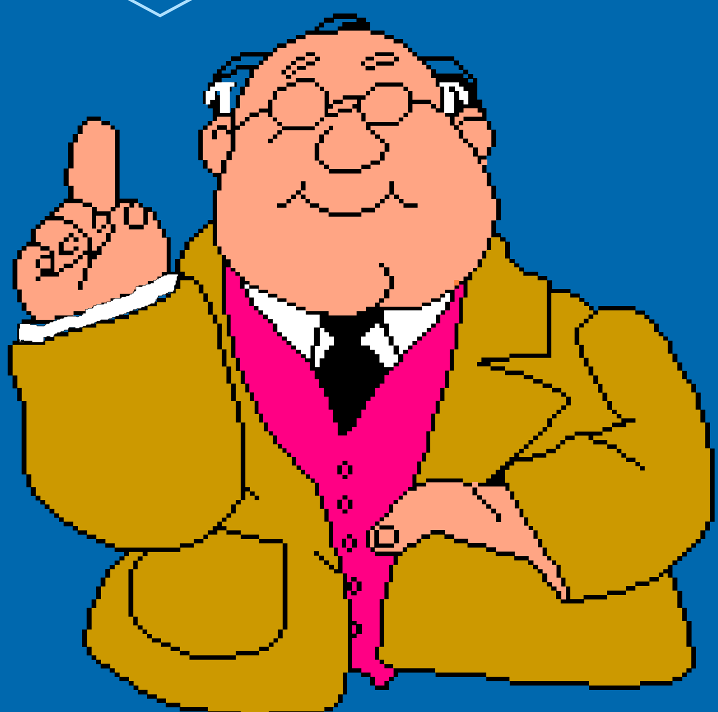
One-Way ANOVA Summary Table

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square (Variance)	F
Treatment	p - 1	SST	MST = SST/(p - 1)	$\frac{MST}{MSE}$
Error	n - p	SSE	MSE = SSE/(n - p)	
Total	n - 1	SS(Total) = SST+SSE		

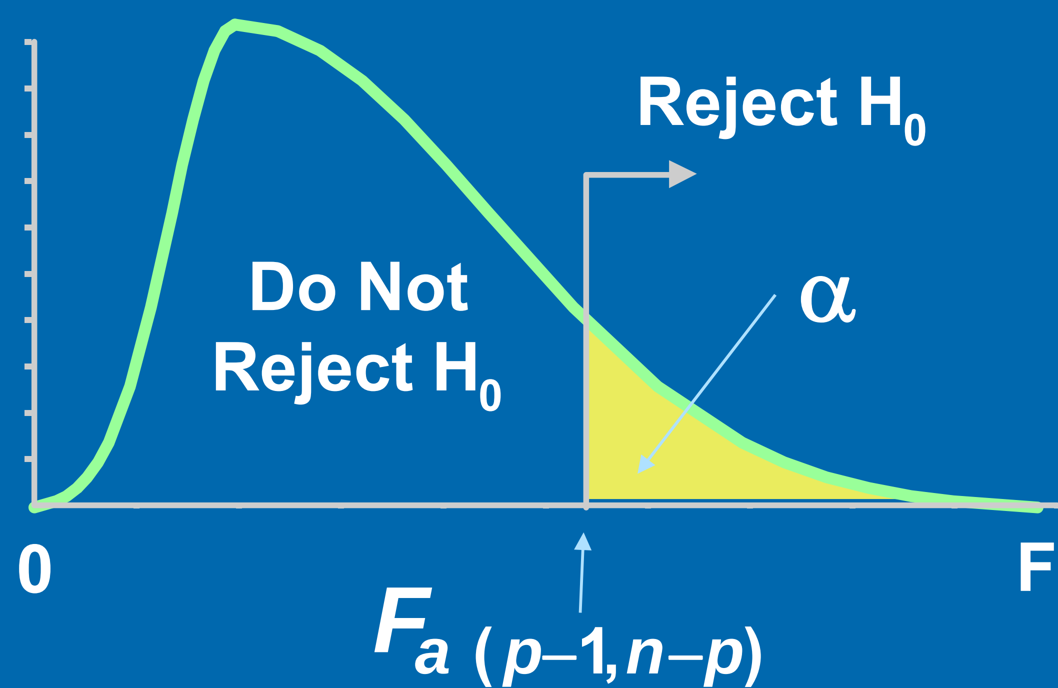
wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

One-Way ANOVA F-Test Critical Value

If means are equal,
***F* = *MST* / *MSE* ≈ 1.**
Only reject large *F*!



© 1984-1994 T/Maker Co.



Always One-Tail!

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

One-Way ANOVA F-Test Example

As a vet epidemiologist you want to see if 3 food supplements have different mean milk yields. You assign 15 cows, 5 per food supplement.

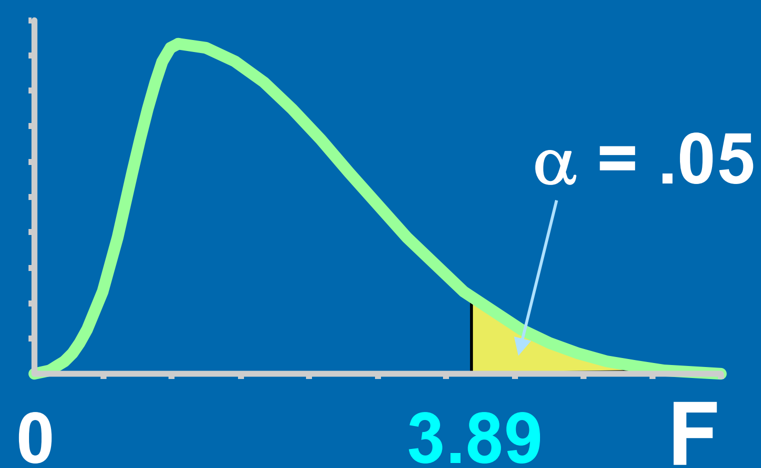
Question: At the .05 level, is there a difference in **mean** yields?

Food1	Food2	Food3
25.40	23.40	20.00
26.31	21.80	22.20
24.10	23.50	19.75
23.74	22.75	20.60
25.10	21.60	20.40

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

One-Way ANOVA F-Test Solution

- H_0: $\mu_1 = \mu_2 = \mu_3$
- H_a: Not All Equal
- $\alpha = .05$
- $\nu_1 = 2$ $\nu_2 = 12$
- Critical Value(s):



Test Statistic:

$$F = \frac{MST}{MSE} = \frac{23.5820}{.9211} = 25.6$$

Decision:

Reject at $\alpha = .05$

Conclusion:

There Is Evidence Pop. Means Are Different

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Summary Table Solution

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square (Variance)	F
Food	3 - 1 = 2	47.1640	23.5820	25.60
Error	15 - 3 = 12	11.0532	.9211	
Total	15 - 1 = 14	58.2172		

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

SAS CODES FOR ANOVA

```
Data Anova;
input group$ milk @@;
cards;
food1 25.40     food2 23.40     food3 20.00
food1 26.31     food2 21.80     food3 22.20
food1 24.10     food2 23.50     food3 19.75
food1 23.74     food2 22.75     food3 20.60
food1 25.10     food2 21.60     food3 20.40
;
run;
```

```
proc anova; /* or PROC GLM */
class group;
model milk=group;
run;
```


wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

OUTPUT - ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	47.16400000	23.58200000	25.60	<.0001
Error	12	11.05320000	0.92110000		
Corrected Total	14	58.21720000			

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Pair-wise comparisons

- Needed when the overall F test is rejected
- Can be done without adjustment of type I error if other comparisons were planned in advance (least significant difference - LSD method)
- Type I error needs to be adjusted if other comparisons were not planned in advance (Bonferroni's and scheffe's methods)

wilopo@yahoo.com

Fisher's Least Significant Difference (LSD) Test

To compare level 1 and level 2

$$t = \left(\bar{y}_1 - \bar{y}_2\right) \Big/ \sqrt{MSE\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Compare this to $t_{\alpha/2}$ = Upper-tailed value or $-t_{\alpha/2}$ *lower*-tailed from Student's t-distribution for $\alpha/2$ and (n - p) degrees of freedom

MSE = Mean square within from ANOVA table

n = Number of subjects

p = Number of levels

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Public Health

Bonferroni's method

To compare level 1 and level 2

$$t = (\bar{y}_1 - \bar{y}_2) \Big/ \sqrt{MSE\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Adjust the significance level α by taking the new significance level α*

$$\alpha^* = \alpha / \binom{p}{2}$$

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Public Health

SAS CODES FOR multiple comparisons

```
proc anova;
class group;
model milk=group;
means group/ lsd bon;
run;
```


wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

SAS OUTPUT - LSD

t Tests (LSD) for milk

NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	12
Error Mean Square	0.9211
Critical Value of t	2.17881 = $t_{.975,12}$
Least Significant Difference	1.3225

Means with the same letter are not significantly different.

t Grouping	Mean	N	group
A	24.9300	5	food1
B	22.6100	5	food2
C	20.5900	5	food3

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

SAS OUTPUT - Bonferroni

Bonferroni (Dunn) t Tests for milk

NOTE: This test controls the Type I experimentwise error rate

Alpha	0.05
Error Degrees of Freedom	12
Error Mean Square	0.9211
Critical Value of t	$2.77947 = t_{1-0.05/3/2,12}$
Minimum Significant Difference	1.6871

Means with the same letter are not significantly different.

Bon Grouping	Mean	N	group
A	24.9300	5	food1
B	22.6100	5	food2
C	20.5900	5	food3

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Randomized Block Design

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Randomized Block Design

1. Experimental Units (Subjects) Are Assigned Randomly within Blocks
   - Blocks are Assumed Homogeneous
2. One Factor or Independent Variable of Interest
   - 2 or More Treatment Levels or Classifications
3. One Blocking Factor

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Randomized Block Design

Factor Levels: (Treatments)	A, B, C, D			
Experimental Units	Treatments are randomly assigned within blocks			
Block 1	A	C	D	B
Block 2	C	D	B	A
Block 3	B	A	D	C
. . .	. . .	. . .	. . .	. . .
Block b	D	C	A	B

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Randomized Block F-Test

1. Tests the Equality of 2 or More (*p*) Population Means

2. Variables
   - One Nominal Independent Variable
   - One Nominal Blocking Variable
   - One Continuous Dependent Variable

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Randomized Block F-Test Assumptions

1. Normality
   - Probability Distribution of each Block-Treatment combination is Normal

2. Homogeneity of Variance
   - Probability Distributions of all Block-Treatment combinations have Equal Variances

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Randomized Block F-Test Hypotheses

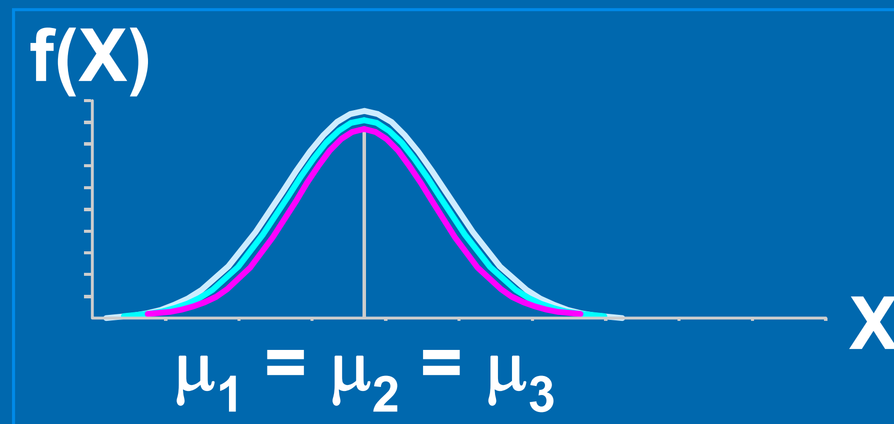
- H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_p$
  - All Population Means are Equal
  - No Treatment Effect
- H_a: Not All μ_j Are Equal
  - At Least 1 Pop. Mean is Different
  - Treatment Effect
  - $\mu_1 \neq \mu_2 \neq \ldots \neq \mu_p$ Is wrong

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Randomized Block F-Test Hypotheses

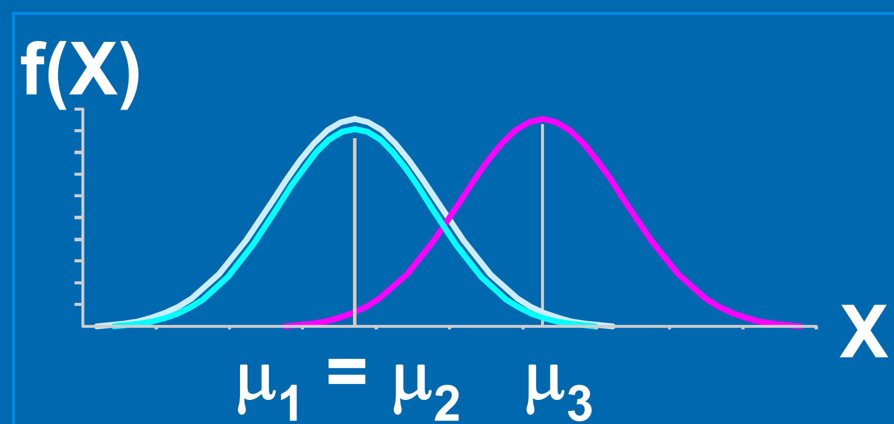
H_0: $\mu_1 = \mu_2 = \ldots = \mu_p$

- All Population Means are Equal
- No Treatment Effect



H_a: Not All μ_j Are Equal

- At Least 1 Pop. Mean is Different
- Treatment Effect
- $\mu_1 \neq \mu_2 \neq \ldots \neq \mu_p$ Is wrong



wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

The F Ratio for Randomized Block Designs

- SS=SSE+SSB+SST

$$F = \frac{\text{MST}}{\text{MSE}} = \frac{SST/(p-1)}{SSE/(n-1-p+1-b+1)}$$

$$= \frac{SST/(p-1)}{SSE/(n-p-b+1)}$$

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

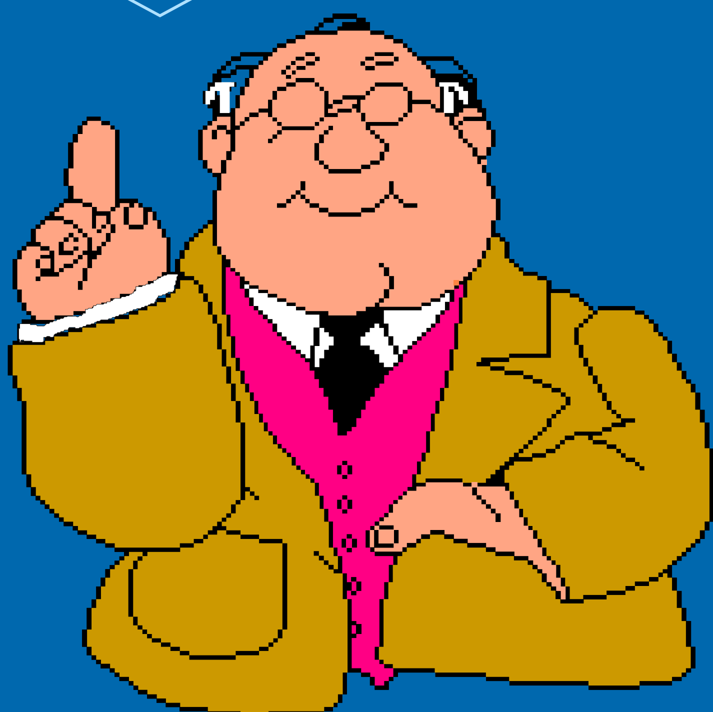
Randomized Block F-Test Test Statistic

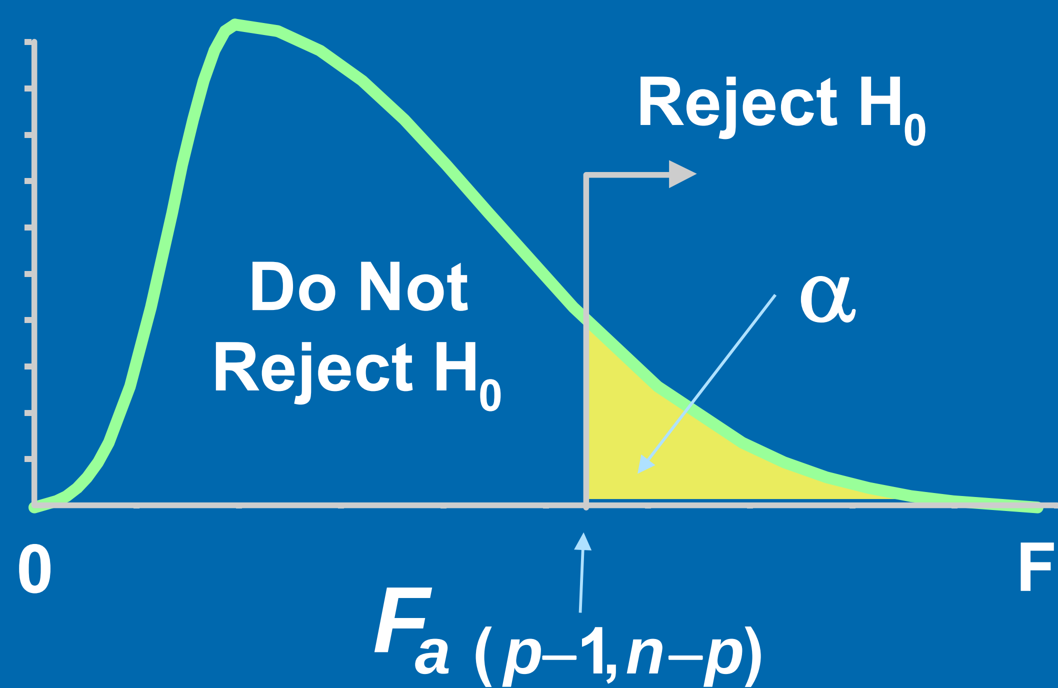
- 1. Test Statistic
  - *F = MST / MSE*
    - *MST* Is Mean Square for Treatment
    - *MSE* Is Mean Square for Error
- 2. Degrees of Freedom
  - $\nu_1 = p - 1$
  - $\nu_2 = n - b - p + 1$
    - *p* = # Treatments, *b* = # Blocks, *n* = Total Sample Size

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Randomized Block F-Test Critical Value

If means are equal,
F = *MST* / *MSE* ≈ 1.
Only reject large *F*!





Always One-Tail!

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Randomized Block F-Test Example

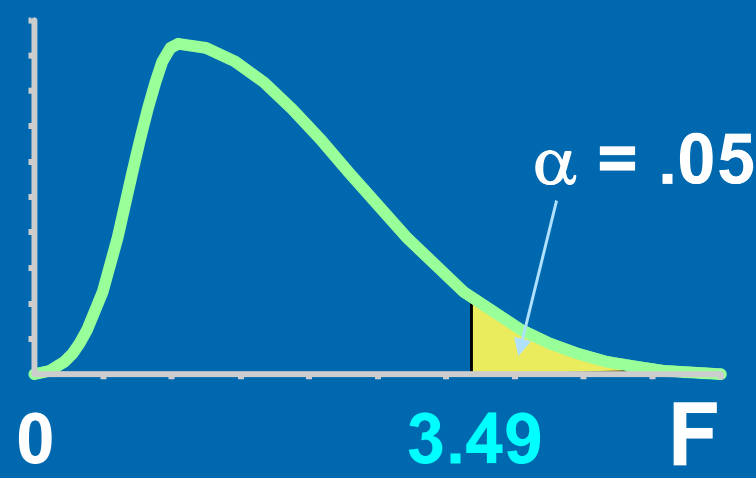
- You wish to determine which of four brands of anasthetic drugs has the longest effects (in second). You randomly assign one of each drugs (A, B, C, and D) to a four location on each of 5 subjects. At the **.05** level, is there a difference in **mean** effects?

	Location at arms			
Block	Left Front	Right Front	Left Rear	Right Rear
Subject 1	A: 42,000	C: 58,000	B: 38,000	D: 44,000
Subject 2	B: 40,000	D: 48,000	A: 39,000	C: 50,000
Subject 3	C: 48,000	D: 39,000	B: 36,000	A: 39,000
Subject 4	A: 41,000	B: 38,000	D: 42,000	C: 43,000
Subject 5	D: 51,000	A: 44,000	C: 52,000	B: 35,000

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Randomized Block F-Test Solution

- H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
- H_a: Not All Equal
- α = .05
- $\nu_1 = 3$ $\nu_2 = 12$
- Critical Value(s):



Test Statistic:

F = 11.9933

Decision:

Reject at α = .05

Conclusion:

There Is Evidence Pop. Means Are Different

wilopo@yahoo.com
Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

SAS CODES FOR ANOVA

```
data block;
input Block$ trt$ resp @@;
cards;
Subject 1        A: 42000 Car1 C: 58000 Car1  B: 38000 Car1 D: 44000
Subject 2        B: 40000 Car2 D: 48000 Car2  A: 39000 Car2 C: 50000
Subject 3        C: 48000 Car3 D: 39000 Car3  B: 36000 Car3 A: 39000
Subject 4        A: 41000 Car4 B: 38000 Car4  D: 42000 Car4 C: 43000
Subject 5        D: 51000 Car5 A: 44000 Car5  C: 52000 Car5 B: 35000
;
run;

proc anova;
class trt block;
model resp=trt block;
Means trt /lsd bon;
run;
```


wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

SAS OUTPUT - ANOVA

Dependent Variable: resp

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	544550000.0	77792857.1	6.22	0.0030
Error	12	150000000.0	12500000.0		
Corrected Total	19	694550000.0			

R-Square	Coeff Var	Root MSE	resp Mean
0.784033	8.155788	3535.534	43350.00

Source	DF	Anova SS	Mean Square	F Value	Pr > F
trt	3	449750000.0	149916666.7	11.99	0.0006
Block	4	94800000.0	23700000.0	1.90	0.1759

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

SAS OUTPUT - LSD

Means with the same letter are not significantly different.

t Grouping		Mean	N	trt
	A	50200	5	C:
	B	44800	5	D:
	B			
C	B	41000	5	A:
C				
C		37400	5	B:

wilopo@yahoo.com

SAS OUTPUT - Bonferroni

Means with the same letter are not significantly different.

Bon Grouping		Mean	N	trt
	A	50200	5	C:
	A			
B	A	44800	5	D:
B				
B	C	41000	5	A:
	C			
	C	37400	5	B:

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Factorial Experiments

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Public Health

Factorial Design

- 1. Experimental Units (Subjects) Are Assigned Randomly to Treatments
  - Subjects are Assumed Homogeneous
- 2. Two or More **Factors** or Independent Variables
  - Each Has 2 or More Treatments (Levels)
- 3. Analyzed by Two-Way ANOVA

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Advantages of Factorial Designs

1. Saves Time & Effort

   e.g., Could Use Separate Completely Randomized Designs for Each Variable

2. Controls Confounding Effects by Putting Other Variables into Model

3. Can Explore Interaction Between Variables

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Two-Way ANOVA

1. Tests the Equality of 2 or More Population Means When Several Independent Variables Are Used

2. Same Results as Separate One-Way ANOVA on Each Variable
   - But Interaction Can Be Tested

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Two-Way ANOVA Assumptions

1. **Normality**
   - Populations are Normally Distributed
2. **Homogeneity of Variance**
   - Populations have Equal Variances
3. **Independence of Errors**
   - Independent Random Samples are Drawn

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Two-Way ANOVA Data Table

Factor	Factor B			
A	1	2	...	b
1	Y_{111}	Y_{121}	...	Y_{1b1}
	Y_{112}	Y_{122}	...	Y_{1b2}
2	Y_{211}	Y_{221}	...	Y_{2b1}
	Y_{212}	Y_{222}	...	YX_{2b2}
:	:	:	:	:
a	Y_{a11}	Y_{a21}	...	Y_{ab1}
	Y_{a12}	Y_{a22}	...	Y_{ab2}

Observation k

Y_{ijk}

Level i Factor A

Level j Factor B

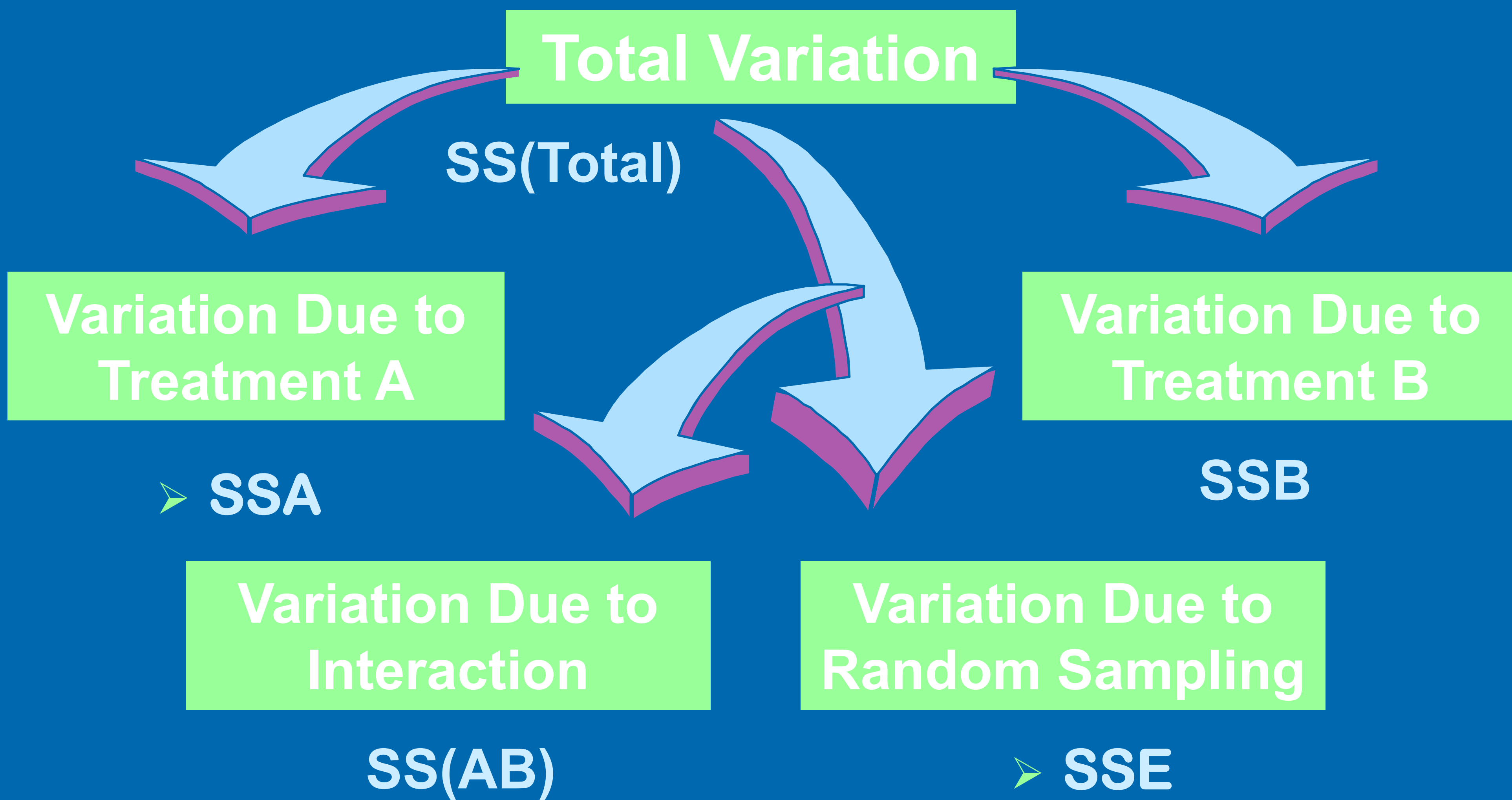
wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Two-Way ANOVA Null Hypotheses

1. No Difference in Means Due to Factor A
   - $H_0: \mu_{1.} = \mu_{2.} = \ldots = \mu_{a.}$
2. No Difference in Means Due to Factor B
   - $H_0: \mu_{.1} = \mu_{.2} = \ldots = \mu_{.b}$
3. No Interaction of Factors A & B
   - $H_0: AB_{ij} = 0$

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Two-Way ANOVA
Total Variation Partitioning



wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Two-Way ANOVA Summary Table

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
A (Row)	a - 1	SS(A)	MS(A)	MS(A) / MSE
B (Column)	b - 1	SS(B)	MS(B)	MS(B) / MSE
AB (Interaction)	(a-1)(b-1)	SS(AB)	MS(AB)	MS(AB) / MSE
Error	n - ab	SSE	MSE	
Total	n - 1	SS(Total)		

Same as Other Designs

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Interaction

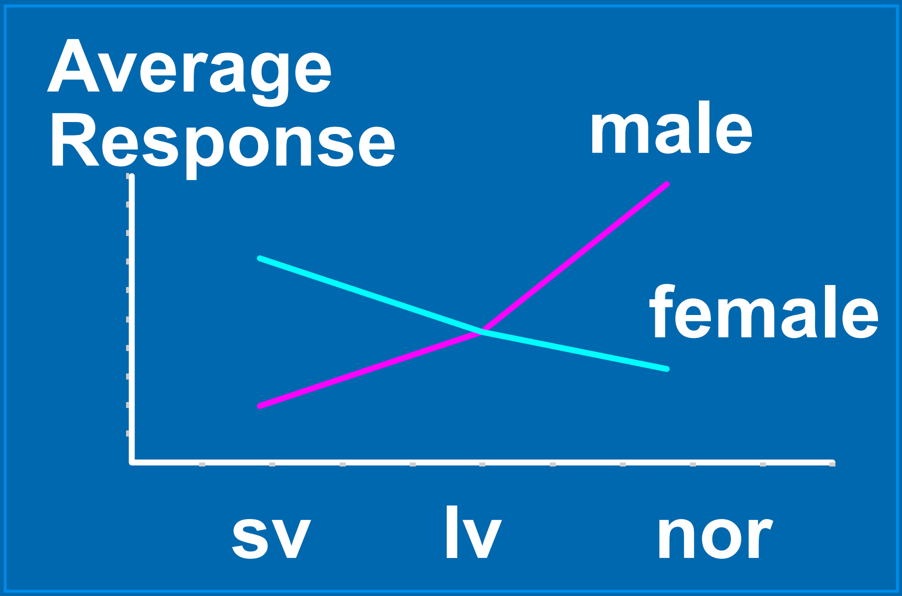
1. Occurs When Effects of One Factor Vary According to Levels of Other Factor
2. When Significant, Interpretation of Main Effects (A & B) Is Complicated
3. Can Be Detected
   - In Data Table, Pattern of Cell Means in One Row Differs From Another Row
   - In Graph of Cell Means, Lines Cross

wilopo@yahoo.com

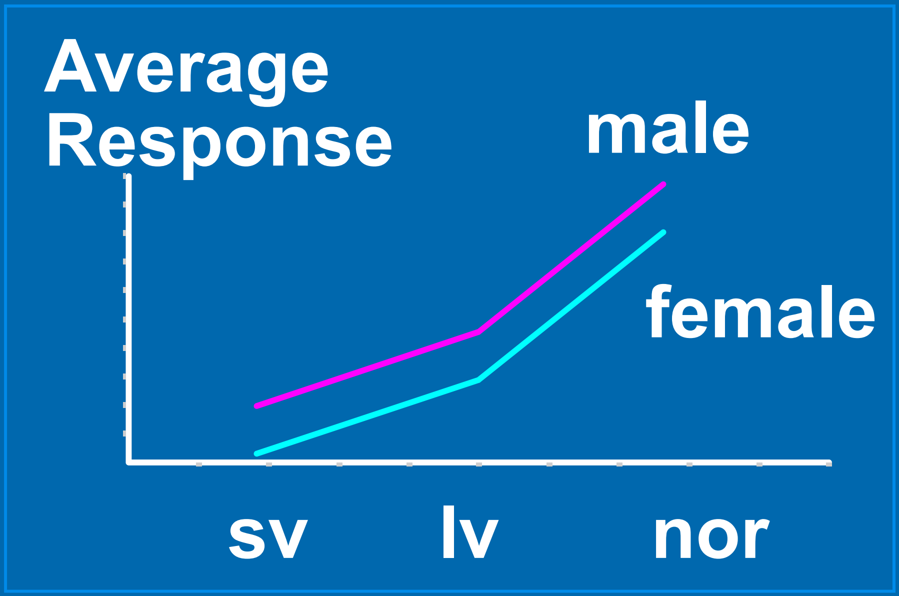
Graphs of Interaction

Effects of Gender (male or female) & dietary group (sv, lv, nor) on systolic blood pressure

Interaction



No Interaction



wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Two-Way ANOVA F-Test Example

Effect of diet (sv-strict vegetarians, lv-lactovegetarians, nor-normal) and gender (female, male) on systolic blood pressure.

Question: Test for interaction and main effects at the .05 level.

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

SAS CODES FOR ANOVA

```
data factorial;
input dietary$ sex$
    sbp;
cards;
sv male   109.9
sv male   101.9
sv male   100.9
sv male   119.9
sv male   104.9
sv male   189.9

sv female 102.6
sv female 99
sv female 83 .6
sv female 99.6
sv female 102.6
sv female 112.6
lv male 116.5
lv male 118.5
lv male 119.5
lv male 110.5
lv male 115.5
lv male 105.2

nor male 128.3
nor male 129.3
nor male 126.3
nor male 127.3
nor male 126.3
nor male 125.3

nor  female 119.1
nor  female 119.2
nor  female 115.6
nor  female 119.9
nor  female 119.8
nor  female 119.7
;
run;
```

SAS CODES FOR ANOVA

```
proc glm;
class dietary sex;
model sbp=dietary sex dietary*sex;
run;

proc glm;
class dietary sex;
model sbp=dietary sex;
run;
```


wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

SAS OUTPUT - ANOVA

Dependent Variable: sbp

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	2627.399667	525.479933	1.96	0.1215
Error	24	6435.215000	268.133958		
Corrected Total	29	9062.614667			

R-Square	Coeff Var	Root MSE	sbp Mean
0.289916	14.08140	16.37480	116.2867

Source	DF	Type I SS	Mean Square	F Value	Pr > F
dietary	2	958.870500	479.435250	1.79	0.1889
sex	1	1400.686992	1400.686992	5.22	0.0314
dietary*sex	2	267.842175	133.921087	0.50	0.6130

Source	DF	Type III SS	Mean Square	F Value	Pr > F
dietary	2	1039.020874	519.510437	1.94	0.1659
sex	1	877.982292	877.982292	3.27	0.0829
dietary*sex	2	267.842175	133.921087	0.50	0.6130

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Linear Contrast

- **Linear Contrast is a linear combination of the means of populations**

- **Purpose: to test relationship among different group means**

$$L = \sum c_j \mu_j \quad \text{with} \quad \sum c_j = 0$$

Example: 4 populations on treatments T1, T2, T3 and T4.

Contrast	T1	T2	T3	T4	relation to test
L1	1	0	-1	0	$\mu_1 - \mu_3 = 0$
L2	1	-1/2	-1/2	0	$\mu_1 - \mu_2/2 - \mu_3/2 = 0$

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

T-test for Linear Contrast (LSD)

- **Construct a *t* statistic involving *k* group means. Degrees of freedom of *t* - test: *df* = *n*-*k*.**

To test H_0: $L = \sum_{j=1}^{k} c_j \mu_j = 0$ Construct $t = \dfrac{L}{\sqrt{s^2 \sum_{j=1}^{k} \dfrac{c_j^2}{n_j}}}$

Compare with critical value $t_{1-\alpha/2,, n-k}$.

Reject H_0 if $|t| \geq t_{1-\alpha/2,, n-k}$.

SAS uses contrast statement and performs an F – test df (1, n-k);

Or estimate statement and perform a t-test df (n-k).

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

T-test for Linear Contrast (Scheffe)

- **Construct multiple contrasts involving *k* group means. Trying to search for significant contrast**

To test H_0: $L = \sum_{j=1}^{k} c_j \mu_j = 0$

Construct $t = \dfrac{L}{\sqrt{s^2 \sum_{j=1}^{k} \dfrac{c_j^2}{n_j}}}$

Compare with critical value.

$$a = \sqrt{(k-1)F_{k-1,n-k,1-\alpha}}$$

Reject H_0 if $|t| \geq a$

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

SAS Code for contrast testing

```
proc glm;
class trt block;
model resp=trt block;
Means trt /lsd bon scheffe;
contrast 'A - B = 0' trt 1 -1 0 0 ;
contrast 'A - B/2 - C/2 = 0' trt 1 -.5 -.5 0 ;
contrast 'A - B/3 - C/3 -D/3 = 0' trt 3 -1 -1 -1 ;
contrast 'A + B - C - D = 0' trt 1 1 -1 -1 ;
lsmeans trt/stderr pdiff;
lsmeans trt/stderr pdiff adjust=scheffe;  /* Scheffe's test  */
lsmeans trt/stderr pdiff adjust=bon;      /* Boneferoni's test */
estimate 'A - B' trt 1 -1 0 0 0;

run;
```


wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Regression representation of Anova

Regression representation of Anova

- **One-way anova:**

$$y_{ij} = \mu_i + e_{ij} = \mu + \alpha_i + e_{ij}$$

$$\sum_{i=1}^{p} \alpha_i = 0$$

- **Two-way anova:**

$$y_{ijk} = \mu_{ij} + e_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

$$\sum_{i=1}^{a} \alpha_i = 0, \sum_{j=1}^{b} \beta_j = 0, \ \sum_{j=1}^{b} \gamma_{ij} = 0 \ \text{for all } i \ \text{ and } \sum_{i=1}^{a} \gamma_{ij} = 0 \ \text{for all } j$$

- **SAS uses a different constraint**

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Regression representation of Anova

- One-way anova: Dummy variables of factor with p levels

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{p-1} x_{p-1} + e$$

$$\text{where} \quad x_i = \begin{cases} 1 & \text{if} \quad \text{level } i \\ 0 & \text{if} \quad \text{otherwise} \end{cases}$$

- This is the parameterization used by SAS

wilopo@yahoo.com
Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health

Conclusion: should be able to

1. Recognize the applications that uses ANOVA

2. Understand the logic of analysis of variance.

3. Be aware of several different analysis of variance designs and understand when to use each one.

4. Perform a single factor hypothesis test using analysis of variance manually and with the aid of SAS or any statistical software.

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Public Health

Conclusion: should be able to

5. Conduct and interpret post-analysis of variance pairwise comparisons procedures.

6. Recognize when randomized block analysis of variance is useful and be able to perform the randomized block analysis.

7. Perform two factor analysis of variance tests with replications using SAS and interpret the output.

sawilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Department of Public Health

Key Terms

- **Between-Sample Variation**
- **Completely Randomized Design**
- **Experiment-Wide Error Rate**
- **Factor**
- Levels
- One-Way Analysis of Variance
- Total Variation
- Treatment
- Within-Sample Variation

wilopo@yahoo.com Universitas Gadjah Mada, Faculty of Medicine, Dept. Biostat. Epi. and Pop. Health