One Factor Linear Model
(Population Model)

FROM BEFORE

$$\boxed{Y_{ij}} = \boxed{\mu} + \boxed{\tau_j} + \boxed{\varepsilon_{ij}}$$

Score on Y for the ith individual in the jth group = Grand Mean + Treatment Effect for Group j + Error

One Factor Linear Model
(Sample Model)

FROM BEFORE

$$\boxed{Y_{ij}} = \boxed{\bar{Y}} + \boxed{t_j} + \boxed{e_{ij}}$$

Score on Y for the ith individual in the jth group = Est. Grand Mean + Estimated Treatment Effect for Group j + Error (Residual)

Inferences about Treatment Effects

FROM BEFORE

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_j = 0$$

$$H_1 : Not \text{ All } \tau_j = 0$$

Statistical Inference with Linear Models

- Analysis of Variance Approach
- General Linear Test Approach *(model comparison)*

5

Inferences about Treatment Effects

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_j = 0$$

$$H_1 : Not \;\text{All}\; \tau_j = 0$$

6

The Fisher-Snedecor Distribution

is distributed as

df for numerator

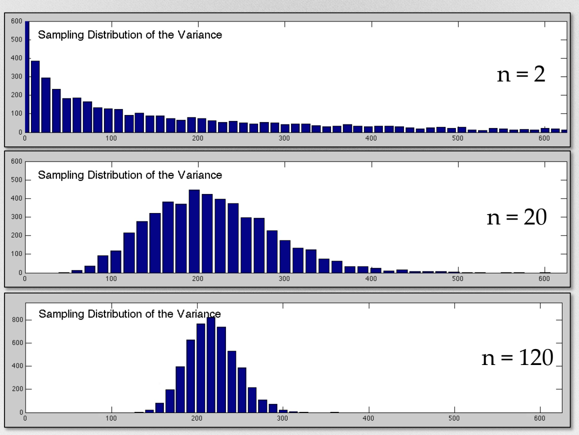
df for denominator

$$\frac{Variance \text{ Estimate 1}}{Variance \text{ Estimate 2}} \sim F_{\upsilon_1, \upsilon_2}$$

Assuming estimates are for the same population variance

7

FROM BEFORE



9

The Fisher-Snedecor Distribution

is distributed as

df for numerator

df for denominator

$$\frac{Variance \text{ Estimate 1}}{Variance \text{ Estimate 2}} \sim F_{\upsilon_1, \upsilon_2}$$

Assuming estimates are for the same population variance

11

The Fisher-Snedecor Distribution

$$\frac{\chi^2_{(\upsilon_1)}}{\chi^2_{(\upsilon_2)}} \sim F_{\upsilon_1,\upsilon_2}$$

is distributed as

df for numerator

df for denominator

Assuming estimates are for the same population variance

The Fisher-Snedecor Distribution

$$\frac{\textit{Variance}\ \text{Estimate 1}}{\textit{Variance}\ \text{Estimate 2}} \sim F_{\upsilon_1,\upsilon_2}$$

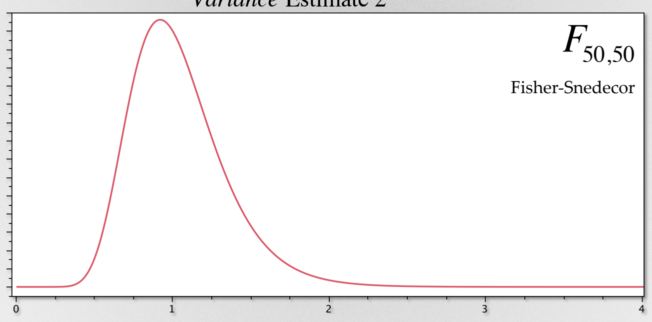
is distributed as

df for numerator

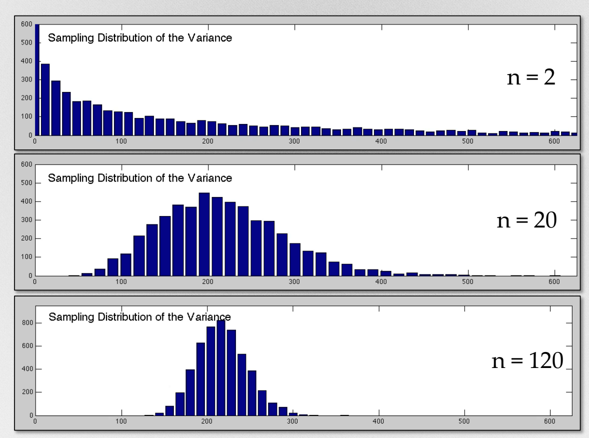
df for denominator

Assuming estimates are for the same population variance

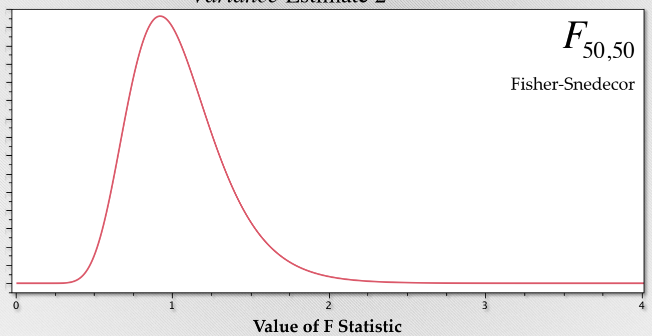
$$\frac{\textit{Variance}\ \text{Estimate 1}}{\textit{Variance}\ \text{Estimate 2}} \sim F_{\upsilon_1,\upsilon_2}$$



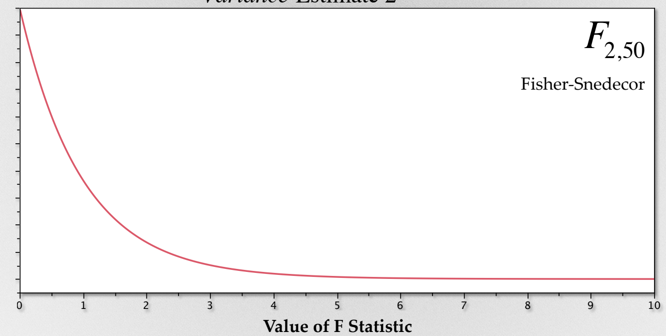
FROM BEFORE



$$\frac{Variance\ \text{Estimate 1}}{Variance\ \text{Estimate 2}} \sim F_{\upsilon_1,\upsilon_2}$$



$$\frac{Variance\ \text{Estimate 1}}{Variance\ \text{Estimate 2}} \sim F_{\upsilon_1,\upsilon_2}$$



The Fisher-Snedecor Distribution

is distributed as

$$\frac{Variance\ \text{Estimate 1}}{Variance\ \text{Estimate 2}} \sim F_{\upsilon_1,\upsilon_2}$$

Assuming estimates are for the same population variance

Inferences about Treatment Effects

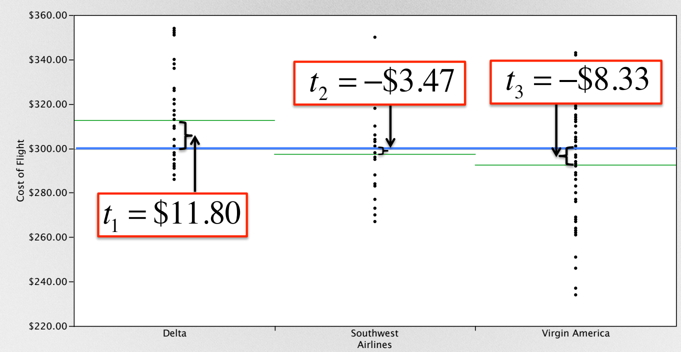
$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_j = 0$$

$$H_1 : Not\ \text{All}\ \tau_j = 0$$

Analysis of Variance Test Statistic

$$F = \frac{\textit{Variance} \text{ of Treatments}}{\textit{Variance} \text{ of Error}}$$

FROM BEFORE



Analysis of Variance Test Statistic

$$F = \frac{\textit{Variance} \text{ of Treatments}}{\textit{Variance} \text{ of Error}}$$

Check Your Understanding

$$F = \frac{\textit{Variance} \text{ of Treatments}}{\textit{Variance} \text{ of Error}}$$

If there is a large true effect of the treatment factor (the τ_j are very different) what kind of values of F should we expect?:

- A • Values larger than 1
- B • Values around 1
- C • I don't know (F this question)

Check Your Understanding

$$F = \frac{\mathit{Variance}\ \text{of Treatments}}{\mathit{Variance}\ \text{of Error}}$$

If there is a large true effect of the treatment factor (the τ_j are very different) what kind of values of F should we expect?:

- A • Values larger than 1
- B • Values around 1
- C • I don't know (F this question)

24

Analysis of Variance Test Statistic

$$F = \frac{\mathit{Variance}\ \text{of Treatments}}{\mathit{Variance}\ \text{of Error}}$$

25

Analysis of Variance Test Statistic

$$F = \frac{\mathit{Systematic\ Variance} + \mathit{Random\ Variance}}{\mathit{Random\ Variance}}$$

$$\mathit{If}\ H_1 \text{ is true: not all } \tau_j = 0$$

26

Analysis of Variance Test Statistic

$$F = \frac{\mathit{Random\ Variance}}{\mathit{Random\ Variance}}$$

$$\mathit{If}\ H_0 \text{ is true: } \tau_1 = \tau_2 = \ldots = \tau_j = 0$$

27

Analysis of Variance Test Statistic

$$F = \frac{Variance\text{ of Treatments}}{Variance\text{ of Error}}$$

Analysis of Variance Test Statistic

$$F = \frac{MS_{treatments}}{MS_{error}}$$

Analysis of Variance Test Statistic

$$F = \frac{MS_{treatments}}{MS_{error}} = \frac{SS_t / df_t}{SS_e / df_e}$$

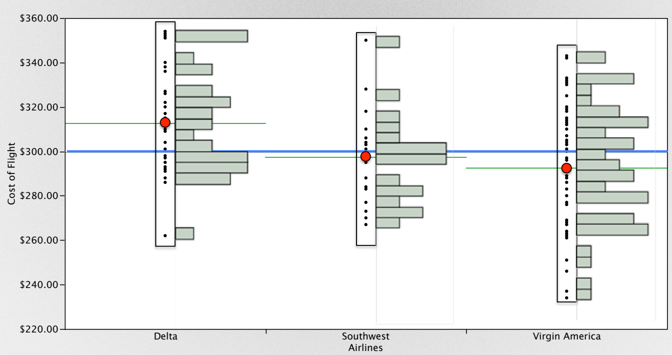
FROM BEFORE

Sums of Squares for Error

$$SS_{error} = \sum_j \sum_i (\boxed{Y_{ij}} - \boxed{\hat{Y}_{ij}})^2$$

The sums of squares for error is the sum of the squared distance from each individuals **actual score** and their **predicted score**

FROM BEFORE



FROM BEFORE

Sums of Squares for Error

$$SS_{error} = \sum_j \sum_i (\boxed{Y_{ij}} - \boxed{\hat{Y}_{ij}})^2$$

The sums of squares for error is the sum of the squared distance from each individuals **actual score** and their **predicted score**

Sums of Squares for Error

$$SS_{error} = \sum_j \sum_i (\boxed{Y_{ij}} - \boxed{\hat{Y}_{ij}})^2$$

The sums of squares for error is the sum of the squared distance from each individuals **actual score** and their **predicted score**

Also called the sums of squares within
SS_{within}

Analysis of Variance Test Statistic

$$F = \frac{MS_{treatments}}{MS_{error}} = \frac{SS_t / df_t}{SS_e / df_e}$$

Sums of Squares Treatment

$$SS_{treat} = \sum_j \sum_i (\boxed{\hat{Y}_{ij}} - \boxed{\bar{Y}})^2$$

The sums of squares treatment is the sum of the squared distances for each individual's **predicted score** from the **grand mean**

Also called the sums of squares between
$SS_{between}$

36

$$SS_{error} = \sum_j \sum_i (\boxed{Y_{ij}} - \boxed{\hat{Y}_{ij}})^2$$

$$+ \; SS_{treat} = \sum_j \sum_i (\boxed{\hat{Y}_{ij}} - \boxed{\bar{Y}})^2 \; =$$

$$SS_{total} = \sum_j \sum_i (\boxed{Y_{ij}} \; - \; \boxed{\bar{Y}})^2$$

37

Sums of Squares Total

$$SS_{total} = \sum_j \sum_i (\boxed{Y_{ij}} - \boxed{\bar{Y}})^2$$

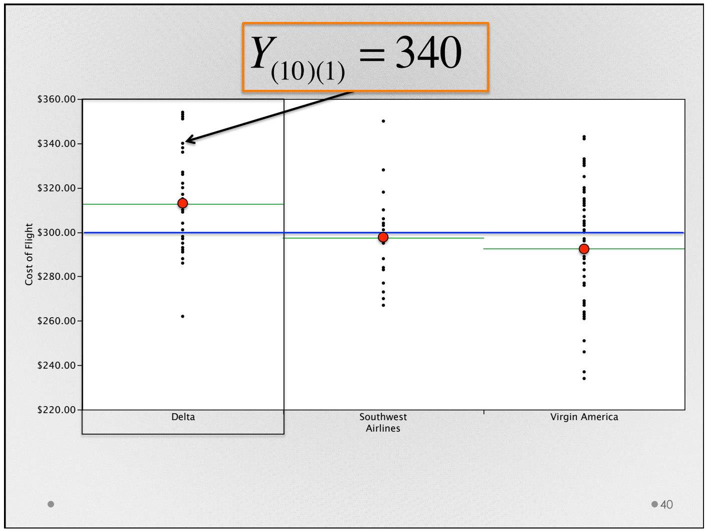
The sums of squares total is the sum of the squared distance for each individual's **actual score** and the **grand mean**

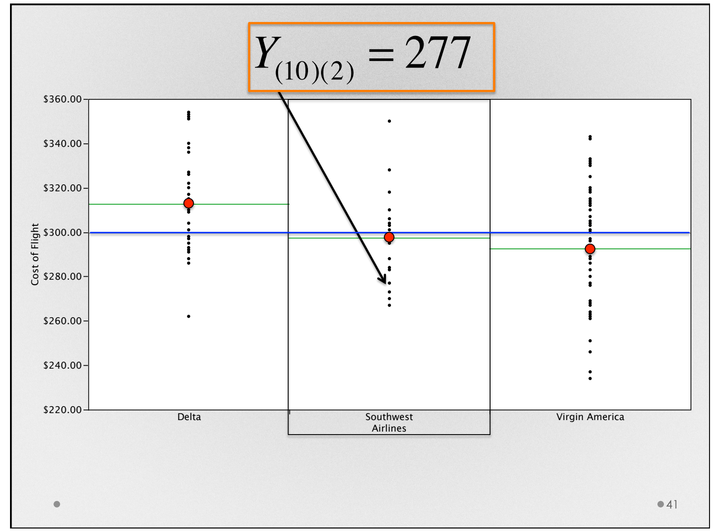
These are the sums of squares you would get if ignored your factor and found SS for the Y variable alone

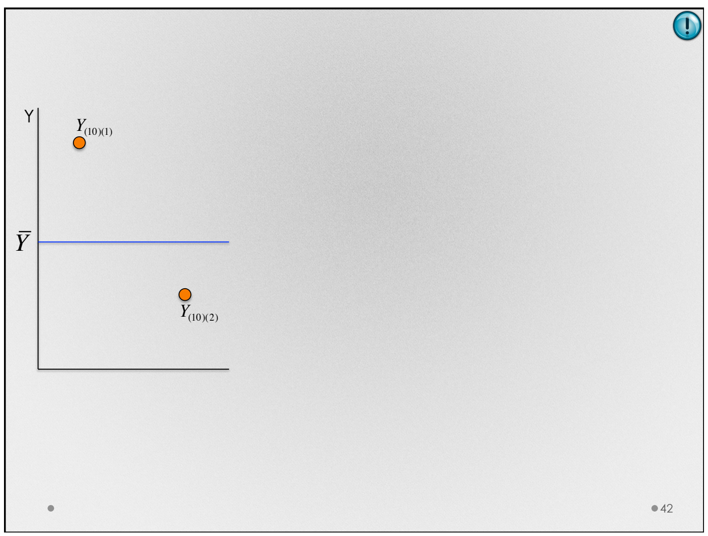
38

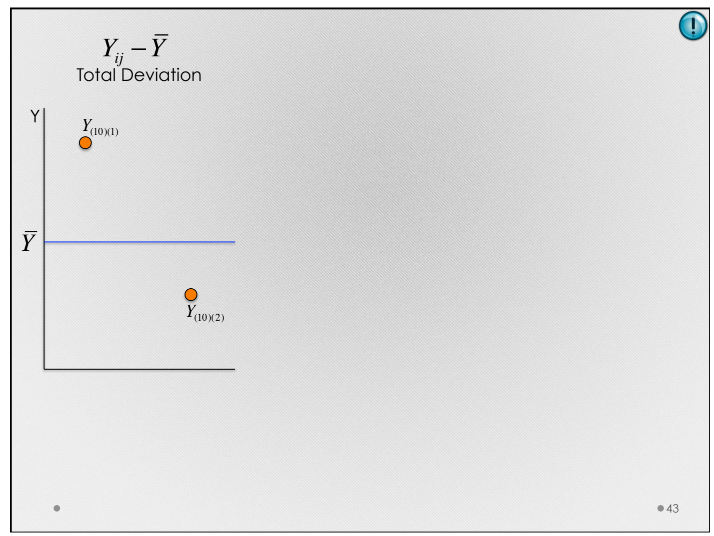
Partitioning the Total Sums of Squares

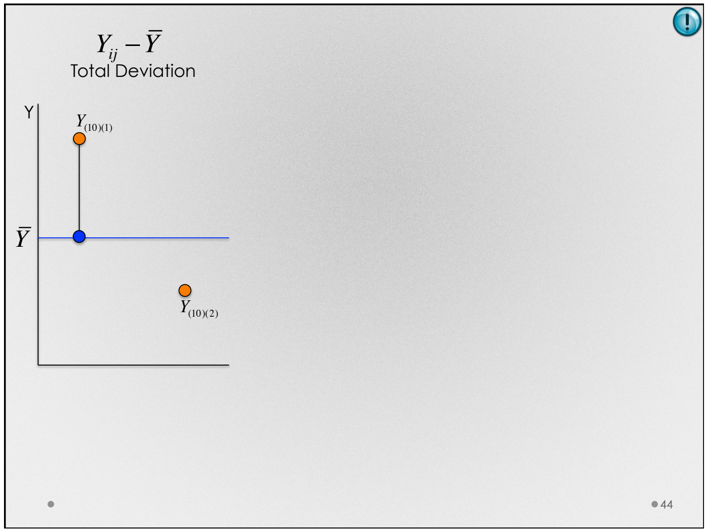
39

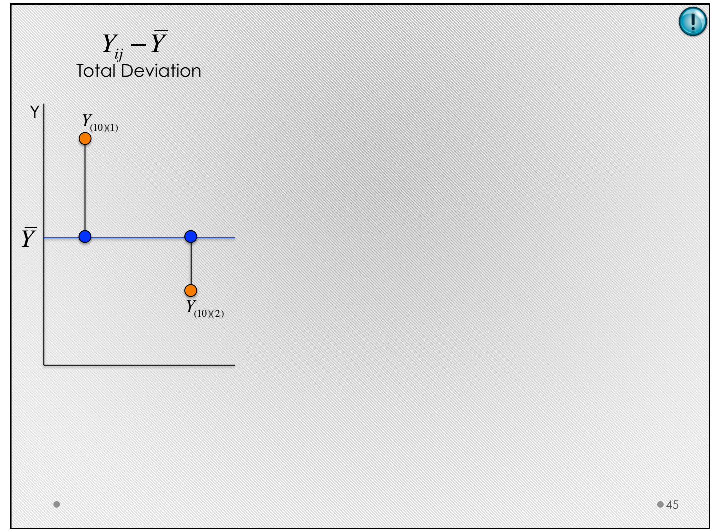

$Y_{(10)(1)} = 340$
$360.00
$340.00
$320.00
$300.00
$280.00
$260.00
$240.00
$220.00
Cost of Flight
Delta
Southwest Airlines
Virgin America
40

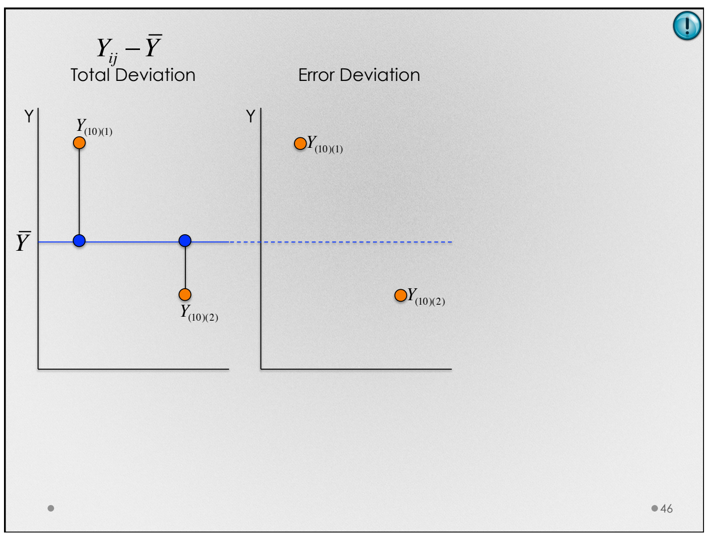

$Y_{(10)(2)} = 277$
$360.00
$340.00
$320.00
$300.00
$280.00
$260.00
$240.00
$220.00
Cost of Flight
Delta
Southwest Airlines
Virgin America
41

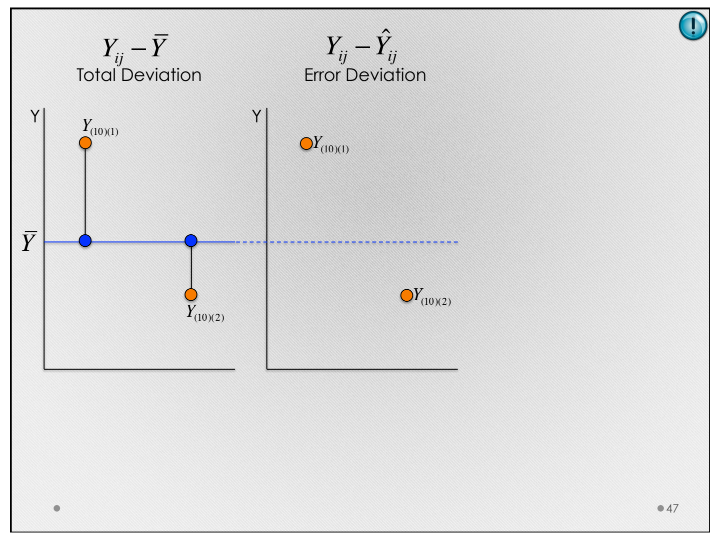

Y
$Y_{(10)(1)}$
$\overline{Y}$
$Y_{(10)(2)}$
42

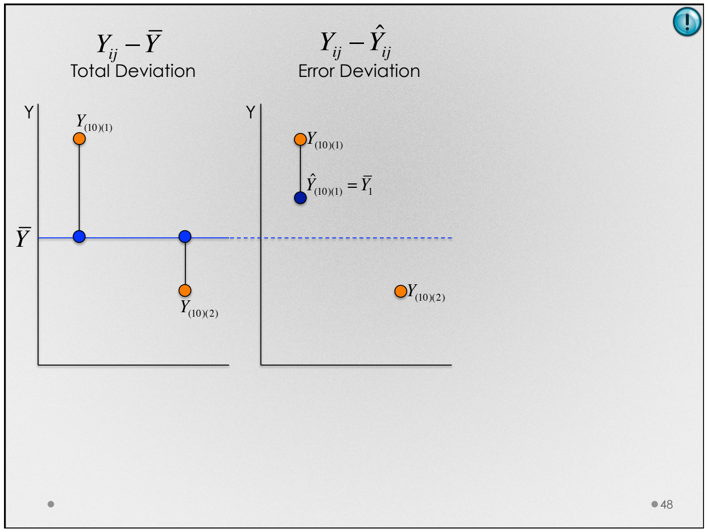

$Y_{ij} - \overline{Y}$
Total Deviation
Y
$Y_{(10)(1)}$
$\overline{Y}$
$Y_{(10)(2)}$
43

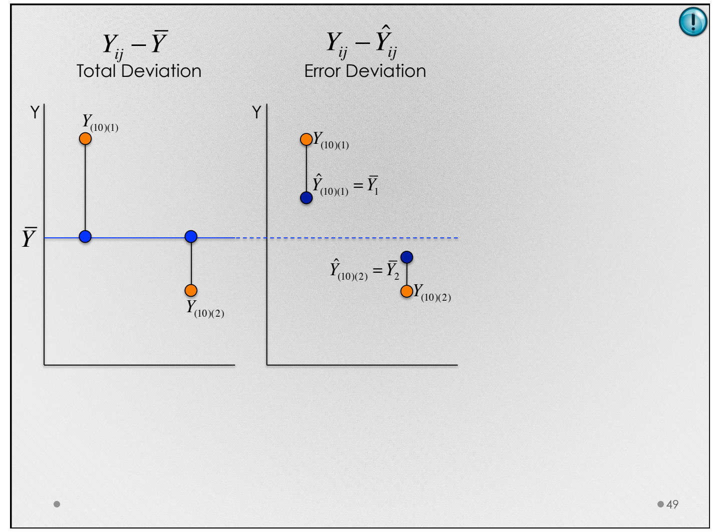

$Y_{ij} - \overline{Y}$
Total Deviation
Y
$Y_{(10)(1)}$
$\overline{Y}$
$Y_{(10)(2)}$
44

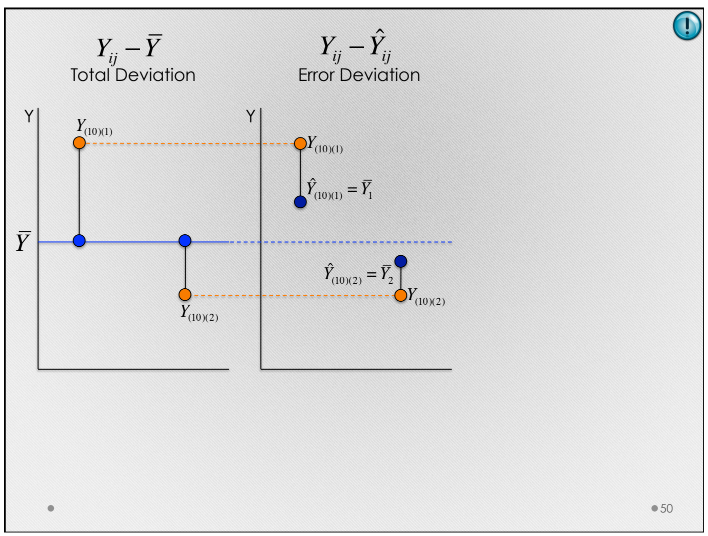

$Y_{ij} - \overline{Y}$
Total Deviation
Y
$Y_{(10)(1)}$
$\overline{Y}$
$Y_{(10)(2)}$
45

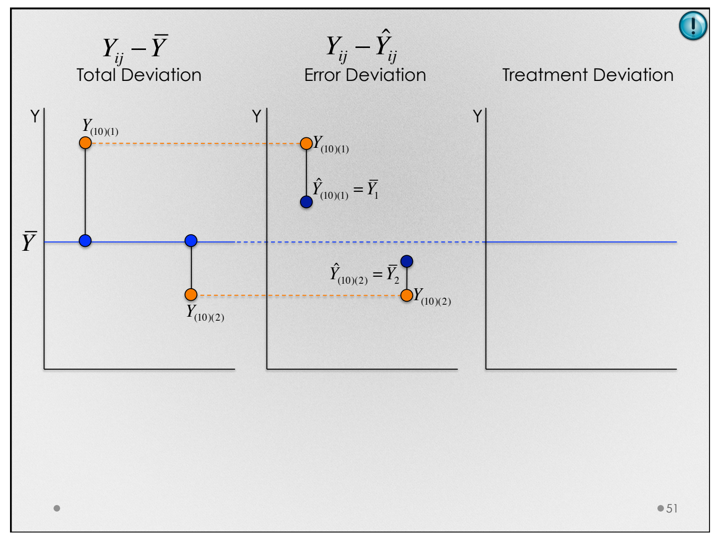

$Y_{ij} - \overline{Y}$
Total Deviation
Error Deviation
Y
$Y_{(10)(1)}$
$\overline{Y}$
$Y_{(10)(2)}$
Y
$Y_{(10)(1)}$
$Y_{(10)(2)}$
46

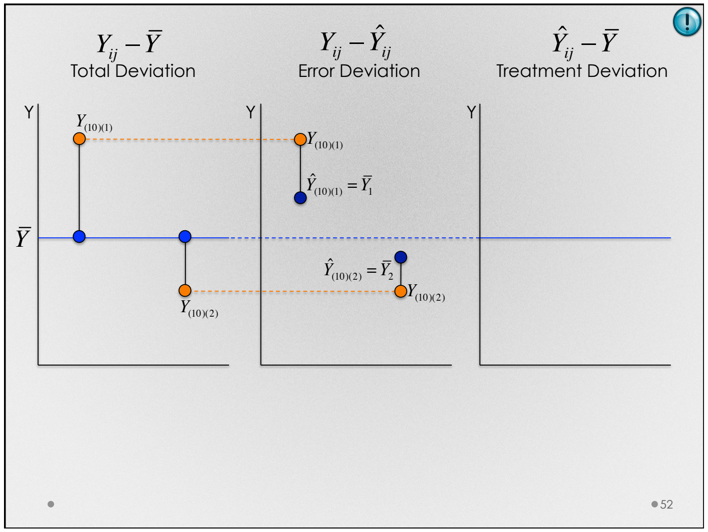

$Y_{ij} - \overline{Y}$
Total Deviation
$Y_{ij} - \hat{Y}_{ij}$
Error Deviation
Y
$Y_{(10)(1)}$
$\overline{Y}$
$Y_{(10)(2)}$
Y
$Y_{(10)(1)}$
$Y_{(10)(2)}$
47

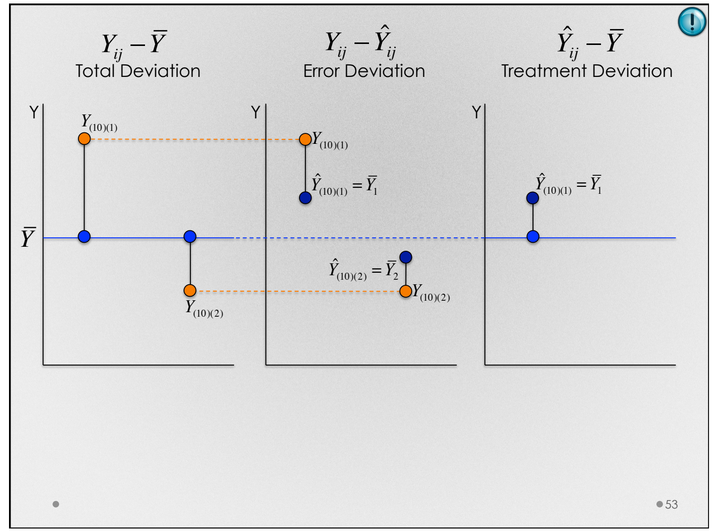

$Y_{ij} - \bar{Y}$
Total Deviation
$Y_{ij} - \hat{Y}_{ij}$
Error Deviation
Y
$Y_{(10)(1)}$
$\bar{Y}$
$Y_{(10)(2)}$
Y
$Y_{(10)(1)}$
$\hat{Y}_{(10)(1)} = \bar{Y}_1$
$Y_{(10)(2)}$
48

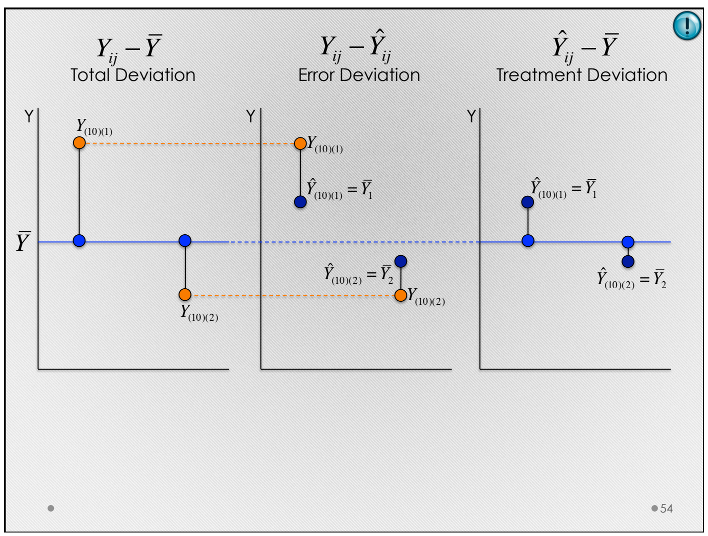

$Y_{ij} - \bar{Y}$
Total Deviation
$Y_{ij} - \hat{Y}_{ij}$
Error Deviation
Y
$Y_{(10)(1)}$
$\bar{Y}$
$Y_{(10)(2)}$
Y
$Y_{(10)(1)}$
$\hat{Y}_{(10)(1)} = \bar{Y}_1$
$\hat{Y}_{(10)(2)} = \bar{Y}_2$
$Y_{(10)(2)}$
49

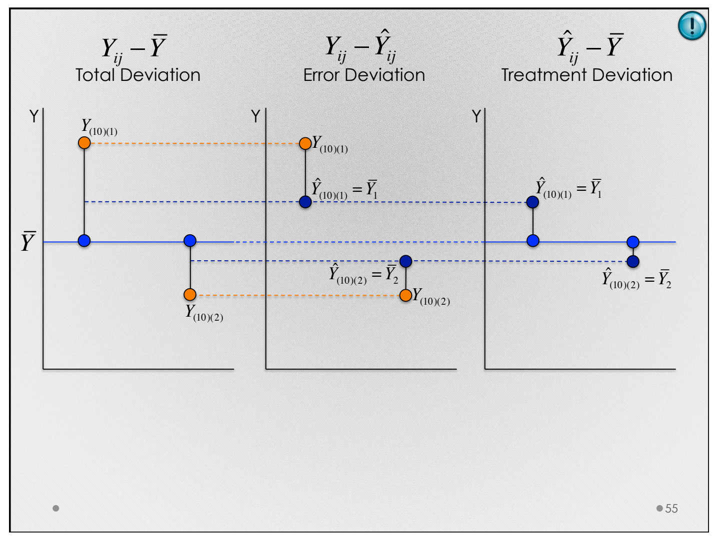

$Y_{ij} - \bar{Y}$
Total Deviation
$Y_{ij} - \hat{Y}_{ij}$
Error Deviation
Y
$Y_{(10)(1)}$
$\bar{Y}$
$Y_{(10)(2)}$
Y
$Y_{(10)(1)}$
$\hat{Y}_{(10)(1)} = \bar{Y}_1$
$\hat{Y}_{(10)(2)} = \bar{Y}_2$
$Y_{(10)(2)}$
50

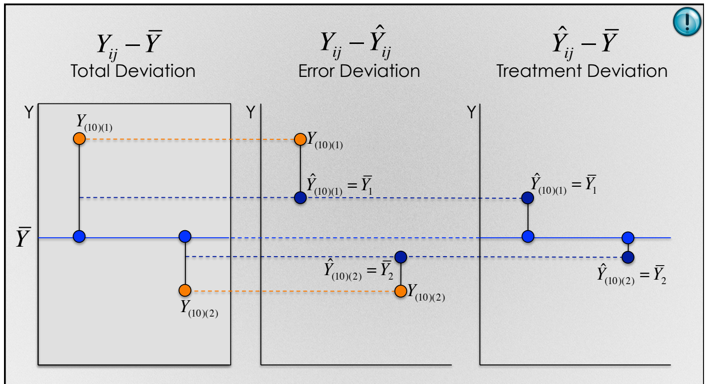

$Y_{ij} - \bar{Y}$
Total Deviation
$Y_{ij} - \hat{Y}_{ij}$
Error Deviation
Treatment Deviation
Y
$Y_{(10)(1)}$
$\bar{Y}$
$Y_{(10)(2)}$
Y
$Y_{(10)(1)}$
$\hat{Y}_{(10)(1)} = \bar{Y}_1$
$\hat{Y}_{(10)(2)} = \bar{Y}_2$
$Y_{(10)(2)}$
Y
51


$Y_{ij} - \bar{Y}$
Total Deviation
$Y_{ij} - \hat{Y}_{ij}$
Error Deviation
$\hat{Y}_{ij} - \bar{Y}$
Treatment Deviation
Y
$\bar{Y}$
$Y_{(10)(1)}$
$\hat{Y}_{(10)(1)} = \bar{Y}_1$
$\hat{Y}_{(10)(2)} = \bar{Y}_2$
$Y_{(10)(2)}$
52


$Y_{ij} - \bar{Y}$
Total Deviation
$Y_{ij} - \hat{Y}_{ij}$
Error Deviation
$\hat{Y}_{ij} - \bar{Y}$
Treatment Deviation
Y
$\bar{Y}$
$Y_{(10)(1)}$
$\hat{Y}_{(10)(1)} = \bar{Y}_1$
$\hat{Y}_{(10)(2)} = \bar{Y}_2$
$Y_{(10)(2)}$
53


$Y_{ij} - \bar{Y}$
Total Deviation
$Y_{ij} - \hat{Y}_{ij}$
Error Deviation
$\hat{Y}_{ij} - \bar{Y}$
Treatment Deviation
Y
$\bar{Y}$
$Y_{(10)(1)}$
$\hat{Y}_{(10)(1)} = \bar{Y}_1$
$\hat{Y}_{(10)(2)} = \bar{Y}_2$
$Y_{(10)(2)}$
54


$Y_{ij} - \bar{Y}$
Total Deviation
$Y_{ij} - \hat{Y}_{ij}$
Error Deviation
$\hat{Y}_{ij} - \bar{Y}$
Treatment Deviation
Y
$\bar{Y}$
$Y_{(10)(1)}$
$\hat{Y}_{(10)(1)} = \bar{Y}_1$
$\hat{Y}_{(10)(2)} = \bar{Y}_2$
$Y_{(10)(2)}$
55



56

$$SS_{error} = \sum_j \sum_i (\boxed{Y_{ij}} - \boxed{\hat{Y}_{ij}})^2$$

$$+ \; SS_{treat} = \sum_j \sum_i (\boxed{\hat{Y}_{ij}} - \boxed{\bar{Y}})^2 \; =$$

$$SS_{total} = \sum_j \sum_i (\boxed{Y_{ij}} \; - \; \boxed{\bar{Y}})^2$$

57

$$SS_{error} = \sum_j \sum_i (\boxed{Y_{ij}} - \boxed{\hat{Y}_{ij}})^2$$

$$+ \; SS_{treat} = \sum_j \sum_i (\boxed{\hat{Y}_{ij}} - \boxed{\bar{Y}})^2$$

$$= \; SS_{total} = \sum_j \sum_i (\boxed{Y_{ij}} - \boxed{\bar{Y}})^2$$

58

Sums of Squares Treatment

$$SS_{treat} = \sum_j \sum_i (\boxed{\hat{Y}_{ij}} - \boxed{\bar{Y}})^2$$

The sums of squares treatment is the sum of the squared distances for each individual's **predicted score** from the **grand mean**

Also called the sums of squares between
$SS_{between}$

59

Sums of Squares Treatment

$$SS_{treat} = \sum_j n_j(\bar{Y}_j - \boxed{\bar{Y}})^2$$

The sums of squares treatment is the sum of the squared distances for each individual's **predicted score** from the **grand mean**

Also called the sums of squares between
$SS_{between}$

Sums of Squares Treatment

$$SS_{treat} = \sum_j n_j(\boxed{\bar{Y}_j - \bar{Y}})^2$$

The sums of squares treatment is the sum of the squared distances for each individual's **predicted score** from the **grand mean**

Also called the sums of squares between
$SS_{between}$

Sums of Squares Treatment

$$SS_{treatment} = \sum_j n_j(\boxed{t_j})^2$$

The sums of squares treatment is the sum of the squared distances for each individual's **predicted score** from the **grand mean**

Also called the sums of squares between
$SS_{between}$

Analysis of Variance Test Statistic

$$F = \frac{MS_{treatments}}{MS_{error}} = \frac{SS_t / df_t}{SS_e / df_e}$$

Degrees of Freedom for the Analysis of Variance

FROM BEFORE

$$SS_{error} = \sum_j \sum_i (\boxed{Y_{ij}} - \boxed{\hat{Y}_{ij}})^2$$

$$+ \; SS_{treat} = \sum_j \sum_i (\boxed{\hat{Y}_{ij}} - \boxed{\bar{Y}})^2$$

$$= \; SS_{total} = \sum_j \sum_i (\boxed{Y_{ij}} - \boxed{\bar{Y}})^2$$

Degrees of Freedom for the Analysis of Variance

Degrees of Freedom for the Analysis of Variance

$$df_{error} + df_{treatment} = df_{total}$$

Sums of Squares Total

$$SS_{total} = \sum_j \sum_i (Y_{ij} - \bar{Y})^2$$

Degrees of Freedom for SS_{total}

$df_{total} = n_{total} - 1$

68

Sums of Squares Treatment

$$SS_{treatment} = \sum_j n_j (t_j)^2$$

Degrees of Freedom for $SS_{treatment}$

$df_t = j-1$

69

Delta

$$t_1 = \$11.80$$

Southwest

$$t_2 = -\$3.47$$

Virgin America

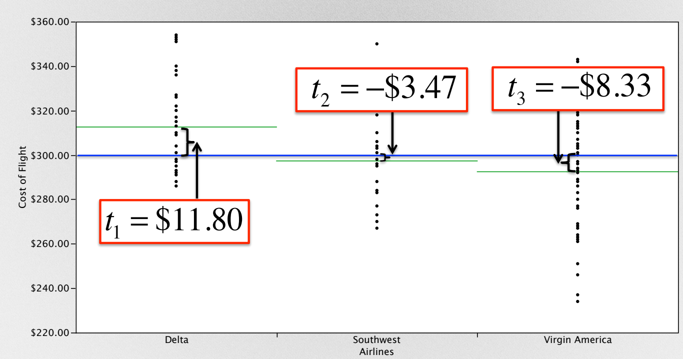
$$t_3 = -\$8.33$$

Restriction

$$\sum_j t_j = 0$$

70

FROM BEFORE



71

Sums of Squares Treatment

$$SS_{treatment} = \sum_j n_j (t_j)^2$$

Degrees of Freedom for $SS_{treatment}$

$df_t = j-1$

Sums of Squares for Error

$$SS_{error} = \sum_j \sum_i (Y_{ij} - \hat{Y}_{ij})^2$$

Degrees of Freedom for SS_{error}

$$df_{error} = \sum_j (n_j - 1)$$

Sums of Squares for Error

$$SS_{error} = \sum_j \sum_i (Y_{ij} - \hat{Y}_{ij})^2$$

Degrees of Freedom for SS_{error}

$$df_{error} = \sum_j (n_j - 1)$$

$$df_{error} = \left[n_{total} - (j-1)\right] - 1$$

$$\begin{array}{ll} df_{error} & n_{total} - (j-1) - 1 \\ +\ df_{treatment} & +\ j-1 \\ \hline =\ df_{total} & =\ n_{total} - 1 \end{array}$$

$$\begin{array}{ll} & df_{error} \\ + & df_{treatment} \\ \hline = & df_{total} \end{array} \qquad \begin{array}{ll} & 100-(3-1)-1 \\ + & 3-1 \\ \hline = & 100-1 \end{array}$$

76

$$\begin{array}{ll} & df_{error} \\ + & df_{treatment} \\ \hline = & df_{total} \end{array} \qquad \begin{array}{ll} & 97 \\ + & 2 \\ \hline = & 99 \end{array}$$

77

Analysis of Variance Test Statistic

$$F = \frac{MS_{treatments}}{MS_{error}} = \frac{SS_t / df_t}{SS_e / df_e} \qquad \begin{array}{l} v_1 = 2 \\ v_2 = 97 \end{array}$$

78

Analysis of Variance Test Statistic

$$F = \frac{MS_{treatments}}{MS_{error}} = \frac{SS_t / 2}{SS_e / 97} \qquad \begin{array}{l} v_1 = 2 \\ v_2 = 97 \end{array}$$

79

The Fisher-Snedecor Distribution

$$\frac{\textit{Variance}\ \text{Estimate 1}}{\textit{Variance}\ \text{Estimate 2}} \underbrace{\sim}_{\text{is distributed as}} F_{\upsilon_1,\upsilon_2}$$

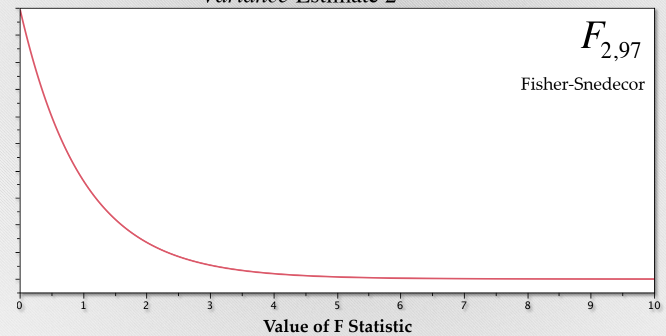
is distributed as

df for numerator

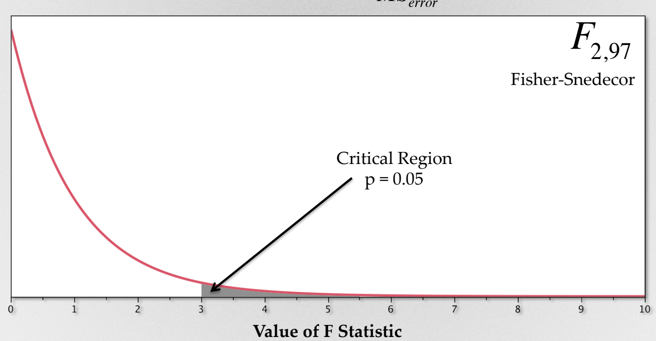
df for denominator

Assuming estimates are for the same population variance

$$\frac{\textit{Variance}\ \text{Estimate 1}}{\textit{Variance}\ \text{Estimate 2}} \sim F_{\upsilon_1,\upsilon_2}$$



$$\textit{If } H_0 \text{ is true } (\tau_1 = \tau_2 = \ldots = \tau_j = 0): \frac{MS_{treatment}}{MS_{error}} \sim F_{\upsilon_1,\upsilon_2}$$



FROM BEFORE

Statistical Inference with Linear Models

- Analysis of Variance Approach
- General Linear Test Approach
  (model comparison)

The General Linear Test

1. Specify a Full (unrestricted) Model and determine amount of "error"
2. Specify a Reduced (restricted) Model and determine amount of "error"
3. Test the reduction in "error"

The General Linear Test

$$F = \frac{\left(\dfrac{\text{Reduction in Error}}{\#\ \text{Added Parameters}}\right)}{\textit{Baseline}\ \text{Error}}$$

Full (unrestricted) Model

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

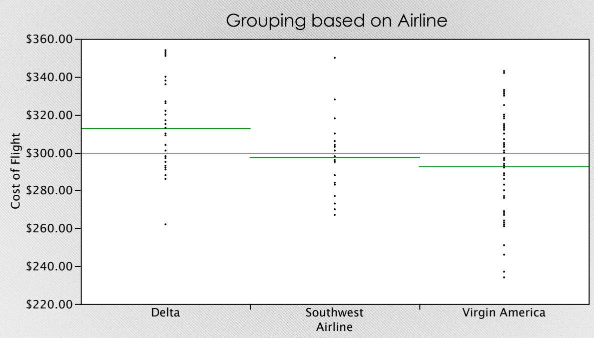
Reduced (restricted) Model

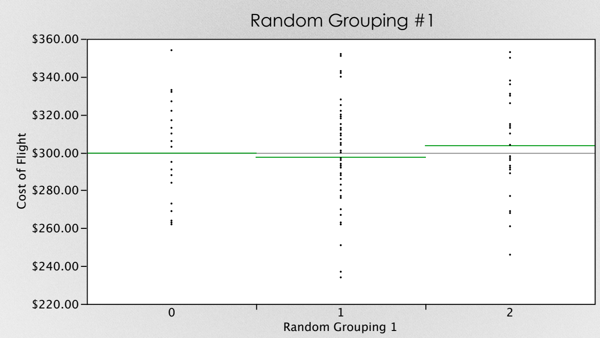
$$H_0 : \tau_1 = \tau_2 = ... = \tau_j = 0$$

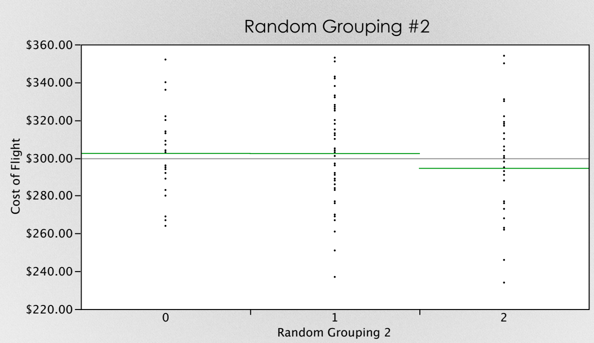
$$Y_i = \mu + \varepsilon_i$$

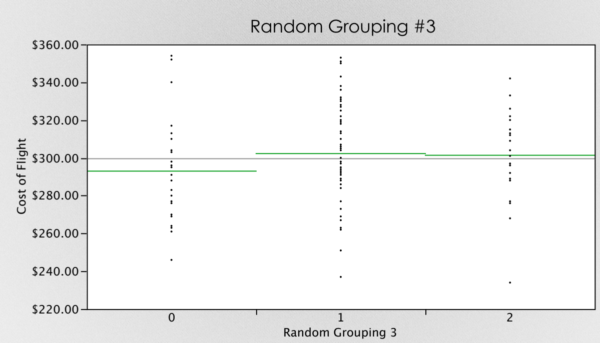
The General Linear Test

$$F = \frac{\left(\dfrac{\text{Reduction in Error}}{\#\ \text{Added Parameters}}\right)}{\textit{Baseline}\ \text{Error}}$$


Grouping based on Airline
$360.00
$340.00
$320.00
$300.00
$280.00
$260.00
$240.00
$220.00
Cost of Flight
Delta
Southwest
Virgin America
Airline

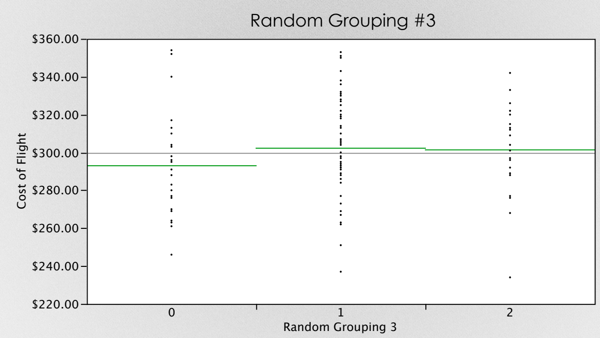

Random Grouping #1
$360.00
$340.00
$320.00
$300.00
$280.00
$260.00
$240.00
$220.00
Cost of Flight
0
1
2
Random Grouping 1

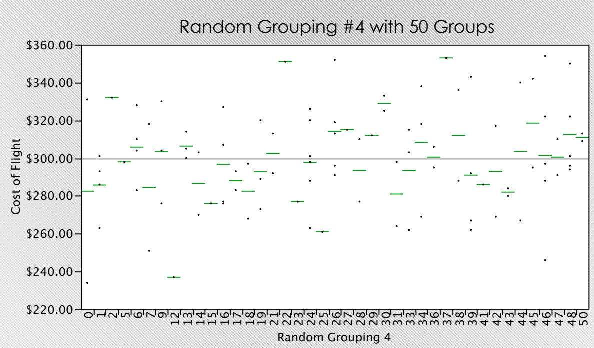

Random Grouping #2
$360.00
$340.00
$320.00
$300.00
$280.00
$260.00
$240.00
$220.00
Cost of Flight
0
1
2
Random Grouping 2


Random Grouping #3
$360.00
$340.00
$320.00
$300.00
$280.00
$260.00
$240.00
$220.00
Cost of Flight
0
1
2
Random Grouping 3

The General Linear Test

$$F = \frac{\left(\dfrac{\text{Reduction in Error}}{\#\ \text{Added Parameters}}\right)}{\textit{Baseline}\ \text{Error}}$$





The General Linear Test

$$F = \frac{\left(\dfrac{\text{Reduction in Error}}{\#\ \text{Added Parameters}}\right)}{\textit{Baseline}\ \text{Error}}$$

The General Linear Test

$$F = \frac{\left(\dfrac{\boxed{\text{Reduction in Error}}}{\text{\# Added Parameters}}\right)}{\textit{Baseline}\ \text{Error}}$$

The General Linear Test

$$F = \frac{\left(\dfrac{SS_e(R) - SS_e(F)}{\text{\# Added Parameters}}\right)}{\textit{Baseline}\ \text{Error}}$$

The General Linear Test

$$F = \frac{\left(\dfrac{SS_e(R) - SS_e(F)}{\boxed{\text{\# Added Parameters}}}\right)}{\textit{Baseline}\ \text{Error}}$$

The General Linear Test

$$F = \frac{\left(\dfrac{SS_e(R) - SS_e(F)}{df_e(R) - df_e(F)}\right)}{\textit{Baseline}\ \text{Error}}$$

The General Linear Test

$$F = \frac{\left(\dfrac{SS_e(R) - SS_e(F)}{df_e(R) - df_e(F)}\right)}{\boxed{\textit{Baseline}\ \text{Error}}}$$

Full (unrestricted) Model

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

Reduced (restricted) Model

$$H_0 : \tau_1 = \tau_2 = ... = \tau_j = 0$$

$$Y_i = \mu + \varepsilon_i$$

Full (unrestricted) Model

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

Reduced (restricted) Model

$$H_0 : \tau_1 = \tau_2 = ... = \tau_j = 0$$

$$Y_i = \mu + \varepsilon_i$$

Full (unrestricted) Model

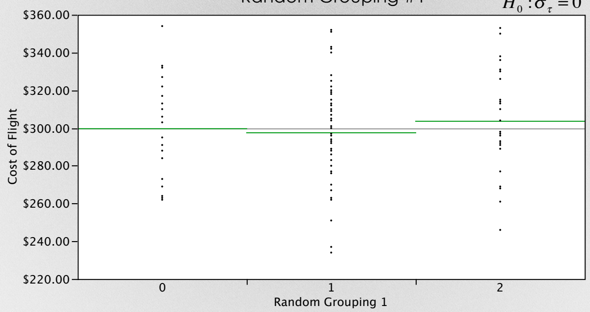
$$Y_{ij} = \mu + \tau_j + \boxed{\varepsilon_{ij}}$$

Reduced (restricted) Model

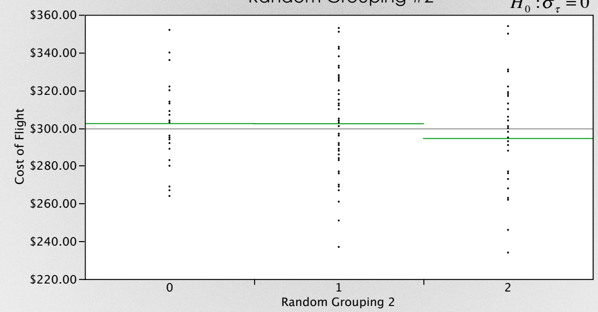
$$H_0 : \tau_1 = \tau_2 = ... = \tau_j = 0$$

$$Y_i = \mu + \varepsilon_i$$

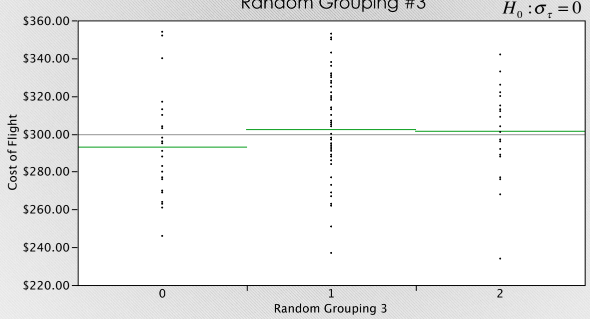
Random Grouping #1 $H_0 : \sigma_\tau = 0$



Random Grouping #2 $H_0 : \sigma_\tau = 0$

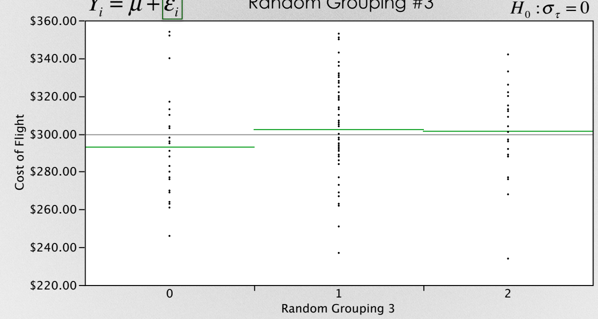


Random Grouping #3 $H_0 : \sigma_\tau = 0$



$$Y_{ij} = \mu + \tau_j + \boxed{\varepsilon_{ij}}$$
$$Y_i = \mu + \boxed{\varepsilon_i}$$

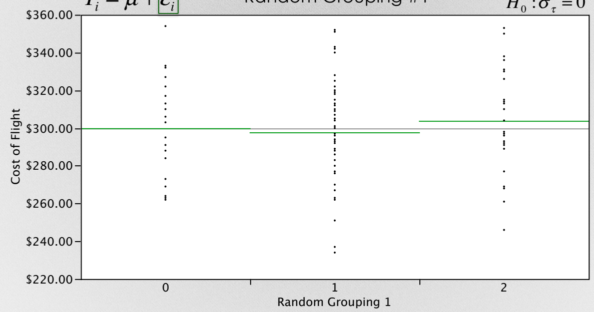
Random Grouping #3 $H_0 : \sigma_\tau = 0$



$$Y_{ij} = \mu + \tau_j + \boxed{\varepsilon_{ij}}$$
$$Y_i = \mu + \boxed{\varepsilon_i}$$

Random Grouping #1

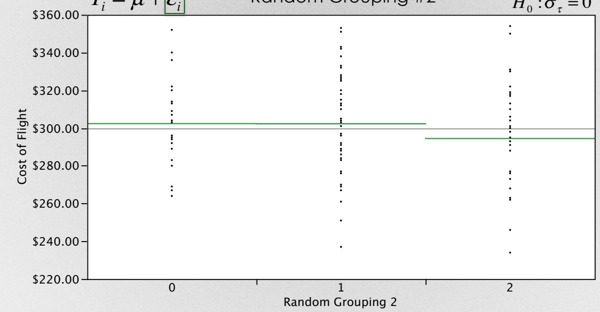
$H_0 : \sigma_\tau = 0$



$$Y_{ij} = \mu + \tau_j + \boxed{\varepsilon_{ij}}$$
$$Y_i = \mu + \boxed{\varepsilon_i}$$

Random Grouping #2

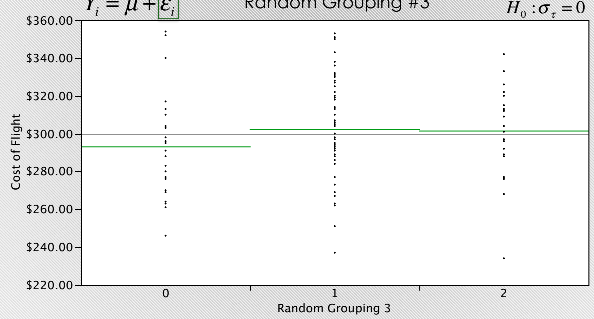
$H_0 : \sigma_\tau = 0$

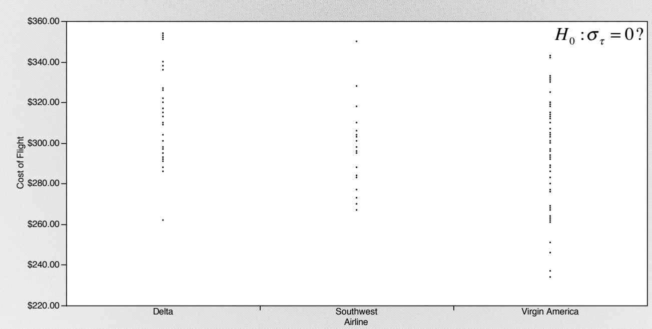


$$Y_{ij} = \mu + \tau_j + \boxed{\varepsilon_{ij}}$$
$$Y_i = \mu + \boxed{\varepsilon_i}$$

Random Grouping #3

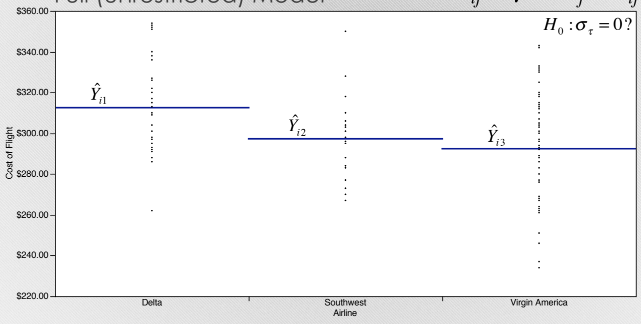
$H_0 : \sigma_\tau = 0$





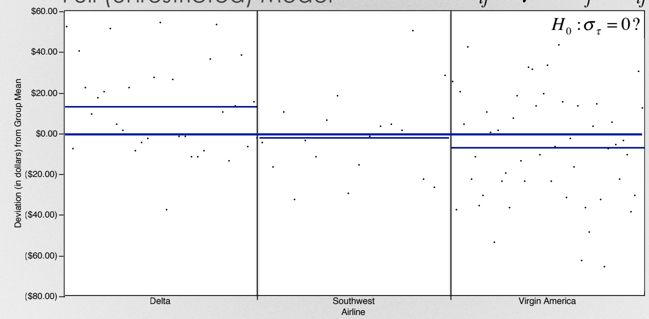
Full (unrestricted) Model

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$



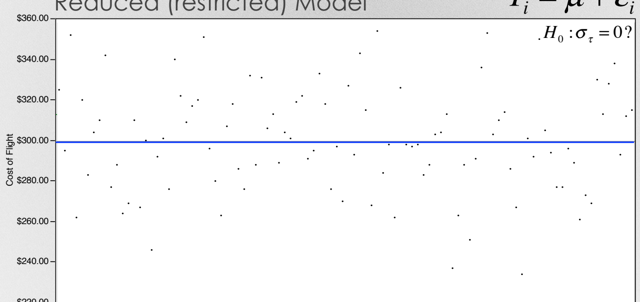
Full (unrestricted) Model

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$



Reduced (restricted) Model

$$Y_i = \mu + \varepsilon_i$$



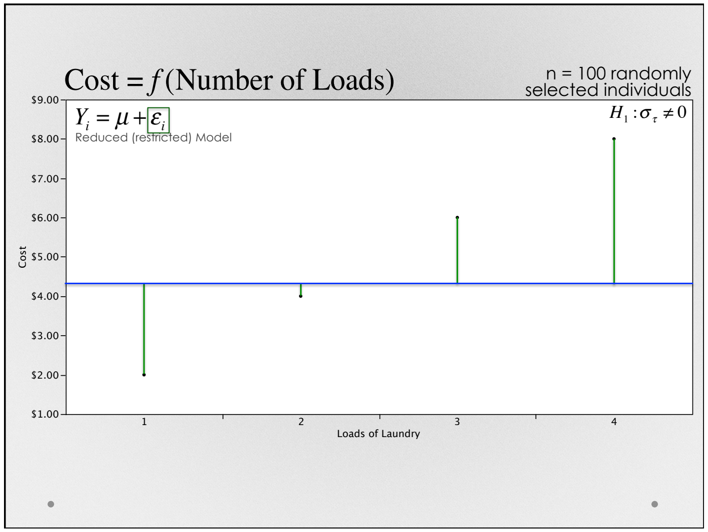
Full (unrestricted) Model

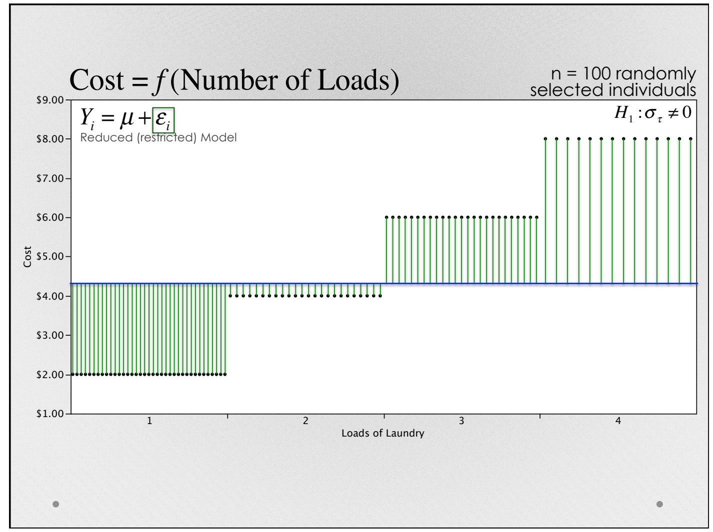
$$Y_{ij} = \mu + \tau_j + \boxed{\varepsilon_{ij}}$$

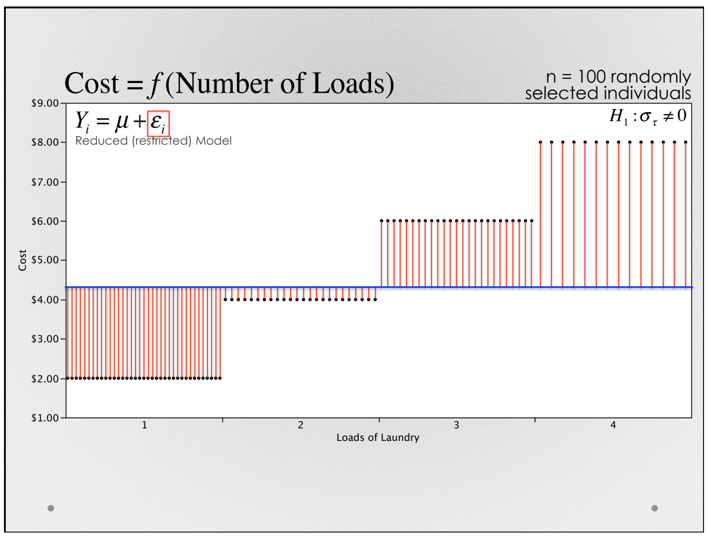
Reduced (restricted) Model

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_j = 0$$

$$Y_i = \mu + \varepsilon_i$$







Full (unrestricted) Model

$$Y_{ij} = \mu + \tau_j + \boxed{\varepsilon_{ij}}$$

Reduced (restricted) Model

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_j = 0$$

$$Y_i = \mu + \varepsilon_i$$

Full (unrestricted) Model

$$Y_{ij} = \mu + \boxed{\tau_j} + \boxed{\varepsilon_{ij}}$$

Reduced (restricted) Model

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_j = 0$$

$$Y_i = \mu + \varepsilon_i$$

Full (unrestricted) Model

$$Y_{ij} = \mu + \boxed{\tau_j} + \boxed{\varepsilon_{ij}}$$

Reduced (restricted) Model

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_j = 0$$

$$Y_i = \mu + \boxed{\varepsilon_i}$$

The General Linear Test

$$F = \frac{\left(\dfrac{SS_e(R) - SS_e(F)}{df_e(R) - df_e(F)}\right)}{\boxed{Baseline \text{ Error}}}$$

The General Linear Test

$$F = \frac{\left(\dfrac{SS_e(R) - SS_e(F)}{df_e(R) - df_e(F)}\right)}{\left(\dfrac{SS_e(F)}{df_e(F)}\right)}$$

The General Linear Test

$$F = \frac{\left(\dfrac{\boxed{SS_e(R)} - SS_e(F)}{\boxed{df_e(R)} - df_e(F)}\right)}{\left(MS_{error}\right)}$$

Full (unrestricted) Model

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

Reduced (restricted) Model

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_j = 0$$

$$Y_i = \mu + \varepsilon_i$$

Sums of Squares Total

FROM BEFORE

$$SS_{total} = \sum_j \sum_i (Y_{ij} - \bar{Y})^2$$

The sums of squares total is the sum of the squared distance for each individual's **actual score** and the **grand mean**

These are the sums of squares you would get if ignored your factor and found SS for the Y variable alone

The General Linear Test

$$F = \frac{\left(\dfrac{\boxed{SS_e(R)} - SS_e(F)}{\boxed{df_e(R)} - df_e(F)}\right)}{\left(MS_{error}\right)}$$

The General Linear Test

$$F = \frac{\left(\dfrac{SS_{total} - \boxed{SS_e(F)}}{df_{total} - \boxed{df_e(F)}}\right)}{\left(MS_{error}\right)}$$

Full (unrestricted) Model

$$Y_{ij} = \mu + \tau_j + \varepsilon_{ij}$$

Reduced (restricted) Model

$$H_0 : \tau_1 = \tau_2 = \ldots = \tau_j = 0$$

$$Y_i = \mu + \varepsilon_i$$

FROM BEFORE

Sums of Squares for Error

$$SS_{error} = \sum_j \sum_i (Y_{ij} - \hat{Y}_{ij})^2$$

The sums of squares for error is the sum of the squared distance from each individuals **actual score** and their **predicted score**

Also called the sums of squares within
SS_{within}

The General Linear Test

$$F = \frac{\left(\dfrac{SS_{total} - \boxed{SS_e(F)}}{df_{total} - \boxed{df_e(F)}}\right)}{\left(MS_{error}\right)}$$

The General Linear Test

$$F = \frac{\left(\frac{\boxed{SS_{total} - SS_{error}}}{\boxed{df_{total} - df_{error}}}\right)}{\left(MS_{error}\right)}$$

FROM BEFORE

$$SS_{error} = \sum_j \sum_i (\boxed{Y_{ij}} - \boxed{\hat{Y}_{ij}})^2$$

$$+ \; SS_{treat} = \sum_j \sum_i (\boxed{\hat{Y}_{ij}} - \boxed{\bar{Y}})^2$$

$$= \; SS_{total} = \sum_j \sum_i (\boxed{Y_{ij}} - \boxed{\bar{Y}})^2$$

FROM BEFORE

Degrees of Freedom for the Analysis of Variance

$$df_{error} + df_{treatment} = df_{total}$$

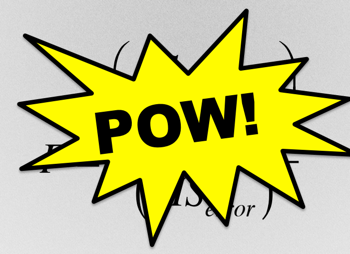
The General Linear Test

$$F = \frac{\left(\frac{\boxed{SS_{total} - SS_{error}}}{\boxed{df_{total} - df_{error}}}\right)}{\left(MS_{error}\right)}$$

The General Linear Test

$$F = \frac{\left(\frac{SS_{treatment}}{df_{treatment}}\right)}{\left(MS_{error}\right)}$$

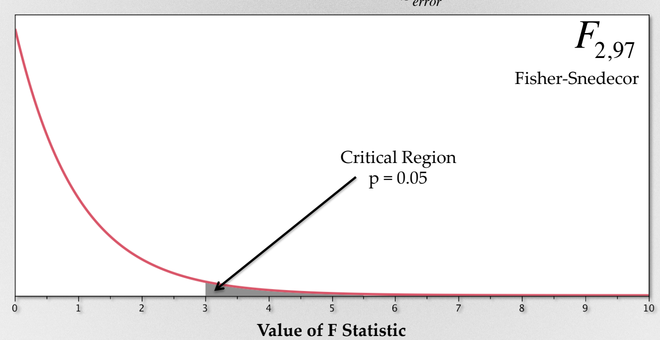
The General Linear Test



The General Linear Test

$$F = \frac{MS_{treatment}}{MS_{error}}$$

If H_0 is true $(\tau_1 = \tau_2 = \ldots = \tau_j = 0)$: $\frac{MS_{treatment}}{MS_{error}} \sim F_{\upsilon_1,\upsilon_2}$



FROM BEFORE

Statistical Inference with Linear Models

- Analysis of Variance Approach
- General Linear Test Approach
  (model comparison)