Stat 411/511

ANOVA F-Test

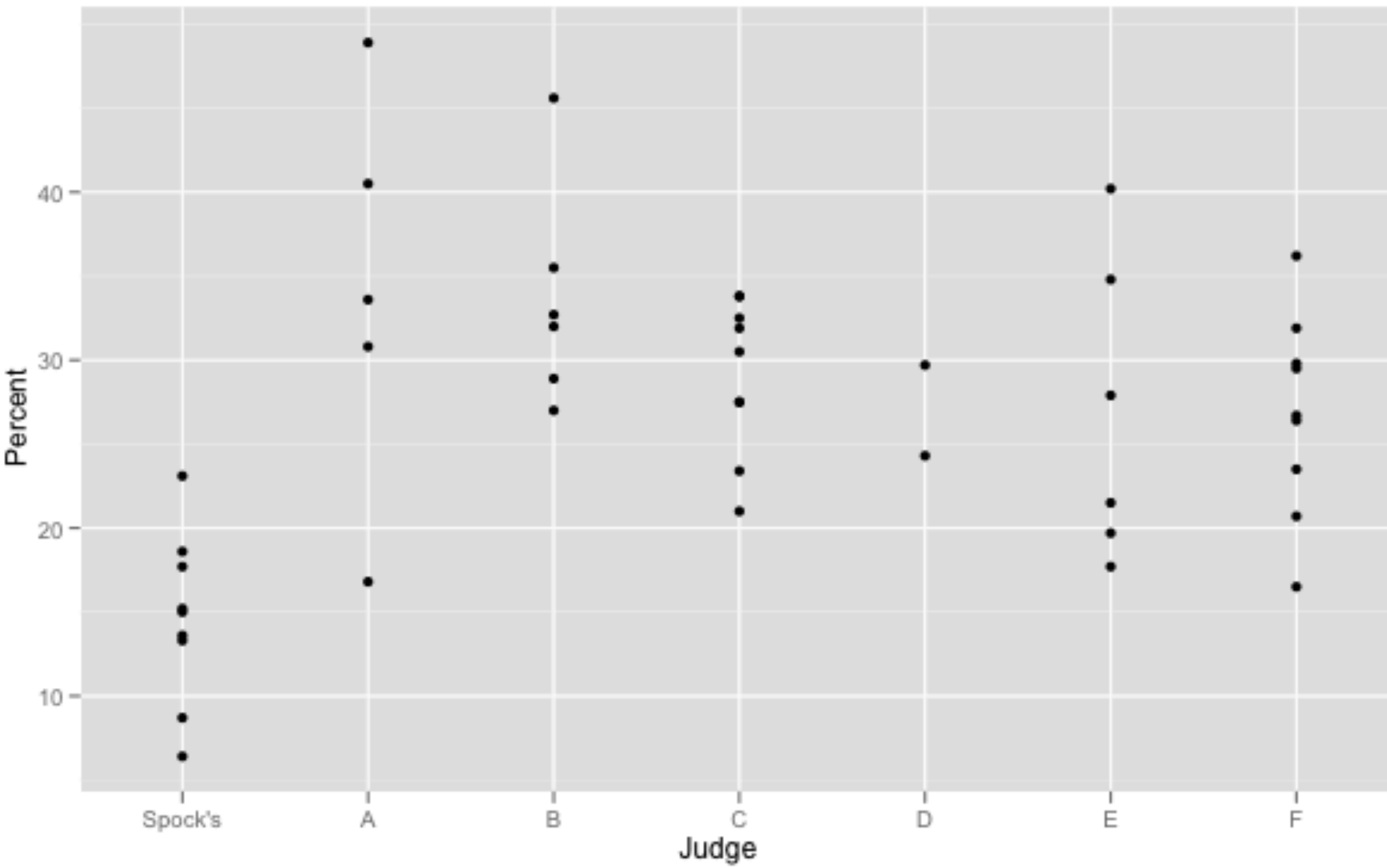
Feb 20 2012

Charlotte Wickham

stat511.cwick.co.nz

```
library(Sleuth2)
library(ggplot2)
qplot(Judge, Percent, data = case0502)
```

Spock Trial case study



A series of tests

A series of tests

Null # 1: All the means are the same.

Initial screening. If we can't reject this null, then there isn't much point proceeding.

A series of tests

Null # 1: All the means are the same.

Initial screening. If we can't reject this null, then there isn't much point proceeding.

Null # 2: The other judges all have the same mean (Spock's is different).

If we fail to reject this null, it is reasonable to compare Spock's average to the average of all of the other judges.

A series of tests

Null # 1: All the means are the same.

Initial screening. If we can't reject this null, then there isn't much point proceeding.

Where you generally always start with multiple groups.

Null # 2: The other judges all have the same mean (Spock's is different).

If we fail to reject this null, it is reasonable to compare Spock's average to the average of all of the other judges.

A series of tests

Null # 1: All the means are the same.

Initial screening. If we can't reject this null, then there isn't much point proceeding.

Where you generally always start with multiple groups.

Null # 2: The other judges all have the same mean (Spock's is different).

If we fail to reject this null, it is reasonable to compare Spock's average to the average of all of the other judges.

Problem specific.

Extra sum of squares F-test

Are any of the means different?

Two models

Two models

Full model: a model that fully describes the data.

Two models

Full model: a model that fully describes the data.

All the means are different

Two models

Full model: a model that fully describes the data.

All the means are different

Restricted model: a restriction of the full model imposed by the null hypothesis.

Two models

Full model: a model that fully describes the data.

All the means are different

Restricted model: a restriction of the full model imposed by the null hypothesis.

Null #1: All the means are the same.

Two models

Full model: a model that fully describes the data.

All the means are different

Restricted model: a restriction of the full model imposed by the null hypothesis.

Null #1: All the means are the same.

Null #2: All the means of the other judges are the same, Spock's is different.

Null #1

Null: the population mean percentage of women are the same for all the judges.

$\mu_{Spock} = \mu_A = \mu_B = \mu_C = \mu_D = \mu_E = \mu_F = \mu$

Alternative: At least one of the judges has a different population mean percentage of women.

Your turn

How could you write Null #2 with mathematical symbols?

Null #2: All the means of the other judges are the same, Spock's is different.

What is the alternative hypothesis?

Null #2

Null: the population mean percentage of women is the same for all the **other** judges.

$\mu_{Spock} = \mu_1, \ \mu_A = \mu_B = \mu_C = \mu_D = \mu_E = \mu_F = \mu_0$

Alternative: At least one of the **other** judges has a different population mean percentage of women.

Fit both models

Fit both models

The **Extra Sum of Squares F-test**, compares how well the models fit, by comparing the sum of squared **residuals** of each model.

Fit both models

The **Extra Sum of Squares F-test**, compares how well the models fit, by comparing the sum of squared **residuals** of each model.

Residual = Y_i - Estimate of μ_i

Fit both models

The **Extra Sum of Squares F-test**, compares how well the models fit, by comparing the sum of squared **residuals** of each model.

Residual = Y_i - Estimate of μ_i

Take each observation and subtract our best guess for it's mean under each model.

Fit both models

The **Extra Sum of Squares F-test**, compares how well the models fit, by comparing the sum of squared **residuals** of each model.

Residual = Y_i - Estimate of μ_i

Take each observation and subtract our best guess for it's mean under each model.

Residual Sum of Squares: Square each residual and add them up

Null #1

Null #1

Restricted model residuals (a single mean):

Null #1

Restricted model residuals (a single mean):

= Observation - best guess for a single mean

Null #1

Restricted model residuals (a single mean):

= Observation - best guess for a single mean

= Observation - $\overline{Y}$

Null #1

Restricted model residuals (a single mean):

= Observation - best guess for a single mean

= Observation - $\bar{Y}$

Full model residuals (a mean of each group):

Null #1

Restricted model residuals (a single mean):

= Observation - best guess for a single mean

= Observation - $\bar{Y}$

Full model residuals (a mean of each group):

= Observation - best guess for group mean

Null #1

Restricted model residuals (a single mean):

= Observation - best guess for a single mean

= Observation - $\bar{Y}$

Full model residuals (a mean of each group):

= Observation - best guess for group mean

= Observation - $\bar{Y}_i$

Null #1

Restricted model residuals (a single mean):

= Observation - best guess for a single mean

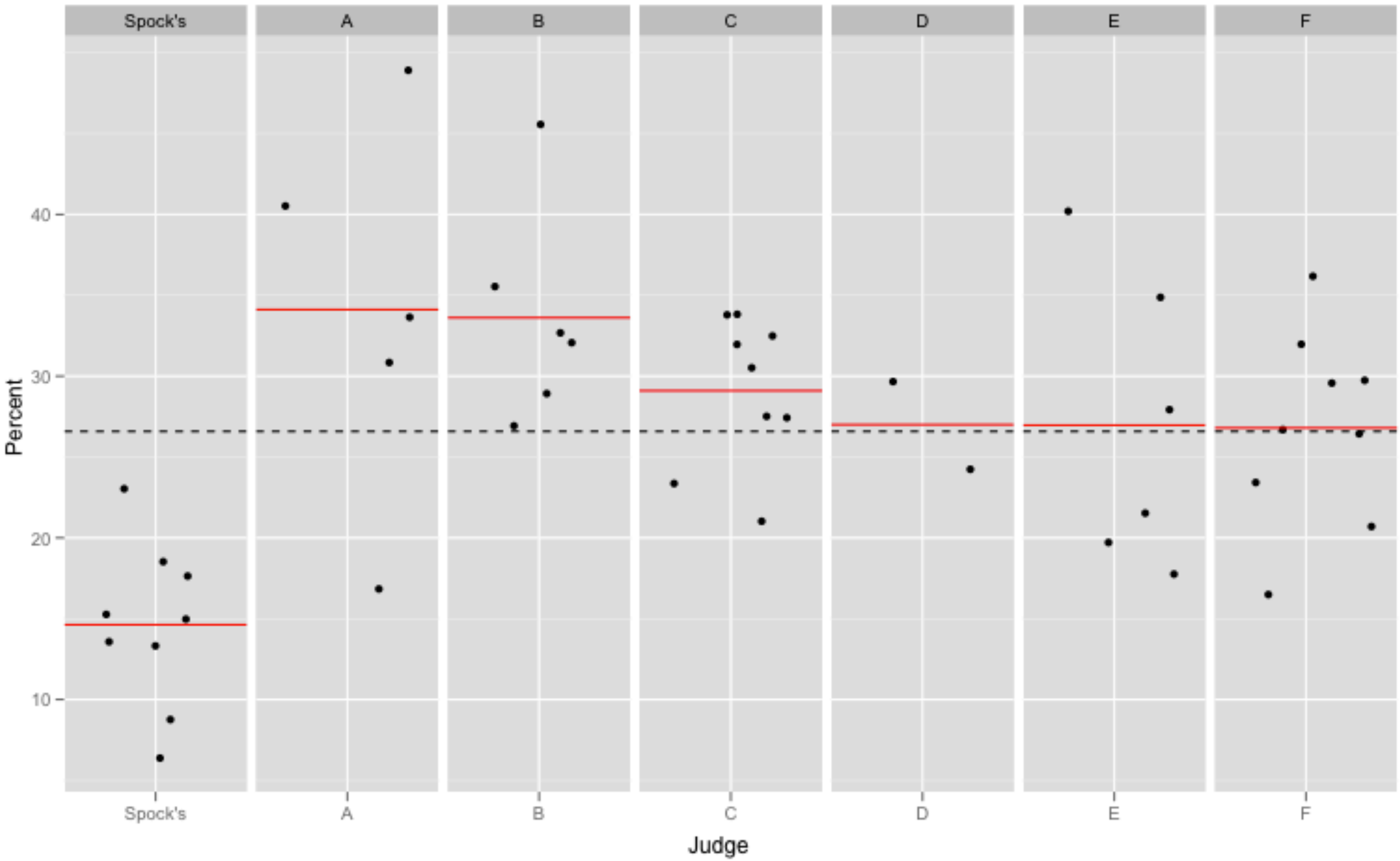
= Observation - $\bar{Y}$

Full model residuals (a mean of each group):

= Observation - best guess for group mean

= Observation - $\bar{Y}_i$ ⟵ Sample average for the group

Null #1: sum of squares illustration



```
> case0502$group_average <- with(case0502, ave(Percent, Judge))
> case0502$overall_average <- with(case0502, mean(Percent))
> head(case0502)
  Percent   Judge group_average overall_average
1     6.4 Spock's      14.62222        26.58261
2     8.7 Spock's      14.62222        26.58261
3    13.3 Spock's      14.62222        26.58261
4    13.6 Spock's      14.62222        26.58261
5    15.0 Spock's      14.62222        26.58261
6    15.2 Spock's      14.62222        26.58261
```

```
> case0502$group_average <- with(case0502, ave(Percent, Judge))
> case0502$overall_average <- with(case0502, mean(Percent))
```

```
> head(case0502)
  Percent   Judge group_average overall_average
1     6.4 Spock's      14.62222        26.58261
2     8.7 Spock's      14.62222        26.58261
3    13.3 Spock's      14.62222        26.58261
4    13.6 Spock's      14.62222        26.58261
5    15.0 Spock's      14.62222        26.58261
6    15.2 Spock's      14.62222        26.58261
```

```
> with(case0502, sum((Percent - overall_average)^2))
[1] 3791.526
```

```
> with(case0502, sum((Percent - group_average)^2))
[1] 1864.445
```

```
> case0502$group_average <- with(case0502, ave(Percent, Judge))
> case0502$overall_average <- with(case0502, mean(Percent))
```

```
> head(case0502)
  Percent   Judge group_average overall_average
1     6.4 Spock's      14.62222        26.58261
2     8.7 Spock's      14.62222        26.58261
3    13.3 Spock's      14.62222        26.58261
4    13.6 Spock's      14.62222        26.58261
5    15.0 Spock's      14.62222        26.58261
6    15.2 Spock's      14.62222        26.58261
```

```
> with(case0502, sum((Percent - overall_average)^2))
[1] 3791.526
```

1 mean

d.f. = n - 1

```
> with(case0502, sum((Percent - group_average)^2))
[1] 1864.445
```

```
> case0502$group_average <- with(case0502, ave(Percent, Judge))
> case0502$overall_average <- with(case0502, mean(Percent))
```

```
> head(case0502)
  Percent   Judge group_average overall_average
1     6.4 Spock's      14.62222        26.58261
2     8.7 Spock's      14.62222        26.58261
3    13.3 Spock's      14.62222        26.58261
4    13.6 Spock's      14.62222        26.58261
5    15.0 Spock's      14.62222        26.58261
6    15.2 Spock's      14.62222        26.58261
```

```
> with(case0502, sum((Percent - overall_average)^2))
[1] 3791.526
```

1 mean

d.f. = n - 1

```
> with(case0502, sum((Percent - group_average)^2))
[1] 1864.445
```

I means

d.f. = n - I

```
> case0502$group_average <- with(case0502, ave(Percent, Judge))
> case0502$overall_average <- with(case0502, mean(Percent))
```

```
> head(case0502)
  Percent   Judge group_average overall_average
1     6.4 Spock's      14.62222        26.58261
2     8.7 Spock's      14.62222        26.58261
3    13.3 Spock's      14.62222        26.58261
4    13.6 Spock's      14.62222        26.58261
5    15.0 Spock's      14.62222        26.58261
6    15.2 Spock's      14.62222        26.58261
```

```
> with(case0502, sum((Percent - overall_average)^2))
[1] 3791.526
```

1 mean

d.f. = n - 1

```
> with(case0502, sum((Percent - group_average)^2))
[1] 1864.445
```

I means

d.f. = n - I

Is this a big enough "reduction" to give evidence the against the null hypothesis?

Extra Sum of Squares =

Extra Sum of Squares =

Residual sum of squares from reduced model **-**

Residual sum of squares from full model =

Extra Sum of Squares =

Residual sum of squares from reduced model **-**

Residual sum of squares from full model =

reduction in residual sum of squares

Extra Sum of Squares =

Residual sum of squares from reduced model **-**

Residual sum of squares from full model **=**

reduction in residual sum of squares

Extra degrees of freedom =

Extra Sum of Squares =

Residual sum of squares from reduced model **-** Residual sum of squares from full model =

reduction in residual sum of squares

Extra degrees of freedom =

degrees of freedom in reduced model **-** degrees of freedom in full model =

Extra Sum of Squares =

Residual sum of squares from reduced model **-** Residual sum of squares from full model =

reduction in residual sum of squares

Extra degrees of freedom =

degrees of freedom in reduced model **-** degrees of freedom in full model =

reduction in degrees of freedom

Extra Sum of Squares =

Residual sum of squares from reduced model **-**
Residual sum of squares from full model =

reduction in residual sum of squares

Extra degrees of freedom =

degrees of freedom in reduced model **-** degrees of freedom in full model =

reduction in degrees of freedom

or how many more parameters does the full model have?

F-statistic

$$\frac{\text{Extra Sum of Squares/Extra degrees of freedom}}{\hat{\sigma^2}_{\text{full}}}$$

F-statistic

$$\frac{\text{Extra Sum of Squares/Extra degrees of freedom}}{\hat{\sigma}^2_{\text{full}}}$$

← Plug in s_p^2

F-statistic

$$\frac{\text{Extra Sum of Squares/Extra degrees of freedom}}{\hat{\sigma}^2_{\text{full}}}$$

← Plug in s_p^2

How much did each extra parameter reduce the "poorness" of fit of the reduced model scaled by the within group variation?

F-statistic

$$\frac{\text{Extra Sum of Squares/Extra degrees of freedom}}{\hat{\sigma}^2_{\text{full}}}$$

← Plug in s_p^2

How much did each extra parameter reduce the "poorness" of fit of the reduced model scaled by the within group variation?

Under the null hypothesis (the reduced model is true) has F-distribution with `I - 1` & `n-I` degrees of freedom

F-statistic

$$\frac{\text{Extra Sum of Squares/Extra degrees of freedom}}{\hat{\sigma}^2_{\text{full}}}$$

← Plug in s_p^2

How much did each extra parameter reduce the "poorness" of fit of the reduced model scaled by the within group variation?

Large values give evidence against the null hypothesis.

Under the null hypothesis (the reduced model is true) has F-distribution with $I - 1$ & $n-I$ degrees of freedom

The ANOVA table

Display 5.10 **p. 127**

Analysis of variance table: a test for equal mean percents of women in venires of seven judges; Spock data

Source of Variation		Sum of Squares	df	Mean Square	F-Statistic	p-value
Between Groups						
Within Groups	Full model	1,864.45	39			
Total	Reduced model	3,791.53	45			

1. *Sum of squared residuals from fitting the full (separate-means) model*
2. *Sum of squared residuals from fitting the reduced (equal-means) model*
3. *degrees of freedom* (*n-I*, *n-1*)
4. *Subtract the "Within" from the "Total "*
5. *A mean square is the ratio of a sum-of-squares to its degrees of freedom*
6. *The F-statistic is the ratio of the Between MS to the Within MS*
7. *The p-value comes from an F-distribution with and df*

***NOTE**: This is s_p^2*

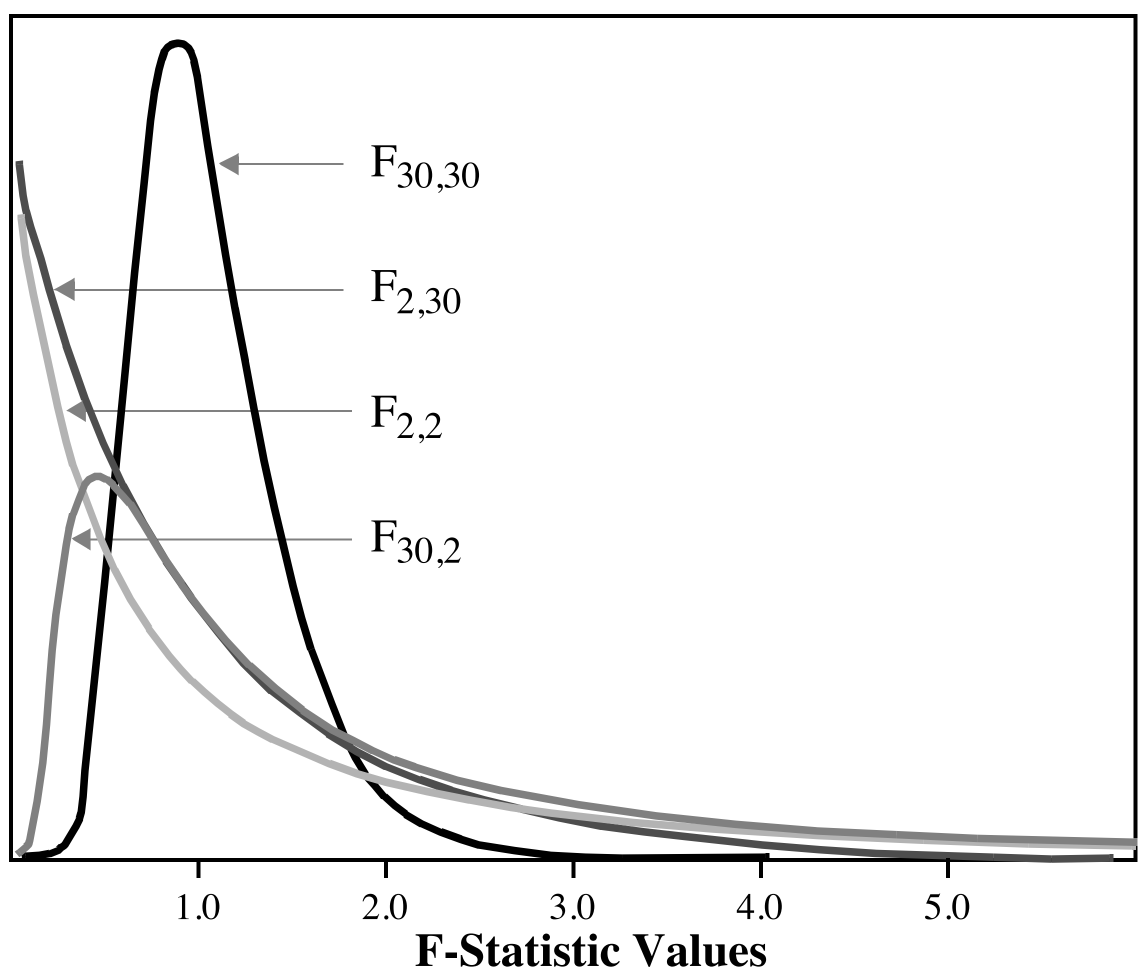
```
> 1 - pf(6.72, 6, 39)
[1] 6.082464e-05
```

=0.00061

The p-value is always one-sided for the F-test

Display 5.9 **p. 126**

Four F-distributions, having different degrees of freedom



Null #1

We have convincing evidence that at least one judge has a different mean percentage of women on their venires.

Null #2

Null: the population mean percentage of women is the same for all the **other** judges.

$\mu_{Spock} = \mu_1, \ \mu_A = \mu_B = \mu_C = \mu_D = \mu_E = \mu_F = \mu_0$

```
> # create new Spock or Not variable
> case0502$two_groups <- ifelse(case0502$Judge == "Spock's", "Spock's", "Other")

> # find two group averages
> case0502$two_group_average <- with(case0502, ave(Percent, two_groups))

> # sum of squared residuals
> with(case0502, sum((Percent - two_group_average)^2))
[1] 2190.903
```

Your Turn: Fill in the rest of the ANOVA table

Display 5.10 **p. 127**

Analysis of variance table: a test for equal mean percents of women in venires of seven judges; Spock data

Source of Variation	Sum of Squares	df	Mean Square	F-Statistic	p-value
Between Groups					
Within Groups	1864.45	39			
Total	2190.90				

1. *Sum of squared residuals from fitting the full (separate-means) model*
2. *Sum of squared residuals from fitting the reduced (means) model*
3. *degrees of freedom* — *n-I*, *n-*
4. *Subtract the "Within" from the "Total "*
5. *A mean square is the ratio of a sum-of-squares to its degrees of freedom*
6. *The F-statistic is the ratio of the Between MS to the Within MS*
7. *The p-value comes from an F-distribution with and df*

***NOTE**: This is s_p^2*

Analysis of variance table: a test for equal mean percents of women in venires of seven judges; Spock data

Source of Variation	Sum of Squares	df	Mean Square	F-Statistic	p-value
Between Groups	326.45	5	65.29	1.37	0.26
Within Groups	1864.45	39	47.81		
Total	2190.90	44			

1 *Sum of squared residuals from fitting the full (separate-means) model*

2 *Sum of squared residuals from fitting the reduced (means) model*

3 *degrees of freedom*

n-I

n-

4 *Subtract the "Within" from the "Total "*

5 *A mean square is the ratio of a sum-of-squares to its degrees of freedom*

6 *The F-statistic is the ratio of the Between MS to the Within MS*

7 *The p-value comes from an F-distribution with and df*

***NOTE**: This is s_p^2*

Null #2

We have no evidence that any of the **other** six judges have a different mean percentage of women on their venires.