Stat 411/511

One Way ANOVA

Nov 4 2015

Charlotte Wickham

stat511.cwick.co.nz

Summary from Weds

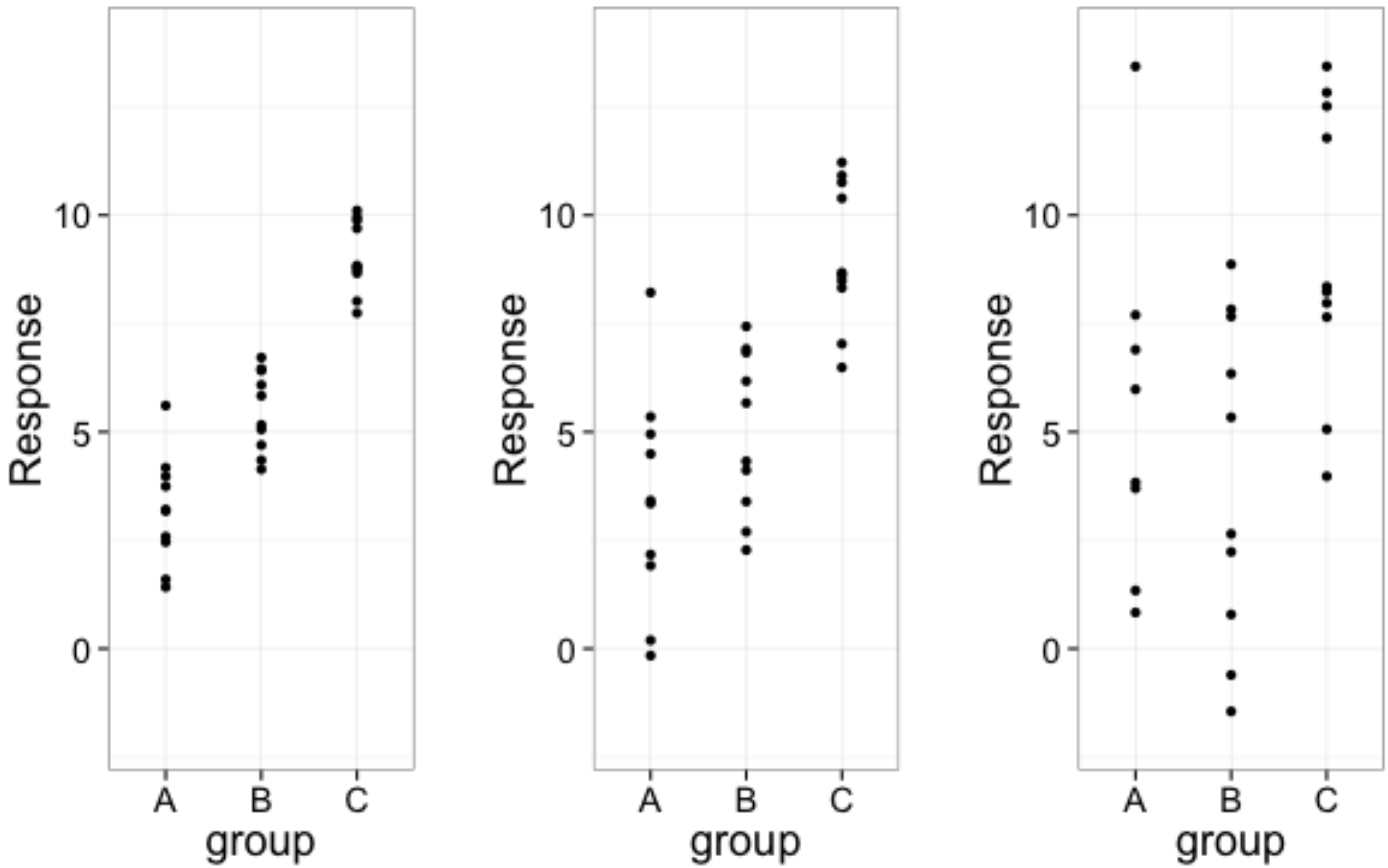
Comparing two groups

A single comparison of two group means

When you have multiple groups:

- Do the usual two-sample t-test but,
- use all groups to get the pooled standard deviation
- the degrees of freedom are n - `I`.

for now you have to do it “by hand”
not using `t.test`



Consider these three datasets.

Which would you say gives more evidence of the groups coming from populations with different means?

A one way ANOVA compares the variation in group averages to the variation of the response within groups

One-way ANOVA

Are any of the means different?

An application of the
extra sum of squares F-test

One-way ANOVA

Null: the population means are the same (OR all treatment effects are the same).

$\mu_1 = \mu_2 = \ldots = \mu_I = \mu$

Alternative: At least one of the population means is different.

This is the traditional null hypothesis in a one-way ANOVA

The extra sum of squares F-test can be used for other null hypotheses too.

Case0502

One-way ANOVA

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

$\mu_{Spock} = \mu_A = \ldots = \mu_F = \mu$

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

An extra Sums of Squares F-test compares **two models**

Full model: a model that fully describes the set of alternatives.

One-way ANOVA: All the means could be different, a.k.a. separate means model

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

One-way ANOVA: All the means are the same, a.k.a. equal means model.

Fit both models

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

The **residual** is the distance between the observation and our estimate for it's mean.

$\text{Residual}_{ij} = Y_{ij}$ - Estimate of μ_i

Sum of squared residuals aka
Residual Sum of Squares (RSS):
For each model, square each residual and add them up

Residuals in a one-way ANOVA

Single mean model residuals

(null/restricted model):

= Observation - best guess for a single mean

= Observation - $\bar{Y}$ ← overall average   d.f. = n - 1

Separate means model residuals

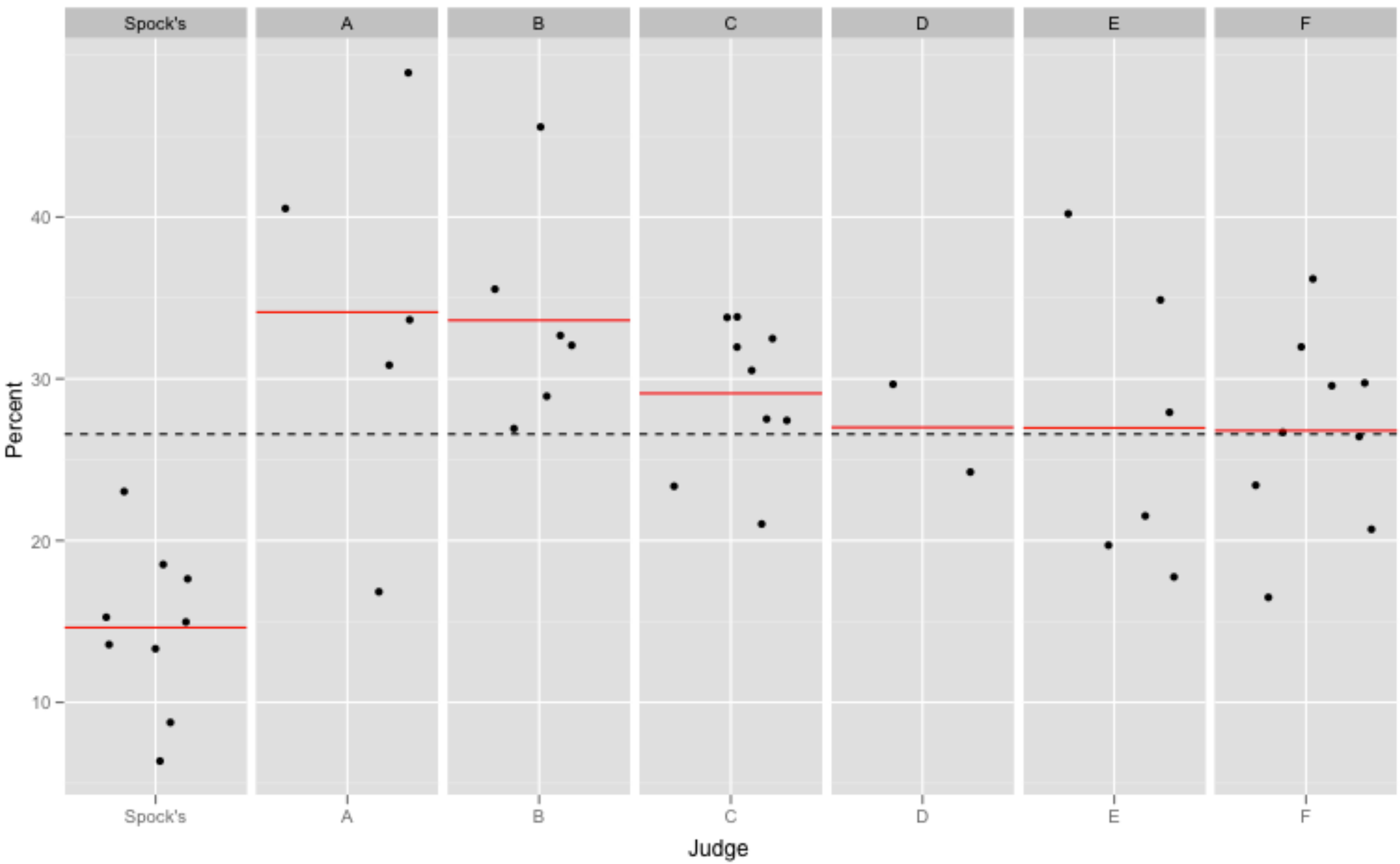
(full model):

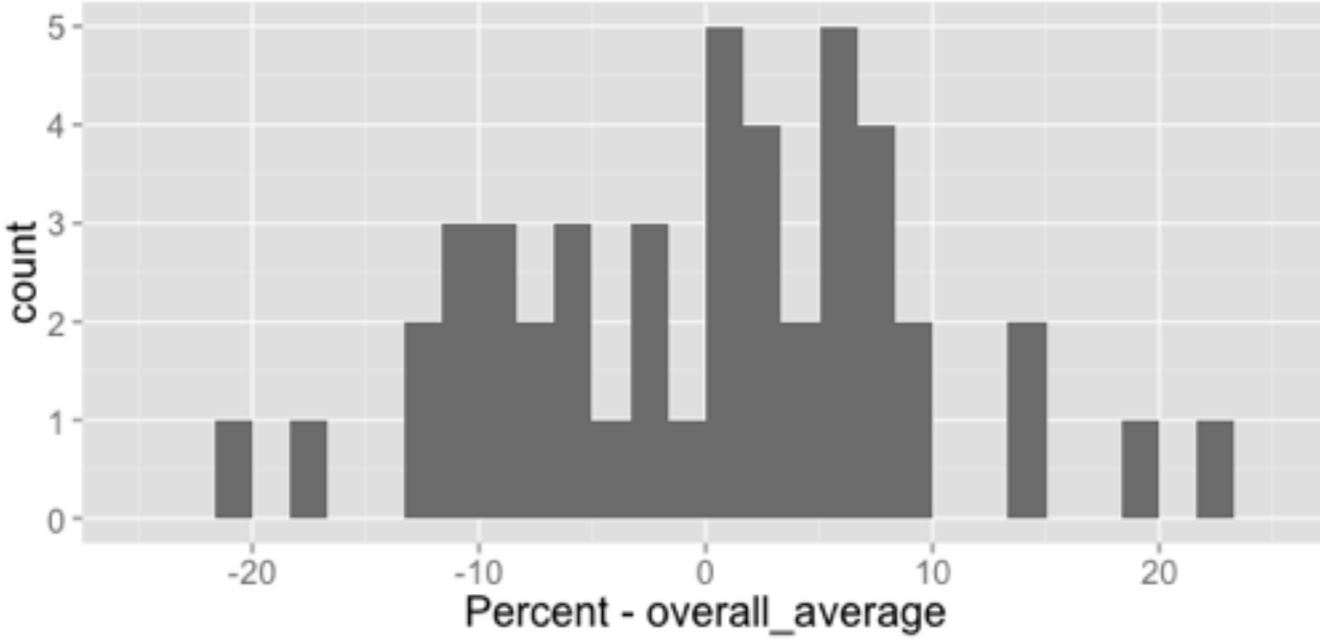
= Observation - best guess for group mean

= Observation - $\bar{Y}_i$ ← group average   d.f. = n - I

With each Residual Sum of Squares there is an associated degrees of freedom =
number of observations - number of parameters for the mean

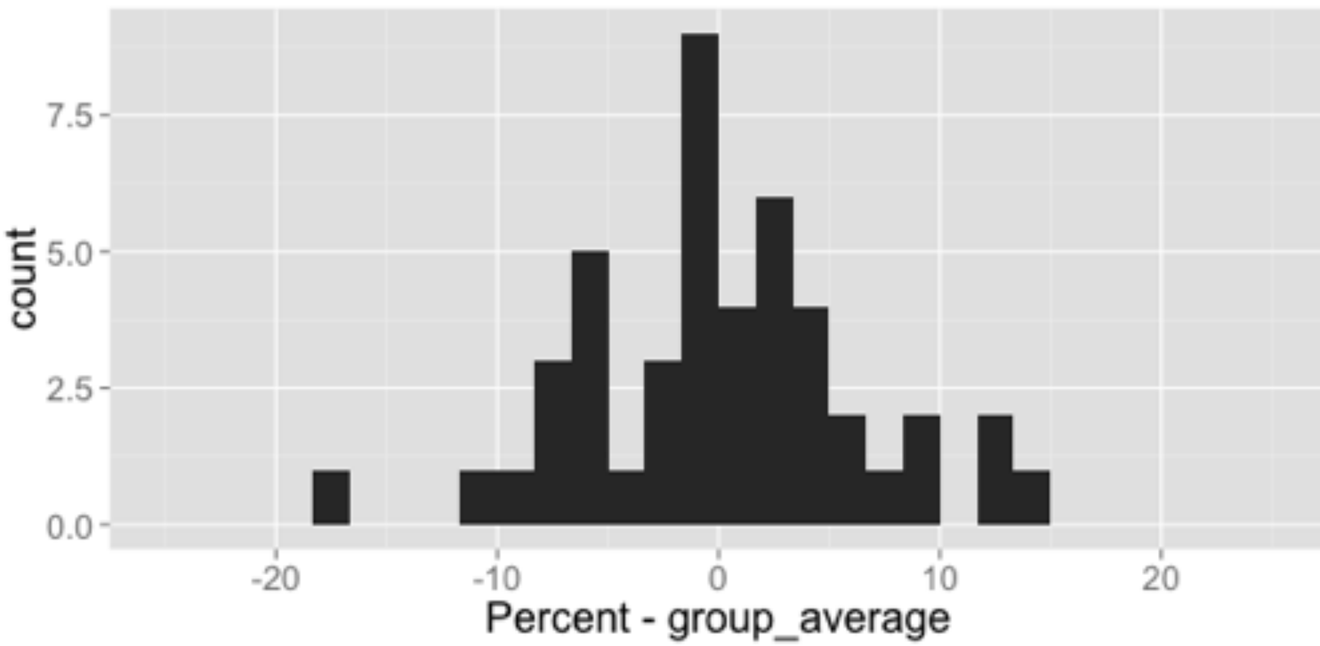
sum of squares illustration





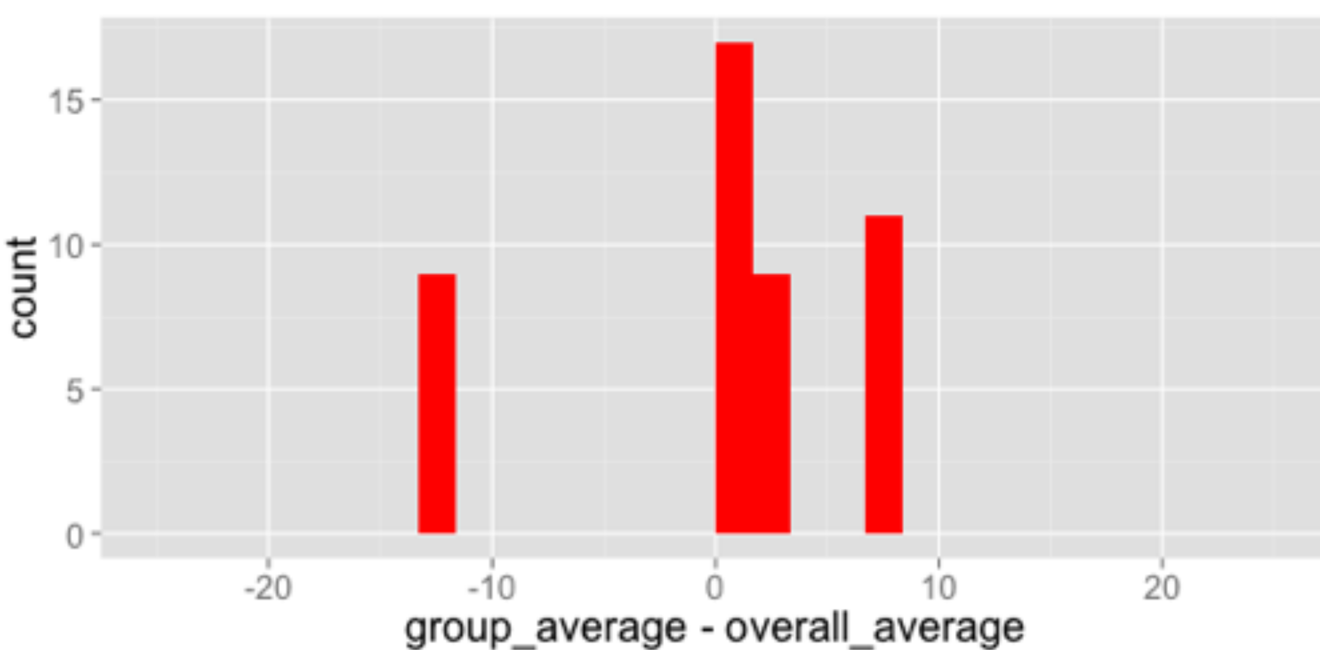
residuals from
equal means model

gives total SS



residuals from
separate means
model

gives within group SS



group averages
compared to overall
average

gives between group SS

= total SS - Within group SS

```
> 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
```

Total sum of squares

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

1 mean

d.f. = n - 1

Within group sum of squares

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

I means

d.f. = n - I

Between group sum of squares

3791.526 - 1864.445 = 1927.081

d.f. = (n-1) - (n - I)

= I - 1

Extra SS F-statistic

for one way ANOVA

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

come back to this form

$$F = \frac{\text{Between group SS/Between group d.f.}}{\text{Within group SS/Within group df}}$$

$$F = \frac{\text{Between group MSS}}{\text{Within group MSS}}$$

Large values give evidence against the null hypothesis.

Under the null hypothesis (equal means model) 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			

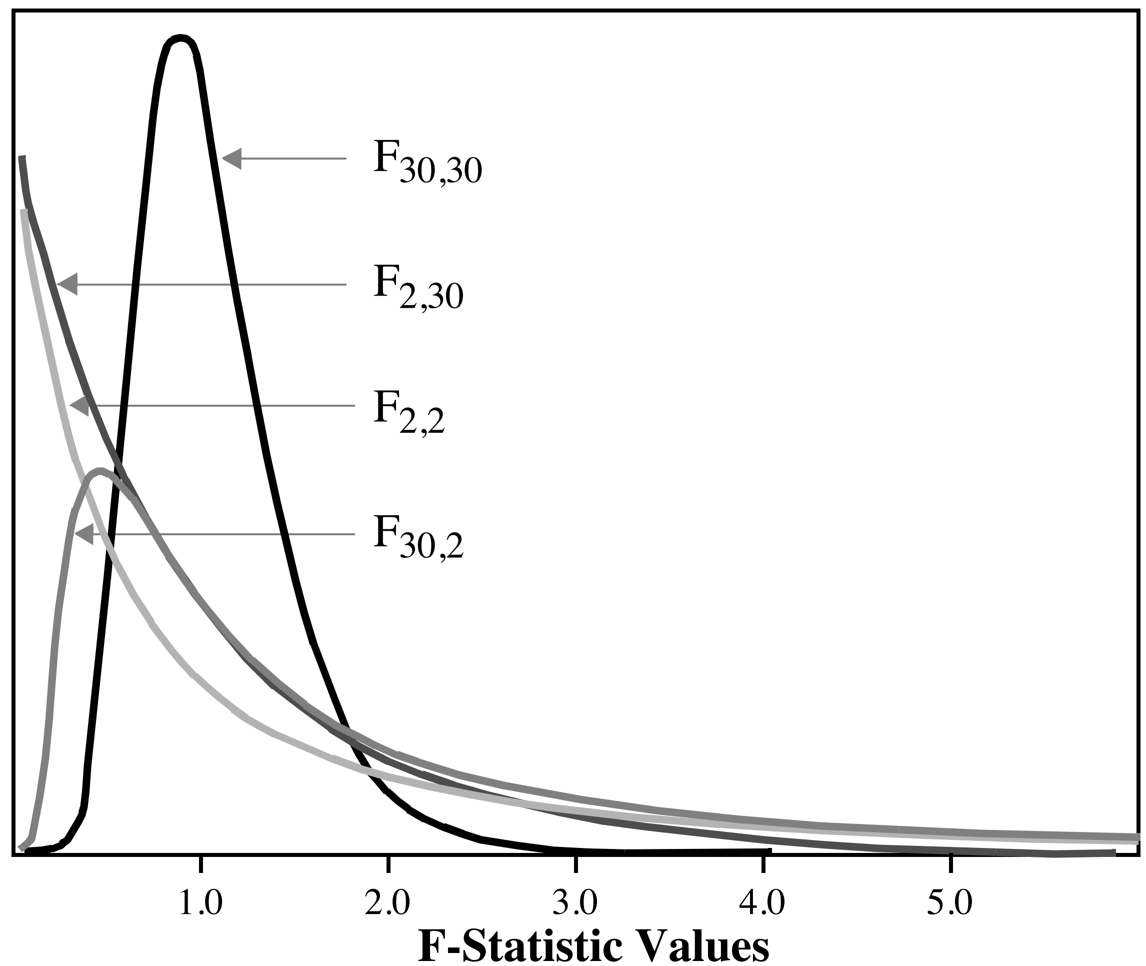
1. *Sum of squared residuals from fitting the full (separate-means) model*
2. *Sum of squared residuals from fitting the reduced (equal-means) model*
3. *degrees of freedom*: *n-I*, *n-1*
4. *Subtract the "Within" from the "Total "*
5. *A mean square is the ratio of a sum-of-squares to its degrees of 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*

Large F stats give evidence against the null hypothesis.

Display 5.9 **p. 126**

Four F-distributions, having different degrees of freedom



To find the p-value for the F-test we always find the area to the right of the test statistic.