EPHE 357

Introduction to Research

Analysis of Variance

Olav Krigolson, PhD
krigolson@uvic.ca

Analysis of Variance

What do you do if you have more than two groups (or time points)?

Basic Logic

The logic is simple, we will conduct a test to disprove the null hypothesis:

$$\mu_1 = \mu_2 = \mu_3$$

We will test the null hypothesis by analyzing the VARIANCES within samples and between the means of the samples.

Omnibus Tests versus Planned Comparisons

Unequally Sized Groups

The variance estimates need complex adjustments to weight information from different groups.

Use a computer to do this!

BUT… using unequal groups makes the analysis of variance much more sensitive to violations of homogeneity of variance

Unequally Sized Groups

So, what is to be done…

1) Make group sizes equal

2) Interpolate “missing” data… group means, bootstrapping, etc

Estimating Population Variance From Variation Within Each Sample

- we do not know the true population variance
- however, we can estimate the population variance from the sample
- note, this estimate comes out the same whether or not the null hypothesis is true, because it is based entirely on variance within each sample.

Estimating Population Variance From Variation Within Each Sample

$$S^2_{within} = \frac{S_1^2 + S_2^2 + S_3^2}{N_{groups}}$$

- this is also called MS_{within}, or mean squares within, or MS_{error}

Why is this "error"

Simply put, because this reflects variance within a group and we are interested in differences between groups…

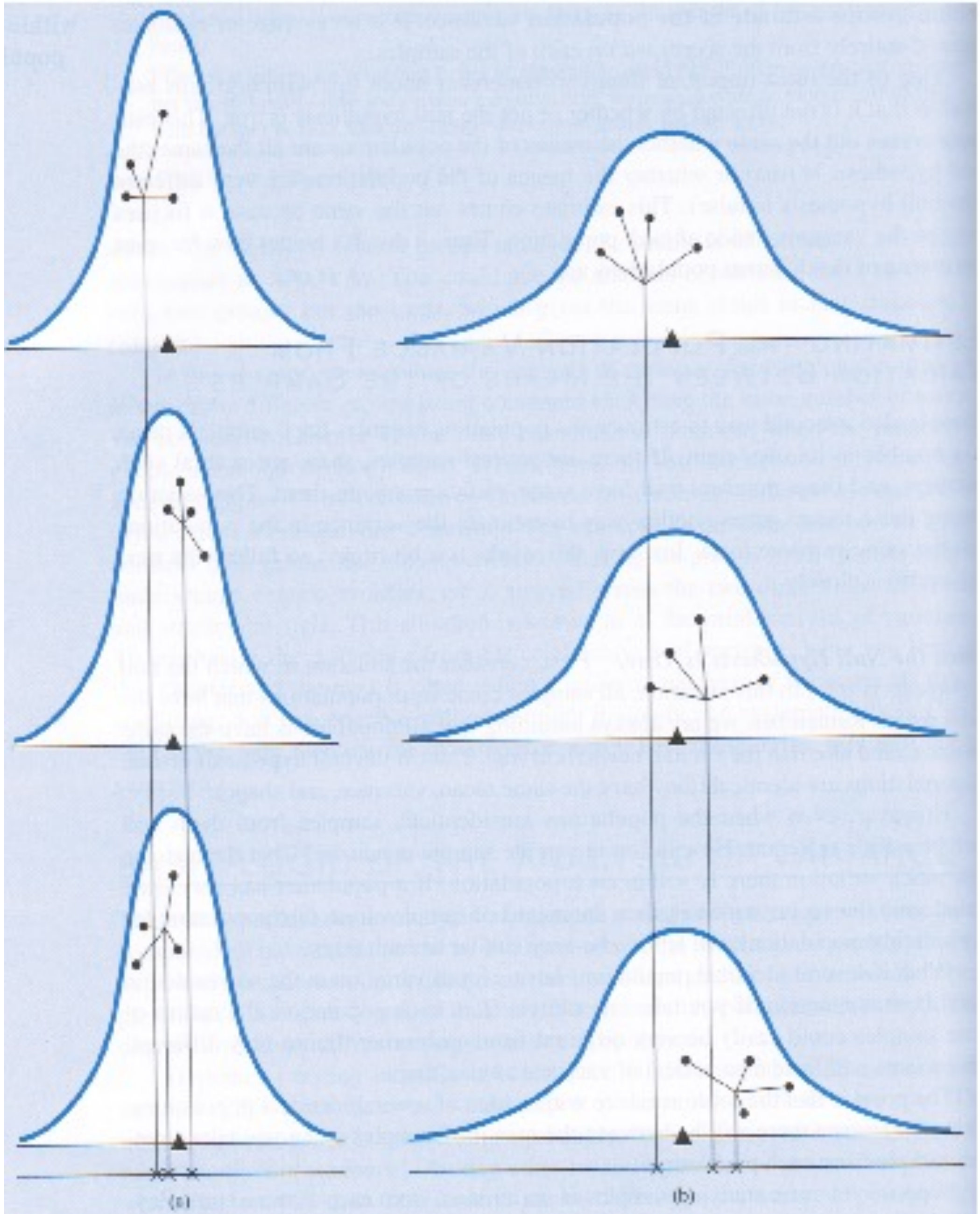
In an ideal world…

Group 1	Group 2	Group 3
8	4	4
8	4	4
8	4	4
8	4	4

Estimating Population Variance From Variation Within Each Sample

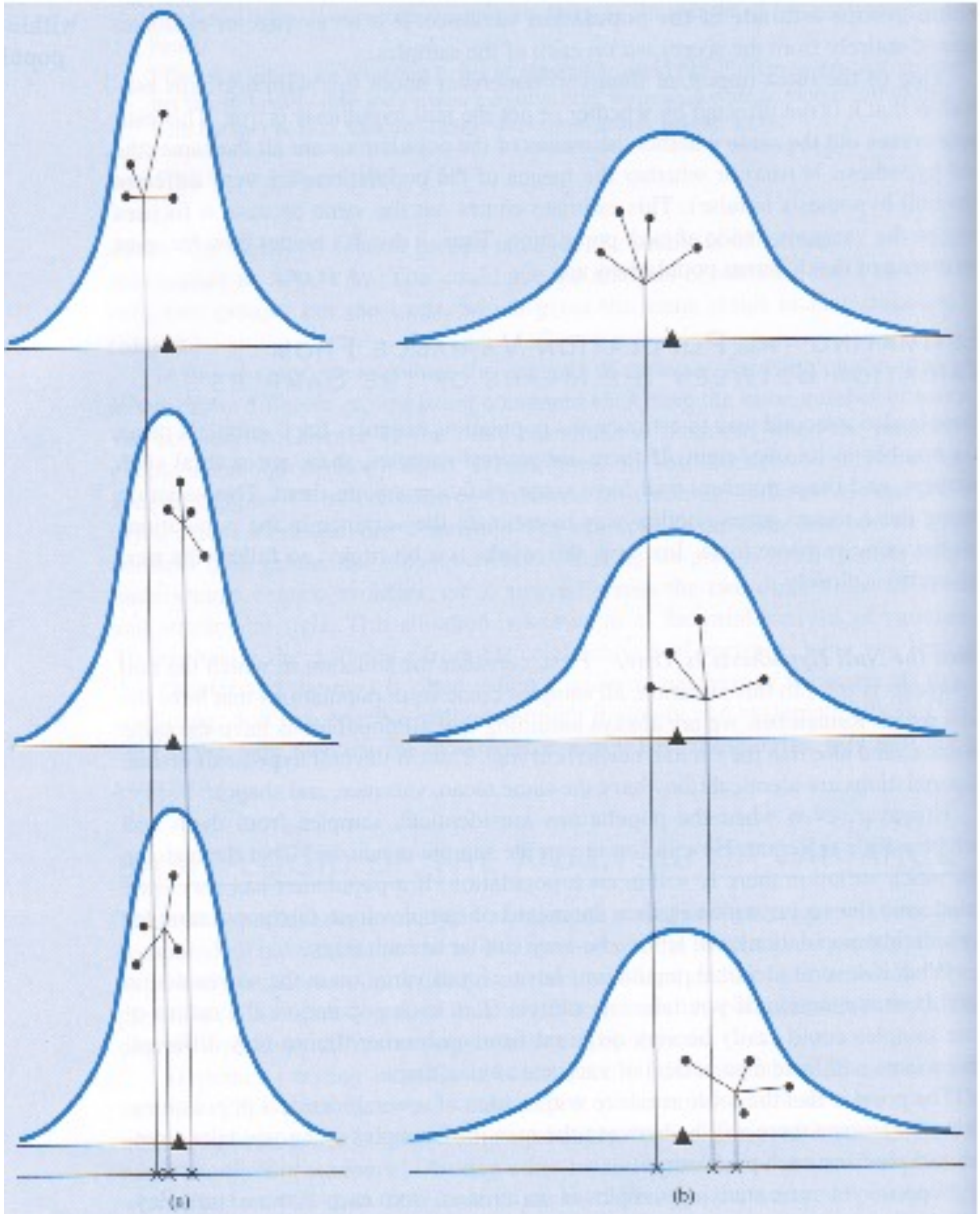
So, MS_{within} or MS_{error} reflects the variability among scores in a sample

Estimating Population Variance From Variation Between the Means of Each Sample



If a population has more variance, then the means of samples taken from that population will be more variable.

This allows us to estimate the variance within each population by examining the variation among the means of our samples.



This would be a <u>between-groups estimate of the population variance</u>.

Estimating Population Variance From Variation Between the Means of Each Sample

if the null hypothesis is true:

between estimate = within estimate

if the null hypothesis is not true:

between estimate > within estimate

Sources of Variation in Within-Group and Between-Group Variance Estimates

	Variation Within Populations	Variation Between Populations
Null Hypothesis is True		
Within-Groups Estimate	X	
Between-Groups Estimate	X	
Null Hypothesis is False		
Within-Groups Estimate	X	
Between-Groups Estimate	X	X

Estimating Population Variance From Variation Between the Means of Each Sample

$$S_M^2 = \frac{\sum (M - GM)^2}{df_{between}}$$

The Estimated Variance of the Distribution of Means

Estimating Population Variance From Variation Between the Means of Each Sample

$$S^2_{between} = (S^2_M)(n)$$

Where n is the number of scores in each group

Note, this is referred to as $MS_{between}$

Finally, because n is in the equation, there is an assumption here that group sizes are equal.

Estimating Population Variance From Variation Between the Means of Each Sample

So, $MS_{between}$ reflects the variability among group means

Recall our rules for hypothesis testing…

1) <u>Check assumptions</u>

2) Calculate a statistic

3) Compare that statistic to the sampling distribution of the means for that statistic

ANOVA Assumptions

Continuous Data

No Outliers

Normality

(recall this goes away with a decent sample size)

Homogeneity of Variance

(rule of thumb, Levene, Bartlett)

Recall our rules for hypothesis testing…

1) Check assumptions

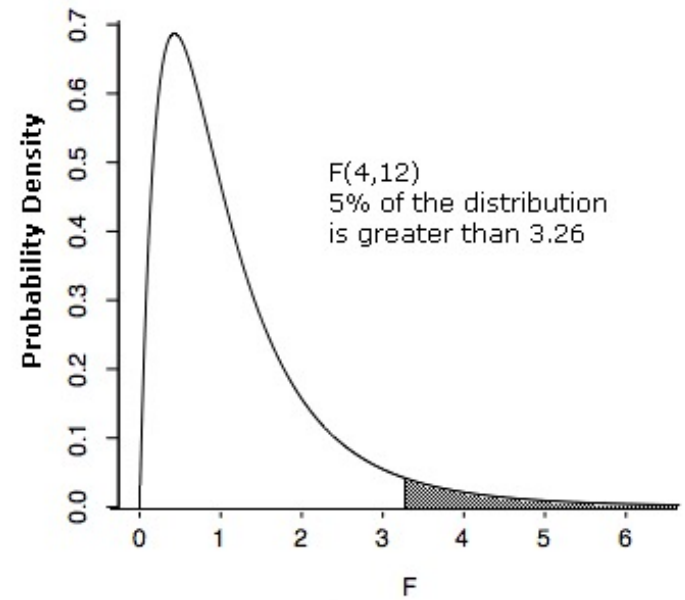
2) <u>Calculate a statistic</u>

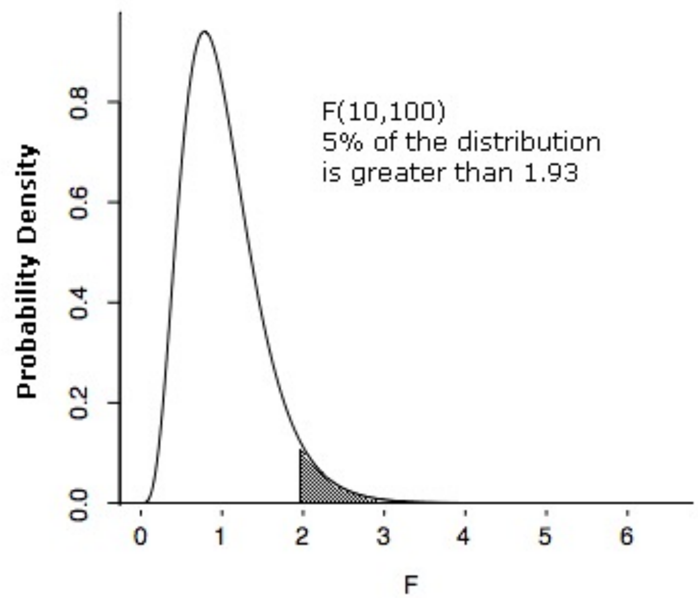
3) Compare that statistic to the sampling distribution of the means for that statistic

The F Statistic

$$F = \frac{MS_{between}}{MS_{within}}$$

The F Distribution





The F distribution has a positive skew. As you can see, the F distribution with 10 and 100 df is much less skewed than the one with 4 and 12 df. In general, the greater the degrees of freedom, the less the skew.

Recall our rules for hypothesis testing…

1) Check assumptions

2) Calculate a statistic

3) Compare that statistic to the sampling distribution of the means for that statistic

The ANOVA Summary Table

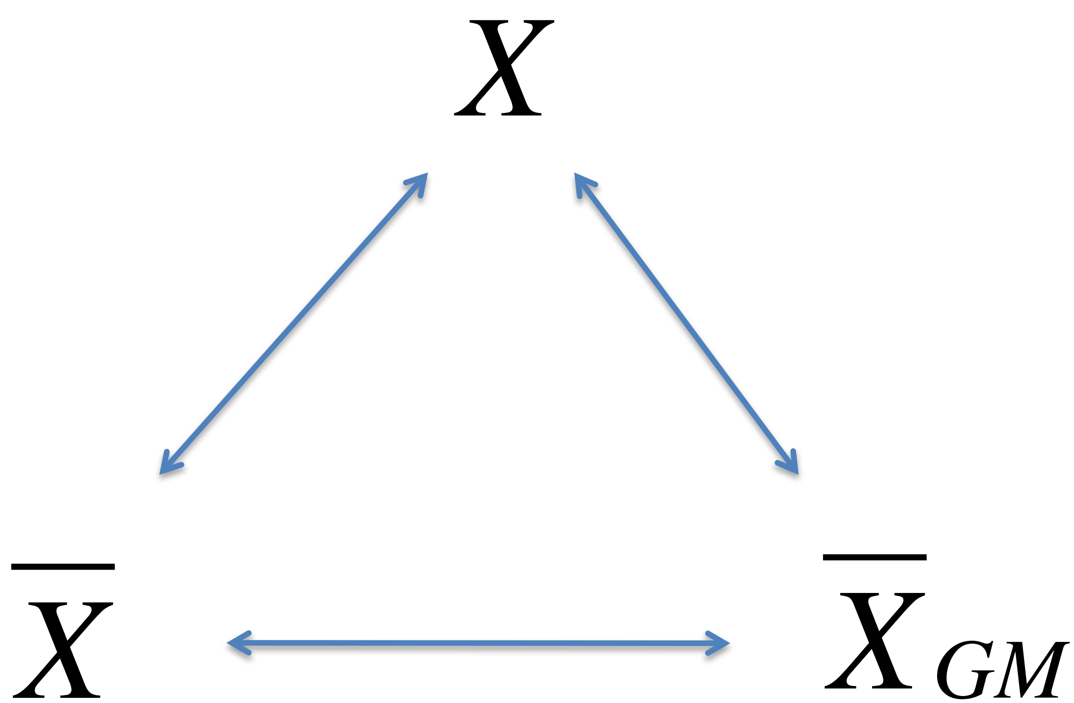
Source	df	SS	MS	F
Between	#	#	#	#
Within	#	#	#	
Total	#	#		

$df_{between}$ = a – 1, where a is the total number of groups

df_{within} = N – a, where N is the total number of scores and a is the total number of groups

df_{total} = N - 1

Score Deviations



Using Sums of Squares to obtain MS_{within} and $MS_{between}$

$$SS_{total} = \sum (X - \overline{X_{GM}})^2$$

Using Sums of Squares to obtain MS_{within} and $MS_{between}$

$$SS_{within} = \sum (X - \overline{X})^2$$

Using Sums of Squares to obtain MS_{within} and $MS_{between}$

$$SS_{between} = n\sum(\overline{X} - \overline{X_{GM}})^2$$

$$SS_{total} = SS_{between} + SS_{within}$$

A Note…

If you are trying to compare SPSS and R results they may differ.

SPSS uses Type III Sums of Squares by default, R uses Type I (which can be changed).

So, the results should be very similar – but a bit different. If they differ greatly, you have made a mistake!

EPHE 591

Post Hoc Comparisons and Contrasts

Field, Chapter 10

Basic Logic

Recall that ANOVA tests the following null hypothesis:

$$\mu_1 = \mu_2 = \mu_3$$

Thus, if it is rejected, all the ANOVA tells you is that there is a difference. We use CONTRASTS or POST HOC comparisons to determine which groups actually differ.

How They Differ

- Post Hoc comparisons compare groups
- Contrasts partition the variance in the model

- Contrasts are used to test hypotheses
- Post Hoc tests are used when you do not have a hypothesis

Posthoc Comparisons

Basic Logic

So why not just run a series of t-tests?

μ_1 vs μ_2 , $\mu_2 = \mu_3$, μ_1 vs μ_3

Because we have to talk about the "family wise error rate" because we are making a series of comparisons.

I’ve heard that I am safe with three comparisons, but not with four…

$\mu_1 = \mu_2 = \mu_3$ we only have one case where we can make a Type I error

This is also true for

$\mu_1 \neq \mu_2 = \mu_3$

$\mu_1 = \mu_2 \neq \mu_3$

$\mu_1 \neq \mu_2 \neq \mu_3$

BUT…

I've heard that I am safe with three groups, but not four…

$\mu_1 = \mu_2$ AND $\mu_3 = \mu_4$

Here we have TWO chances to make a Type I error, thus our family wise error rate is actually 0.1, not 0.05

Thus, we should adjust for the INFLATION of Type I error.

Fisher's Least Significant Difference Test

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{MS_{error}(\frac{1}{n_1} + \frac{1}{n_2})}}$$

Also known as a protected t-test
(we have replaced the estimate of pooled variance with MS_{error})

But, this is in reality still just a t-test so it is very liberal (i.e., not your best choice).

Other Common Posthoc Tests

Bonferroni: Essentially adjusts alpha to correct for multiple comparisons and then t-tests are used.

$alpha_{new}$ = alpha / number of tests

Other Common Posthoc Tests

Sidak's: Similar to Bonferroni but is more conservative (harder to find differences)

Other Common Posthoc Tests

Tukey's: Reflects an adjustment to the actual formulae used to generate the t statistic, and corrects for multiple comparisons this way.

For a large number of comparisons, Tukey's Test is more powerful than Bonferonni. For a small number of comparisons, vice versa.

Other Common Posthoc Tests

Hochberg's GT2 and Gabriel's Pairwise Comparisons Tests are similar to Tukey's Test.

Hochberg's Tests is generally more liberal.

Gabriel's Pairwise Tests are good when cell sizes are unequal.

Other Common Posthoc Tests

Dunnett's Pairwise Tests compare all of the group means against a control mean (the first of the last). Mostly used when comparing clinical populations to a control group.

Other Common Posthoc Tests

Tamhane's T2, Dunnett's T3, Games-Howell, and Dunnett's C are all tests to be used when the variances are unequal.

Other Common Posthoc Tests

Waller-Duncan Test: A Bayesian (probability) approach.

Scheffe Test: Generally the most conservative, examines all linear combinations of variables.

Simple Effects Analysis

Simple Effects Analysis denotes only using t-tests, but using a logical comparison. IE, look at the plot and do not compare things that do not make sense. If Group 2 > Group 1, and Group 3 > Group 2, why compare Group 1 and Group 3?

Repeated Measures ANOVA

Review

Univariate Analysis of Variance

Group A

Group B

Group C

Repeated Measures Analysis of Variance

Condition A

Condition B

Condition C

Repeated Measures Analysis of Variance

Day 1

Day 2

Day 3

Basic Logic of RM ANOVA

<u>Hypothesis Testing</u>

$H_o : \upsilon_1 = \upsilon_2 = \upsilon_3$

$H_1 : \upsilon_1 \neq \upsilon_2 \neq \upsilon_3$ (at least one difference)

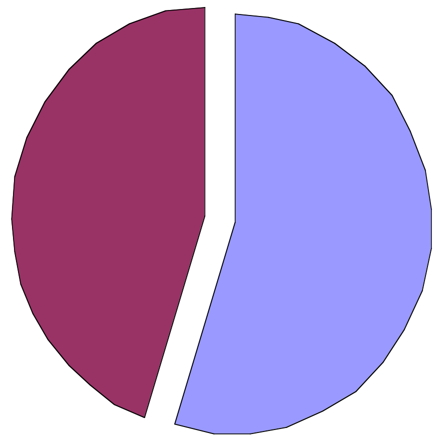
Basic Logic of RM ANOVA

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

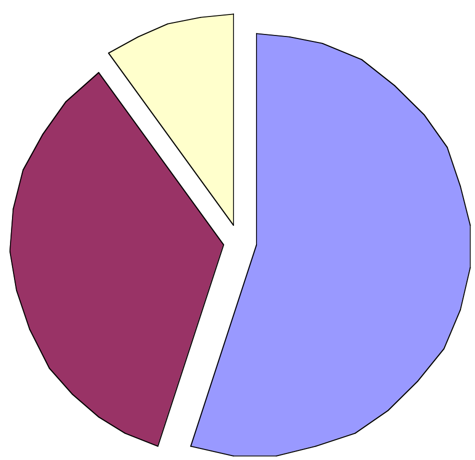
Variance explained by treatment

Variance explained by error

Between Subjects ANOVA



Within Subjects ANOVA (Repeated Measures)



Recall, between subjects ANOVA

$$SS_{total} = SS_{between} + SS_{within}$$

Deviations of group means from the grand mean

Deviations of subject scores from the cell mean

$$SS_{total} = SS_{conditions} + SS_{subjects} + SS_{error}$$

Deviations of condition MEANS from the grand mean

Deviations of subject MEANS from the grand mean