One-way ANOVA, I

9.07
4/15/2004

Review

- Earlier in this class, we talked about two-sample z- and t-tests for the difference between two conditions of an independent variable
  - Does a trial drug work better than a placebo?
  - Drug vs. placebo are the two conditions of the independent variable, "treatment"

Multiple comparisons

- We often need a tool for comparing more than two sample means

What's coming up

- In the next two lectures, we'll talk about a new parametric statistical procedure to analyze experiments with two or more conditions of a single independent variable
- Then, in the two lectures after that, we'll generalize this new technique to apply to more than one independent variable

ANalysis Of Variance = ANOVA

- A very popular inferential statistical procedure
- It can be applied to many different experimental designs
  - Independent or related samples
  - An independent variable with any number of conditions, or *levels*
  - Any number of independent variables
- Arguably it is sometimes over-used. We'll talk more about this later.

An example

- Suppose we want to see whether how well people perform a task depends upon how difficult they believe the task will be
- We give 15 easy math problems to 3 groups of 5 subjects
- Before we give them the test, we tell group 1 that the problems are easy, group 2 that the problems are of medium difficulty, and group 3 that the problems will be difficult
- Measure # of correctly solved problems within an allotted time.

How do we analyze our results?

- We could do 3 t-tests:
  - H_0: $\mu_{easy} = \mu_{medium}$, H_0: $\mu_{medium} = \mu_{difficult}$, H_0: $\mu_{difficult} = \mu_{easy}$
- But this is non-ideal
  - With α=0.05, the probability of a Type I error in a single t-test is 0.05
  - Here, we can make a Type I error in any of the 3 tests, so our *experiment-wise error rate* is $(1-0.95^3) = 0.14$
  - This is much larger than our desired error rate
  - Furthermore, the 3 tests aren't really independent, which cranks up p even more

- We perform ANOVA because it keeps the *experiment-wise error rate* equal to α

ANOVA

- ANOVA is the general-purpose tool for determining whether there are *any* differences between means
- If there are only two conditions of the independent variable, doing ANOVA is the same as running a (two-tailed) two-sample t-test.
  - Same conclusions
  - Same Type I and Type II error rates

Terminology

- Recall from our earlier lecture on experimental design:
- A *one-way* ANOVA is performed when there is only one independent variable
- When an independent variable is studied by having each subject only exposed to one condition, it is a *between-subjects factor*, and we will use a *between-subjects ANOVA*.
- When it is studied using related samples (e.g. each subject sees each condition), we have a *within-subjects factor*, and run a *within-subjects ANOVA*.

One-way, between-subjects ANOVA

- We talk about this for starters.
- The concepts behind ANOVA are very much like what we have talked about in terms of the percent of the variance accounted for by a systematic effect.
- One thing this means is we will be looking for a significant difference in *means*, but we'll do it by looking at a ratio of *variances*.

Assumptions of the one-way, between-subjects ANOVA

- The dependent variable is quantitative
- The data was derived from a random sample
- The population represented in each condition is distributed according to a normal distribution
- The variances of all the populations are homogenous (also referred to as the *sphericity* assumption)
- It is not *required* that you have the same number of samples in each group, but ANOVA will be more robust to violations of some of its other assumptions if this is true.

ANOVA's hypotheses

- ANOVA tests only two-tailed hypotheses
- H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$
- H_a: not all μ's are equal

Typical strategy

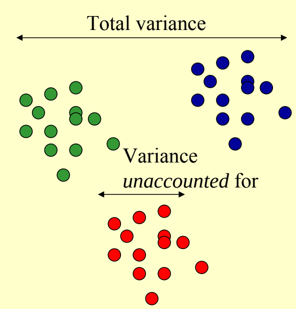
- Run ANOVA to see if there are any differences. If there are, do some additional work to see which means are significantly different:
  - *Post-hoc comparisons*
  - Note that you perform post-hoc comparisons *only* when ANOVA tells you there are significant differences between at least two of the means.
- An exception: if there are only two means to begin with, and ANOVA tells you there is a difference in means, you already know that the two means must differ – no need to do any additional work.

Analysis of variance

- ANOVA gets its name because it is a procedure for analyzing variance
- Though we are interested in testing for a difference in *means*, we can do so by analyzing *variance*
- This has to do with what we've talked about before: proportion of the variance accounted for by an effect

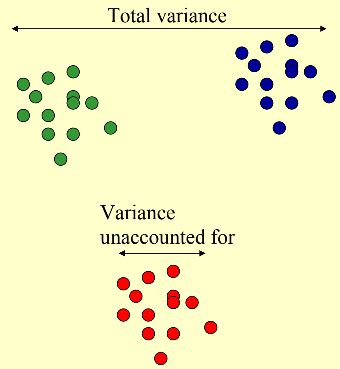
How much do I reduce my uncertainty about the response, if I know the condition?

- In other words, what proportion of the variance is accounted for by the systematic effect?
- ("The effect": the means of the red, blue, and green groups are significantly different)



Keeping the variance within each group the same, the bigger the difference in means, the greater the proportion of the variance accounted for

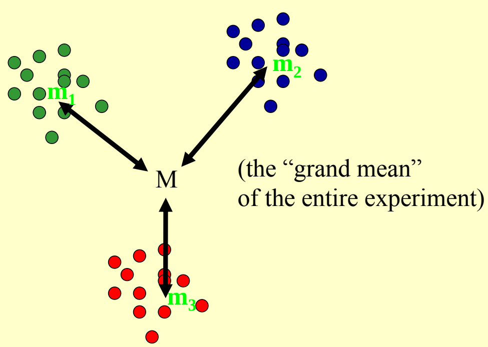
- So, while we're interested in a difference in means, we can get at it by looking at a ratio of variances – the proportion of variance accounted for



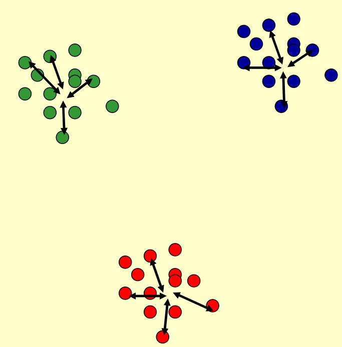
Partitioning the variance

- Before, when we talked about proportion of the variance accounted for, we partitioned the variance in the data this way:
  - Total variance = (variance not accounted for) + (variance accounted for)
- As shown in the previous picture, the variance *not* accounted for is essentially the variance *within groups*. So, the more traditional description of the partitioning of the variance is:
  - Total variance = (variance within groups) + (variance between groups)

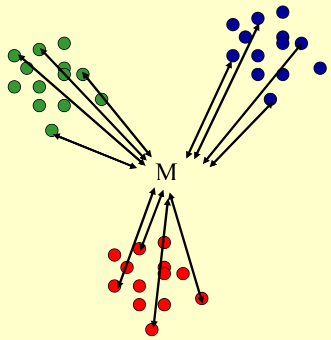
MS_{bn}



MS_{wn}



MS_{total}



Within- and between-group variance

- Essentially, the total variance in the data comes from two sources:
  - Scores may differ from each other even when the participants are in the same condition. This is *within-group* variance. It is essentially a measure of the basic variation or *noise* in the system.
  - Scores may differ because they are from different conditions. This is the *between-groups* variance. This is essentially the *signal* in the system.
- ANOVA is about looking at the *signal* relative to the *noise*

ANOVA and the signal-to-noise ratio

- We want to see if the between-group variance, the *signal*, is comparable to the within-group variance, the *noise*.
  - If the signal is comparable to the noise, don't reject H_0
  - If the signal is large relative to the noise, reject H_0 in favor of H_a

- From the sample data, we will calculate each of these variances (between & within groups)
- But rather than calling them variances, we will call them *mean squares* (short for *mean square deviations*)
  - Mean square within groups
  - Mean square between groups

Mean square within groups

- A measure of the "noise"
- Symbol: MS_{wn} or MS_{error}
- MS_{wn} is like the average variability within each condition (level of a factor)
  - We assumed that the variance is the same in each population, so we estimate the variance in each condition, and then pool them to get MS_{wn}, an estimate of σ^2_{error}

Mean square between groups

- A measure of the "signal", i.e. how much the means for different levels of a factor differ from each other
- Symbol: MS_{bn}
- An estimate of the differences in scores between the different conditions (different levels in a factor)
- How much does the mean of each level differ from the overall mean?

The relationship between MS_{wn} and MS_{bn} when H_0 is true

- H_0 true -> all scores from the same population, regardless of condition
- Means in each condition differ only by chance
- The same sort of process that leads to different means in the different conditions also leads to difference in the scores within a population
- So if H_0 is true, MS_{wn} should be very similar to MS_{bn}
  - MS_{bn} estimates the "noise" in the population just as MS_{wn} does, if H_0 is true

The relationship between MS_{wn} and MS_{bn} when H_0 is false

- H_0 false -> changing conditions causes mean scores to change
- Treatment variance = differences between scores due to a systematic effect
- To some extent, our observed differences in means will also be due, in part, to inherent variability in the scores (noise)
- MS_{bn} is influenced by both treatment variance and noise. It estimates $\sigma^2_{error} + \sigma^2_{treatment}$

The relationship between MS_{wn} and MS_{bn} when H_0 is false

- When H_0 is false, MS_{bn} will be larger than MS_{wn}

Consider what happens to the ratio of MS_{bn} to MS_{wn}

- Let $F_{obt} = MS_{bn} / MS_{wn}$
  - An estimate of $(\sigma^2_{error} + \sigma^2_{treatment}) / \sigma^2_{error}$
- H_0 true ->

  F_{obt} -> $(\sigma^2_{error} + 0) / \sigma^2_{error} = 1$
- H_0 false ->

  F_{obt} -> $(\sigma^2_{error} + \sigma^2_{treatment}) / \sigma^2_{error} > 1$

  The larger the difference in means due to different conditions, the larger F_{obt} will be

The F-distribution

- ANOVA uses a new (to us) distribution, and an associated new test statistic:
  - The F distribution
  - The F statistic
- As usual, we'll compute F_{obt} from the data, and compare this with F_{crit}, to see whether or not to reject the null hypothesis

The F test

- The F test is used to compare two or more means.
- It is used to test the hypothesis that there is (in the population from which we have drawn our 2 or more samples) (a) no difference between the two or more means
- Or equivalently (b) no relationship between membership in any particular group and score on the response variable.

The F distribution

- The sampling distribution of the values of F_{obt} that occur when the null hypothesis is true:
  - There is no difference between the means of the different populations (represented by the different conditions of the experiment)
  - So our samples are all taken from the same population

Approximating the F distribution

- Suppose you have k conditions in your experiment, and n_i samples from each condition
- In MATLAB, you could take n_i samples from a normal distribution, corresponding to each of the k conditions. Then compute F_{obt}
  - Take all samples from the *same* distribution! The F distribution is the distribution assuming the null hypothesis is true, i.e. that all k populations are the same.
- Do this a bunch of times, to get the sampling distribution for F_{obt}

Degrees of freedom

- Like the t- and χ^2-distributions, the F-distribution actually consists of a family of curves of slightly different shape, depending upon the degrees of freedom
- The F-distribution, however, has *two* values of d.f. that determine its shape:
  - d.f. in the numerator (between-groups)
  - d.f. in the denominator (within-groups)

F distribution: Keep $df_{numerator}$ and $df_{denominator}$ straight! If you swap them you get a very different F curve!

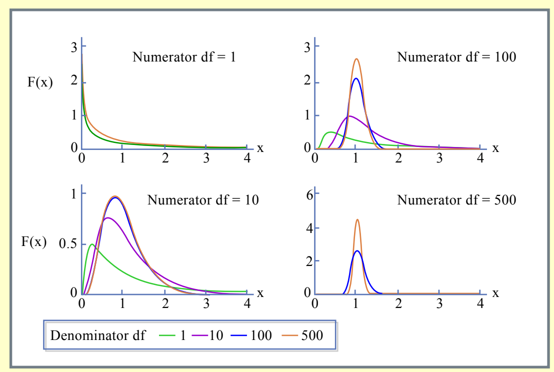


Figure by MIT OCW.

Properties of the F distribution

- The mean of the distribution is 1
  - There is assumed to be no difference between the different conditions, so on average MS_{bn} will equal MS_{wn}, and F will equal 1
- F_{obt} indicates the possibility of a systematic effect of condition only when it is > 1, so we are only interested in the upper tail of this distribution

$F_{obt} = MS_{bn}/MS_{wn}$: Computing MS_{bn} and MS_{wn}

- Both MS's are variances
- Note the form of the equation for the variance we're already familiar with:

$$s_x^2 = \frac{\sum (x - m_x)^2}{n-1}$$

Sum of squares (SS) in the numerator

Degrees of freedom in the denominator

- This is the general form for a variance, i.e. a mean square (MS).

Computing MS_{bn} and MS_{wn}

- So, first we'll compute the *sum of squares*, SS_{bn} and SS_{wn}
- Then we'll figure out the number of degrees of freedom, df_{bn} and df_{wn}
- Finally, MS = SS/df
  - $MS_{bn} = SS_{bn}/df_{bn}$
  - $MS_{wn} = SS_{wn}/df_{wn}$
- Then, we'll compute $F = MS_{bn}/MS_{wn}$

ANOVA summary table

- Report your results in this form on your homework.

Source	Sum of squares		df		Mean square	F	P
Between	SS_{bn}	/	df_{bn}	=	MS_{bn}	= F_{obt}	p-value
Within	SS_{wn}	/	df_{wn}	=	MS_{wn}		
Total	SS_{tot}		df_{tot}				

It's probably easiest to see the calculations with an example

- 3 groups of 5 subjects given 15 math questions each.
- Group 1 told the questions would be easy, group 2 told they would be of medium difficulty, and group 3 told they would be difficult
- Here's the data

of questions answered correctly

Factor: perceived difficulty

Level 1: easy	Level 2: medium	Level 3: difficult
9	4	1
12	6	3
4	8	4
8	2	5
7	10	2

Recall an alternate, computational formula for SS

$$SS = \sum (x - m_x)^2 = \sum x^2 - \frac{(\sum x)^2}{N}$$

We're going to use this version of the formula when we do ANOVA by hand. If you use MATLAB on your homework, use whatever equation is easiest for you.

So, first of all, what is SS_{total}?

$$SS_{tot} = (\sum x^2)_{tot} - \frac{(\sum x)_{tot}^2}{N_{tot}}$$

- This is the sum of squares if you treat the whole experiment as one big sample.
- So $(\Sigma x^2)_{tot}$ is the sum of all the x^2s, and $(\Sigma x)_{tot}$ is the sum of all the x's
- We're going to need the sums of the x's and x^2s for each of the conditions separately, too, so let's compute these sums for each column in the previous table

Factor: perceived difficulty

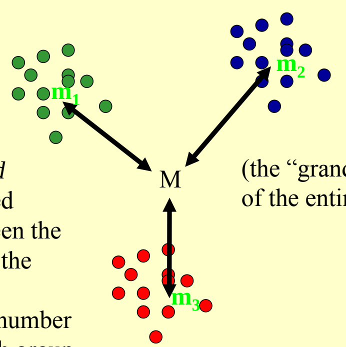
Level 1: easy	Level 2: medium	Level 3: difficult	
9	4	1	
12	6	3	
4	8	4	
8	2	5	
7	10	2	*Totals*
$\Sigma x=40$ +	$\Sigma x=30$ +	$\Sigma x=15$ =	$\Sigma x=85$
$\Sigma x^2=354$ +	$\Sigma x^2=220$ +	$\Sigma x^2=55$ =	$\Sigma x^2=629$
$n_1=5$ +	$n_1=5$ +	$n_1=5$ =	$N=15$

$SS_{tot} = 629 - (85)^2/15 = 147.33$

So, fill SS_{tot} into our ANOVA table

Source	Sum of squares	df	Mean square	F	P
Between	SS_{bn}	df_{bn}	MS_{bn}	F_{obt}	p-value
Within	SS_{wn}	df_{wn}	MS_{wn}		
Total	147.33	df_{tot}			

SS_{bn}



SS_{bn} is a *weighted* sum of the squared differences between the group means and the grand mean, M. Weighted by the number of samples in each group.

(the "grand mean" of the entire experiment)

Next, compute SS_{bn}

$$SS_{bn} = \sum_{i=1}^{\#condits} n_i (m_i - M)^2 = \sum n_i m_i^2 - M^2 N$$

- This is equivalent to the equation in your handout
- So, it looks like we have some means to compute…

Factor: perceived difficulty

Level 1: easy	Level 2: medium	Level 3: difficult	
9	4	1	
12	6	3	
4	8	4	
8	2	5	
7	10	2	*Totals*
$\Sigma x=40$	$\Sigma x=30$	$\Sigma x=15$	$\Sigma x=85$
$\Sigma x^2=354$	$\Sigma x^2=220$	$\Sigma x^2=55$	$\Sigma x^2=629$
$n_1=5$	$n_1=5$	$n_1=5$	$N=15$
$m_1=8$	$m_1=6$	$m_1=3$	$M\approx5.67$

$SS_{bn} = 5*(64+36+9) - 15*(5.67)^2 \approx 63.33$

So, fill SS_{bn} & SS_{wn} into our ANOVA table

Source	Sum of squares	df	Mean square	F	P
Between	63.33	df_{bn}	MS_{bn}	F_{obt}	p-value
Within	84.00	df_{wn}	MS_{wn}		
Total	147.33	df_{tot}			

SS_{wn} is easy – it's just $SS_{tot} - SS_{bn} = 84$
This is just "total variance = sum of component variances".
So, we can fill that one in, too.

What are the degrees of freedom?

- Total degrees of freedom = N–1
  - Usual story – we're computing the variance of all the data, but we lost a degree of freedom in computing the mean.
- Degrees of freedom between groups = k–1
  - k = number of levels in the factor = # condits
  - We're essentially computing the variance of k numbers m_i, but we lose a degree of freedom because $\Sigma n_i(m_i - M) = 0$
- $Df_{wn} = df_{tot} - df_{bn} = N-k$

Filling in nearly the rest of the table

Source	Sum of squares	df	Mean square	F	P
Between	63.33 /	2 =	31.67	= 4.52	p-value
Within	84.00 /	12 =	7.00		
Total	147.33 /	14 =			

For one-way ANOVA, F_{obt} is always placed on the "Between" row, as is the p-value. This is convention, essentially because "Between" is the "signal" in our signal-to-noise ratio. "Between" is essentially what we're testing – is the signal large enough for this to be a real effect?

Now we just need to find F_{crit}

- To do this, look in an F-table, with degrees of freedom = (2, 12) = (bn, wn) = (numerator, denominator)
- I'll have electronic versions of an F-table for you by the end of the day.

What your table will look like:

df (within) = denominator	α	df(between) = numerator 1	2	3	…
11	.05	4.84	3.98	3.59	
	.01	9.65	7.20	6.22	
12	.05	4.75	3.88	3.49	
	.01	9.33	6.93	5.95	
13	.05	4.67	3.80	3.41	
	.01	9.07	6.70	5.74	

$F_{obt} = 4.52$
-> $p<0.05$

Results, and reporting them

- So, it seems that there is a significant effect on math scores of how easy people have been told the problems will be.
- $F(2, 12) = 4.52$, $p<0.05$
  - $F(df_{bn}, df_{wn}) = F_{obt}$, p<p-value

Results, and reporting them

- We are confident that there's a real effect in the population, but we don't know whether each increase in perceived difficulty produces a significant drop in performance.
  - Perhaps there's only a difference between "easy" and "difficult".
  - A significant F_{obt} just means that *at least one* of our of differences is significant.

Next time

- Next time we'll talk about how to determine which pairs are significantly different, if you get a significant result from the ANOVA.
- We'll also talk about how, in some cases, you can deal with more than 2 levels of the independent variable without doing an ANOVA.
- If there's time, we'll also talk about the one-way, within-subjects ANOVA.