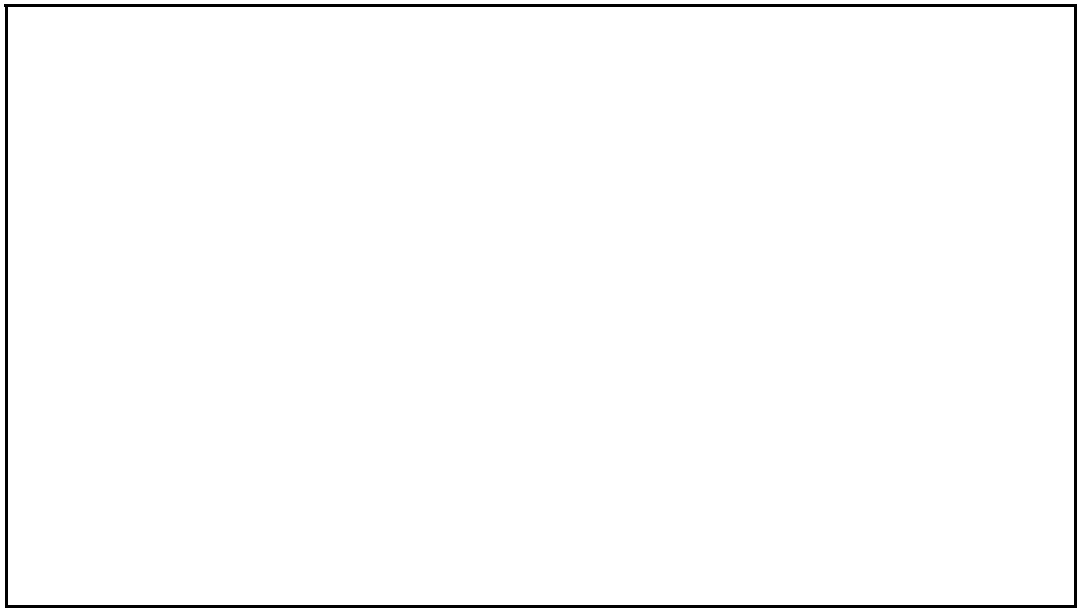
ONE-WAY ANOVA



0:00 / 12:01 1.0x

SPEAKER: MICHAEL J. MAHOMETA, Ph.D.

When it comes to comparing means - two means -

we can turn to the t-test family.

But we won't always have data that has just two groups.

What if your data had three groups or four groups or even more?

Could you still run the analysis?

Sure.

One option is to run multiple t-tests and

that would acutally be bad -

as we'll come to see.

The better option to to run a new type of test

- a test about the variance of data.

Help

And that test is called the Analysis of Variance.

Before we begin, let's start with a definition of insulin resistance.

Basically, the body produces insulin but doesn't use it effectively.

Now consequently, this means that glucose builds up in the blood

and can eventually lead an individual to be

a Type II diabetic or a pre-diabetic.

Back to our question.

Now when we first read this, we might think: exercise, OK.

We can have two groups - those men who exercised and those that did not.

And we can use a simple t-test.

We'll be using data based on a recent article from the European

Journal of Applied Physiology.

Investigators were interested in the effect of a 24 week exercise program

on Insulin Resistance.

But they wanted to further compare the

intensity of the exercise

- so they effectively had three groups: no exercise,

exercise at moderate intensity, and exercise at high intensity.

Help

Now the cool thing about this research was

that the energy expenditure for the exercise groups

was held constant - at 400 kilocals per session.

So although intensity was different, the expenditure was the same.

Here's what the data looks like - each group

has a score that represents the change in Insulin Resistance from baseline

to the end of the 24 week program.

And here are the means and Standard Deviations

along with the n's of each group.

Now our first stab at this data might be to compare

the three means using multiple t-tests.

But as mentioned, that would be BAD.

Why?

Well, remember that thing called Type I error?

It's the likelihood of saying that something is significant

when in fact it's not (effectively putting an innocent man in jail).

That error rate exists for EACH test we run.

And in the normal use of it, it assumes NEW data.

Help

But if we use three, separate, independent samples t-tests,

here we'll be using the SAME data, and inflating our Type I error rate.

Now by how much?

Well we can use this formula to find out.

We find that if we use three independent samples t-test to look at this data,

we've almost a 15% chance of saying something is significant when

in actuality, it's not - erroneously Rejecting our Null Hypothesis.

So what can we do?

We turn to something called Analysis of Variance or ANOVA.

Now ANOVA is a framework of testing, that can handle multiple situations.

The flavor we'll be examining today is the One-way ANOVA.

The phrase One-way means that we have ONE factor - a categorical variable -

impacting the means of a quantitative variable.

To formalize, if we have one categorical grouping

variable (an independent variable) and one quantitative or numerical variable

(as our dependent measure), and we wish to compare the means of said groups,

and the number of those groups is greater than two,

we turn to the one-way ANOVA.

Now before we jump in, I want to cover the basic idea

of what going on with this new test.

Now here's some hypothetical data based on our study in graph form.

Now looking at just the means, can we tell that there's a difference?

Now if you've been paying any attention at all, your response should be

- we need more information - we NEED a way

to tell if these raw differences are out of the ordinary,

and usually we do that by looking at the error of the groups.

We did it with our z test; we did it with our t tests.

So let's add some hypothetical data to our graph:

I'm showing the individual subjects and the

box plot for the group.

Now we still have the mean of the groups as a read diamond.

Now, do we think that there's a difference with this new information?

Our brains should say NO.

But how did we get there?

The typical thing that happens when we internally process,

we start by saying "these means are not very far apart from one another,

as compared to the error that actually exists."

And we get that "natural error" as an idea

of spread or deviation around the group means.

Now because the error in our graph seems to be high

compared to the group differences, we say

that there looks to be no difference across the groups.

But now look at another graph one that is

based on the error observed in the article

- do you think the means are different now?

You should be saying YES.

Well why?

Well we should be doing the same thing that we did before.

We compare the mean differences to the spread around those means.

Now since the "error" has DECREASED, the comparative difference to the means

now seem greater.

Help

And that's basically what the ANOVA is doing

- it's analyzing the variance (or error) from multiple sources.

Let's see this in action: Here's the 1st graph with its corresponding ANOVA

table.

Notice the Sums of Squares for the between group and the Sums of Squares

for the within group.

We know that the Mean Square and the F-statistic come from these -

the Sums of Squares values - with F being the ratio of the Mean Square

between to the Mean Square within.

Now, here's the second "significant" graph and its table.

Notice anything?

The Sums of Squares between didn't change - the error between groups

stayed exactly the same.

BUT the Means Square within DECREASED - and has affected our F-statistic.

What was low before, is now higher.

So now, the ratio of error between groups to the error within groups

is higher - and we think (visually) that we have a difference.

Help

As a general rule, the higher the F-ratio

statistic, the more likely that there will

be a significant difference across the groups.

Now let's see how to use this F-ratio in a hypothesis test.

First, our assumptions:

All observations are independent - no one is measured twice.

They are randomly selected from the population from which they come.

And the population distribution of each group is approximately normal.

And, the variances for each group are approximately equal to one another.

Now on to the hypothesis test:

Step one: our Hypothesis Statement.

We can use something like this in symbolic form,

or we can use plain English: Our Null is that the means

of all groups (or factor levels) are equal.

Our alternative is that at least one group (or

one factor level)

is different from the others.

Step two: our alpha level.

As alway, it's easiest to assume a given 0.05 alpha level - the standard.

Step three: the analysis.

Now we COULD run this by hand if we want,

using the following formulas for each source Sums of Squares.

But let's just look at the table and see how to USE it instead

Here are the Sums of Squares for the two sources plus the total.

For the Sums of Squares between we have 2 degrees of freedom

because we have three groups, so our Mean Square is

our Sums of Squares between divided by the degrees of freedom for between.

Our Sums of Squares within has 36 degrees of freedom

and gives us our Mean Square within.

So, the ratio of between error to within error -

as measured by our F-ratio statistic is 6.854.

Step four, our conclusion: Now, just like our chi-square distribution -

we can't get a negative value for F, so our

distribution looks a little weird.

Here's the F-distribution for 2 and 36 degrees of freedom,

with the critical level of 3.26.

Now since our F-statistic is beyond this critical value,

we Reject our Null Hypothesis.

Now here's our output from the aov() function in R.

Here the p-value is under our alpha, so again, we Reject our Null Hypothesis.

It turns out that at least one group is different from the others

- but which one?

Well, to answer that question, we turn to our Post-hoc tests.

Now there are two common post-hoc tests for running after an ANOVA.

The Bonferroni - where we run all of the t-tests and then correct our alpha

using this formula to account for the Type I error inflation.

But this correction is really conservative - imagine 5 groups -

that means we have ten possible post-hoc t-tests.

So our correct alpha that we would need to be below for

significance would be 0.005 not just 0.05.

And that might be really had to get to.

The other common post-hoc is the Tukey Honestly Significant Difference

post-hoc or Tukey HSD.

This is my preferred post-hoc as it less conservative than the Bonferroni

when the number of groups is high.

And we can run this also in R.

If we do, we see that we have a significant difference

between the control and the two exercise groups respectively

- but not between the two exercise groups.

The intensity doesn't matter when reducing insulin resistance

- as long as the expenditure is comparable.

This means that people can see good effects from moderate exercise,

and don't have to use intense exercise especially

if they are in a high risk cardiovascular health ANOVA.

Help

1. ANOVA is an appropriate statistical measure when we want to:

(1 point possible)

- determine whether the hypothesized mean in one population is the same for two or more other populations.
- compare the means of three or more populations at once. ✔
- estimate the size of the difference between two or more group means.
- compare the distributions of two or more categorical variables.

Hide Answer

2. In ANOVA, we calculate an F statistic. The F statistic is the ratio of:

(1 point possible)

- the difference in group means compared to the expected group difference in means.
- the variation in one group compared to the variation in the other.
- the variation within groups to the total variation.
- the variation between groups to the variation within groups. ✔

Help

Hide Answer

3. If the null hypothesis for an ANOVA test is $\mu_A = \mu_B = \mu_C$, what is the appropriate alternative hypothesis?

(1 point possible)

- $\mu_A \neq \mu_B \neq \mu_C$
- At least one of the means is different. ✔
- $\mu_A \neq \mu_B$ *or* $\mu_A \neq \mu_C$

Hide Answer

4. The source table below presents the results from an ANOVA comparing four treatment conditions with n=25 participants in each condition. Compare all the missing values. *Hint: Start with degrees of freedom*. (Round to 3 decimal places where needed.)

Source	SS	df	MS	F-statistic	F-critical
Between	[SS_B]	[df_B]	19	F-stat	2.699
Within	[SS_W]	[df_W]	[MS_W]		
Total	117	[df_T]			

Help

(7 points possible)

4a. SS_B =

Answer: 57

4b. SS_W =

Answer: 60

4c. df_B =

Help

Answer: 3

4d. df_W =

Answer: 96

4e. df_T =

Answer: 99

4f. MS_W =

Answer: .625

4g. F-stat=

Answer: 30.400

Hide Answer

(1 point possible)

4h. We should _____ the null hypothesis.

Reject

Hide Answer

(1 point possible)

4i. Assume your F-statistic is significant, suggesting that at least one treatment condition is different from the others. How many post-hoc group comparisons will you need to run?

Remember the formula for group comparisons: $\frac{k(k-1)}{2}$

- 6 ✔
- 15
- 10
- 12

Help

Hide Answer

(1 point possible)

4j. Using the Bonferroni correction, what significance level should you use for each post-hoc hypothesis test if you want an overall significance level of 0.05?

- .050
- .100
- .008 ✔
- .003

Hide Answer

(1 point possible)

4k. What was the overall risk of making a Type I error in this ANOVA? *(Round to 2 decimal places.)*

Help

Answer: .05

Hide Answer

(1 point possible)

4l. What would the risk of a Type I error have been if you had run multiple t-tests instead? Assume $\alpha = 0.05$ for each test. *(Round to 2 decimal places)*

Hint: $1 - (1 - \alpha)^C$

Answer: .26

Hide Answer

Help

literature, math, medicine, music, philosophy, physics, science, statistics and more. EdX is a non-profit online initiative created by founding partners Harvard and MIT.

Help

Terms of Service and Honor Code

Privacy Policy (Revised 10/22/2014)

edX Blog

Donate to edX

Jobs at edX

LinkedIn

Google+