Slide 1 – Title Slide

Hello and welcome to Week 8, Part 4 of EGM101: ANOVA. In this lesson, we'll learn about a way that we can do hypothesis testing with more than two samples.

Slide 2 – What happens if we have >2 samples?

So far, we've seen how we can use hypothesis tests to answer questions about whether our samples come from a population with different mean values. If we only have one sample, we can use the one-sample t-test to compare the sample mean to the hypothesized population mean. If we have two samples, we've seen different forms of the two sample t-test, depending on whether our samples are paired or independent.

What if we have 3 or more samples, though? We could conduct pairwise two-sample tests, but this raises a different problem – multiple tests increases the likelihood that we make a Type-I error. Instead of conducting multiple two-sample tests, we can instead use a technique called analysis of variance, frequently shortened to ANOVA.

Slide 3 – The F-distribution

ANOVA relies on something called the F-distribution, which is a ratio of two "chi-square" distributions – we'll talk more about the chi-square distribution later this week.

Like Student's t-distribution, the F-distribution depends on degrees of freedom – unlike the t-distribution, we actually have two separate degrees of freedom. The first, df1 or nu 1, is equal to k minus 1, where k is the number of groups that we have. The second, df2, is equal to n minus k, where n is the sum of the individual sample sizes.

The graph here shows the variation of the F-distribution for different combinations of degrees of freedom. At low values of both df1 and df2, the distribution looks something like an exponential distribution; as we increase the degrees of freedom, you can see how the distribution changes from a skewed distribution to become more symmetrical as the "peak" of the distribution gets taller and narrower.

Slide 4 – One-way ANOVA: Assumptions

Before we can use ANOVA, we have a number of assumptions that need to be met. The first is that the dependent variable is continuous – we can't use ANOVA if we have discrete variables.

Next, the independent variable must be categorical – that is, we're separating the data into different distributions based on some category. This could be based on ranges of a particular value, or it could be based on the month or season that the measurements were recorded, or it could be based on the body of water that we sampled from.

We also assume that the populations are normally-distributed – meaning that ANOVA is a parametric test. Finally, we also assume that the samples that we are using are independent, random samples, and that the different populations have equal variances. In practice, this means that the ratio of the different sample variances is between 0.5 and 2, like we saw previously with the unpaired samples t-test.

As long as these different assumptions are met, we can use ANOVA – if not, we need to look into alternative tests.

Slide 5 – Mean squares

Before we get to the actual test, we need to introduce a few more things, starting with something called the treatment mean square. To calculate the treatment mean square, we first calculate the treatment sum of squares, SS treatment – this is the sum of the squares of the difference between each sample mean, xbar i, and and the overall mean of the data, multiplied by the sample size of each group. We then divide the treatment sum of squares by k – 1, where k is the number of groups.

The treatment mean square depends on the differences between the groups – essentially, it's the sample variance of the different sample means, with k – 1 degrees of freedom.

The other term we need to introduce is the error mean square, calculated using the formula shown here. To calculate the error mean square, we first calculate the error sum of squares, SS error – this is the sum of the sample variance of each group, multiplied by the size of each group minus 1. We then divide the error sum of squares by n – k, where n is the sum of the sample sizes of each group.

The error mean square depends on the differences within the groups – essentially, it's the pooled estimate of the population variance, with n – k degrees of freedom.

Slide 6 – The One-way ANOVA Identity

In addition to the treatment sum of squares and the error sum of squares, we can also calculate the total sum of squares for the samples, SS total, using the formula shown here. In effect, this tells us the total variation among all of our sample data – it's very similar to the calculation for the variance, with the exception that we're not taking the average by dividing by the number of degrees of freedom.

The total sum of squares is also equal to the sum of the treatment and error sums of squares, a fact that is known as the One-way ANOVA identity. Effectively, this shows us that we can partition the total sum of squares into the treatment and error sums of squares – but more importantly, it means that we don't have to calculate all three of these things separately – we only have to calculate two of them, and we can use this identity to calculate the other one.

Slide 7 – The F-statistic

With that, we arrive at the F-statistic, which is what we use for the ANOVA test. The F-statistic, calculated as the ratio of the treatment mean square to the error mean square, compares the variation between the sample means to the variation within the individual samples.

Larger values of the F-statistic mean that we have more variation between the groups than we have within the groups – in other words, it's more likely that the samples have different population means. As we discussed on the previous slide, we have two degrees of freedom for the F-statistic – the treatment mean square in the numerator, has k – 1 degrees of freedom, while the error mean square, in the denominator, has n – k degrees of freedom, where n is the sum of all the individual sample sizes.

Before moving on, you should note that the F-statistic is not so different from the coefficient of determination. Remember that the coefficient of determination is calculated as the ratio of the explained variability to the total variability, whereas the F-statistic can be thought of as the ratio of the explained variance to the unexplained variance.

Slide 8 – One-way ANOVA

We will wrap up by looking at the one-way ANOVA test. In the one-way ANOVA test, the null hypothesis is that that group means are all equal – that is, there is no difference in the group means.

The alternative hypothesis is that the population means are not equal for some pair of groups. On the boplot here, we can see what these two scenarios look like. If the null hypothesis is correct, shown in the top panel, then the differences in the sample means for each group are due to random variability in the samples, and not because of an actual difference between the populations that each sample is drawn from. The differences between the mean values are very small in comparison to the dispersion in the samples, represented by the size of the boxes.

If the null hypothesis is not correct, though, then the differences between the groups are larger than the dispersion of the samples – that is, the differences between the samples is unlikely to be due to random variation in the samples.

ANOVA will only be able to tell us if there is a difference between at least one pair of groups, though – if we want to determine which groups are different, or how they are different, then we need to do additional tests, called "post hoc" tests because we do them "after the fact", to figure out where the differences are.

The actual test procedure is very similar to what we saw for our other hypothesis tests. First, we have to calculate the F-statistic for our observations, then we compare that value of the F-statistic to the critical value (or p-value) for the F-distribution with the degrees of freedom equal to df1 and df2.

Finally, the one-way ANOVA test is always a right-tailed test – that is, we're only testing whether the observed F-statistic is larger than the critical F-statistic.

Slide 9 – Summary

In this lesson, we've seen how ANOVA helps us determine if there is a difference in means between multiple samples or groups.

We do this by comparing the variance between the groups to the variance within the groups – if the variance between the groups is greater than the variance within the groups, the chances are that at least one of the groups comes from a population with a different population mean.

Finally, we discussed how ANOVA will only tell us if there is a difference between the groups – it won't tell us which group is different, nor will it tell us what the difference is. For that, we need to do additional testing.

Slide 10 – Additional resources

You can read more about the topics we've discussed here in the textbooks – Illowsky and Dean, Chapter 13; and Weiss, Chapter 13.

I've included links to two additional articles here on how F-tests work in ANOVA, and about using post hoc tests with ANOVA. I've also added links to two videos from Khan Academy, about calculating the total sum of squares, and calculating the SS_{error} and $SS_{treatment}$.

That's all for this lesson – I hope you found it interesting, and if you have any questions, please don't hesitate to e-mail me or post in the discussion forum on blackboard. Bye!