Slide 1 – Title Slide

Hello and welcome to Week 8, Part 3 of EGM101: Parametric tests. In this lesson, we'll learn about hypothesis testing using parametric tests.

Slide 2 – What is a parametric test?

To begin, we should probably define what a parametric test is. In a parametric test, we assume that the population can be modeled using a probability distribution that has fixed parameters – for example, the normal distribution, which we model using a fixed population mean and standard deviation. We've already seen an example of this in the previous lesson when we looked at using a Z-test to determine whether our evidence suggested that the population mean was actually lower than our assumed mean of 150 cm.

In the rest of this lesson, we'll look at examples using something called Student's t-distribution – and don't worry, we'll introduce what this is before we get there.

The fact that we're talking about parametric tests implies that we have some other kind of test, and we do. Non-parametric tests are statistical tests where we don't assume a particular probability distribution for the population.

Slide 3 – When do I use parametric tests?

Most often, we use parametric tests when we are working with normally-distributed, continuous data. Remember that the Central Limit Theorem tells us that with a random sample of sufficient size, we can treat the distribution of sample means as being normally-distributed, no matter what the population looks like. Whether or not we can use a parametric test for non-normally distributed data might also depend on the type of test we're using, - when in doubt, check the data.

We also use parametric tests when our observations are independent, and when the variance of the samples is approximately the same. In general, we prefer to use parametric tests because they tend to have more power – that is, the chances of a true positive are much higher with parametric tests, compared to non-parametric tests.

Slide 4 – Degrees of Freedom

Before we go much further with parametric tests, we need to talk about freedom. Remember that the sample mean and sample standard deviation are estimates of the population mean and the population standard deviation – and specifically, we calculate the standard deviation differently, depending on whether or not we're working with a sample or a population.

The reason for this is something called degrees of freedom – that is, the number of independent pieces of information that are used to calculate an estimate. We often denote the degrees of freedom using the lowercase Greek letter, nu.

To illustrate how degrees of freedom works, let's assume that I have a very simple wardrobe, and no sense of style: I have exactly 7 t-shirts, and I want to wear a different t-shirt every day of the week. So, let's say that on Monday, I choose my blue t-shirt, because it's my favorite. On Tuesday, I choose my green t-shirt, because it's also my favorite. On Wednesday, I choose my red t-shirt, on Thursday, my yellow t-shirt, on Friday, my black t-shirt, and on Saturday, my brown t-shirt. This means that on Sunday, I have to wear my white t-shirt – it's the only one left.

In all, then, I have only made 6 independent choices – the white t-shirt is not an independent choice, because it depends on all of the other choices I've made earlier in the week.

Another definition for "degrees of freedom" is the number of observations that are free to vary when we make our estimate. To show how this works, let's look at how we calculate a sample mean. Let's say that we have a dataset of 10 observations, with a sample mean of 6. When we calculate this sample mean, only 9 of the values are "free" to vary – if we sum up the first 9 values, we get 57, which means that the last value has to be equal to 3, in order for the sample mean of 10 values to equal 6.

Looking at a second example, if our dataset looks like this and we have a sample mean of 10, the first 9 values sum to 91, which means that the last value has to be equal to 9, in order for the sample mean of 10 values to equal 10.

In practice, for a sample mean, the number of degrees of freedom, nu, is equal to the number of values, n, minus 1. That is, for a given sample mean of n values, only n – 1 values are actually free to vary.

Slide 5 – The Student t-Distribution

This brings us to the distribution that we will most frequently use for parametric tests – Student's t-distribution. Student's t-distribution is very similar to a normal distribution, but it takes degrees of freedom into account – that is, the probability density varies based on the number of degrees of freedom. As you can see from the graph here, the peak of Student's t-distribution is slightly shorter than the normal distribution, and the tails are slightly larger.

You should also be able to see that as the number of degrees of freedom increases, the distribution gets closer to the normal distribution – after nu increases to about 20 or so, the distribution is essentially the same as the normal distribution.

We use this distribution when we have a small sample size – below about 20 or so, or when we don't know the population standard deviation, which is most of the time. If we have a large-enough sample size and we know the population standard deviation, we can use the Z-test that we saw in the previous lesson; otherwise, we should use the t-distribution.

Slide 6 – Student's t-Test for a single mean

Student's t-Test for a single mean looks pretty much exactly the same as the z-test did – but, instead of the normal distribution, we use the t-distribution with n – 1 degrees of freedom.

Looking again at our example from the previous lesson, we have a sample of 30 salmon with a mean length of 142 cm and a sample standard deviation of 20 cm. Once again, we want to know whether the evidence supports the hypothesis that the population mean is less than 150 cm.

Our hypotheses are the same as before: the null hypothesis is that the population mean is equal to 150 cm, and the alternative hypothesis is that the population mean is less than 150 cm.

We calculate the t-statistic using this formula: t is equal to the difference between the sample mean and the assumed population mean, divided by the standard error of the mean, calculated using the sample standard deviation.

When we plug in the values that we have, we get a value for t of -2.19; the critical value for a t-distribution for nu equals 29 and a significance level of 0.05 is -1.699. Because t is less than the critical value (or, alternatively, the calculated p-value using this distribution and our t-value is less than alpha), we reject the null hypothesis, and we can say that based on the evidence, the population mean is most likely less than 150 cm.

Slide 7 – Two samples

So far, we've looked at examples with only a single sample, where we compare the sample mean to an assumed population mean. The test procedure for two samples is very similar, but with some important differences. First, the comparison is between the difference of the means of two samples and two populations. Instead of a single sample mean and population mean, we're comparing the difference between the two sample means and the difference between the two population means. To calculate the test statistic, we still divide by the estimate of the standard error of the mean – what form this takes depends on our samples, though.

If we have paired samples – for example, we're examining the differences before and after some treatment, the samples are dependent – that is, the members of each sample are related to each other, because they are paired together. Paired samples are often used for observing changes in a single population over time, where we have multiple observations for each member of the population, or at least the sample.

If we have unpaired samples, on the other hand, the samples are independent – they're not connected in any way. This means that each individual in one sample is independent of every individual in the other sample. This type of test is useful for comparing different populations – for example, populations from different geographic regions.

Slide 8 – Paired samples

The formula for the test statistic for paired samples looks like this – we have our sample difference, minus by the hypothesized difference, divided by an estimate for the standard error of the difference. For paired differences, the value that we use here, sd, is the standard deviation of the differences. That is, we calculate the difference between each of our paired values, and calculate the sample standard deviation of those differences.

To illustrate an example of paired samples, let's say that we have a sample of 25 salmon that we tagged and weighed before their migration, and then again after their migration, because we think that the fish will have lost a lot of weight as a result of the migration.

We see that before the migration, the sample mean was 3.5 kg, and after the migration, the sample mean was 3.25 kg, and we had a standard deviation of differences equal to 0.9 kg. Our hypotheses are stated as follows: the null hypothesis is that the two population means are equal – that is, there is no difference between the mean weight before and after the migration. The alternative hypothesis is that the weight is larger before the migration, or that the difference between the population means is greater than zero.

So, our test question is as follows: at the 0.05 significance level, is there a decrease in the population mean? First, we can calculate the test statistic as follows: because the difference in the population means under the null hypothesis is 0, the test statistic simplifies to be the difference between the sample means, divided by the standard error of the differences. Plugging in the numbers that we have, we get a value for the test statistic of 1.389.

Another way of stating the alternative hypothesis is that the difference between the population means is greater than zero, which means that we're going to use a right-tailed test. Because we have a sample size of 25, the number of degrees of freedom is 25 – 1 equals 24, so that's the t-distribution that we need to use. And, for that t-distribution, the p-value for 1.389 is 0.089; the critical value is 1.711. Because our test statistic is less than the critical value, or because the p-value is greater than our chosen alpha, we say that we fail to reject the null hypothesis – that is, our observations do not provide sufficient evidence to conclude that the population mean has changed.

Slide 9 – Unpaired samples

For unpaired samples, the calculation is slightly more complicated, because the form of the standard error of the mean depends on the comparison between the sample variances. If the variances are "similar", that is, if the ratio of the two sample variances (or the sample standard deviations) is between ½ and 2, then the test statistic is calculated according to this formula, using something called the "pooled variance", which is calculated using this formula. In this case, the number of degrees of freedom is equal to the sum of the two sample sizes, minus two.

If the variances aren't similar – if one of the sample standard deviations is more than two times the other – then we use the formulas shown here, a form of the t-test known as Welch's t-test. When the variances aren't equal, the degrees of freedom becomes a bit more complicated to calculate, and I'm not going to put the formula on the slide.

As always, you don't need to memorize or know these formulas, but you should understand when to use which form – it depends on whether the two sample variances or standard deviations are similar or not.

Slide 10 – Independent samples

To illustrate an example of a t-test with independent samples, we'll use an example where we want to see whether wastewater discharge has an affect on fish health, using the mean lengths of two samples drawn from different parts of a stream – the first, a sample of 18 fish taken from upstream of where the wastewater discharge enters the stream, and the other, a sample of 14 fish taken from downstream of where the wastewater discharge enters the stream.

The table here shows the different sample means and standard deviations – because the sample standard deviations are similar, we can use the pooled variance, or pooled standard deviation, instead of Welch's t-test.

The hypotheses are stated as follows: the null hypothesis is that the two population means are equal – there is no difference between the means. The alternative hypothesis is that the two means are not equal – this example is using a two-tailed test, as all we're trying to determine is whether there is a difference or not.

The test question is then: at a significance level of 0.01, is there a difference in length between the two populations? This is a two-tailed test. If we plug the numbers from the table into the formulas here, we get the following value for t. The number of degrees of freedom is the sum of the two sample sizes minus two, and the p-value for this t value and that distribution is 0.004. The critical value for that distribution and our chosen alpha value is 2.749. Because t is greater than the critical value, or because p is less than alpha, we reject the null hypothesis, and conclude that there is a difference between the population means.

Slide 11 – Summary

In this lesson, we've discussed how parametric tests are when we assume something about the distribution of the population – most often, that the population follows a normal distribution.

If we know the population standard deviation and we have a sufficently large sample, we can use something called a z-test; usually, though, we end up using a form of student's t-test, because we either have a small sample size, or because we don't know the population standard deviation.

As we have seen, the calculation of the test statistic depends on how many samples we have – either one or two; it also depends on whether those samples are independent or not, and whether the population variances are similar.

And finally – make sure that you focus on understanding why you would use a particular form of the test, rather than memorizing the formulas.

Slide 12 – Additional resources

You can read more about the topics we've discussed here in the textbooks – Illowsky and Dean, Chapters 9 and 10; Caswell, Chapter 15; and Weiss, Chapters 9 and 10.

I’ve also included a link to two articles – one is about degrees of freedom in statistics, and the other discusses the differences between non-parametric and parametric tests in more detail. Finally, there are links to videos from Khan Academy about the differences between one- and two-tailed tests, and between Z-statistics and T-statistics.

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!