Let us look at how we can run hypothesis test when we do not have information on the population standard deviation.

>>

Let us look at an example. Imagine that we are testing the IQs of a group of people, for example, say students at Jigsaw. We find that the average IQ is 104 for a random sample of 40 Jigsaw students with a standard deviation of 12 in the sample. In general, IQ tests are designed in a way that average score for the population turns out to be 100. Now the other question is that can we confidently state that this group of people, which is Jigsaw students, are smarter than the general population? We notice here that we do not have the population standard deviation. We have the population mean score, which is 100, but we don't have access to the population standard deviation.

It turns out that if the population standard deviation is not available, we can use the sample standard deviation as an acceptable substitution for the population standard deviation. If the population standard deviation is not known, the distribution of sample means will follow a T-Distribution with a mean equal to the population mean and standard deviation equal to sample standard deviation divided by the square root of the sample size.

Remember, you still need to divide the standard deviation by the square root of the sample size, because we know the variation in the sample means is inversely proportional to sample size.

>>>

In this particular example, what we need to calculate is how likely is it that simply by random chance we ended up with an average of 104 in the sample when the actual population average is 100.

- The null hypothesis – Jigsaw students IQ is the same as the general population
- Alternate Hypothesis – Jigsaw students IQ is higher than the general population

Let us say that we want to be very certain – so 95% certain. In that case, we will use a significance level of 5%.

Because this is a T-Distribution, we will calculate test statistic, which is

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

, which is 2.10

We use this 2.10 in excel for the distribution function formula for T-Distribution and remember we will do 1 – T Distribution (2.10, 39, TRUE). The reason we do 1 minus is that in any hypothesis test, we are looking for outcomes more extreme than observed. More extreme is greater than 104 and therefore to get the P value of greater than 104, we have to do 1 minus cumulative probability of less than or equal to 104. If you do this in excel, it turns out that the P value is 0.02.

Therefore, what is our conclusion? We will reject the null hypothesis. When we reject the null hypothesis, what exactly are we concluding? We are concluding that the Jigsaw students' IQ is indeed greater than the general population.

>>

In this section, we will look at directional hypothesis test. When we run a hypothesis, we set up null hypothesis and then alternate hypothesis. A null hypothesis is always the hypothesis that there is really no difference between the sample and the population and any difference we observe is simply because of random chance variation.

Alternate hypothesis is negation of the null hypothesis. If we look at the IQ example that we had reviewed, the null hypothesis would have been that the IQ of Jigsaw students is the same as the general population.

What about the alternate hypothesis? The negation of the alternate hypothesis can happen in two ways. We could say that IQ of Jigsaw students is greater than the general population or the IQ of Jigsaw students is different from the general population. Both of these are negations of Null.

If I want to set up my alternate hypothesis as the IQ of Jigsaw students is greater than the general population, my null hypothesis ideally will be that

IQ of Jigsaw students is less than or equal to the general population. Or If I set up my null as IQ of Jigsaw student is the same as general population, I could set up an alternate as IQ of Jigsaw students is different from the general population. Which one of these is appropriate? It turns out that it depends on the business context.

>>>

If we have a strong reason to believe that, the IQ of Jigsaw students can only be greater than the general population, then we should use what is called One Tailed test or an alternate hypothesis test that this testing the IQ of Jigsaw students is greater than the general population.

To give you another example, let us say that I am a manufacturer of Cereals and I am testing packaged weight of my cereal products. If, for example, I am looking at a packaged weight of 500 grams and the actual weight of that product is less than 500 grams, you could be legally liable for under selling product. So if we were doing a quality control test for this particular manufacturer of cereal products, what would be checking for is - Is the packaged weight less than the advertised weight? We may not be worried at least from a legal standpoint of the other alternative which is – is the package weight greater than what was advertised? So there I would use what is called One Tail test.

>>>

What is a One Tail test? When you have an alternate hypothesis that says that the sample outcome is greater than the population outcome or the sample outcome is less than the population outcome. But if we don't have any strong reason to believe that the sample outcome has to be higher or lower than the expected population mean, we should use what is called a Two Tail test.

For example, if we have a manufacturing process that is generating auto parts and the weight of the auto parts has to be 0.8 lbs. Now when we run quality check, we will want to make sure that the weight is exactly 0.8 lbs. It should not be higher and it should not be lower. So what I will do in terms of setting up my alternate hypothesis - the alternate hypothesis will be that the process produces weights that are different from 0.8 lbs. They could be higher or they could be lower, but they are not 0.8 lbs. That is a Two Tail test.

Another example of a Two Tail test. If I am testing the efficacy of a drug and I really cannot be sure if it is better or worse than the existing protocol, use a Two Tail test. What is the implication of One Tail Vs Two Tail? In fact, why are they called One Tail or Two Tail?

>>

If we are testing a One Tail test, which means that the alternative hypothesis says that the sample outcome is expected to be higher than the population outcome or lower than the population outcome, then essentially, we are saying the sample averages will be on one side of the distribution or one Tail of the distribution. If higher, then the right side of the distribution; if lower, then the left side of the distribution and so, that is why this is called One Tail test.

On the other hand, if the alternate hypothesis is set up as sample mean will be different from the population mean, we are essentially saying, it could be higher or it could be lower, but we are not sure. Therefore, we expect the sample mean to be on either side of the distribution.

>>>

What is the implication in terms of the actual execution of a hypothesis test? It turns out that everything else about the hypothesis test stays the same. We have a null hypothesis; we will set up an alternate hypothesis depending on our business situation to be One Tailed of Two Tailed. We will figure out what is the appropriate test distribution to use and we calculate a P value. So until that point, everything is exactly the same.

The difference is that if we are using a Two Tailed test, we will compare the calculated probability of the sample outcome, being observed because of random chance variance to $\alpha/2$. The reason we are doing that is, remember α is the level of significance. For a Two Tailed test, we are saying we want to 95% confident. The random chance probability has to be less than 5%. But because we are doing a Two Tail test, we don't know whether we are looking at the right side of the distribution of the left side of the distribution. So we divide the α, the level of significance by 2. So the 5% becomes 2.5% on each side of the distribution, because distribution of a T-Test or a normal distribution is symmetric.

Therefore, what we do, we compare the P value to α/2. If α is 5%, we compare the P value to a 0.025 cut off value. If α is 10%, you will compare the P value to a 5% cut off. If we are doing a one Tail test, then there is no difference. You compare P to α and if P is less than α, then we reject the null hypothesis. So remember, the only difference in term of execution of a hypothesis test, when you are doing a Two Tail test is that you have to compare the calculated P value to α/2. Everything else about the hypothesis test stays the same.

If we think about this, which is the stricter test – One Tail or Two Tail test? A Two Tail test is much stricter because remember, you need more level of confidence to reject the null hypothesis for a Two Tail test relative to a One Tail test. Because for a Two Tail test, we want the rejection to be $P < \alpha/2$. Whereas for a One Tail test, we will reject P, if it is less than α.

In most business situations, we will end up doing One Tail test because most often, we have an idea of what we expect the sample out to be. For example, I have worked on process improvement project. After executing process improvement, I want to check whether the process has improved or not. It is unlikely that the process will deteriorate or I am doing some drug testing, most often we have an idea that if we use this sort of protocol, there will be an improvement in the outcome for the patient.

So very often in business situations, we end up doing One Tail test. But of course, there are some situations where a Two Tail test is more appropriate. If you really have no good reason to believe that, the sample outcome has to be higher or lower than the population outcome, you should be doing a Two Tail test. If there is a strong reason to believe that the sample outcome can only be higher or only be lower, then we should use One Tail test.

>>