MAT 222 — Spring 2019
Pooled *t*-Tests

Background

Recall when wanting to compare two means, μ_1, μ_2, we look at their difference: $\mu_1 - \mu_2$. The sign of this difference gives us some information:

$$\begin{cases} \mu_1 - \mu_2 > 0, & \mu_1 \text{ is greater than } \mu_2 \\ \mu_1 - \mu_2 = 0, & \mu_1 \text{ is equal to } \mu_2 \\ \mu_1 - \mu_2 < 0, & \mu_1 \text{ is less than } \mu_2 \end{cases}$$

Of course, the greater the magnitude of $\mu_1 - \mu_2$, the more μ_1 is greater/less than μ_2. For example, suppose we wanted to test the effectiveness of a new medication for those with chronic high blood pressure. If people taking the medication have an average lower blood pressure than those that do not take the medication, we could claim that the medication is effective. Notice that this is exactly comparing μ_1 with μ_2 – the average blood pressure of those on the medication to those not on the medication. To compare these averages, we examine $\mu_1 - \mu_2$. If $\mu_1 - \mu_2 > 0$, i.e. $\mu_1 > \mu_2$, then we might think the medication somehow *increases* blood pressure. If $\mu_1 - \mu_2 = 0$, i.e. $\mu_1 = \mu_2$, we might think the medication has no effect. While if $\mu_1 - \mu_2 < 0$, i.e. $\mu_1 < \mu_2$, we might think that the medication decreases blood pressure. However, there is an issue. We do not know μ_1, μ_2! So how are we to compare two unknown means?

We could create a simple random sample of individuals, one group which would receive the medication and one which would receive a placebo. Call the individuals receiving the medication Group 1 and the individuals which receive the placebo Group 2. Now call the average blood pressure of those in Group 1 (those on the medication) $\overline{x}_1$ and call the average blood pressure of those in Group 2 (those on the placebo) $\overline{x}_2$. So we have 'guesses' for μ_1, μ_2, the means from our groups, $\overline{x}_1$ and $\overline{x}_2$, respectively. This gives us an approximation for $\mu_1 - \mu_2$, namely $\overline{x}_1 - \overline{x}_2$.

We can now test the hypothesis that $H_0 : \mu_1 = \mu_2$ (the medication has no effect) against the hypothesis $H_a : \mu_1 < \mu_2$ (the medication reduces blood pressure). Assuming that the blood pressure of those on the medication has the $N(\mu_1, \sigma_1)$ distribution, i.e. normal distribution with mean μ_1 and standard deviation σ_1 and the blood pressure of those on the placebo has the $N(\mu_2, \sigma_2)$ distribution, the (two–sample) z statistic

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

where n_1 is the number of individuals in Group 1 and n_2 is the number of people in Group 2 (notice $\mu_1 - \mu_2 = 0$ as $\mu_1 = \mu_2$ by the null hypothesis). But again, we have a problem. We do not know σ_1 and σ_2. Again, we approximate σ_1, σ_2 using our samples. Approximate the standard deviation of Group 1, σ_1, by the standard deviation of Group 1 (call this s_1). Approximate the standard deviation of Group 2, σ_2, by the standard deviation of Group 2 (call this s_2). This gives us the two–sample t statistic:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

This is a bit misnamed as this statistic *does not* have a t–distribution. But we can still approximate this distribution by using a t–distribution with degrees of freedom equal to the smallest of the samples degrees of freedom. For instance, if we used a sample of 20 people for Group 1, i.e. $n_1 = 20$, and a sample of 17 people for Group 2, i.e. $n_2 = 17$, then the degrees of freedom for our 't–distribution' above would use the group of size 17 to get degrees of freedom 16.

This method allows us to make hypothesis test comparisons of two means as well as build confidence intervals for the difference of the two means. There are better approximations when we have more information about the standard deviations of the original samples. Notice above we did not know μ_1, μ_2 (then again we never do know this for hypothesis testing, this is the point of hypothesis testing) and we did not know σ_1, σ_2. However, if we knew more about σ_1, σ_2, we could create a better approximation for this distribution. This is exactly the idea of a pooled two–sample t statistic.

Pooled t–Procedures

Again consider the situation of the previous section, where we have two normal distributions: $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$, again with $\mu_1, \mu_2, \sigma_1, \sigma_2$ all unknown. If we knew more about σ_1, σ_2, we could create more accurate predictions about $\mu_1 - \mu_2$. After all, the more information, the more you can say! Suppose that there is reason to believe σ_1 is approximately σ_2 or better yet that $\sigma_1 = \sigma_2$. Then there is no need to distinguish between σ_1, σ_2 because they are the same! This allows us to give σ_1, σ_2 a single 'name', say σ. By mathematical nonsense, the *variance* of $\mu_1 - \mu_2$ is

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

Taking the square root gives us the standard deviation for $\mu_1 - \mu_2$ as

$$\sqrt{\sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Now we may know that $\sigma_1 \approx \sigma_2$ or $\sigma_1 = \sigma_2$, but we still do not know what this common value is! Once again, we will have to approximate σ.

Now both the standard deviation of Group 1, s_1, and the standard deviation of Group 2, s_2, approximate σ (since they both approximate σ_1, σ_2, which are both equal to σ). We can then 'average' s_1, s_2 to get a better approximation of σ. However, there is an issue (though at this stage are we really surprised there is a problem!?). Group 1 and Group 2 have different sizes so a simple average will not do. For instance, if 10,000 people voted for a tax rate of 9% but another 500 people voted for a tax rate of 11%, you would not implement a tax rate of 10% to be 'fair' – far more people voted for the 9% rate. You need to take into account the sample sizes. This is exactly why we cannot simply average s_1, s_2. Instead, we pool the samples together and average them by their 'weight'. [This is why it is called a pooled t–procedure.]

To see how this is done, take the sample of the voting again. Instead of thinking of these individuals as belong to the 9% or 11% groups, we pool them together into one large group of people that voted for one tax rate or the other. Now 10,000 people voted for the 9% rate and 500 for the 11% rate, making 10,500 total people in our sample. Recalling that an average is the sum of the values divided by the total number of values, this would (normally) give us

$$\frac{10,000 \cdot 9 + 500 \cdot 11}{10000 + 500}$$

But we are estimating variances and standard deviations, where we 'remove an individual' (recall in the formula for standard deviation, it is $n-1$ not n). So removing an individual from each of the groups gives 9,999 people for the 9% tax rate and 499 people for the 11% tax rate. This makes for a total of $9,999+499=10,498$ total individuals (note this is the same as $10,500-2$). Then our weighting is

$$\frac{9,999 \cdot 9\% + 499 \cdot 11\%}{10,500-2} = 9.095\%$$

Now applying this same idea to the variances s_1^2, s_2^2, this gives us an 'average', which we shall call s_p^2 for the pooled standard deviation,

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

To find our new approximation for the standard deviation, s_p, one only need take the square root. So instead of using $\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ for the standard deviation as we did before, we have a better approximation:

$$s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Again, we use the t–distribution. The degrees of freedom is n_1+n_2-2 – the 'denominator' in s_p (we do not need to subtract one from this since we already have from each group to obtain this number).

This finally gives us a confidence interval formula for $\mu_1-\mu_2$ and

$$(\overline{x}_1-\overline{x}_2) \pm t^* s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

and (pooled two–sample) t statistic

$$t = \frac{\overline{x}_1-\overline{x}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

A Few Things to Note

- Notice just as for the two–sample t procedure, σ_1, σ_2 must be unknown. However unlike the two–sample t procedure, we need to know σ_1, σ_2 are equal or approximately equal.

- Since the standard deviations are unknown, pooled two–sample t procedures can be risky as one does not know for sure that the unknown standard deviations are equal or approximately equal.

- Just like the two–sample t procedure for our approximations to work, we need...

$$\begin{cases} n_1+n_2<15, & \text{data close to normal.} \\ 15 \leq n_1+n_2<40, & \text{anytime except in case of outliers or strong skewness.} \\ 40 \leq n_1+n_2, & \text{anytime, even in presence of outliers or strong skewness.} \end{cases}$$

- The pooled two–sample t procedures are fairly robust against non–normality and unequal standard deviations (when the sample sizes are large). When the same sizes n_1, n_2 are very different in size, the pooled two–sample t procedures become very sensitive to unequal standard deviations. In this case, pooled t procedures should only be used if the sample sizes are large.

Example

A cellphone company has created a new battery design to compete with a rival company whose phones have longer battery life than their previous model phones. To examine this, the company's engineers test 16 of their new batteries and find a mean of 11.2 hours with standard deviation 2.3 and 13 of their competitor's batteries and find an average battery life of 10.8 hours with standard deviation 2.5. Assuming that the average battery life for both phones is normally distributed with equal standard deviations (since they use a similar design).

(a) Find a 90% confidence interval for the difference in average battery life.

(b) Find the t statistic and p–value for hypothesis that these engineers have designed a battery with longer battery life than their competitor.

(c) State your conclusion at the $\alpha = 0.05$ level.

Solution: Throughout, let $\overline{x}_1, s_1$ denote the mean and standard deviation for the engineers' new battery, respectively, and $\overline{x}_2, s_2$ denote the mean and standard deviation for the competitor's battery, respectively. Let n_1, n_2 denote the corresponding sample sizes. We need to use a t–procedure since the standard deviations of the populations are unknown (though we do know the *sample* standard deviations) and the original distributions are normal (so the size of $n_1 + n_2$ does not matter). We use a pooled t procedure because we believe the standard deviation for the two batteries is equal.

(a) The confidence interval is given by the values

$$(\overline{x}_1 - \overline{x}_2) \pm t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

We need only find the individual variables. First, $\overline{x}_1 - \overline{x}_2 = 11.2 - 10.8 = 0.4$. The degrees of freedom here is $n_1 + n_2 - 2 = 16 + 13 - 2 = 27$. Since we are using a 90% confidence interval (with degrees of freedom 27), we have $t^* = 1.703$. Furthermore,

$$\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \sqrt{\frac{1}{16} + \frac{1}{13}} = \sqrt{0.139423} = 0.373$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(16-1)2.3^2 + (13-1)2.5^2}{16 + 13 - 2} = \frac{154.35}{27} = 5.717$$

so that $s_p = \sqrt{5.717} = 2.391$. Then...

$$(\overline{x}_1 - \overline{x}_2) \pm t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 0.4 \pm 1.703(2.391)0.373 = 0.4 \pm 1.52$$

Therefore, we are 90% certain that the difference in battery life of the phones, $\mu_1 - \mu_2$, is between -1.12 and 1.92 hours, i.e.. there is a 90% the difference in battery life is in the interval

$[-1.12, 1.92]$. Put another way, there is a 90% chance that on average the company's battery last about 1 hour and 15 minutes less to approximately 2 hours longer than their competitor's battery.

(b) We are testing:

$$\begin{cases} H_0: & \mu_1 = \mu_2 \\ H_a: & \mu_1 > \mu_2 \end{cases}$$

To calculate the t statistics, we already have calculated all the necessary values:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} = \frac{0.4}{2.391(0.373)} = 0.449$$

Using the t table with degrees of freedom 27, this gives p–value greater than 0.25. [In fact, the exact p–value is 0.6715.]

(c) Since $\alpha = 0.05$ and $p \geq 0.25$, there is not enough evidence to suggest that the new battery has a longer battery life than their competitors battery.