Math 316

Section 7.3 Further Discussion of the t Test

Recall how we found a confidence interval for μ using the sample mean, etc.:

$$\mu = \bar{y} \pm t_{\alpha/2} \cdot SE_{\bar{Y}}$$

where

$$SE_{\bar{Y}} = \frac{s}{\sqrt{n}}.$$

We can find a confidence interval for the difference of two populations means $\mu_1 - \mu_2$:

$$\mu_1 - \mu_2 = (\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2} \cdot SE_{\bar{Y}_1 - \bar{Y}_2}$$

where

$$SE_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{SE_1^2 + SE_2^2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

In particular, we're interested in whether or not the confidence interval

$$(\bar{y}_1 - \bar{y}_2) - t_{\alpha/2} \cdot SE_{\bar{Y}_1 - \bar{Y}_2} < \mu_1 - \mu_2 < (\bar{y}_1 - \bar{y}_2) + t_{\alpha/2} \cdot SE_{\bar{Y}_1 - \bar{Y}_2}$$

includes the value of 0. If it does, then it is possible that $\mu_1 - \mu_2 = 0$, i.e. $\mu_1 = \mu_2$.
If 0 is not in the interval, then we conclude that $\mu_1 - \mu_2 \neq 0$, i.e. $\mu_1 \neq \mu_2$.

Let's look at HW 7.2.7 and its three variations from our previous class handout, but let's find the confidence intervals now. For the original version of the HW 7.2.7:

$$t_{\alpha/2} \cdot SE_{\bar{Y}_1 - \bar{Y}_2} = t_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 2.145\sqrt{\frac{2404^2}{8} + \frac{2217^2}{12}} \approx 2282.$$

Case	Change	Confidence interval	Confidence interval	Reject H_0?
1	Original case	-1470 ± 2282	$(-3752, 812)$	No
2	Larger difference $y_1 - y_2$	$\mathbf{-2810} \pm 2282$	$(\mathbf{-5092}, \mathbf{-528})$	Yes
3	Smaller s_1 and/or s_2	$-1470 \pm \mathbf{1279}$	$(\mathbf{-2749}, \mathbf{-191})$	Yes
4	Larger n_1 and/or n_2	$-1470 \pm \mathbf{1372}$	$(\mathbf{-2832}, \mathbf{-108})$	Yes

Confidence level affects the interval width. We'll again revisit HW 7.2.7 with different levels of confidence. In that problem we had $df \approx 14$ and $SE_{\bar{Y}_1 - \bar{Y}_2} = 1040$.

Conf.	α	$t_{\alpha/2} \cdot \ SE\ =$	Confidence interval	Confidence interval	Reject H_0?
80%	$.20$	$\mathbf{1.345} \cdot 1064 = \mathbf{1431}$	$-1470 \pm \mathbf{1431}$	$(-2901, -39)$	Yes
95%	$.05$	$2.145 \cdot 1064 = 2282$	-1470 ± 2282	$(-3752, 812)$	No
99%	$.01$	$\mathbf{2.977} \cdot 1064 = \mathbf{3167}$	$-1470 \pm \mathbf{3167}$	$(-4637, 1697)$	No

So higher confidence (good) means larger $t_{\alpha/2}$, which means a wider interval (bad)—it's a trade-off—which means we are less likely to conclude that $\mu_1 - \mu_2 \neq 0$ (i.e. $\mu_1 \neq \mu_2$), since a wider confidence interval is more likely to include 0.

When rejecting the null hypothesis and accepting the alternative hypothesis—or not—there are four possible situations and outcomes (this is Table 7.3.2).

Situation	What we should do	What we actually do	Did we err?
Population means are approximately the same	Don't reject H_0	Don't reject H_0	Nope
		Reject H_0, accept H_A	Type I error
Population means are different	Reject H_0, accept H_A	Don't reject H_0	Type II error
		Reject H_0, accept H_A	Nope

A Type I error can occur if the two population means are approximately the same, but due to the fact that we can get different samples from (or similar) populations with the same mean, we ended up with samples that seem to be from different populations. (Their test statistic t_s is large.) For example, if we had two populations with the same mean and standard deviation as given for **Example 5.2.5**, we might still end up with the samples in **(b)** and **(c)**. These two samples, although from populations that have the same mean, appear to be from populations that have different means. For those two samples we find

$$t_s = \frac{538 - 445}{\sqrt{\frac{119^2}{25} + \frac{113^2}{25}}} = 2.83$$

(which, with $df \approx 48$, gives a P value of .007—that is, there is a 0.7% likelihood of two samples being this different if the populations from which they come have the same mean). A Type 2 error can occur if the populations' means really are different, but due to chance the samples appear to be approximately the same. For example, suppose that for two populations

$$\mu_1 = 525 \text{ and } \sigma_1 = 15 \qquad \mu_2 = 475 \text{ and } \sigma_2 = 20,$$

but that in their respective samples we have

$$\bar{y}_1 = 505 \text{ and } s_1 = 17 \qquad \bar{y}_2 = 495 \text{ and } s_2 = 19.$$

These samples appear to come from populations that are quite similar, similar enough that we would *not* conclude that they are different. For example, if both samples had sample sizes $n_1 = n_2 = 10$, we would have the (not very large) value of

$$t_s = \frac{495 - 505}{\sqrt{\frac{17^2}{25} + \frac{18^2}{25}}} = 1.24.$$

We would end up *not* concluding that $\mu_1 \neq \mu_2$. Remember that *not* concluding that $\mu_1 \neq \mu_2$ does *not* mean that we *are* concluding that $\mu_1 = \mu_2$.

See the paragraph after Example 7.3.4 for a brief discussion of what Type I and Type II errors might mean in a real life situation. HW 7.3.8 asks you to determine what each of these errors would mean in a different real life situation.

When you read the *Significance Level versus P-value* paragraph on page 245, keep in mind that α is how much uncertainty we can accept, and P is (in a way) our level of certainty that $\mu_1 = \mu_2$. Recall the more precise definition of P:

> P is the likelihood of getting two samples that are this different, or even more different (as measured by t_s), if it were actually true that $\mu_1 = \mu_2$.

For example, if $\alpha = 0.05$, then we can accept a likelihood of up to 5% of making a Type I error. If $P = 0.037 < 0.05$ we would reject the null hypothesis. $P = 0.037$ means we are only about 3.7% certain the two population means are the same, so in deciding that they are different (by rejecting the null hypothesis), there is a $P = .037$ likelihood we are making a Type I error. The power of a test is the likelihood of not making a Type II error. See the short discussion starting at the bottom of page 247. We'll further discuss power in Section 7.7.

Section 7.4 Association and Causation

A couple of words we'll see later. Their standard definitions:

- Confound: confuse, mix up.
- Spurious: fake, false, misleading, deceptive, not being what it purports to be.

Two types of studies:

- In an experiment, the researcher intervenes in or manipulates the experimental conditions.
- In an observational study, the researcher merely observes an existing situation.

One big question we often have to deal with is what actually causes the differences in whatever it is we are measuring. In **Example 7.4.3**, they describe how low baby birth weights are associated with the smoking during pregnancy. By now we understand that smoking is bad for you in every way (including during pregnancy), but years ago it was yet not clear that smoking harmed unborn babies. Careful studies were needed to show that.

Association is not causation. In **Example 7.4.5** they continue their discussion about how one variable (such as how much the woman drinks) might confound or confuse the effect of the other variable (such as how much the woman smokes).

While our modern understanding of smoking by now makes that cause (smoking) and effect (lower birth weight) pretty obvious, it could also be that there is some other cause of the lower birth weights, such as poor diet, poor prenatal care, alcohol consumption, etc., all of which might be effects of simply being poor. So perhaps women who smoke tend to be poorer, which results in poorer diet and prenatal care, and more drinking, which are the true causes of low birth weight. What you would need is a study with two groups of women in which the average diet, prenatal care, and anything else that you might expect to affect birth weight are the same, but in which one group the women smoke and in the other group they don't smoke. See the top of page 254 for some discussion on this. It's tough to come up with exactly that situation. As I've said before, the easier part of collecting and analyzing data is generally the analysis. The harder part is creating a good study from which we get the data.

Not only can the effect of one variable confound or confuse the effect the other, there may be a variable that seems to have an effect on something but in reality does not. (This is a little different that the smoking and drinking which probably both have some negative effect on birth weight, it's just that they are confounding variables because it is not clear how much it is the smoking and how much it is the drinking that results in lower birth weight.) **Example 7.4.7** describes how some once believed that ultrasounds might cause (rather than just giving us images of) birth defects in unborn children. A possible current version of this issue is whether mammograms (in our attempt to detect breast cancer) are actually increasing the likelihood of breast cancer (which some people believe).

See the book's thoughts on <u>experimental units</u>. Also, a <u>randomization distribution</u> is simply all of the possible samples we could have from a particular collection of data. For example, if our population were the numbers 1, 2, 3 and 4, and we were taking samples of size 2, the randomization distribution would consist of the six samples 1 & 2, 1 & 3, 1 & 4, 2 & 3, 2 & 4 and 3 & 4 (along with their respective means and standard deviations). **Example 5.1.3 and Table 5.1.1** also give an example which gives all possible samples.

If time, we'll look at **HW 7.4.2**.

Statistical ideas and method are very useful, but must be treated with caution, else the saying "there are lies, there are damn lies, and then there are statistics" will actually become reality. I mention this in relation to causation and association.