Math 316

Section 7.1 Hypothesis Testing: The Randomization Test

Read this on your own. I think Section 7.1 makes more sense after reading and understanding Section 7.2. You might try reading it before 7.2 and again afterwards.

Section 7.2 Hypothesis Testing: The t Test

This is our first time doing hypothesis testing. Hypothesis (from Google): "A supposition or proposed explanation made on the basis of limited evidence as a starting point for further investigation." In this section we look at the hypothesis that *two population means are different*. For example, we might hypothesize that the average Pepperdine student height is different from the average LMU student height.

First, an old idea. Then a related new idea.

- Old idea: For a population with mean μ and standard deviation σ, the z-value for a sample mean $\bar{y}$ is

$$z = \frac{\bar{y} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\bar{y} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

  measures how far away the *sample* mean is from the *population* mean, relative to the standard deviation (and sample size matters as well).

- New idea: For two samples from two different populations, the test statistic

$$t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{SE_{\bar{Y}_1 - \bar{Y}_2}} = \frac{\bar{y}_1 - \bar{y}_2}{SE_{\bar{Y}_1 - \bar{Y}_2}} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

  measures how different the two samples means are from each other. That is, t_s tells us how far away their difference is from 0, relative to their standard deviations (and samples sizes matter as well), i.e. relative to the standard error. (As described below, we typically start by assuming that the two populations have the same mean, which would mean that we would expect that the two sample means $\bar{y}_1$ and $\bar{y}_2$ are the same, i.e. that their difference would be 0).

  Since these t_s values are new to you, you're wondering how large t_s needs to be in order to consider it extreme, kind of like how when you first learned about z values you weren't sure what a large z value was. It turns out that the values of t_s and z are similar. For example, $z = 2$ would be pretty extreme and similarly $t_s = 2$ would be pretty extreme.

Thoughts on t_s and P:

- Test statistic t_s measures how different the two samples are. The larger t_s, the more different the samples are and the more likely it is that we will decide that the two populations from which the samples come have different means. Table 4 helps us decide how extreme t_s: we use t_s to find the corresponding P-value.
- Assuming the two samples come from populations with equal means (in general, if the null hypothesis H_0 were true), the P-value is the probability that we would get two samples that are this (or even more) different, as measured by t_s. That is, the P-value is the fraction of the t_s values larger than the given t_s value. Visually, the P-value is the area in the tails beyond t_s.

Let's work **HW 7.2.7.** We will have found that $t_s = -1.38$ and $P = 0.1892$. Meaning: if the two population means really were the same, then there would be a probability of 0.1892 of getting a test statistic of ± 1.38 or larger (in magnitude), i.e. there is a .1892 probability that we would have $|t_s| \geq 1.38$. (For now, we don't care so much whether $t_s > 0$ or $t_s < 0$. We are just interested in its size.) So if, based on these samples, we decide to conclude that the population means are actually *different*, there is essentially a 18.92% likelihood that we are making a mistake. We have to decide how much potential error we can live with. This is the value of α. In HW 7.2.7 $\alpha = 0.05$. Anytime we decide to conclude, based on the two samples, that the two populations actually have different means, there is always some chance we are making a mistake.

Let's summarize the basic steps of this process:

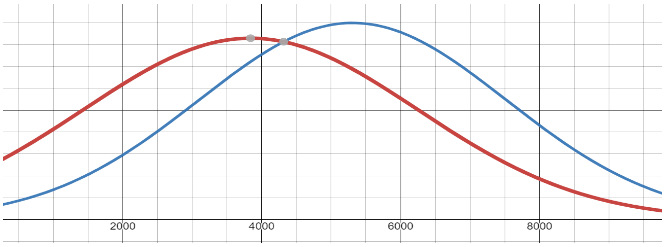
- Decide what the null hypothesis H_0 and alternative hypothesis H_A are.
- Decide what level α of uncertainly is acceptable. (If we wanted to be 95% confident, then we would have $\alpha = 5\% = .05$.)
- Compute the test statistic t_s, and find the corresponding P-value using Table 4 or technology, such as Excel (e.g. $=\mathrm{TDIST}(1.38, 14.268, 2)$).
- If $P < \alpha$ then reject H_0 and accept H_A. Otherwise, we've not definitively shown anything. In particular, by not rejecting H_0, we are not actually accepting it. See the "subtle issue" at the top of this handout page 4.

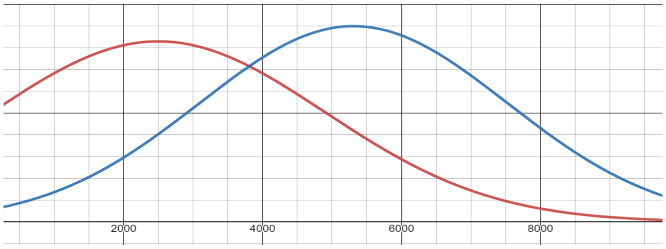
Using the online **Compare two sample means Excel worksheet**, let's work **HW 7.2.7** again with other values. In each case $\alpha = 0.05$.

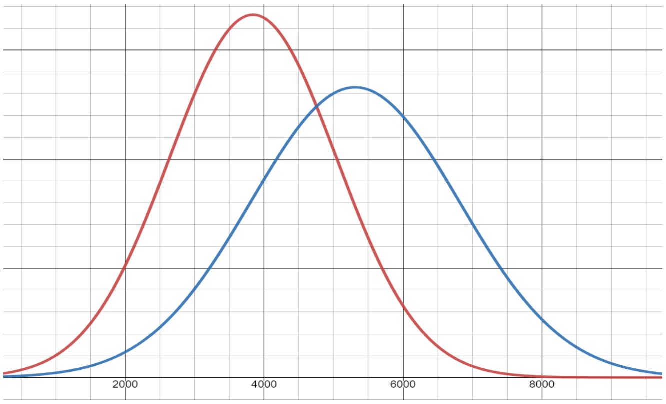
Case	**Original**		**Larger $y_1 - y_2$**		**Smaller s_1, s_2**		**Larger n_1, n_2**	
Samples	S1	S2	S1	S2	S1	S2	S1	S2
n	8	12	8	12	8	12	**20**	**30**
$\bar{y}$	3840	5310	**2500**	5310	3840	5310	3840	5310
s	2404	2217	2404	2217	**1200**	**1500**	2404	2217
SE	850	640	850	640	**424**	**433**	**538**	**405**
$Comb.$	1064		1064		**606**		**673**	
t_s	-1.38		**-2.64**		**-2.42**		**-2.18**	
df	14.3		14.3		**17.3**		**38.5**	
P	.189		**.019**		**.027**		**.035**	
Reject H_0?	No		**Yes**		**Yes**		**Yes**	

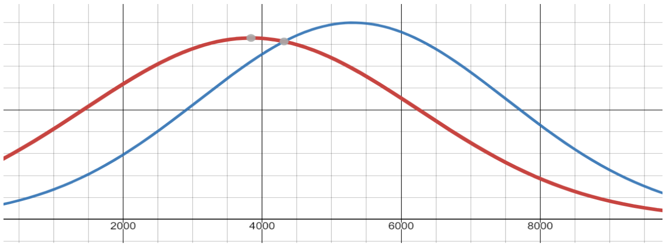
Remember: P is the likelihood we would get two samples that are this (or even more) different, as measured by t_s, *if the two population means were actually the same*.

Visually it's all about overlap of samples. What do these four cases look like?









Some basic intuition:

t_s is larger $\Rightarrow P$ is smaller

$\Rightarrow$ It is more likely that we will conclude the two population means are different.

Next, a subtle issue. "Innocent until proven guilty." In trying a person in court, there are two possibilities: (1) there is sufficient evidence to convict or (2) there might be some evidence, but not enough to convict. Option (2) does not mean that we have proven that the person is innocent—the person might actually be guilty—it simply means that there is not enough evidence to convict. Similarly, in hypothesis testing, there are two possibilities:

1. There is sufficient evidence to conclude that the two means are different: reject the null hypothesis (population means are the same) and accept the alternative hypothesis (population means are different); or
2. The samples might be different, but not different enough to conclude that the two means are different, so don't reject the null hypothesis.

Option 2 does *not* mean that we have proven that the two means are *not* different. That is, we have *not* proven that they are the same: the two population means might actually be different. It simply means that there is not enough evidence to show that they are different at the desired level of confidence. See "Note carefully…" in the middle of page 235.

A couple of miscellaneous notes:

- See the note on Using Tables versus Using Technology on Book Page 235. Bracketing means that we have a range (a lower and an upper bound) for the value of P, rather than the exact value. For example, in HW 7.2.7, using Table 4 we found that $.10 < P < .20$. We "bracketed" the value of P.
- In Section 7.5 we will look at trying to show that the mean of one population is *greater than* (or *less than*), rather than simply *different from*, the mean of another population.