David Youngberg
BSAD 210—Montgomery College

LECTURE 25: COMPARING TWO POPULATIONS II

I. Unknown but Symmetric σ

a. When we don't know σ, two things happen:
   i. We will use the t-distribution rather than the normal distribution
   ii. A minimum sample size of 30 is needed ***or*** both populations need to be normally distributed.

b. Suppose we don't know what σ is but we have good reason to believe it should be the same across samples.
   i. A common example of this is when you pull two different samples from the same group and then you do different things to each group.
      1. Examples: lab animals, customers, students

c. Dr. Betty Ortega would like to know which painting—a small dog painting or a large dog painting—her patients prefer to see in her waiting room.
   i. One week she hangs the small dog painting and has 30 patients.
   ii. The next week she hangs the large dog painting and has 35 different patients.
   iii. She then has them rate the waiting room on a scale of 0 to 10 (10 being high). (Note she's ignoring any patients that came both weeks; she wants to make sure the samples are independent.) Here are the results:

	x-bar	**s**	**n**
Small	7.8	0.8	30
Large	7.5	0.5	35

   iv. Note that the sample standard deviations are different; that's okay. We have every reason to believe that if the same people who saw the small dog painting saw the large dog painting instead, they would have a similar consensus. Remember: this is a sample.

d. Because we assume the population standard deviations to be the same, we must "pool" the standard deviations. We call this value, s_p.

e. Using the equation below, we calculate the t-score; let's start with the pooled variance (recall the variance is the standard deviation squared.

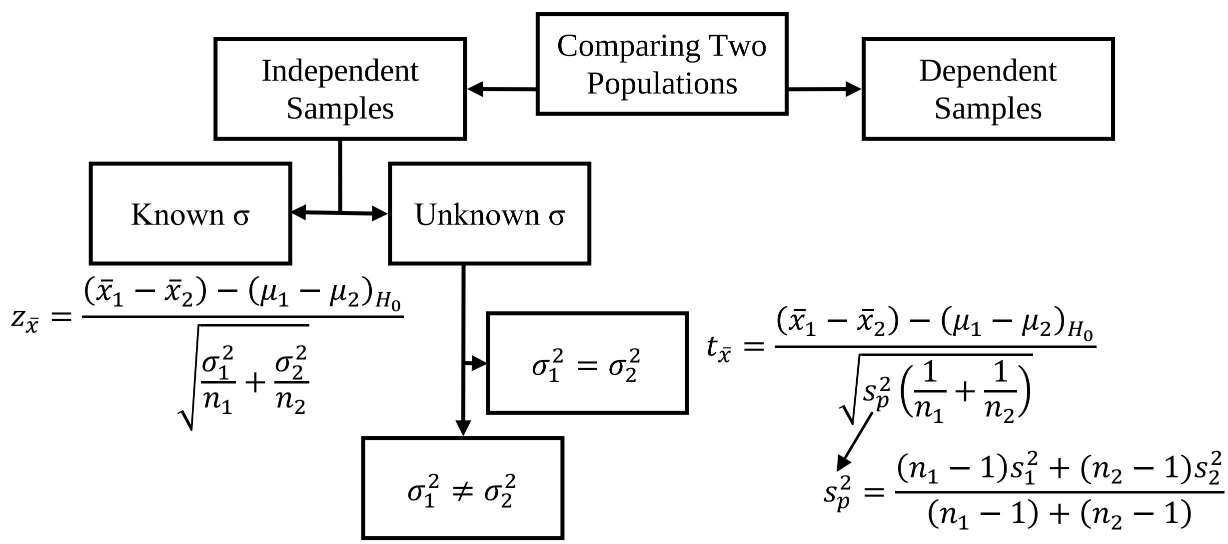
$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)} = \frac{(29)0.8^2 + (34)0.5^2}{29+34} = \frac{18.56+8.5}{63} = 0.43$$

$$t_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{7.8-7.5}{\sqrt{0.43\left(\frac{1}{30}+\frac{1}{35}\right)}} = \frac{0.3}{0.163} = 1.84$$

f. Our degrees of freedom is equal to all our observations (65 total) minus two (because we have two samples) for a total of 63.

g. Since it's a two-tailed test, we are significant at the 90% level (at 60 df, t=1.671) but not at the 95% level (at 60 df, t=2.000).

   i. You could argue that we have 63 degrees of freedom, not 60, and you're correct. But at 70 df, t=1.994; we still wouldn't make 95% confidence.



II. Unknown and Asymmetric σ

a. Sometimes we can't claim the population standard deviation is the same. For example, a cat painting is more likely to be more controversial than a dog painting. Let's revisit Dr. Ortega now choosing between the more popular dog painting and a cat painting:

	x-bar	**s**	**n**
Dog	7.8	0.8	17
Cat	8.5	1.3	26

i. Note we've dropped below our 30 sample size minimum. We must assume a normal distribution. If it turns out we're wrong—say we look at the distribution and it's not normal—we can't use this equation. We don't really know what's going on.

b. We begin with a simpler equation:

$$t_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} = \frac{7.8 - 8.5}{\sqrt{\left(\frac{0.8^2}{17} + \frac{1.3^2}{26}\right)}} = \frac{-0.7}{0.3204} = -2.18$$

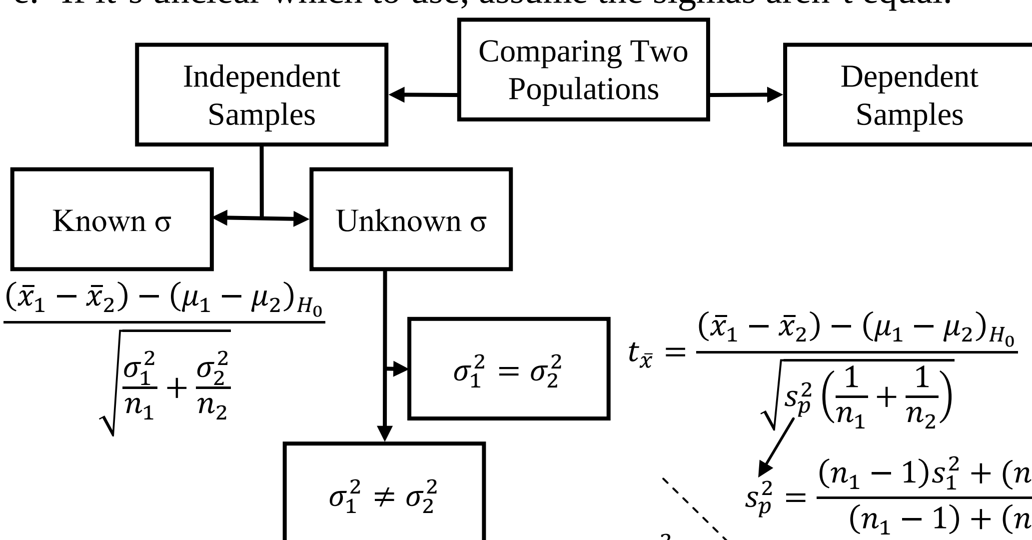
c. But there's a hitch: the degrees of freedom equation is much more complicated:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{\left(\frac{0.8^2}{17} + \frac{1.3^2}{26}\right)^2}{\frac{\left(\frac{0.8^2}{17}\right)^2}{17 - 1} + \frac{\left(\frac{1.3^2}{26}\right)^2}{26 - 1}} = \frac{0.01054}{0.00026} = 40.91$$

i. Always round down with df; our df=40.

d. With df of 40, the t-score for 95% confidence is 2.021; we have statistical significance.

e. If it's unclear which to use, assume the sigmas aren't equal.



$$z_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\boxed{\sigma_1^2 = \sigma_2^2} \quad t_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

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

$$\boxed{\sigma_1^2 \neq \sigma_2^2}$$

$$t_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} \quad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$