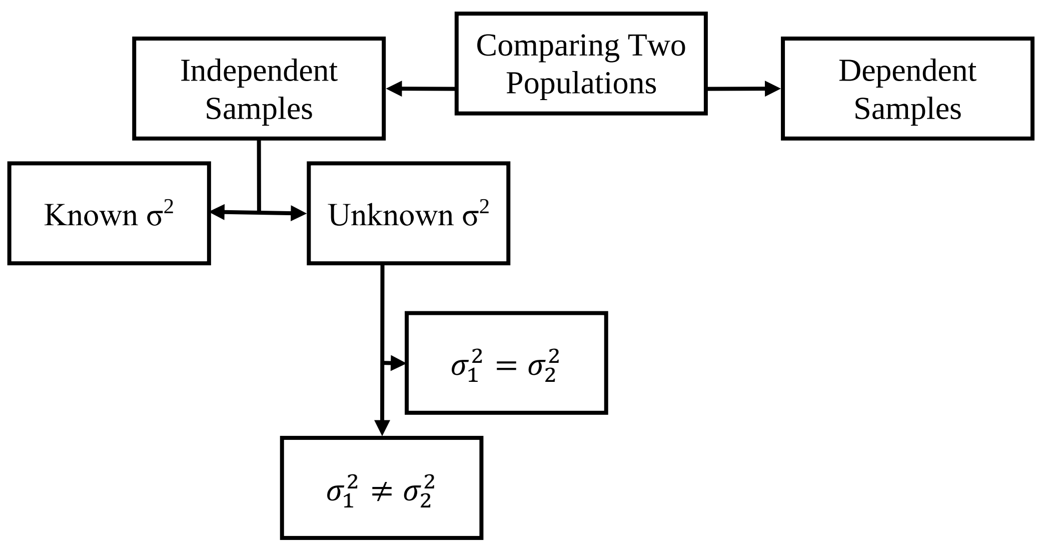
David Youngberg
BSAD 210—Montgomery College

LECTURE 18: COMPARING TWO POPULATIONS I

I. On Comparing Populations
   a. We don't really compare two different populations directly; as usual, we take a sample from each population.
   b. But we call it comparing populations because these samples represent that bigger picture. There's really no point to limiting our analysis to just the sample. We want to know what's going on for the population.
   c. There's a lot of complexity that seeps into this sort of analysis. Two summarize what we'll be working with, here's a diagram:



      i. When samples are *independent*, each observation for one sample does not change with the observations for the other sample.
         1. Example: the safety ratings of two different kinds of cars. If one trial for a Honda is good, that wouldn't alter the trial of a Ford good or bad. Ford's safety is independent from Honda's.
      ii. *Dependent* samples are when the observations of one sample change based on the observations of the other sample.
         1. Example: comparing the effectiveness of a diet. The more you weigh before a diet, the more you should weigh after a diet, even if the diet is effective. In other words,

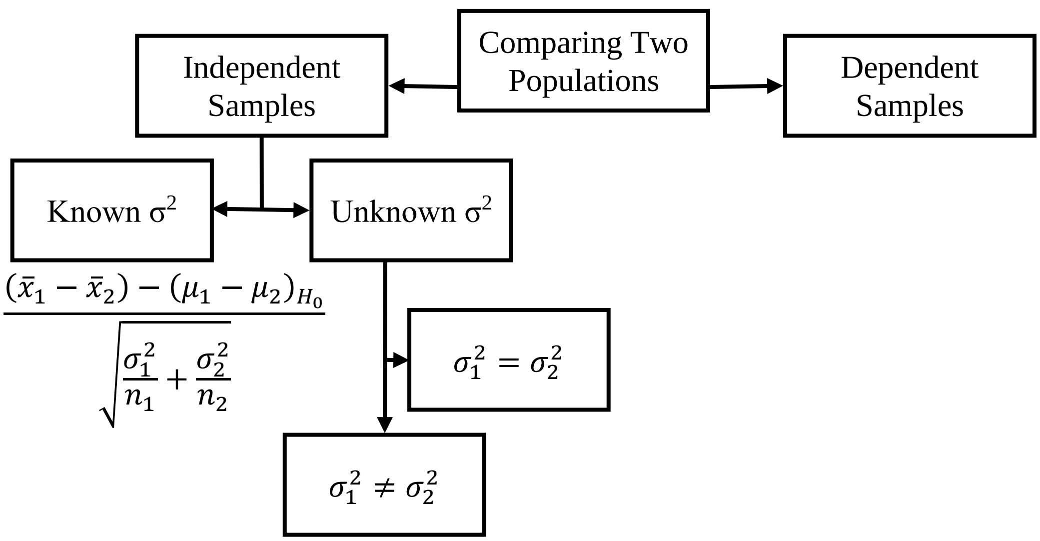
what you're really interested in is the change in weight between each observation.

iii. The population standard error can either be known or unknown. Known σ^2 typically relate to samples drawn from a population with a pre-existing body of empirical data.

II. Known σ^2

a. Let's begin with the simplest case: the variance (standard deviation squared) of both populations is known.

b. This occurs when there's a large body of pre-existing data on the population in question. Here's the equation:



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

III. Example

a. Suppose our company has a new candy bar and marketing came up with two different displays. Which display works better? We put each display in different stores to find out.

b. Our null hypothesis is that the difference doesn't matter: $\mu_1 = \mu_2$; our alternative is that it does: $\mu_1 \neq \mu_2$.

i. We used Display 1 in 50 stores and got $400K in sales.

ii. We used Display 2 in 25 stores and got $410K in sales.

iii. We know from previous new candy bars, the population standard deviation for each is $20K.

$$z_{\bar{x}} = \frac{(400 - 410) - 0}{\sqrt{\frac{20^2}{50} + \frac{20^2}{25}}} = \frac{-10}{\sqrt{\frac{400}{50} + \frac{400}{25}}} = \frac{-10}{\sqrt{8 + 16}} = \frac{-10}{4.899} = -2.04$$

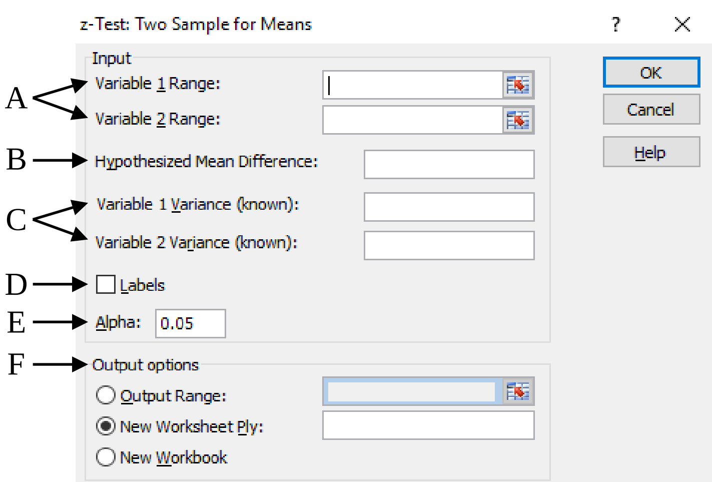
iv. At 95% confidence (z=1.96), the results are statistically significant.

IV. A Note About Excel

a. Excel has all the different tests we're talked about programmed in.
b. ***However*** these tests require the raw data as the commands are built so that Excel will calculate the means, standard deviations, and sample sizes itself.
c. If you only have the relevant values without the raw data (which sometimes happens), you have to calculate the t or z values yourself.
d. Use the data found in Dataset 4. This is a database of every MC professor's rating at Rate My Professor, provided they had at least 25 ratings. The data was gathered in July 2014.

V. Excel: Known Variance

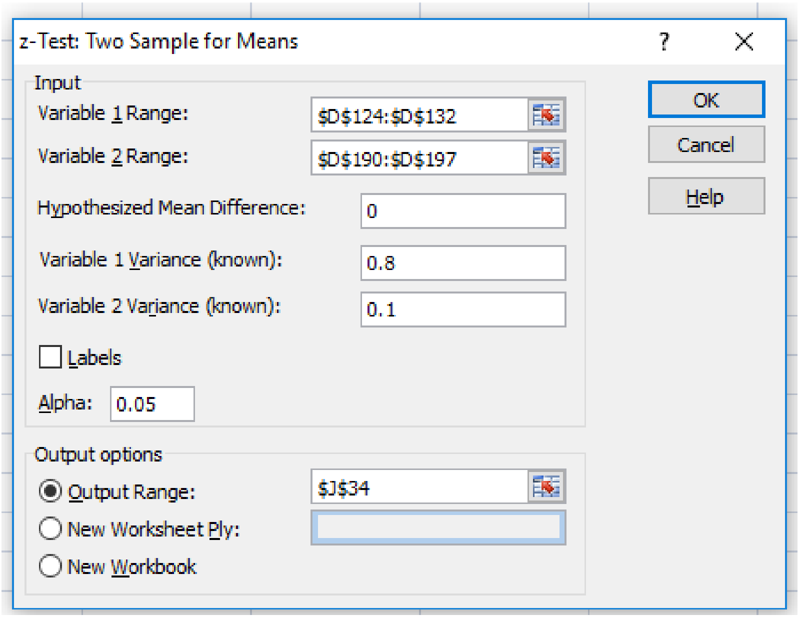
a. Go to Data >>> Data Analysis >>> z-Test: Two Sample for Means and click OK.
    i. You'll get a window that looks like this.



Key

A: The raw data for each of your samples. Select the field and then highlight the data.
B: The null hypothesized difference between sample means. Typically zero.
C: Variance is standard deviation squared. Also make sure that which variable is #1 and which is #2 is consistent with **A**.
D: Select this and Excel will assume the first row of your data is a label for each variable.
E: The significance level at which you're testing.
F: Where the results will display. I typically choose the first option and select a cell away from the raw data.

b. Let's use the data to determine which discipline has the higher rating: history or psychology.
    i. Note I've grouped this data by discipline; all the history Quality ratings and all the psychology Quality ratings are easy to find.
    ii. I don't have data on the population variances so we'll make them up for this example. Suppose history had a population variance of 0.8 and psychology has a population variance of 0.1.



c. You'll get an output like this (here Variable 1 is history and Variable 2 is psychology):

z-Test: Two Sample for Means		
	Variable 1	*Variable 2*
Mean	3.63333	4.3625
Known Variance	0.8	0.1
Observations	9	8
Hypothesized Mean Difference	0	
z	-2.28998	
P(Z<=z) one-tail	0.01101	
z Critical one-tail	1.64485	
P(Z<=z) two-tail	0.02202	
z Critical two-tail	1.95996	

i. First, note that psychology has a higher mean; but is this difference statistically significant?

ii. Also note the one- and two-tailed critical values; they are same as on our table for that α because we know the population standard deviation/variance.

   1. Since our null hypothesis is that the averages are equal, we should focus on the two-tailed version. This is typically what you will focus on for this kind of test.

iii. Z is the calculated value: about -2.290; the absolute value is greater than critical value. This difference is statistically significant.

VI. Unknown but Symmetric σ^2

a. When we don't know σ^2, 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. Note if the sample size is less than 30, we simply assume a normal distribution.
b. Suppose we don't know what σ^2 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.

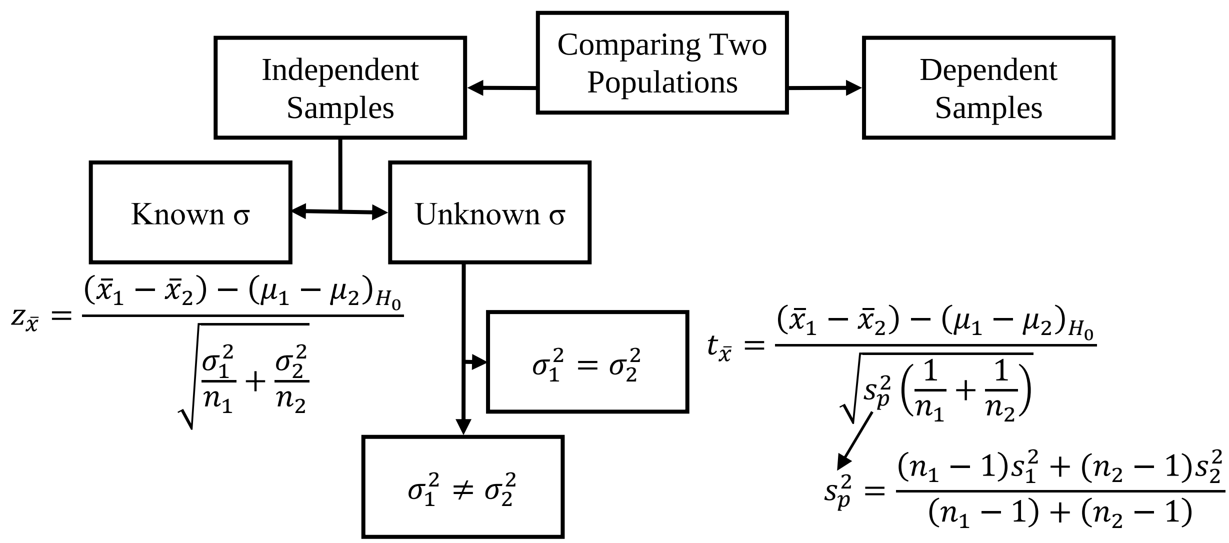
$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{(29)0.8^2 + (34)0.5^2}{29 + 34} = \frac{18.56 + 8.5}{63} = 0.43$$

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

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.

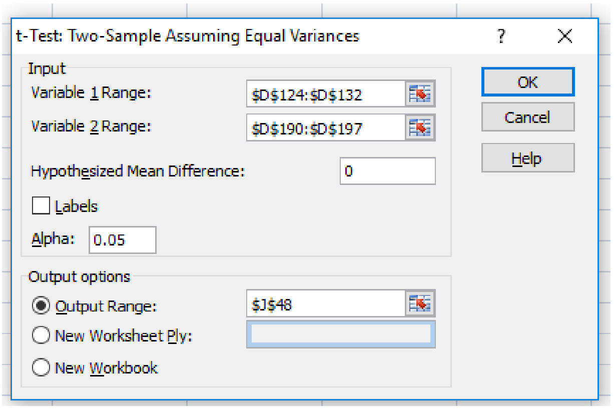
g. Since it's a two-tailed test, we are significant at the 90% level (at 60 df, t=1.671) but not at the 95% level (at 60 df, t=2.000).

   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.



VII. Excel: Unknown and Equal Variance

a. Now let's do the same thing but select "t-Test: Two-Sample Assuming Equal Variances".



t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	3.63333	4.3625
Variance	0.9425	0.13696
Observations	9	8
Pooled Variance	0.56658	
Hypothesized Mean Difference	0	
df	15	
t Stat	-1.99359	
P(T<=t) one-tail	0.03235	
t Critical one-tail	1.75305	
P(T<=t) two-tail	0.06471	
t Critical two-tail	2.13145	

b. Because we have less information (we didn't know the population variances), the critical values increases; it's harder to find statistical significance.

c. In this case, the absolute value of the t stat is not greater than the critical value. The difference between these two averages is probably by chance.