

## LECTURE 19: COMPARING TWO POPULATIONS II

### I. Unknown and Asymmetric $\sigma^2$

- 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
<b>Dog</b>	7.8	0.8	17
<b>Cat</b>	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.

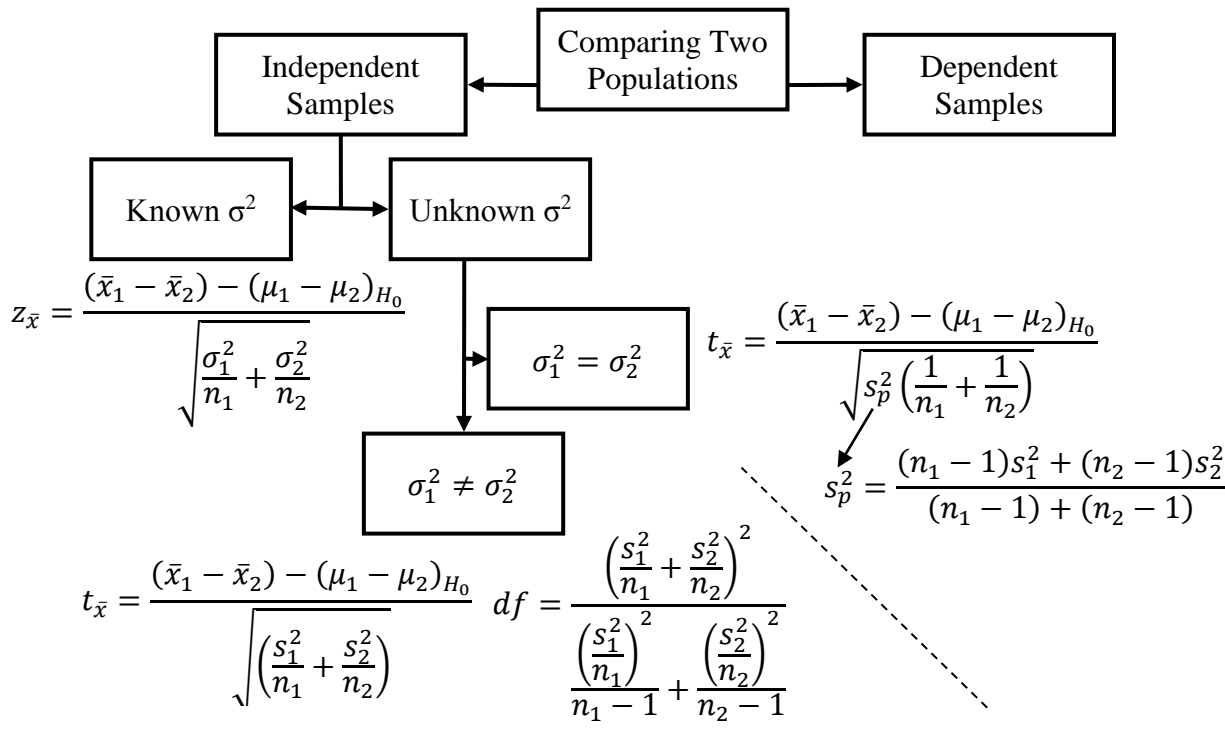
- b. We begin with a simpler equation:

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

- 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.



II. Excel: Unknown and Unequal Variance

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

	Variable 1	Variable 2
Mean	3.63333	4.3625
Variance	0.9425	0.13696
Observations	9	8
Hypothesized Mean Difference	0	
df	11	
t Stat	-2.08894	
P(T<=t) one-tail	0.03038	
t Critical one-tail	1.79588	
P(T<=t) two-tail	0.06075	
t Critical two-tail	2.20099	

- b. Note that with unequal population variances, the critical value increases because we have even less information than when we assumed equal variances.
- c. The absolute value of the t-stat may increase or decrease; here it happened to increase.
- d. Again, the difference is not statistically significant.

### III. Dependent samples

- a. Now we turn to when each observation in one sample is related to an observation in the other sample: a *dependent sample*.
- b. A dependent sample occurs when you can match two different observations from two different samples based on a common subject. This is called a *matched-pair test*.
  - i. Examples: customer behavior before and after a promotion; a person's weight before and after a diet; store performance when an additional person is hired; etc.
  - ii. Note many of these examples involve the same subject with before and after measurements. But it could also involve testing two different kinds of subjects where each observation in one sample has a natural connection to exactly one observation in the other sample.
  - iii. Example: You might have one sample of husbands' rating of their marriage and another sample of those husbands' wives' rating of their marriage. By looking at the difference between each pair, you can use this data to answer the question of who has a higher opinion of their marriage: men or women?
- c. In this analysis, each observation in a sample has a "partner" in the other sample. We're interested in the difference between each pair, called *d*.
  - i.  $d = x_1 - x_2$
- d. Adding all these differences and dividing by the sample size results in *d-bar*, the average matched-pair difference.
  - i. This means that the number of observations from each sample must be equal.

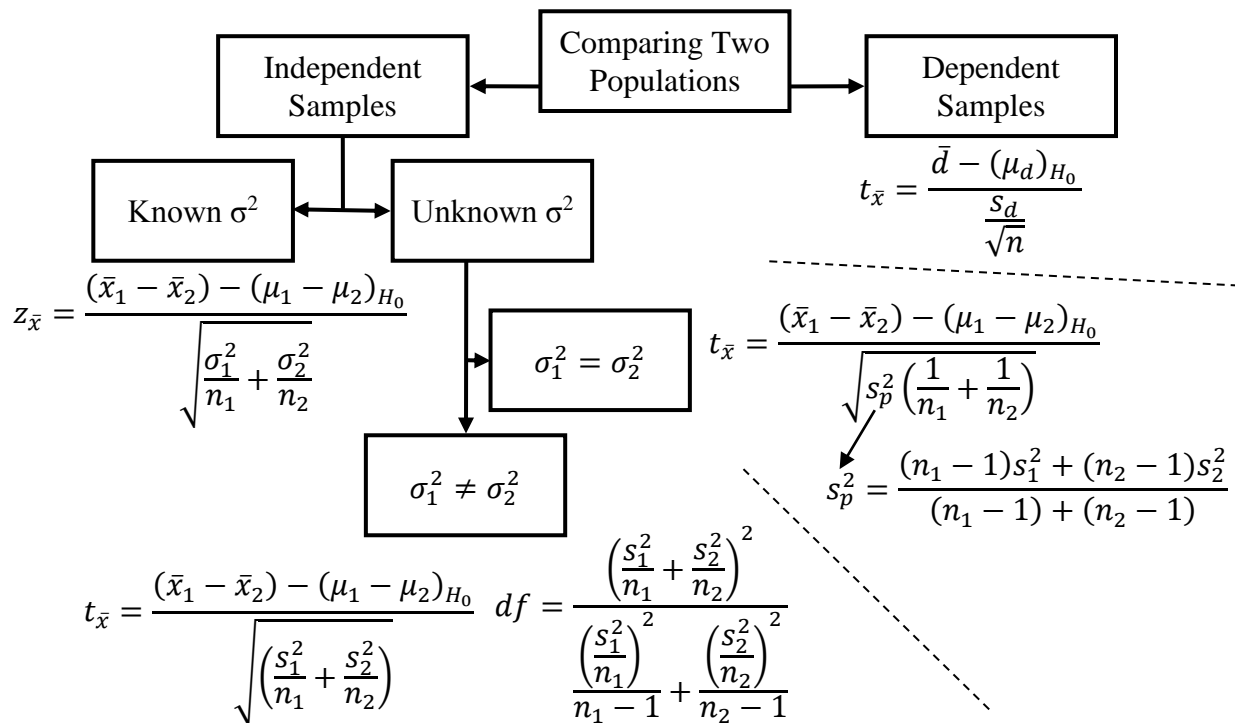
$$t_{\bar{d}} = \frac{\bar{d} - (\mu_d)_{H_0}}{\frac{s_d}{\sqrt{n}}}$$

- ii. Where  $s_d$  is the standard deviation of the differences; and
  - iii. *d-bar* is the average difference of the matched-pairs.
- e. Suppose you're testing a weight-loss pill. You weigh 16 randomly selected people. The average weight change is -2.6 pounds and the standard deviation is 5. Note this is a one-tailed test.

$$t_{\bar{x}} = \frac{\bar{d} - (\mu_d)_{H_0}}{\frac{s_d}{\sqrt{n}}} = \frac{-2.6}{\frac{5}{\sqrt{16}}} = \frac{-2.6}{\frac{5}{4}} = \frac{-2.6}{1.25} = -2.08$$

- f. With 15 degrees of freedom, the critical value at  $\alpha=0.05$  is 1.753 so it's statistically significant. It's barely shy the 2.131 value at  $\alpha=0.025$ .

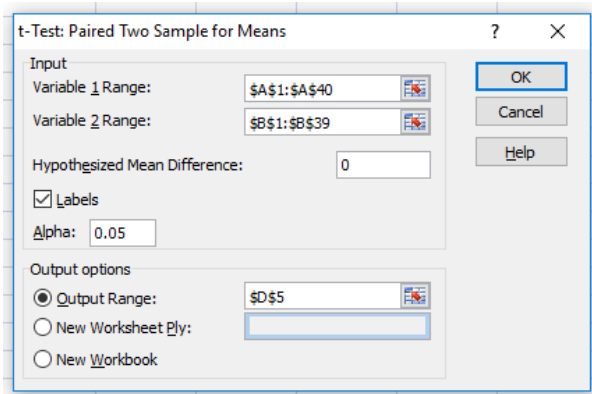
#### IV. Finished Diagram



#### V. Excel: T-test: Paired Two Sample for Means

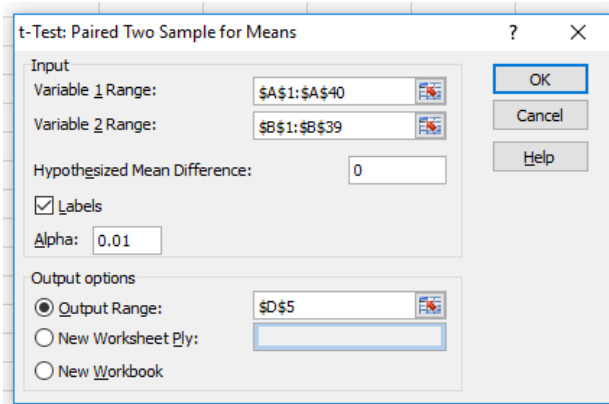
- In the second tab, under "Homework Data," are the scores from two different homeworks from a class I taught several years ago. Each row represents one student's score on the first homework (H1) and that student's score on the second homework (H2).
- Which homework did students do better on? While we can compare averages, it's possible those differences are random. Again, the CLT rears its head.
- So we use a paired test. But why not just use one of the previous tests?
  - We would if one group of students did H1 and a different group did H2.
  - But it's easy to see why it's more likely differences in means would be by chance since average intelligence, time for homework, motivation, etc. could vary widely across groups.

- iii. For this data, though, many differences between those who completed H1 and those who completed H2 aren't there. The same people are taking the homework. Yes, time for homework might differ, but probably not as much as if the students were completely different people. Other traits like intelligence and interest level are probably unchanged.
  - iv. In other words, we have more information so our calculations should give us a bit of a break.
- d. Here's the test:



t-Test: Paired Two Sample for Means		
	H1	H2
Mean	66.15385	54.48718
Variance	1018.449	1499.52
Observations	39	39
Pearson Correlation	0.564885	
Hypothesized Mean Difference	0	
df	38	
t Stat	2.175304	
P(T<=t) one-tail	0.017949	
t Critical one-tail	1.685954	
P(T<=t) two-tail	0.035897	
t Critical two-tail	2.024394	

- i. Note I'm using Labels here.
- e. Since the absolute value of the t stat is greater than the critical value, we can claim statistical significance!
- i. Try doing this same test again but with Two-Sample Assuming Unequal Variances; you won't find statistical significance because the test assumes we don't have information that we actually have. The test you choose matters a lot!
- f. Try doing the same test but with 99% confidence rather than 95%:



t-Test: Paired Two Sample for Means		
	H1	H2
Mean	66.15385	54.48718
Variance	1018.449	1499.52
Observations	39	39
Pearson Correlation	0.564885	
Hypothesized Mean Difference	0	
df	38	
t Stat	2.175304	
P(T<=t) one-tail	0.017949	
t Critical one-tail	2.428568	
P(T<=t) two-tail	0.035897	
t Critical two-tail	2.711558	

- g. We no longer have statistical significance.
- VI. =T.TEST
- a. Excel also has the T.TEST function.
    - i. The first two inputs are the arrays.
    - ii. The third input is the number of tails (one or two tails).
    - iii. The last input is the test type: 1 for paired, 2 for equal variances, and 3 for unequal variances.
  - b. The output is the lowest possible alpha you could have and still be statistically significant. This is called the p value.
  - c. For example, let's solve the previous question about the homework. Type "`=T.TEST(A2:A40,B2:B40,2,1)`" and press ENTER.
    - i. You should get about 0.035897.
    - ii. In other words, these values are different at a significance level of 96.4103% ( $1 - 0.035897$ ). That means it's significant at 95% but not at 99% (what we found).
    - iii. Note that you *never* explicitly state something is significant at 96.4103% because you'd sound weird. While we'll report alpha scores all the time, statisticians talk in terms of established thresholds: 90%, 95%, 99%, and 99.9%.
  - d. It's tempting to use the T.TEST command but keep in mind:
    - i. If you find out they're different, you won't know which one is larger. You'll have to make separate mean calculations; but the Data Analysis option tells you the averages as part of its output.
    - ii. The T.TEST command defaults to a hypothesized mean of zero; if you want to assume a different hypothesis, you can't use this command.
- VII. Choosing the Right Test
- a. Choosing the right test can be quite easy. It's usually obvious if the samples are dependent or independent samples and even more obvious if you have, or don't have, the population standard deviation.
  - b. The tricky part comes with independent samples and unknown variance. Should you assume the variances are the same or are different?
    - i. Assume the variances are the same when circumstances between the samples are similar *and* the samples are pooled from the same population or similar populations.
    - ii. Assume the variances are different when the samples underwent very different circumstances *or* the samples came from very different populations.
    - iii. When it's still not clear, assume different population variances.