Math 247: Comparing Two Population Means (Section 9.5)

Dependent Samples vs. Independent Samples

Suppose you're going to do your Senior Project on using meditation to reduce stress. How would you do this?

Using the same group of people	Using two different groups of people
Paired t-Test: Used for these **Dependent** samples: • Before-and-After studies • Twin studies • Any situation where there is a natural pairing of the data values.	**Two Sample t-Test:** Used for **Independent** samples This is the most common situation in studies where you have two independent groups you're comparing, like the banded and unbanded penguins in Dr. Saraux's banding experiment.

Dependent Samples: Paired t-Test

Meditation and Resting Heart Rate: Suppose a random sample of 10 people is drawn to study the effect of meditation on resting heart rate. The entire group is assigned meditation for 20 minutes each day over a 6-week period. The raw data is given below.

Subject	1	2	3	4	5	6	7	8	9	10
Before (bpm)	74	67	83	70	77	83	70	75	72	88
After (bpm)	70	67	72	65	78	75	60	70	64	86
D = Difference										

Find the difference in heart rate for each subject.

If meditation didn't actually have a beneficial effect, we'd expect the difference in heart rate to be _____, on average.

Looking at this data, does it appear that meditation actually helped reduce heart rate?

How can we tell if the reduction was *significant*?

		N	Mean	StDev
Here are the summary statistics for the two sets of data:	**Before**	10	75.90	6.81
	After	10	70.70	7.56
	Difference	10	5.20	4.08

Comment on the variability in the data for the Before and After groups vs. in the Difference data.

Before-and-After study design is useful because it _____

_____

Use StatCrunch to perform the computational step of a hypothesis test on the data **and** to find the Confidence Interval for the difference in means, using

(a) the Paired t-Test,
(b) the Two-Sample t-Test.

The data is on the wrightmath.info website. The results are below for reference.

Interpret the results of the hypothesis tests and the CI's.

(a) Paired T hypothesis test:

$\mu_D = \mu_1 - \mu_2$: Mean of the differences between Heartrate Before and Heartrate After

$H_0 : \mu_D = 0$ $H_A : \mu_D \neq 0$

Hypothesis test results:

Difference	Mean	Std. Err.	DF	T-Stat	P-value
Heartrate Before - Heartrate After	5.2	1.289272	9	4.033284	0.003

Interpret Hypothesis Test:

Paired T confidence interval:

$\mu_D = \mu_1 - \mu_2$: Mean of the difference between Heartrate Before and Heartrate After

95% confidence interval results:

Difference	Mean	Std. Err.	DF	L. Limit	U. Limit
Heartrate Before - Heartrate After	5.2	1.289272	9	2.2834642	8.1165358

Interpret CI:

(b) Two sample T hypothesis test:

μ_1 : Mean of Heartrate Before μ_2 : Mean of Heartrate After

$\mu_1 - \mu_2$: Difference between two means

$H_0 : \mu_1 - \mu_2 = 0$ $H_A : \mu_1 - \mu_2 \neq 0$ (without pooled variances)

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	5.2	3.216278	17.805913	1.6167756	0.1235

Interpret Hypothesis Test:

95% confidence interval results:

Difference	Sample Diff.	Std. Err.	DF	L. Limit	U. Limit
$\mu_1 - \mu_2$	5.2	3.216278	17.805913	-1.5624328	11.962433

Interpret CI:

Follow-up: Comment on the importance of choosing the appropriate hypothesis test when analyzing data.

Independent Samples. If our two samples are not associated (i.e., they're not paired or dependent in some other way), then we use a Two Sample t-Test for Means (similar to the Two-Proportion z-test we used for proportions, back in Section 8.4.)

In comparing two populations, remember that individual values will vary and there may be overlap between the populations, but ON AVERAGE there can still be a significant difference between the populations.

Consider heights of women and heights of men:

Are some individual women taller than some individual men?

On average, are men taller than women?

Two Sample t-test for Means (unpooled variance)

Hypotheses:

H_0 : $\mu_1 - \mu_2 = 0$ There is ZERO difference, on average, between the two populations with regard to the variable of interest.

H_a : $\mu_1 - \mu_2 \neq 0$ There IS a difference, on average, between the two populations with regard to the variable of interest.

(Note: The alternative hypothesis above is for a _____-Tailed Test. We could also use a One-tailed test.)

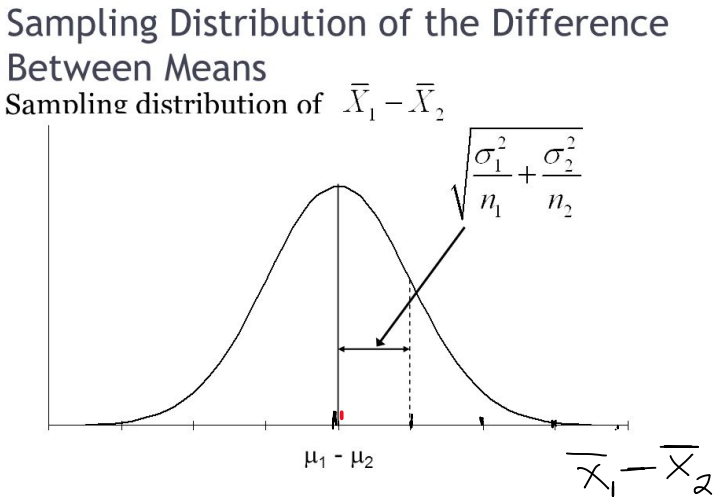
Conditions:

1. **Random Samples and Independence <u>within</u> samples.** We have two random samples from the <u>two</u> populations. Each observation is independent from all others.
2. **Independence <u>between</u> samples.** The two samples themselves are independent from each other. The individuals in one sample are in no way associated with the individuals in the other sample.
3. **Sample Size Considerations.** Either the sample size <u>in each sample</u> is 25 or more ("large" samples) OR if, n < 25, ("small" samples) the underlying populations are <u>each</u> normal.

(Central Limit Theorem) for the Difference of Means: Two Independent Samples If the conditions above are met, then the Sampling Distribution of the <u>differences</u> will follow the t-distribution*.

*Since we don't usually know the <u>population</u> standard deviation, the calculations involved in finding the Test Statistic will use the <u>sample</u> standard deviation. <u>StatCrunch will do these calculations for us</u>!

Standard Error (unpooled):



Test statistic:

By hand, we'd estimate degrees of freedom as df = least of $n_1 - 1$ and $n_2 - 1$

(This is just for reference…StatCrunch will do the compute step for us, and will use a more complex, and more accurate, way to find df.)

Example: Handedness and typing speed. Is handedness associated with how fast people type? Two random samples of left-handed students and of right-handed students who completed an online typing class are given a typing speed test (words per minute), and the results are compared. The 16 students in the right-handed sample had an average typing speed of 55.8 words per minute (wpm) with standard deviation of 5.7 wpm. The 9 students in the sample of left-handed students had an average speed of 59.3 wpm with a standard deviation of 4.3 wpm.

List and label the given values (Parking Lot!)

Looking at the sample data, does it appear that there's a difference in how fast people type, on average, based on whether they're left-handed or right-handed?

Is this difference *significant*?

Conduct a hypothesis test to determine if there is a significant difference in typing speed in right-handed and left-handed people. Use a .05 level of significance.

1.

2.

3. Use StatCrunch for this step.

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	-3.5	2.0211555	20.795052	-1.7316827	0.0981

4.