Notes 11.2 Two-sample *t* inference

In the last section, we did inference on one sample and one sample matched pairs when we didn't know the population mean OR population standard deviation. Technically, the matched pairs procedures were based on **dependent sampling**, which occurs when the two samples are not only the same person, but also can be related, like husband-wife, siblings, similar characteristics, etc. Now we are going to compare results from two separate **independent samples**, and we will perform inference on their difference. We will once again use the *t* procedures since we will not know the population standard deviations σ_1 or σ_2.

Examples of two-sample problems:

- Drug research studies involving one group on medication and another group on placebo.
- Studies to determine differences between gender groups (like mathematical aptitude, social insight, etc.)
- Experiments to determine which program customers prefer, like two different incentive programs offered by a bank to its customers to increase credit card usage.

The assumptions for two-sample problems are the same for one-sample problems except we must assume that the two samples are independent as well, one sample has no influence on the other. Believe it or not, the samples can even be of **different sizes.**

Assumptions

- Random samples from two **independent** populations
- Given normality, (by the Central Limit Theorem using **sum** of sample sizes, or asses using data)
- Population size is 10 times **larger sample size**
- Degrees of freedom = n – 1 of **smaller sample size** or use calculator
- Must define μ_1 and μ_2.
- Standard deviations unknown, use s_1 and s_2 to approximate.

Confidence Interval

$$\bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Test of Significance

$H_0: \mu_1 = \mu_2$	$H_a: \mu_1 < \mu_2, \quad H_a: \mu_1 \neq \mu_2, \quad H_a: \mu_1 > \mu_2$	
or	*or*	*Choose one.*
$H_0: \mu_1 - \mu_2 = 0$	$H_a: \mu_1 - \mu_2 < 0, \quad H_a: \mu_1 - \mu_2 \neq 0, \quad H_a: \mu_1 - \mu_2 > 0$	

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Example 1: Does increasing the amount of calcium in our diet reduce blood pressure? Examination of a large sample of people revealed a relationship between calcium intake and blood pressure. The relationship was strongest for black men. Such observational studies do not establish causation. Researchers therefore designed a randomized comparative experiment.

The subjects in part of the experiment were 21 healthy black men. A randomly chosen group of 10 of the men received a calcium supplement for 12 weeks. The control group of the other 11 men received a placebo pill that looked identical. The experiment was double-blind. The response variable is the decrease in systolic (heart contracted) blood pressure for a subject after 12 weeks, in millimeters of mercury. An increase in blood pressure appears as a negative response.

Take Group 1 to be the calcium group and Group 2 the placebo group. Here are the data for the 10 men in Group 1 (calcium).

7	**-4**	**18**	**17**	**-3**	**-5**	**1**	**10**	**11**	**-2**

and for the 11 men in Group 2 (placebo).

-1	**12**	**-1**	**-3**	**3**	**-5**	**5**	**2**	**-11**	**-1**	**-3**

Calculate the summary statistics for each group (and write them down) I recommend putting Group 1 in L1 and Group 2 in L2.

Group 1

```
1-Var Stats
 x̄=5
 Σx=50
 Σx²=938
 Sx=8.743251366
 σx=8.294576541
↓n=10
```

Group 2

```
1-Var Stats
 x̄=-.2727272727
 Σx=-3
 Σx²=349
 Sx=5.900693334
 σx=5.626090344
↓n=11
```

The calcium group shows an average drop in blood pressure, $\bar{x}_1 = 5.000$, while the placebo group had almost no change, $\bar{x}_2 = -0.273$. The question now is, ***is this outcome good evidence that calcium decreases blood pressure in the entire population of healthy black men more than the placebo does?***

1st step: Identify the population(s) of interest and the parameter(s) you want to draw conclusions about. State the hypothesis in both symbols AND words.

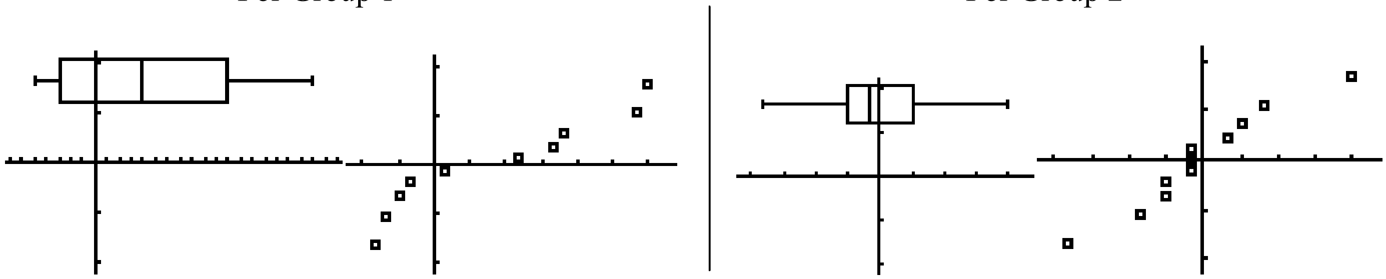
- We are interested in the mean decrease in blood pressure for the population of healthy black men taking calcium μ_1 for 12 weeks from that of healthy black men taking placebo μ_2.
    - $H_0 : \mu_1 = \mu_2$, the calcium shows no effect on lowering blood pressure.
    - $H_a : \mu_1 > \mu_2$, the calcium improves blood pressure by lowering blood systolic pressure.

2nd step: Choose the appropriate inference procedure and check the necessary conditions.

- Since we have two samples and don't know the population standard deviations, we'll use a **two-ample *t*-test**.
- Assumptions
    - The question says the experiment was a randomized comparative experiment, so will regard the calcium and placebo groups as two separated random samples.

- We'll assume the population of healthy black men is at least 110. We'll use 9 degrees of freedom.
- The sum of the two groups is only 21, so we can't assume that the distribution of differences of means is normal. We often won't be able to set up L3 for the differences for $\mu_1 - \mu_2$ and then check the box plot for outliers, histogram for distribution shape and/or the normal probability plot, since the sample sizes are different. Instead, check each distribution separately. If the two separated population distributions are both normal, then the distribution of $\bar{x}_1 - \bar{x}_2$ is also normal.

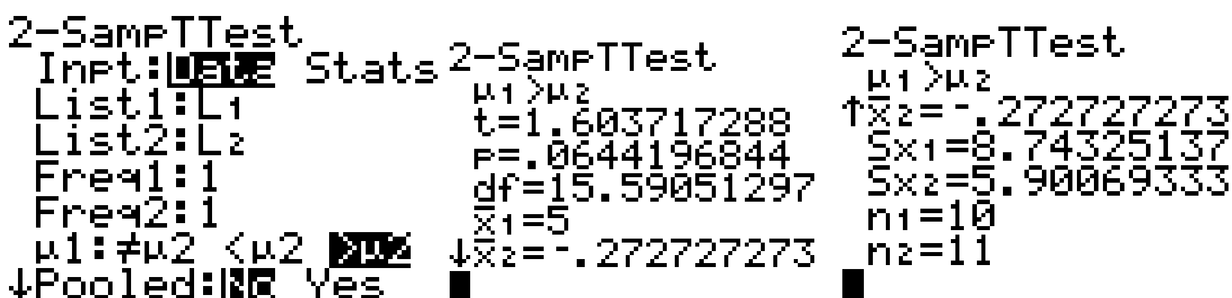
For Group 1 | For Group 2

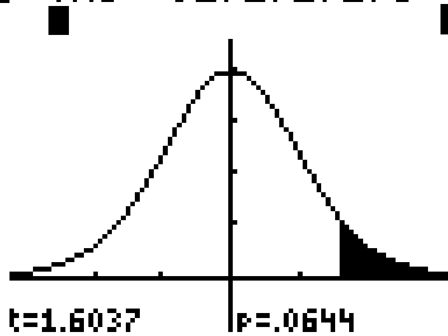


Neither group shows outliers and both normal probability plots are roughly linear so both are roughly normal.

3rd step: Carry out the inference test on the calculator and write down all your information.

- Note: if you choose your degrees of freedom to be one less than the smaller sample, choose the UNPOOLED test. This is called the **conservative method**.





- Degrees of freedom = 9, t-value $= 1.604$, p-value $= 0.064$

4th step: Interpret your results in the context of the problem.

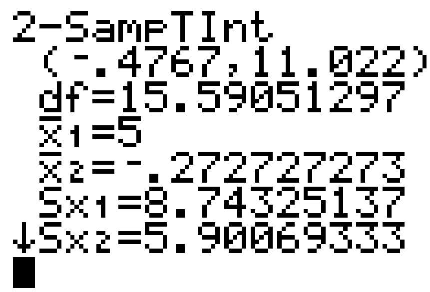
- The experiment found evidence that calcium reduces blood pressure. However, since our p-value is greater than an alpha-value of 0.05, the evidence falls a bit short of being statistically significant at the 5% level. We fail to reject the null at the 5% level.

Example 2: Using the data from Example 1, construct a 90% confidence interval for the true estimate of the difference in the mean decreases in blood pressure from the calcium and placebo populations.

Step 1: State the inference procedure and check the conditions

- We don't need to recheck our conditions since we already have. We will construct a **two-sample *t* interval.** The calculator uses a formula for calculating a more accurate degrees of freedom. We will let it do this, then report the value it gives us.

Step 2: Construct the interval using the calculator, then copy the information to your page.



- NOTE: the pooled degrees of freedom would be 20, but for the unpooled, your calculator is programmed to via a formula to calculate more accurate, fractional degrees of freedom. Don't concern yourself with the calculator's "magical" powers. Just go with it.
- A 90% confidence interval for the true difference in mean decrease in blood pressure is $(-0.477, 11.022)$ with 15.591 degrees of freedom. I am 90% confident that true difference in systolic pressure between is calcium and placebo is pills is between 11 points higher or half a point lower in those who take calcium. Because the interval crosses zero, we would not be able to reject the null hypothesis for a two-sided alternative at the 0.01 level.