AP Statistics – Lesson Notes – Chapters 24 and 25: Comparing Means

When we have 2 samples for numerical data (means) we have two analysis methods

- **2 Independent Sample Hypothesis Test / Confidence Interval**
- **Matched Pairs Hypothesis Test / Confidence Interval**

To see how this works, we will look at an example medical experiment, and first analyze using the 2 Independent Sample method.

The experiment: We are comparing a new cholesterol medicine for reducing cholesterol levels to a current medicine.

This experiment will be conducted using a group of patients with high cholesterol and there are many sources of variation in how the treatment will work on each patient:

- There are factors which are different for each patient (initial cholesterol level, age, weight) which might influence the effectiveness of the treatment.

- The treatment itself will have some effect on cholesterol level due to its actual effectiveness and this effectiveness may be different for each patient (it may be more effective for some patients and less effective for others).

- There exact amount that the treatment improves cholesterol will vary randomly from patient to patient, at least a little bit just because there is always random variation in any real-world scenario.

   We will include something to model each of these in our experiment

Here are the subjects and their individual data before the experiment:

Subject Number	Age	Weight	Initial Cholestrol
1	58	255	299
2	45	190	183
3	42	138	191
4	60	217	318
5	52	151	329
6	51	159	197
7	43	105	228
8	52	167	270
9	42	175	310
10	56	290	338
11	56	241	268
12	41	154	229
13	46	153	209
14	50	130	285
15	42	230	303
16	53	252	321
17	49	165	187
18	58	264	350
19	50	156	252
20	42	170	205

Now, we will conduct the experiment using a standard randomized design...

Group 1
10 patients → New treatment $\bar{x}_{new}$

Subjects
20
cholesterol
patients
RANDOM

compare mean
reduction in
cholesterol

Group 2
10 patients → Current treatment $\bar{x}_{current}$

To randomly assign patients to the groups, I ran randInt(1,20) repeatedly and put the number of the patient selected into group 1 (ignoring repeats) until I had 10 patients randomly selected for group 1...

Here are the subjects sorted randomly into their groups...

	Subject Number	Age	Weight	Initial Cholestrol
Group 1				
	8	52	167	270
	3	42	138	191
	9	42	175	310
	1	58	255	299
	20	42	170	205
	17	49	165	187
	10	56	290	338
	13	46	153	209
	11	56	241	268
	15	42	230	303
Group 2				
	2	45	190	183
	4	60	217	318
	5	52	151	329
	6	51	159	197
	7	43	105	228
	12	41	154	229
	14	50	130	285
	16	53	252	321
	18	58	264	350
	19	50	156	252

Let's compare the mean cholesterol reduction for the groups

Cholesterol Reduction

|	Subject Number	Age	Weight	Initial Cholestrol	How much better	Individual Effect	Random Effect randNorm(0,2)	Total Reduction
Group 1								
current treatment	8	52	167	270	0	59	1.85	61
	3	42	138	191	0	19	-2.38	17
	9	42	175	310	0	79	-0.12	79
	1	58	255	299	0	69	3.78	73
	20	42	170	205	0	29	-2.82	26
	17	49	165	187	0	10	-1.2	9
	10	56	290	338	0	99	0.55	100
	13	46	153	209	0	30	-0.11	30
	11	56	241	268	0	50	0.1	50
	15	42	230	303	0	70	-5.67	64
Group 2								
new treatment	2	45	190	183	10	9	-0.4	19
	4	60	217	318	10	80	3.3	93
	5	52	151	329	10	90	-1.3	99
	6	51	159	197	10	20	3.87	34
	7	43	105	228	10	39	5.05	54
	12	41	154	229	10	40	-0.24	50
	14	50	130	285	10	60	1.65	72
	16	53	252	321	10	89	-2.04	97
	18	58	264	350	10	100	0.06	110
	19	50	156	252	10	49	-0.12	59

+   +   = result

total cholesterol reduction is a combination of how much better than current the actual drug really is, plus the patient's individual response to the drug, plus a small amount of random variation

Enter the current treatment data in L1 and the new treatment data into L2 and use 1-Var stats twice to find the mean cholesterol reduction between the drugs.

Do you think the new drug is better?

Comparing the results

From 1-Var Stats:

$$\overline{X} = 50.9$$
$$s = 29.0$$

current treatment — total cholesterol reduction

$$\overline{X} = 68.7$$
$$s = 30.5$$

new treatment — total cholesterol reduction
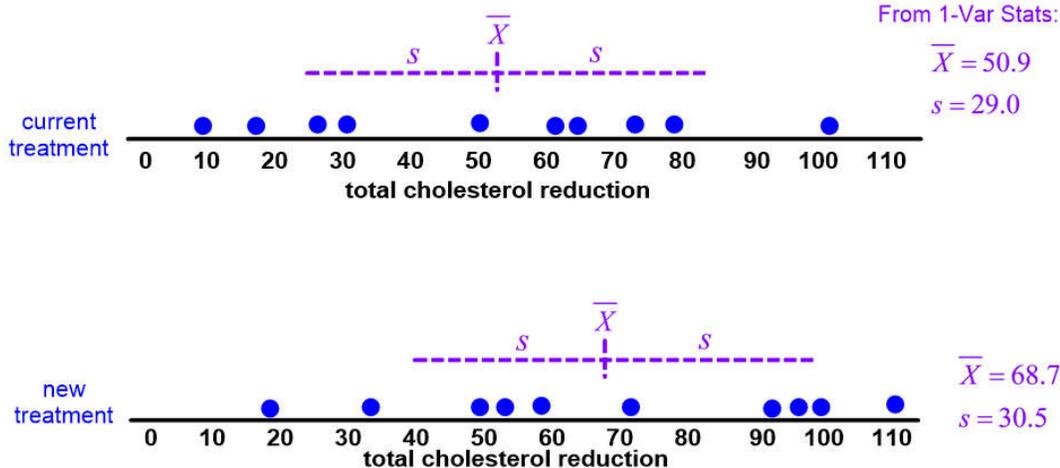
Is the new drug better?

The question is, is this difference in means statistically significant with this much natural variation?

Hard to say. The new drug mean is 17.8 above the current drug mean, but this is around half of the standard deviation. Probably not statistically significant.

The other method: Matched-pairs

The new drug did have a higher mean reduction in cholesterol than the old drug, but this difference was swamped out by the patient to patient variation. If we could find a way to reduce the amount of patient to patient variation, that might help us see the true effect of the new drug.

To remove sources of variation inherent in the subjects we can use a **matched-pair experiment design**. In this case, we find some aspect of the subjects themselves which might have an effect on how well the cholesterol drug works, and we group the subjects into pairs where we match as closely as possible on this aspect of the subjects.

Here, we know 3 things about the subjects:

- Their age
- Their weight
- Their initial cholesterol level

Since this drug is mainly about affecting cholesterol, this suggests maybe initial cholesterol level might be an important thing to 'control' - to try to remove the effect of the individual patient to patient variations in initial cholesterol level from the analysis.

The way we do this is by reordering the patient list by this factor, initial cholesterol level:

Subject Number	Age	Weight	Initial Cholestrol
2	45	190	183
17	49	165	187
3	42	138	191
6	51	159	197
20	42	170	205
13	46	153	209
7	43	105	228
12	41	154	229
19	50	156	252
11	56	241	268
8	52	167	270
14	50	130	285
1	58	255	299
15	42	230	303
9	42	175	310
4	60	217	318

Then we **pair up** the subjects into pairs so that the subjects **match** on this aspect (here initial cholesterol level).

Pair	Subject Number	Age	Weight	Initial Cholestrol
1	2	45	190	183
1	17	49	165	187
2	3	42	138	191
2	6	51	159	197
3	20	42	170	205
3	13	46	153	209
4	7	43	105	228
4	12	41	154	229
5	19	50	156	252
5	11	56	241	268
6	8	52	167	270
6	14	50	130	285
7	1	58	255	299
7	15	42	230	303
8	9	42	175	310
8	4	60	217	318
9	16	53	252	321
9	5	52	151	329
10	10	56	290	338
10	18	58	264	350

The other method: Matched-pairs

The way randomness is used is to randomly assign one of the subjects in each pair to the new treatment (and the other subject in that pair gets the current treatment)...

		Pair	Subject Number	Age	Weight
C	current	1	2	45	190
	new	1	17	49	165
C	new	2	3	42	138
	current	2	6	51	159
C	new	3	20	42	170
	current	3	13	46	153
C	new	4	7	43	105
	current	4	12	41	154
C	current	5	19	50	156
	new	5	11	56	241
C	new	6	8	52	167
	current	6	14	50	130
C	current	7	1	58	255
	new	7	15	42	230
C	new	8	9	42	175
	current	8	4	60	217
C	new	9	16	53	252
	current	9	5	52	151
C	current	10	10	56	290
	new	10	18	58	264

Here I used a random number generator - if an even number occurred, the second subject in the pair got the new treatment

Repeat the test by giving the drug to get the total reductions...

	Pair	Subject Number	Initial Cholestrol	How much better	Individual Effect	Random Effect randNorm(0,2)	Total Reduction	Difference (new-current) in each pair
current	1	2	183	0	9	-0.4	9	
new	1	17	187	10	10	-1.2	19	10
new	2	3	191	10	19	-2.38	27	
current	2	6	197	0	20	3.87	24	3
new	3	20	205	10	29	-2.82	36	
current	3	13	209	0	30	-0.11	30	6
new	4	7	228	10	39	5.05	54	
current	4	12	229	0	40	-0.24	40	14
current	5	19	252	0	49	-0.12	49	
new	5	11	268	10	50	0.1	60	11
new	6	8	270	10	59	1.85	71	
current	6	14	285	0	60	1.65	62	9
current	7	1	299	0	69	3.78	73	
new	7	15	303	10	70	-5.67	74	1
new	8	9	310	10	79	-0.12	89	
current	8	4	318	0	80	3.3	83	6
new	9	16	321	10	89	-2.04	97	
current	9	5	329	0	90	-1.3	89	8
current	10	10	338	0	99	0.55	100	
new	10	18	350	10	100	0.06	110	10

...but now within each pair we compute the **difference** in cholesterol reducing by taking the value for 'new' and subtracting the value for 'current'

Find the mean of these differences - do you think that the new drug is better now?

For a matched-pairs test the result is more likely to be significant...

differences in total cholesterol reduction (new - current)

From 1-Var Stats:

$\overline{X} = 7.8$

$s = 3.9$
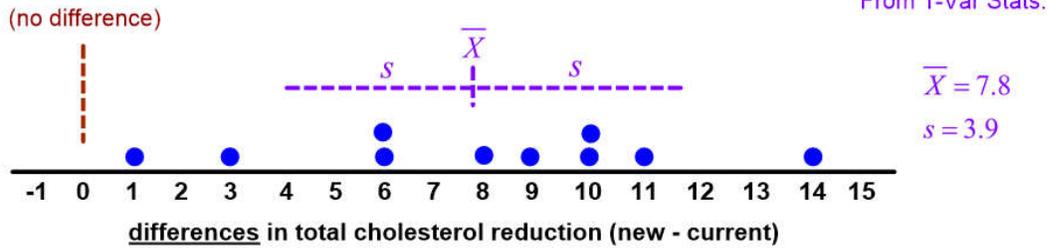
This data shows that the new drug is likely better than the current drug. In every pair of patients, the difference (new - current) was positive (new drug better).
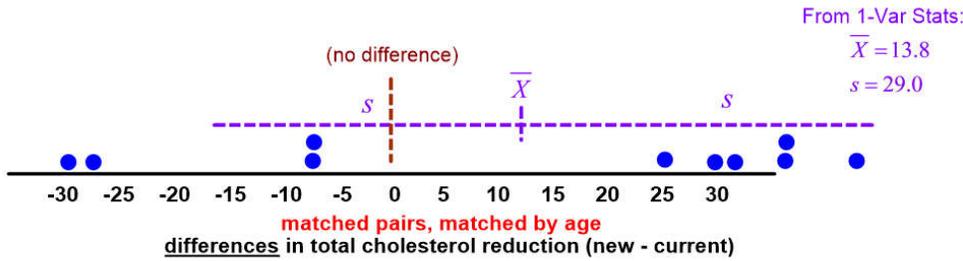
The reason this works is that by pairing up subjects based upon initial cholesterol level, we effectively removed that factor from the experiment, which dramatically lowered the standard deviation (7.8 for the matched-pairs differences compared to around 30 for the standard unmatched experiment).

...but only if you match on something that affects the outcome

Let's say we thought age was the aspect of the subjects that was most likely to affect cholesterol reduction instead of initial cholesterol level. If we instead matched the subjects into pairs by their age...

	Pair	Subject Number	Age	How much better	Individual Effect	Random Effect randNorm(0,2)	Total Reduction	Difference (new-current) in each pair
new	1	12	41	10	40	-0.24	50	
current	1	3	42	0	19	-2.38	17	33
current	2	15	42	0	70	-5.67	64	
new	2	9	42	10	79	-0.12	89	25
new	3	20	42	10	29	-2.82	36	
current	3	7	43	0	39	5.05	44	-8
current	4	2	45	0	9	-0.4	9	
new	4	13	46	10	30	-0.11	40	31
new	5	17	49	10	10	-1.2	19	
current	5	19	50	0	49	-0.12	49	-30
current	6	14	50	0	60	1.65	62	
new	6	6	51	10	20	3.87	34	-28
current	7	8	52	0	59	1.85	61	
new	7	5	52	10	90	-1.3	99	38
new	8	16	53	10	89	-2.04	97	
current	8	11	56	0	50	0.1	50	47
new	9	10	56	10	99	0.55	110	
current	9	1	58	0	69	3.78	73	37
current	10	18	58	0	100	0.06	100	
new	10	4	60	10	80	3.3	93	-7

...then performed a matched-pair experiment and looked at the differences...

From 1-Var Stats:
$\overline{X} = 13.8$
$s = 29.0$

(no difference)

s $\overline{X}$ s

matched pairs, matched by age
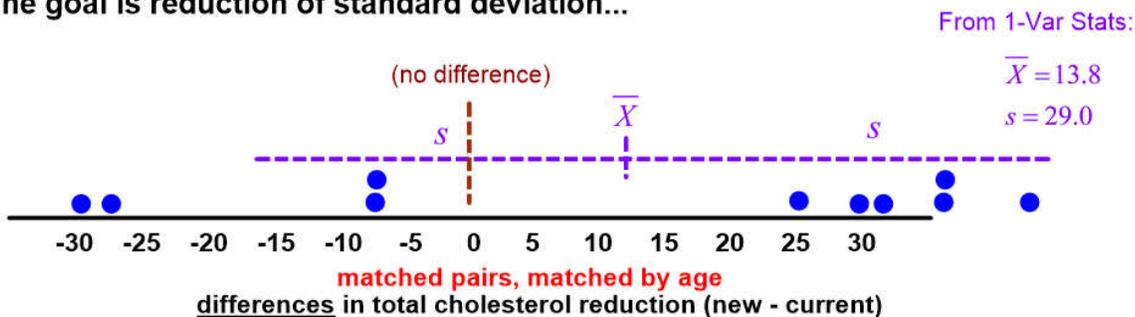differences in total cholesterol reduction (new - current)

The results are all over the place. There are large variations from pair to pair, with close to half the pairs having the new drug perform worse than the current drug.
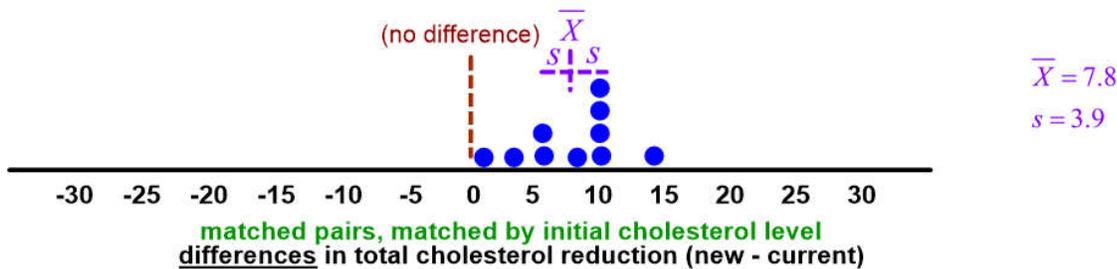
Matching by age didn't work here because age isn't a factor which affects the cholesterol reduction much.

We still have the more important variation in initial cholesterol level still varying from patient to patient, and unaccounted for in the analysis.

The goal is reduction of standard deviation...

From 1-Var Stats:
$\overline{X} = 13.8$
$s = 29.0$

(no difference)

s $\overline{X}$ s

matched pairs, matched by age
differences in total cholesterol reduction (new - current)

Compare matching by age (above) to matching by initial cholesterol level (below). Only matching by initial cholesterol level reduced the standard deviation.

(no difference) $\overline{X}$

s | s

$\overline{X} = 7.8$
$s = 3.9$

matched pairs, matched by initial cholesterol level
differences in total cholesterol reduction (new - current)

There are two methods of analyzing two sample mean data

- **2 Independent Sample Hypothesis Test / Confidence Interval**
- **Matched Pairs Hypothesis Test / Confidence Interval**

But we can't just arbitrarily pick one...we <u>must</u> use the method that matches the type of data we are provided in a problem.

Next, we'll look at the steps to perform a hypothesis test or a confidence interval calculation, and the conditions were these are different, depending up what type of data we have: if the data in the two samples are matched in some way, this must be stated as a condition, and if the two samples are independent, this must be stated as a condition. These conditions affect how the standard error is calculated.

1 Sample mean	2 Independent Sample means "Difference of Means"	Matched-Pair "Mean of Differences"
Historically, US adults have a mean cholesterol level of 190. The cholesterol levels of a random sample of US adults are taken. *Does this sample give us reason to believe that cholesterol levels are now higher than the historic value?* **T-Test**	A drug company has developed a new drug for reducing cholesterol levels. To test this, the company gives the drug to subjects and a placebo to subjects each randomly selected and compares the mean cholesterol reduction between the groups. **2-Sample T-Test**	A drug company has developed a new drug for reducing cholesterol levels. To test this, the company matches pairs of subjects by initial cholesterol level and randomly assigns the drug to one subject and placebo to the other in each pair, then computes the difference (drug - placebo) to see if there is a positive difference **T-Test on the differences**

The data...

Cholesterol level
209
200
208
192
194
212
188
206
216
182

Cholesterol reduction

drug	placebo
21	68
25	79
101	21
49	81
70	100
60	38
70	31
98	68
21	50
107	30

Difference (D) in cholestrol reduction (drug - placebo)

pair 1: 17 - 8 = 9
pair 2: 31 - 22 = 9
pair 3: 39 - 29 = 10
pair 4: 50 - 43 = 7
pair 5: 57 - 48 = 9
pair 6: 68 - 61 = 7
pair 7: 78 - 72 = 6
pair 8: 92 - 75 = 17
pair 9: 99 - 92 = 7
pair 10: 108 - 99 = 9

The statistics...

$$\overline{X} = 200.7$$
$$s = 11.27$$

$$\overline{X}_{drug} = 62.2 \qquad \overline{X}_{placebo} = 56.6$$
$$s_{drug} = 33.12 \qquad s_{placebo} = 26.37$$
$$\overline{X}_{drug} - \overline{X}_{placebo} = 62.2 - 56.6 = 5.6$$

$$\overline{X}_{difference,\,D} = 9.0$$
$$s_{difference,\,D} = 3.09$$

Hypotheses...

$$H_0 : \mu = 190$$
$$H_A : \mu > 190$$

$$H_0 : \mu_{drug} = \mu_{placebo}$$
$$H_0 : \mu_{drug} > \mu_{placebo}$$
$$\left(\begin{matrix} H_0 : \mu_{drug} - \mu_{placebo} = 0 \\ H_0 : \mu_{drug} - \mu_{placebo} > 0 \end{matrix} \right)$$

define D = the difference in individual cholesterol levels, D = drug - placebo

$$H_0 : \mu_D = 0$$
$$H_0 : \mu_D > 0$$

Conditions...

- **SRS or representative sample**

- **n < 10% of population**

- **sample Nearly Normal**

- **SRS or representative sample**

- **n$_1$ < 10% of population$_1$**
  n$_2$ < 10% of population$_2$

- **both samples Nearly Normal (must check both groups separately)**

- **Groups vary independently**

- **SRS or representative sample**

- **n < 10% of population**

- **differences Nearly Normal (must check only differences)**

- **Pairs are matched on some aspect of the subjects**

1 Sample mean	**2 Independent Sample means**	**Matched-Pair "Mean of Differences"**

Degrees of freedom...

$$df = n - 1$$
$$= 10 - 1$$
$$= 9$$

$$df = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2} \right)^2}{\dfrac{1}{n_1 - 1}\left(\dfrac{s_1^2}{n_1} \right)^2 + \dfrac{1}{n_2 - 1}\left(\dfrac{s_2^2}{n_2} \right)^2}$$

$$df = n - 1$$
$$= 10 - 1$$
$$= 9$$

$$= \frac{\left(\dfrac{33.12^2}{10} + \dfrac{26.37^2}{10} \right)^2}{\dfrac{1}{10-1}\left(\dfrac{33.12^2}{10} \right)^2 + \dfrac{1}{10-1}\left(\dfrac{26.37^2}{10} \right)^2}$$

$$= 17.14$$

Unfortunately, the way degrees of freedom is determined for 2 independent samples is quite complicated, and requires using the Welch-Satterthwaite equation. Note that this often produces non-integer *df* values.

But in practice, we will always get our *df* for this case by using tests in our calculator. (If you need to computer a confidence interval or hypothesis inference test by hand, you can perform a non-pooled 2-SampTTest and the calculator results will display this *df* value).

The Standard Error...

$$SE_{\bar{X}} = \frac{s}{\sqrt{n}}$$
$$= \frac{11.27}{\sqrt{10}}$$
$$= 3.564$$

$$SE_{\bar{X}_{drug} - \bar{X}_{placebo}} = \sqrt{\frac{s_{drug}^2}{n_{drug}} + \frac{s_{placebo}^2}{n_{placebo}}}$$

$$= \sqrt{\frac{33.12^2}{10} + \frac{26.37^2}{10}}$$

$$= 13.388$$

$$SE_D = \frac{s_D}{\sqrt{n_D}}$$
$$= \frac{3.09}{\sqrt{10}}$$
$$= 0.977$$

The standard error for 2 independent samples from the combining variances.

Note that the standard error for matched pairs is smaller than the standard error for independent samples.

This is because the natural variation from subject to subject within the sample is removed by the matching process. As a result, a matched-pairs test, if it is appropriate to use, is usually a much stronger test (higher power, easier to see small effect sizes as significant).

1 Sample mean	**2 Independent Sample means**	**Matched-Pair "Mean of Differences"**

The Test Statistic...

$$t = \frac{\overline{X} - \mu_0}{SE_{\overline{X}}}$$
$$= \frac{200.7 - 190}{3.564}$$
$$= 3.002$$

$$t = \frac{\left(\overline{X}_{drug} - \overline{X}_{placebo}\right) - 0^*}{SE_{\overline{X}_{drug} - \overline{X}_{placebo}}}$$
$$= \frac{(62.2 - 56.6) - 0}{13.388}$$
$$= 0.418$$

$$t = \frac{\overline{X}_D - 0^*}{SE_{\overline{X}_D}}$$
$$= \frac{9.0 - 0}{0.977}$$
$$= 9.212$$

(* actually the null hypothesis difference, but this is usually zero)

The p-value...

(df = 9 t-distribution)

0   **3.002**
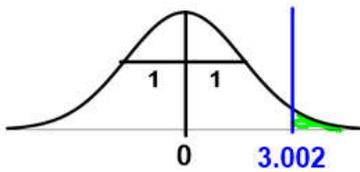
$$p-value = tcdf\underset{\text{lower upper df}}{(3.002, 999, 9)}$$
$$= 0.0075$$

(df = 17.14 t-distribution)

0   **0.418**
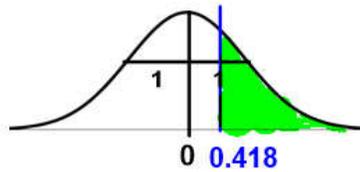
$$p-value = tcdf\underset{\text{lower upper df}}{(0.418, 999, 17.14)}$$
$$= 0.3406$$

(df = 9 t-distribution)

0   **9.212**
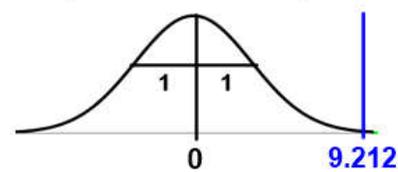
$$p-value = tcdf\underset{\text{lower upper df}}{(9.212, 999, 9)}$$
$$= 3.6 * 10^{-6} \approx 0$$

The conclusion...

With $\alpha = .05$, $p-value = 0.0075$ is low, so we <u>reject</u> H_0. We <u>do</u> have sufficient statistical evidence to conclude that the mean cholesterol level has increased.

With $\alpha = .05$, $p-value = 0.3406$ is high, so we <u>fail to reject</u> H_0. We <u>do not</u> have sufficient statistical evidence to conclude that the mean cholesterol reduction is higher for the drug than for placebo.

With $\alpha = .05$, $p-value \approx 0$ is low, so we <u>reject</u> H_0. We <u>do</u> have sufficient statistical evidence to conclude that cholesterol reduction is higher for the drug than for placebo, on average.

P-value with a calculator...

Perform a T-Test in Ti-84 using: data in L1
$$\mu_0 = 190$$
$$\mu > \mu_0$$
Result: t = 3.001
$$p = .0075$$
$$df = 9$$
(*df* not provided, but include it)

Perform a 2-SampTTest in Ti-84 with drug data in L1 and placebo data in L2 using:
$$\mu_1 > \mu_2$$
no pooling
Result: t = .4183
$$p = .3405$$
$$df = 17.14$$

Perform a T-Test in Ti-84 on the differences in L3 using:
$$\mu_0 = 0$$
$$\mu > \mu_0$$
Result: t = 9.207
$$p = 3.5 * 10^{-6}$$
$$df = 9$$
(*df* not provided, but include it)
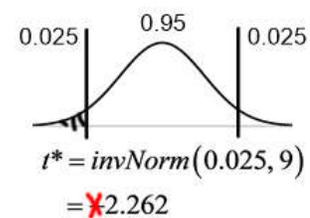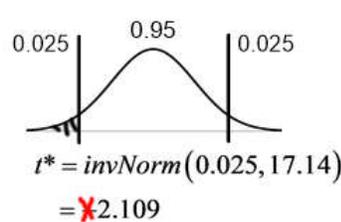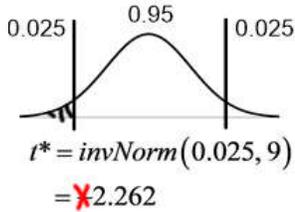
Confidence Intervals

| 1 Sample mean 95% Confidence Interval | 2 Independent Sample means 95% Confidence Interval | Matched-Pair "Mean of Differences" 95% Confidence Interval |

Check conditions: Same as for hypothesis tests, confidence intervals also require conditions.

Critical value (t*) for selected confidence level...

0.025 | 0.95 | 0.025	0.025 | 0.95 | 0.025	0.025 | 0.95 | 0.025
$t^* = invNorm(0.025, 9)$ $= \cancel{\times}2.262$	$t^* = invNorm(0.025, 17.14)$ $= \cancel{\times}2.109$	$t^* = invNorm(0.025, 9)$ $= \cancel{\times}2.262$

The formula...

$$CI = \overline{X} \pm (t^*)SE_{\overline{X}}$$
$$= (200.7) \pm (2.262)(3.564)$$
$$= 200.7 \pm 8.062$$
$$= (192.6, 208.8)$$

$$CI = \left(\overline{X}_{drug} - \overline{X}_{placebo}\right) \pm (t^*)SE_{\overline{X}_{drug} - \overline{X}_{placebo}}$$
$$= (62.2 - 56.6) \pm (2.109)(13.388)$$
$$= 5.6 \pm 28.235$$
$$= (-22.6, 33.8)$$

$$CI = \overline{X}_D \pm (t^*)SE_D$$
$$= (9.0) \pm (2.262)(0.977)$$
$$= 9.0 \pm 2.210$$
$$= (6.8, 11.2)$$

The conclusion...

We are 95% confident that the mean cholesterol level for all adults is between 192.6 and 208.8.

We are 95% confident that the mean cholesterol reduction for all adults taking the drug is between 22.6 points below and 33.8 above the mean for those taking a placebo.

We are 95% confident that, for all adults, the drug will reduce cholesterol by between 6.8 and 11.2 points more than placebo, on average.

Confidence intervals with a calculator...

```
TInterval
Inpt:Data Stats
List:L1
Freq:1
C-Level:.95
Calculate
```
```
TInterval
(192.63,208.77)
x̄=200.7
Sx=11.2748491
n=10
```

```
2-SampTInt
Inpt:Data Stats
List1:L1
List2:L2
Freq1:1
Freq2:1
C-Level:.95
Pooled:No Yes
```
```
2-SampTInt
(-22.63,33.831)
df=17.14035393
x̄1=62.2
x̄2=56.6
Sx1=33.1219967
Sx2=26.3742298
```

```
TInterval
Inpt:Data Stats
List:L3
Freq:1
C-Level:.95
Calculate
```
```
TInterval
(6.7887,11.211)
x̄=9
Sx=3.091206165
n=10
```

Perform a T-Interval in Ti-84 using: data in L1

$$C - Level : .95$$

Result: (192.63, 208.77)
df = 9

Perform a 2-SampTInt in Ti-84 with drug data in L1 and placebo data in L2 using:

$$C - Level : .95$$

no pooling

Result: (-22.63, 33.831)
df = 17.14

Perform a T-Interval in Ti-84 on the differences in L3 using:

$$C - Level : .95$$

Result: (6.7887, 11.211)
df = 9

For 2 independent sample case, which data you enter into which list doesn't matter...

...but the calculator always calculates List1 - List2...

(-22.63, 33.831)

(-33.831, 22.63)

...so you are allowed to swap these values (and change their signs) if that makes it easier to write up the conclusion.

Pooling - Just don't do it for means (almost)

When we compared the difference of *proportions*, we said there were times you might consider *pooling* the data for purposes of computing the Standard Error. The reason this was acceptable was that, for proportions:

Our $H_0 : p_1 - p_2 = 0$ assumes that the proportions are equal, and the SE therefore depends upon something which we assume is equal:

$$SE_{diff,pooled} = \sqrt{\frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_1} + \frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_2}}$$

But for the difference of *means*, our $H_0 : \mu_1 - \mu_2 = 0$ does not allow us to say anything about SE:

$$SE_{\overline{X_1}-\overline{X_2}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Assume the means of the two distributions are equal does not mean that the standard deviations are necessarily equal.

Our formula for Standard Error for difference of means... $SE_{\overline{X_1}-\overline{X_2}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

...could be changed to use a *pooled* standard deviation:

$$SE_{\overline{X_1}-\overline{X_2},pooled} = \sqrt{\frac{s_{pooled}^2}{n_1} + \frac{s_{pooled}^2}{n_2}}$$ where: $s_{pooled} = \sqrt{\frac{(n-1)s_1^2 + (n-1)s_2^2}{(n_1-1)+(n_2-1)}}$

But for this to be reasonable, we have to **know** that the two independent sample standard deviations are approximately equal: $s_1 \approx s_2$

For sampling distributions of sample means $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ a null hypothesis assuming $\mu_1 = \mu_2$ does nothing that helps us assert this (unlike in the proportions case).

So, in practice, we can almost never justify using pooled methods for means.
(And if we happen to have $s_1=s_2$ the more general SE formula will give approximately the same result anyway.)

A faster way to compare means - The Tukey Test

There is one additional method we can use to quickly compare the means of two samples if we have all of the sample data. It was discovered, and named for, John Tukey (who also invented stem-and-leaf diagrams, boxplots, coined the word 'bit', and is responsible for the mathematics behind Fast Fourier Transforms (pitch detection)).

The Tukey Test:
1) Check condition: The overall max and min must be in different data sets.
2) Determine the max and min value in each data set.
3) Count how many values in 'high' group are higher than the max of the 'low' group (a tie=1/2).
4) Count how many values in the 'low' group are lower than the min of the 'high' group (a tie=1/2).
5) Add these values and compare results:

$$sum \geq 7, \quad p-value < .05$$

$$sum \geq 10, \quad p-value < .01$$

$$sum \geq 13, \quad p-value < .001$$

The Tukey test is not as precise as more formal methods and not universally accepted, but is a good way to quickly approximate p-value without a calculator.

Data set 1	Data set 2
275	260
300	250
260	175
300	130
255	200
275	225
290	240
300	

Data set 1   255   -   300   *6.5*

Data set 2   130   -   260   *6*

Here, sum=6+6.5=12.5 which means p-value is less than .01 ...we would reject H_0.