§9.2 & 9.3 Comparing Means

9.2 Large Independent Samples 2SampTtest
2SampTInterval

Two groups Men & Women $H_0: \mu_1 = \mu_2$
Placebo Treatment
Class1 Class2

9.3 Matched Pairs TTest on $L_3 = L_1 - L_2 = d$
$H_0: \mu_d = 0$

One Group Before & After
Test Twice Test1 & Test2

§9.2 Test Means from Large Independent Samples.

Sample Data

$\bar{X}_1 = 6.7$ $\bar{x}_2 = 6.3$

$n_1 = 123$ $n_2 = 114$

$s_1 = .21$ $s_2 = .23$

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$ or $\neq$ or $<$

CV: $t^* = \text{invt}(1-\alpha, df)$ Simple df = Smaller of $n_1 - 1$ & $n_2 - 1$

$t^* = \text{invt}(.95, 113) = 1.658$

PE: $\bar{x}_1 - \bar{x}_2 = 6.7 - 6.3$ hrs $= .4$

TS: $t = \dfrac{\bar{x}_1 - \bar{x}_2 - \overbrace{(\mu_1 - \mu_2)}^{0}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ ← Use 2SampTtest

Pooled $\boxed{NO}$

P-value = Tcdf $(-9999, TS, \underset{228}{113}) = 1.06 \times 10^{-32} \approx 0$

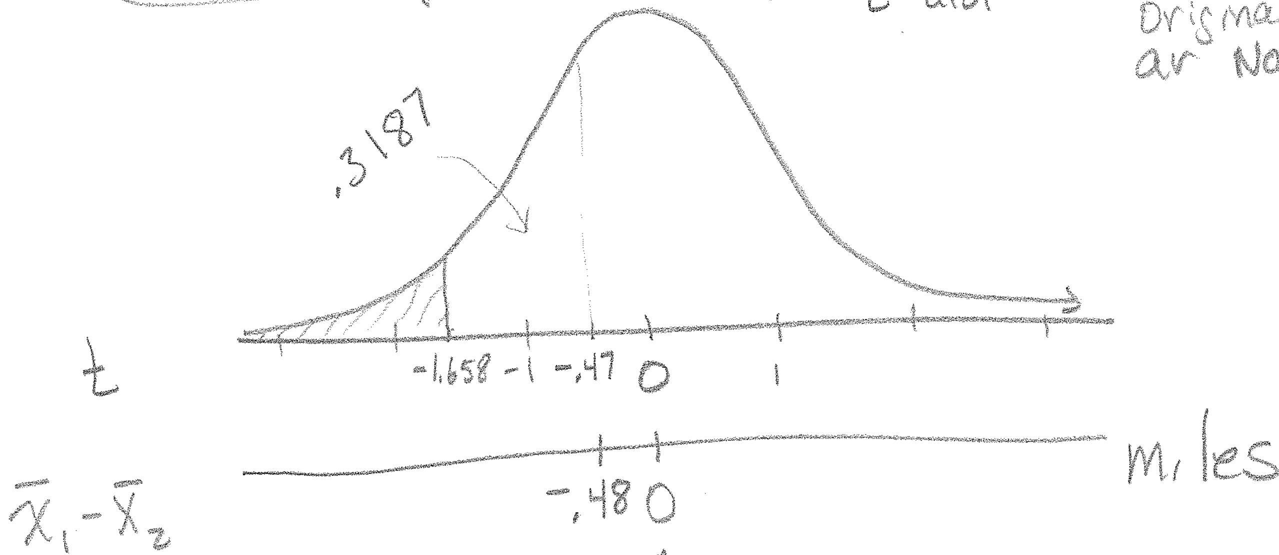
<u>Conclusion</u> Reject H_0

TISE We <u>Reject</u> that the mean # of hr slept by students is equal to the mean for Non students

So we can Support H_1

TISE to <u>Support</u> that the mean # of hrs slept by Students is greater than the # hrs by Non students

Draw Graph Sampling Distribution is t-dist $n > 30$ or Original populations are Normal



$H_0: \mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0$

Women	Men	
$\bar{x}_1 = 8.56$ mi	$\bar{x}_2 = 9.04$	$H_0: \mu_1 = \mu_2$
$s_1 = 9.04$	$s_2 = 6.35$	$H_1: \mu_1 < \mu_2$
$n_1 = 118$	$n_2 = 118$	$\alpha = .05$

CV: $t = \text{invt}(\alpha, df)$

$t = \text{invt}(.05, 117)$

2SampTtest

$\boxed{\text{TS: } t = -.47}$ ← Not significant → $\boxed{\text{CV: } t = -1.658}$

p-value = .3187 > .05

df = 209

PE: $\bar{x}_1 - \bar{x}_2 = 8.56 - 9.04 = -.48$ miles

We fail to Reject H_0

TINSE to Reject that mean distance to school is the same for men & women.

TINSE to Support that mean dist to school for women is less than men.

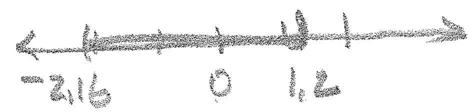
Confidence Intervals

Use 2SampT Interval

$(-2.16, 1.20)$

The <u>difference</u> between the mean dist. to School for Men & women is between

-2.16 and 1.20 miles.



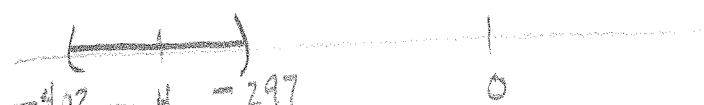
Since Zero is in this interval

So $\mu_1 - \mu_2 = 0$ is possible

So we can't say there is a Significant Difference.

<u>Hrs of sleep</u>

$\mu_N - \mu_S \neq 0$



Meaning of P-value Sentence

PE

If H_0 is True then there is a p-value chance of Sample or more extreme.

If the mean distance to School is the Same for men and women then there is a 31.87% chance of seeing women with mean difference distance of .48 miles less than Men or more.

(A difference of .48 miles or More between men & women)

§9.2 and 9.3 Comparing Means

9.2 Means are Independent

Large Groups of Men & Women

Placebo & Treatment

Same test Class 1 & Class 2

9.3 Means are Matched Pairs

Same class ~~Test~~ Test 1 & Test 2

Before & After

Same group Tested Twice

9.2 Independent

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$ or $<$ or $>$

df = Smaller of $n_1 - 1$ and $n_2 - 1$

(Calculator uses complicated formula - p.475 use if you want)

P-value & TS given by 2SampTTest

CV: $t^* = \text{invT}(\text{area}, df)$

H_1: $<$, $>$, $\neq$

Area $= \alpha$, $1-\alpha$, $1-\alpha/2$

Ex Miles To School

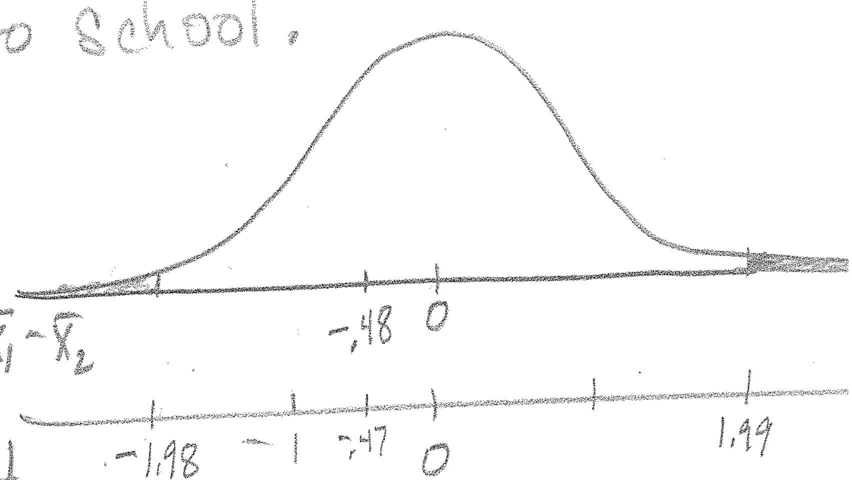
Summary Statistics

Women	Men
$n_1 = 118$	$n_2 = 118$
$\bar{x}_1 = 8.563$	$\bar{x}_2 = 9.043$
$s_1 = 9.043$	$s_2 = 6.346$

Question At the $\alpha = .05$ level of significance Test the claim that Men and Women have the same mean distance to school.

$H_0: \mu_1 = \mu_2$ $\mu_1 - \mu_2 = 0$

$H_1: \mu_1 \neq \mu_2$



CV: $t^* = \text{invT}(1 - \alpha/2, df)$

$t^* = \text{invT}(1 - .05/2, 117)$

$t^* = 1.980$

PE: $\bar{x}_1 - \bar{x}_2 = 8.563 - 9.043$

$\boxed{\bar{x}_1 - \bar{x}_2 = -.48}$

2SampTTest

TS: $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \boxed{-.472 = t\text{:TS}}$

$\boxed{\text{p-value} = .637 > .05 = \alpha}$

$df = 209.7$

fail to reject H_0

Conclusion fail to reject $H_0: \mu_1 = \mu_2$

→ We can not reject that Men & Women live the same mean distance from school

fail to Support $H_1: \mu_1 \neq \mu_2$

We can not Support the Mean distance to School for Women is Not equal to the Mean distance for men.

Do Not Write

Men & women live the same Mean distance

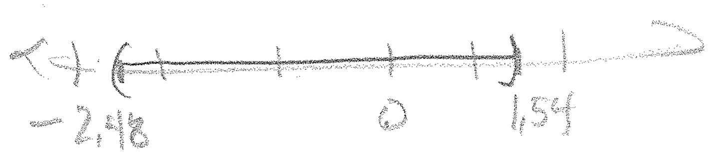
Confidence Interval

2-SampT Interval

$CL = 1 - \alpha$ Two Tailed

$CL = 1 - 2\alpha$ One Tailed Test

$(-2.485, 1.525)$



Zero is in CI $\Rightarrow$ fail to show a significant difference

$-2.5 < \mu_1 - \mu_2 < 1.5$

the difference in pop. Means might be Zero

Class Project Part 2 3 Sara Jones

Means Comparison Distance to School Math 15 Jones Your name

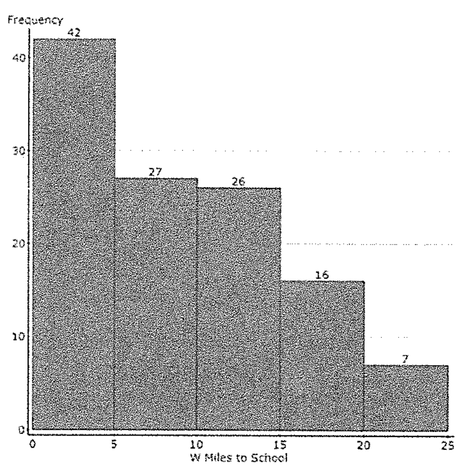
Summary statistics:

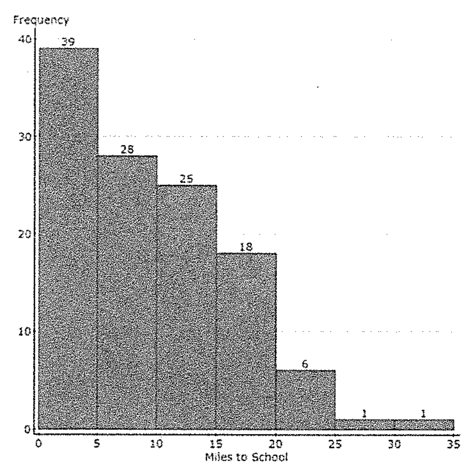
Column	n	Mean	Variance	Std. dev.	Median
Women Miles to School	118	8.563	35.0732	5.922262	7.5
Men Miles to School	118	9.043	40.2748	6.346242	9

$n_1 = 118$ $\bar{x}_1 = 8.56$ $s_1 = 5.92$

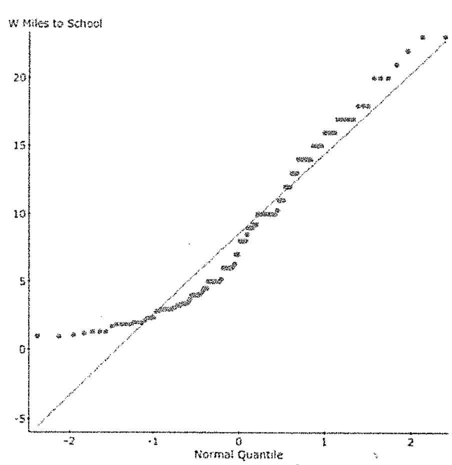
Requirements

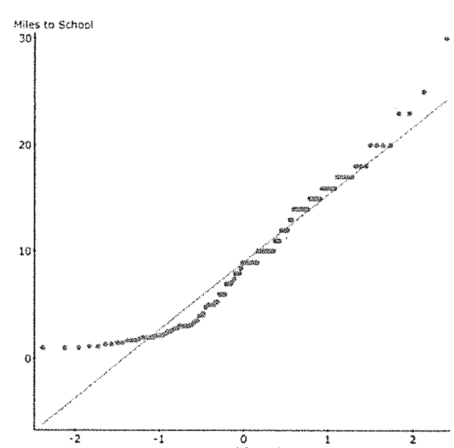
The requirements to make confidence intervals and do hypothesis test with two means are that the sample is a SRS and the Population is normal or the sample size is greater than 30. Our sample is probably not a simple random sample, but it was collected from the population of students at SRJC in a way that would not introduce undue bias. The population is not normal. It is skewed to the right as we can see in the histograms of the sample data. However our sample size is n=118 which is greater than 30. Hence the requirements for the test have been satisfied.





We also notice that the QQPlots do not look like straight lines.





Confidence Intervals

One sample T confidence interval:
μ : Mean of variable

95% confidence interval results:

Variable	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
W Miles to School	8.5631356	0.54518836	117	7.4834186	9.6428526
Miles to School	9.0432203	0.5842189	117	7.886205	410.200235

$7.5 < \mu_W < 9.6$ miles

1) We are 95% confidence that the mean distance to school for women is between 7.5 and 9.6 miles.

$7.9 < \mu_M < 10.2$ miles

2) We are 95% confidence that the mean distance to school for men is between 7.9 and 10.2 miles.

95% confidence interval results:

Difference	Sample Diff.	Std. Err.	DF	L. Limit	U. Limit
$\mu_1 - \mu_2$	-0.48	0.799	232.89	-2.05445	1.094281

$-2.1 < \mu_w - \mu_m < 1.1$ miles

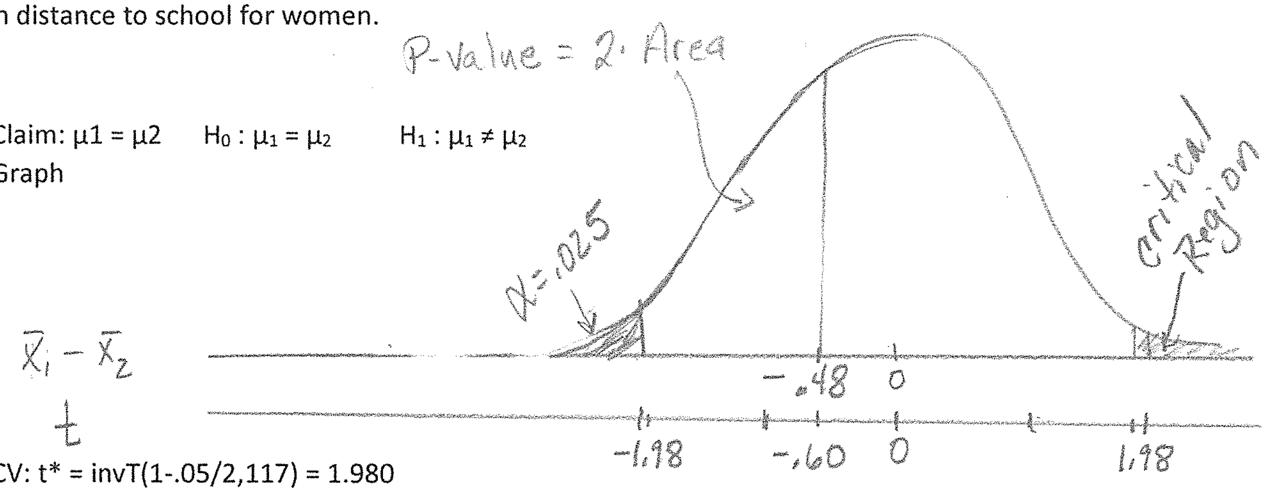
3) We are 95% confident that the difference in the mean distance to school for men and women is between -2.1 and 1.1 miles. Since 0 is in this interval we cannot support that there is a significant difference in the distance to school for male and female students.

Hypothesis Test

Question

At the 0.05 level of significance test the claim that the mean distance to school for men is the same as the mean distance to school for women.

1) Claim: $\mu 1 = \mu 2$ $\quad H_0 : \mu_1 = \mu_2$ $\quad H_1 : \mu_1 \neq \mu_2$
2) Graph



3) CV: $t^* = \text{invT}(1-.05/2,117) = 1.980$

Hypothesis test results:

Difference	Sample Diff.	Std. Err.	DF	T-Stat	P-value
$\mu_1 - \mu_2$	-0.48	0.799	232.89	-0.600791	0.5486

4) TS: $t=-0.60$ p-value = 0.54
5) Fail to reject $H_0 : \mu_1 = \mu_2$

Final Conclusion

P-Value Method: p-value = .5486 > 0.05 = α Fail to reject H_0: $\mu_1 = \mu_2$

If we assume that the mean distance to school is the same for the populations or all SRJC men and women, then there is a 55% chance of seeing a difference in the sample means of .48 miles or larger. Since it is likely that we would get this large a difference is sample means by chance, we cannot reject that the population means are the same.

Traditional method $|TS| < |CV|$ Fail to reject H_0: $\mu_1 = \mu_2$

Our test statistic $t = -0.60$ is between the critical values $t^* = -1.98$ and $t^* = 1.98$, therefore it is not in the critical region and it does not correspond to a statistically significant sample difference in means. We cannot support that the population of SRJC women have a different mean distance to school than the men.

Confidence Interval Method 0 in interval Fail to reject H_0: $\mu_1 = \mu_2$

We are 95% confident that the difference in the mean distance to school for men and women is between -2.1 and 1.1 miles. Since 0 is in this interval we cannot support that there is a significant difference in the distance to school for male and female students.

§9.2 & 9.3 Comparing Two Means

9.2 Means are Independent
Two unrelated groups, treatment and control (placebo)
Two classes on same test,

9.3 Means are Matched pairs
Two Data points from same group
Test 1 & Test 2 from same class
Before & After Diet / Drug / therapy

$H_0: \mu_1 - \mu_2 = 0$ or $\mu_1 = \mu_2$
$H_1: \mu_1 < \mu_2$, $\mu_1 > \mu_2$, $\mu_1 \neq \mu_2$

df = Smaller $(n_1 - 1, n_2 - 1)$ for us to calculat CV:
P-value given by 2-SampTTest uses the ugly formula for df. $\boxed{P475}$

Test the indicated claim about the means of two populations. Assume that the two samples are independent and that they have been randomly selected.

8) A researcher wishes to determine whether people with high blood pressure can reduce their blood pressure by following a particular diet. Use the sample data below to test the claim that the treatment population mean $\mu 1$ is smaller than the control population mean $\mu 2$. Test the claim using a significance level of 0.01. a) (5 Pts) State the null and alternate hypothosis.

Treatment Group	Control Group
$n1 = 85$	$n2 = 75$
$\overline{x}1 = 189.1$	$\overline{x}2 = 203.7$
$s1 = 38.7$	$s2 = 39.2$

$H_0: \mu_1 = \mu_2 \rightarrow \mu_1 - \mu_2 = 0$

$H_1: \mu_1 < \mu_2$

2-SampTTest

b) (5 Points) Graph and shade the critical region. Find the critical value.

c) (5 Points) Find the point estimate for the difference between the population means and it's test statistic. Label this on your graph. State the initial conclusion. 155 = -2.35

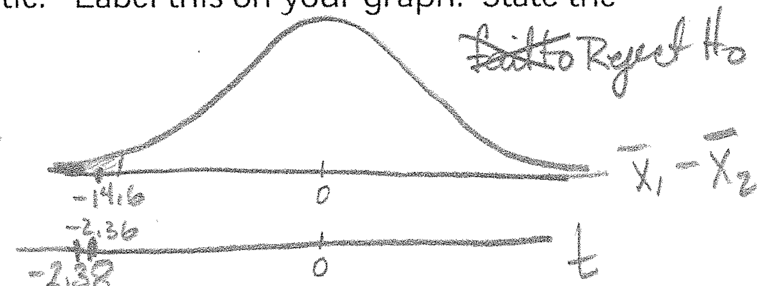
The two df Give a different Answer

CV: $t^* = \text{invt}(.01, 74) = -2.38$

PE: $\overline{x}_1 - \overline{x}_2 = 189.1 - 203.7 = -14.6$

TS: $t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = -2.36$

~~Fail to~~ Reject Ho



d) (5 points) Find the p-value, draw a new graph and label this area. Explain the meaning of this p-value.

P-value = .0096 < .01 = α ⇒ Reject Ho

Use →

there is less than a 1% chance that we would see this much lower Blood pressure readings from random sample variation.

e) (5 Points) Clearly state your final conclusion in language addressing the original claim.

Since the likelyhood of seeing this large a difference by random sampling variation is so small we conclude ~~HSE~~ to support that the Diet is effective at lowering blood pressure.

Construct the indicated confidence interval for the difference between the two population means. Assume that the two samples are independent and that they have been randomly selected.

9) A researcher wishes to determine whether people with high blood pressure can reduce their blood pressure by following a particular diet. Use the sample data below to construct a 98% confidence interval for $\mu_1 - \mu_2$ where μ_1 and μ_2 represent the mean for the treatment group and the control group respectively.

2-SampT Interval

$\bar{X}_1$, Not x_1

Treatment Group	Control Group
$n_1 = 85$	$n_2 = 75$
$\bar{x}_1 = 189.1$	$\bar{x}_2 = 203.7$
$s_1 = 38.7$	$s_2 = 39.2$

Test $\alpha_T = .01$ with one Tail
$\Rightarrow CL = .98 = 1 - 2\alpha$ for a Confidence Interval

CV: $t = \text{invT}(\alpha, df) = \text{invT}(.01, df) = \text{invT}(1 - .02/2, df)$

$\alpha_c = .02$

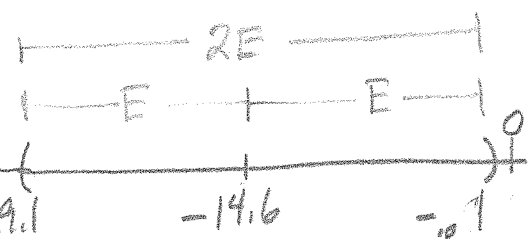
a) (2 Points) What point estimate of the difference in the population means, does this survey give?

PE: $\bar{X}_1 - \bar{X}_2 = -14.6$

b) (3 Points) What is the margin of error?

$$E = t_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 14.5$$

$$E = \frac{UB - LB}{2} = \frac{-.1 - (-29.1)}{2} = \frac{29}{2} = 14.5$$



c) (5 Points) Find the confidence interval.

-14.6 ± 14.5

$-29.11 < \mu_1 - \mu_2 < -.089$

$(\bar{X}_1 - \bar{X}_2) - E < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + E$

PE ± E

e) (5 Points) Interpret the meaning of this confidence interval.

Since 0 is Not in interval we are 98% confident that ~~there~~ is a difference in blood pressure after going on diet.

When there is a difference Say what it is
Blood pressure is lowered by between 29.1 and .1 points.

Determine whether the samples are independent or consist of matched pairs. Explain.

3) The effect of caffeine as an ingredient is tested with a sample of regular soda and another sample with decaffeinated soda.

Two groups of people would be independent.

Same person on Two Days would be dependent

4) The accuracy of verbal responses is tested in an experiment in which individuals report their heights and then are measured. The data consist of the reported height and measured height for each individual.

Dependent = Matched Pairs

Determine whether the samples are independent or dependent.

5) The effectiveness of a headache medicine is tested by measuring the intensity of a headache in patients before and after drug treatment. The data consist of before and after intensities for each patient.

6) The effectiveness of a new headache medicine is tested by measuring the amount of time before the headache is cured for patients who use the medicine and another group of patients who use a piacebo drug.

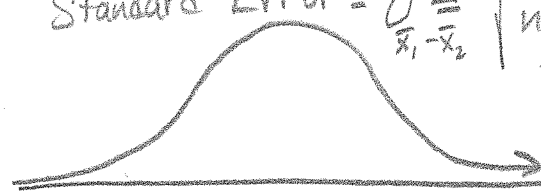
7) Suppose you wish to test a claim about the mean of the differences from dependent samples or to construct a confidence interval estimate of the mean of the differences from dependent samples. What are the requirements?

Standard Error = $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$

① SRS

② $n_1 > 30$ & $n_2 > 30$

or Normal populations



$\bar{x}_1 - \bar{x}_2$

③ σ_1 & σ_2 are unknown & Not equal. (Part I only)

t

$df = (n_1 - 1$ or $n_2 - 1)$ Smaller of

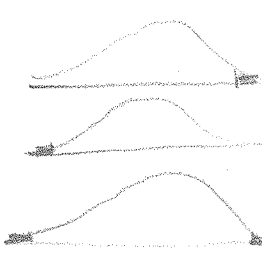
P-values = tcdf(TS, 999) Rt tail

(−999, TS) Lt Tail

2 · Above 2-tail

df = 2SampT test

CV: $t^* = \text{invt}(1-\alpha,\ \alpha,\ 1-\alpha/2,\ df)$ Rt, Lt, ± 2-tail



2

11) Below are the test scores for two tests in a class. Enter them into Excel. Then copy and paste them with Variable Names into StatCrunch.

1. Generate Summary Statistics and make Histograms, Normal Probability Plots and Side-by-Side Box Plots for the data sets. Paste them into your excel spread sheet.
2. Compare the displays. Comment on their similarities in center, spread, and shape.

Student	Test 1	Test 2
1	72	84
2	66	61
3	77	72
4	76	75
5	93	85
6	84	62
7	62	68
8	87	89
9	84	76
10	90	78
11	81	95
12	55	20
13	74	51
14	72	61
15	79	58
16	94	91
17	71	62
18	75	87
19	79	91
20	75	48

3. Create two 95% confidence interval for the difference between the means. First assume that the data sets are independent and then assume that they are matched pairs. Use **Calc->Estimate->2-sample t-interval for μ1-μ2** and **Calc->Estimate->Paired t-interval.** Clearly state the meaning of each confidence interval and determine which is the appropriate one to use and why.

4. The teacher claims that the students did better on the first test. Perform both hypothesis tests at the .05 level of significance. Use **Calc->Test->2-sample t test for μ1-μ2** and **Calc->Test->Paired t-test.** Write a statement about the conclusion of your t-test. Indicate which is the appropriate test to use and why.

Example does Diet Reduce Blood pressure

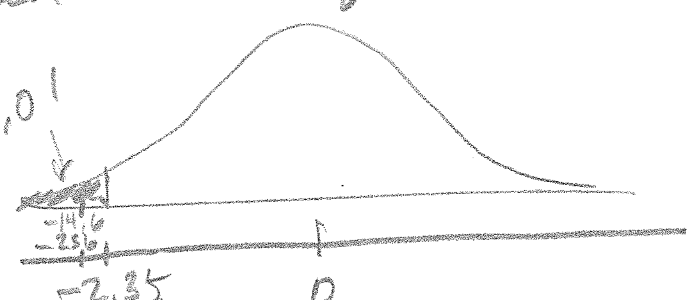
Summary Stats

Treatment	Control	
$n_1 = 85$	$n_2 = 75$	df = Smaller of $n_1 - 1$ & $n_2 - 1$
$\bar{X}_1 = 189.1$	$\bar{X} = 203.7$	$= 74$
$S_1 = 38.7$	$S_2 = 39.2$	$\alpha = .01$

Test claim that mean of Treatments is smaller than mean of control.

① $H_0: \mu_1 = \mu_2$

$H_1: \mu_1 < \mu_2$



② CV: $\text{Int}(.01, \cancel{74}\ 155) = -2.35$

$t = -2.38$

③ TS: $t = -2.365$

P-Value = .00963

df = 155

PE: $\bar{X}_1 - \bar{X}_2$

$= 189.1 - 203.7$

$= -14.6$

Reject H_0

the blood pressure of the Treatment group is significantly lower than the control group.

4/22/14

M15

§ 9.3 & 9.4 Hypothosis Tests and conf. dence Intervals to Compare Two Population Means.

Hypothosis Test to two Independent Means.

PE: $\bar{X}_1 - \bar{X}_2$

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 < \mu_2,\ \mu_1 > \mu_2$ or $\mu_1 \neq \mu_2$

CV: $t = \text{InvT}(\text{arealeft}, df)$

Left	α
Rt	$1-\alpha$
Two	$1-\alpha/2$

TS: $t = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ 2Samp-T-Test Independent Samples

Final Conclusion: a difference in population Means
1. fail to show
2. there is a significant difference in pop. Means.

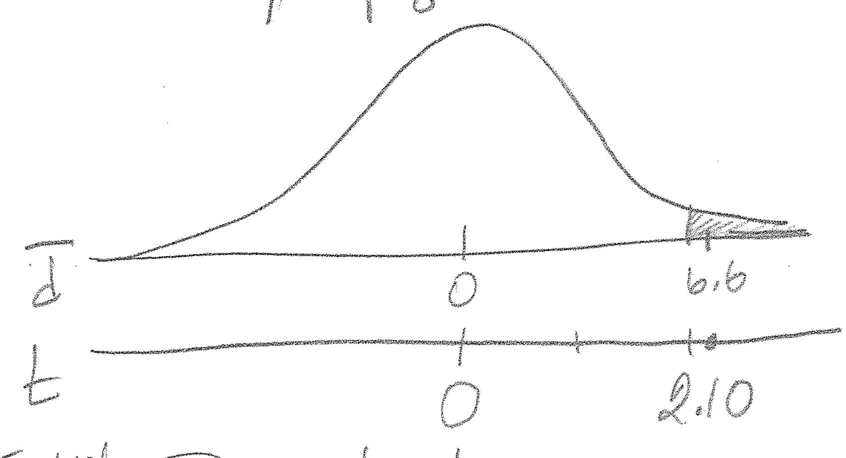
Independent
Test 1 = Test 1 from fall 2013
Test 2 = test 1 from Spring 2013

Dependent = Matched pairs

Test 1 = Fall 2013 Test 1
Test 2 = Fall 2013 Test 2
then the data is paired by Student.
we do a match pairs Test on the set of differences. $d_1 = x_{1,1} - x_{1,2}$

$H_0: \mu_d = 0$
$H_1: \mu_d \neq 0$

→ TS: $t = 2.104$
P-value = .0489 < α
PE: $\bar{d} = 6.6$
↑
Mean of differences between the paired Samples.
Calculator
$\bar{d}$ = mean of $L_3 = L_1 - L_2$



Initial Reject H_0

Final Conclusion

We can show they did Not do the same on both Test. They did Better on Test 1 than on Test 2.

$$\bar{d} = T1 - T2 > 0$$

$$\Rightarrow T1 > T2$$

because sample Data is in Right Tail.

Computer Practice:
Goals: Understand how to use computers to do hypothesis test and make confidence intervals to test differences between means.

Below are the test scores for two tests in a class. Enter them into Excel. Then copy and paste them with Variable Names into Data Desk.

1. Generate Summary Statistics and make Histograms, Normal Probability Plots and Side-by-Side Box Plots for each of the data sets. Paste them into your excel spread sheet.
2. Compare the displays. Comment on their similarities in center, spread, and shape.

Student	Test 1	Test 2
1	72	84
2	66	61
3	77	72
4	76	75
5	93	85
6	84	62
7	62	68
8	87	89
9	84	76
10	90	78
11	81	95
12	55	20
13	74	51
14	72	61
15	79	58
16	94	91
17	71	62
18	75	87
19	79	91
20	75	48

3. Create two 95% confidence interval for the difference between the means. First assume that the data sets are independent and then assume that they are matched pairs. Use **Calc->Estimate->2-sample t-interval for µ1-µ2** and **Calc->Estimate->Paired t-interval.**

4. Clearly state the meaning of the confidence interval and determine which is the appropriate one to use and why.

5. Perform both hypothesis tests at the .05 level of significance. Use **Calc->Test->2-sample t test for µ1-µ2** and **Calc->Test->Paired t-test.**

6. Write a statement about the conclusion of your t-test and which is the appropriate test to use and why.

Exploring Data Sets

- Find descriptive statistics $\bar{x}, s, n$
- Create Boxplots one same scale (Side by Side)
  Compare.
- Histograms
- Outliers

Matched Pairs is generally better than two independent samples