

9.1

Statistical Inference

Sample Data



Sample Statistic
(\bar{x} or \hat{p})

} Point Estimator



Make reasonable conclusion
about population parameter



Confidence Interval



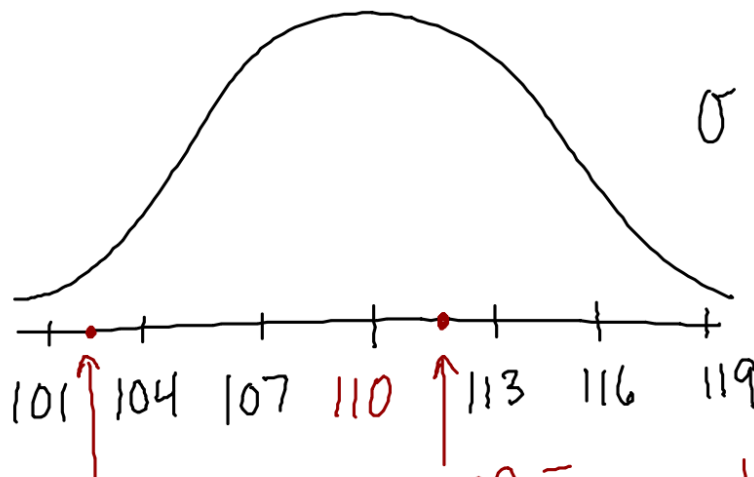
Hypothesis Test

General Process

- 1) Guess the population proportion / mean
- 2) Use a sample proportion (\hat{p}) or sample mean (\bar{x}) to determine if guess is reasonable

Principal Croushore says average IQ of LN students is 110 ($\sigma = 12$)... SRS of 16 students

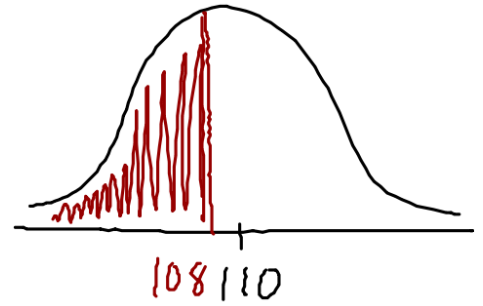
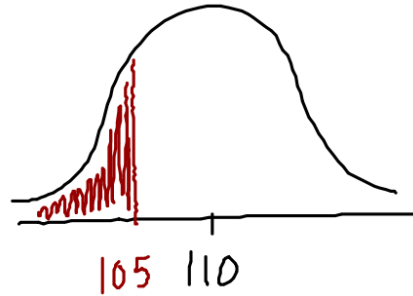
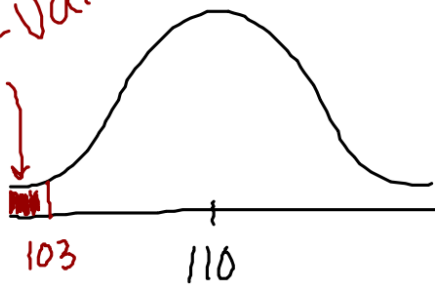
$$\sigma = \frac{12}{\sqrt{16}} = 3$$



If $\bar{x} = 103$ then
 $\mu = 110$ is not reasonable

If $\bar{x} = 112$, then $\mu = 110$
is reasonable

P-value



In General

$P\text{-value} < 5\% \rightarrow$ statistical significance
to reject

$5\% < P\text{value} < 50\% \rightarrow$ not statistically significant
(guess is reasonable)

Types of Hypotheses

Null Hypothesis (H_0)

Crousore thinks average IQ is 110

$$H_0 : \mu = 110$$

Crousore thinks 90% of seniors going to college

$$H_0 : p = .90$$

Alternative / Research Hypothesis (H_a)

Reporter thinks avg IQ is more than 110

$$H_a: \mu > 110$$

Reporter thinks avg IQ is less than 110

$$H_a: \mu < 110$$

Reporter thinks avg IQ is not 110

$$H_a: \mu \neq 110$$

one
sided

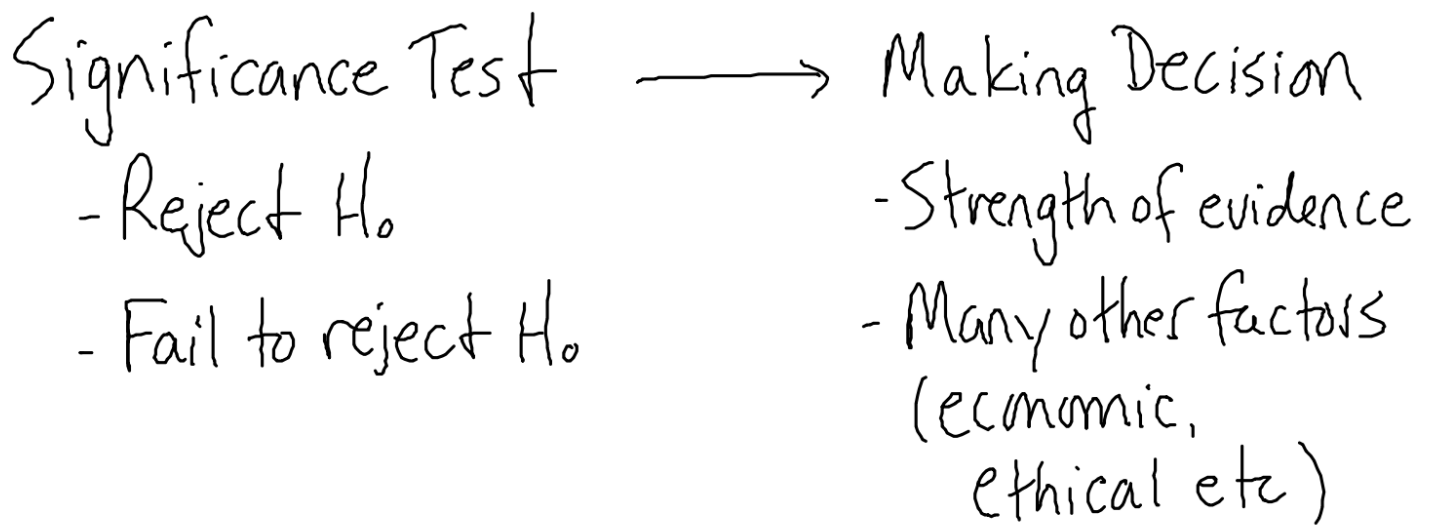
Two Sided
Hypothesis

Significance level (α Level)

- A fixed value used to determine statistical significance
- If $\alpha = .05$ then P-value must be less than .05 to reject H_0 (accept H_a)
- If $\alpha = .01$ then P-value must be less than .01 to reject H_0 (accept H_a)

- If P-value $> \alpha$ level
then we fail to reject H_0
- We never "accept" H_0
- We simply have evidence that H_0
is reasonable (not guilty vs guilty)

Making Decisions



	H_0 True	H_0 False	
Reject H_0 (Accept H_a)	Type I Error $P = \alpha$ level	Power $(1 - \beta)$	Ability of a test to detect a different H_0
"Accept" H_0 (Reject H_a)	✓	Type II Error $P = \beta$ risk	Depends on true value of parameter and α level

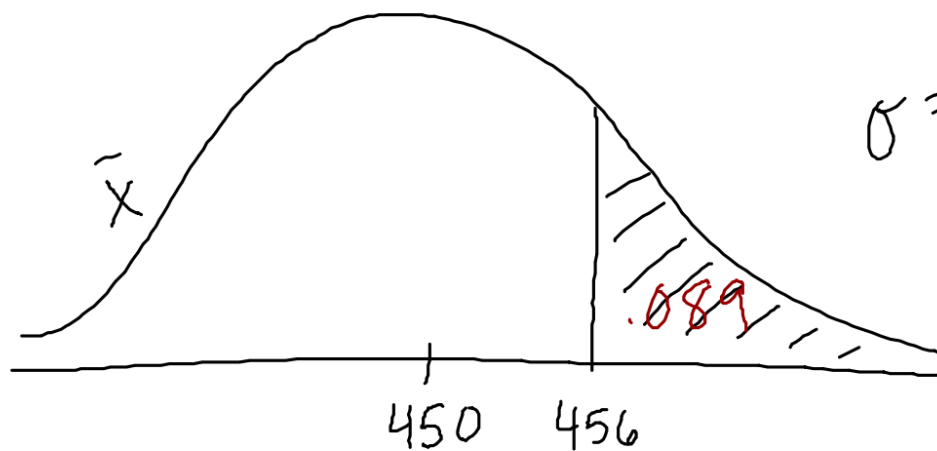
Increase Power

- 1) Increase α level (from 1% to 5%)
- 2) Increase sample size

A reporter claims that if every HS senior in Indiana took the SAT then the average Math score would be no more than 450

SRS ($n=500$) $\rightarrow \bar{X} = 456, \sigma = 100$

$$H_0: \mu = 450 \quad H_a: \mu > 450$$

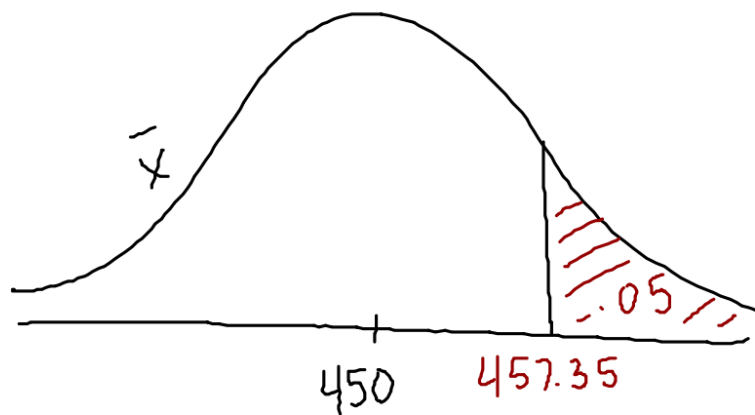


$$\sigma = \frac{100}{\sqrt{500}} = 4.4$$

Since $p > .05$, we fail to reject H_0 and conclude the mean is 450.

BUT, what if $M = 460$... then we have made a Type II error. Determine the $P(\text{Type II Error})$ for $M = 460$ and $\alpha = .05$.

1) Determine critical \bar{X} (at $\alpha = .05$)



$$2) P(\text{Type II Error})$$

$$= P(\bar{X} < 457.35 \text{ when } \mu = 460)$$

$$= P\left(Z < \frac{457.35 - 460}{4.47}\right)$$

$$= P(Z < -.5928)$$

$$= .28$$

Assuming the true mean is 460 (and not 450):

	H_0 True	H_0 False
Reject H_0	Type I Error $P = .05$	Power $*$ $P = .72$
Fail To Reject H_0	✓	Type II Error $P = .28$

* At $\alpha = .05$, the probability of rejecting H_0 when $\mu = 460$ is .72

Sec 9.2

One-Proportion (Sample) Z-Test

Used to determine if a hypothesized population proportion (p) is reasonable based on a sample proportion (\hat{p})

Steps

P State population parameter
($p = ?$, $M = ?$)

H State null/alternative hypotheses

A State assumptions required

1) Random

2) Normal Sampling Distribution

$$np_0 \geq 10 \quad n(1-p_0) \geq 10$$

3) Independent

$$N > 10n$$

T Perform test

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

→ Table A,
Calculator or
Compare

S State conclusion in context

1- PROPORTION Z-TEST

This test is used to determine if a hypothesized population proportion (p) is reasonable based on a sample proportion (\hat{p})

P. 553

A potato-chip producer selects a random sample of 500 potatoes from a truck shipment and determines that 47 have blemishes.

At $\alpha = .05$, is there evidence that more than 8% of the shipment's potatoes have blemishes?

P) IDENTIFY POPULATION PARAMETER:

p = proportion of all potatoes in shipment with blemishes

H) STATE HYPOTHESES:

$$H_0: p = .08$$

$$H_a: p > .08$$

A) VERIFY CONDITIONS REQUIRED FOR TEST:

✓a) Random

Random sample used

✓b) Normal Sampling Distribution

$$np_0 = (500)(.08) = 40 \geq 10 \quad \checkmark$$

$$n(1-p_0) = (500)(.92) = 460 \geq 10 \quad \checkmark$$

✓c) Independent

Number of potatoes > 10 (500) > 5000 ?

T) PERFORM TEST USING

$$\hat{p} = \frac{47}{500} = .094$$

a) TABLE A:

i) Calculate z test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.094 - .08}{\sqrt{\frac{(.08)(.92)}{500}}} = 1.15$$

} Table A or calculator

ii) Determine area under curve (P-value)



P-value $\approx .1251$

Since $1.15 < 1.645$
the P-value $> .05$

b) CALCULATOR:

STAT → **TESTS** → 1 Prop Z Test → pValue = .1243

S) STATE CONCLUSION:

At $\alpha = .05$, we fail to reject H_0 ...there is not enough evidence to conclude that the proportion of blemished potatoes in the truck is more than 8%

CONFIDENCE INTERVAL (Use PAIS):

After checking for normal distribution [$n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$], a 90% confidence interval for the proportion of blemished potatoes in this truck is:

STAT → TEST → 1-Prop Z Int = (.073, .115)

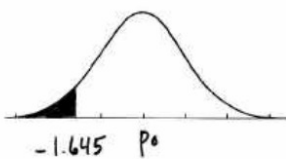
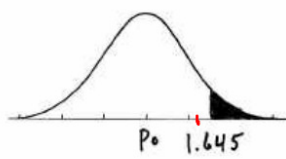
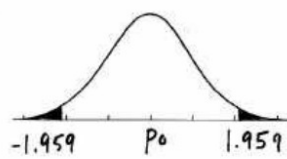
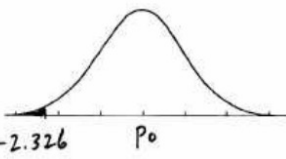
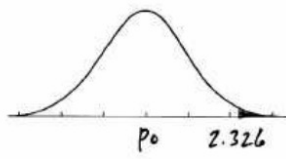
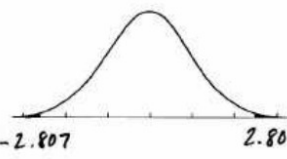
We are 90% confident that between 7.3% and 11.5% of potatoes in this shipment are blemished (which reinforces the conclusion from the hypothesis test since as few as 7.3% may actually be blemished).

CRITICAL Z-VALUES

One Tailed

One Tailed

Two Tailed

α - level	$H_a: p < p_0$	$H_a: p > p_0$	$H_a: p \neq p_0$
.05			
.01			

2nd > VARS (DISTR) > InvNorm (% , 0, 1)

Sec 9.3

(One-Sample) T Test

Used to determine if a hypothesized population mean (μ) is reasonable based on a sample mean (\bar{x})

Steps

P Population parameter

H State hypotheses

$$H_0: \mu = \square$$

$$H_a: \mu < \square$$

$$H_a: \mu > \square$$

$$H_a: \mu \neq \square$$

A

State assumptions

1) Random

2) Normal Sampling Distribution

a) Population normal

b) Sample size $(n) > 30$

c) Check data 

x

3) Independent ($N > 10n$)

T Perform Test

$$t = \frac{\bar{X} - M_0}{\frac{s}{\sqrt{n}}} \longrightarrow \text{Calculator (tcdf)} \\ \text{or Table B}$$

S State conclusion in context

ONE SAMPLE T-TEST

This test is used to determine if a population mean (μ) is reasonable based on a sample mean (\bar{x}).

Researchers believe that women (18-24) get less than the RDA of calcium (1200mg/day).

To test this hypothesis at the $\alpha = .05$ significance level, an SRS of 38 women between the ages of 18 and 24 years estimated their daily intakes of calcium (in mg):

808	882	1062	970	909	802	374	416	784	997	651	716
438	1420	1425	948	671	696	1156	684	1933	748	1203	2433
1050	976	572	403	626	774	1253	546	1325	446	465	1269
1255	1100										

P) STATE POPULATION PARAMETER:

M = the average amount of daily calcium intake for women (18-24 yo)

H) STATE HYPOTHESES:

$H_0: M = 1200 \text{ mg}$ $H_a: M < 1200 \text{ mg}$

A) VERIFY CONDITIONS REQUIRED FOR TEST:

✓ a) Random

Random sample used

✓ b) Normal population or large sample size or justification for normality after omitting outliers

Sample size large (CLT)

c) Independent

Number of women (18-24) > 10 (38) > 30 ✓

T) PUT DATA INTO LIST AND

a) USE TABLE B:

- i) Determine mean (\bar{x}) and standard deviation (s)

$$\bar{X} = 926 \quad S = 427.30$$

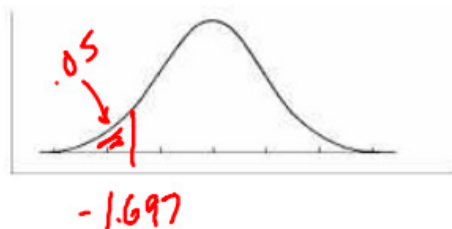
- ii) Calculate t statistic

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{926 - 1200}{\frac{427.30}{\sqrt{38}}} = -3.95$$

- iii) Determine degrees of freedom

$$df = n - 1 = 38 - 1 = 37$$

- iv) Determine critical t -value and P-value



Since $-3.95 < -1.697$
P-value $< .05$

$$\boxed{\text{DISTR}} \rightarrow t\text{cdf}(-100, -3.95, 37) = .00016$$

b) USE CALCULATOR

$$\boxed{\text{STAT}} \rightarrow \text{TESTS} \rightarrow \text{T Test} \begin{cases} \nearrow t = -3.95 \\ \searrow p = .00016 \end{cases}$$

S) STATE CONCLUSION:

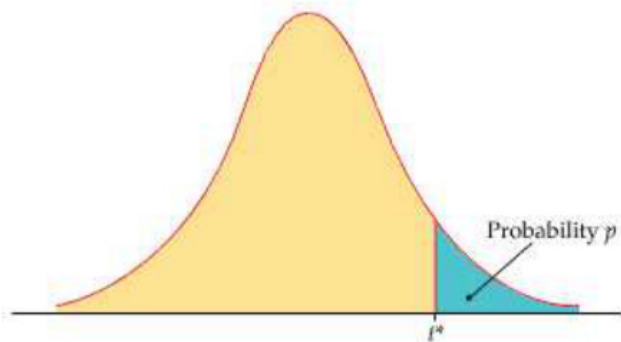
At $\alpha = .05$, there is sufficient evidence ($p = .00016$) to reject H_0 and conclude the average amount of calcium intake for women (18-24) is less than the RDA of 1200 mg.

CONFIDENCE INTERVAL (Use PAIS):

A 90% confidence interval for the mean daily intake in calcium can be found using:

STAT → TESTS → T Interval = (809, 1043)

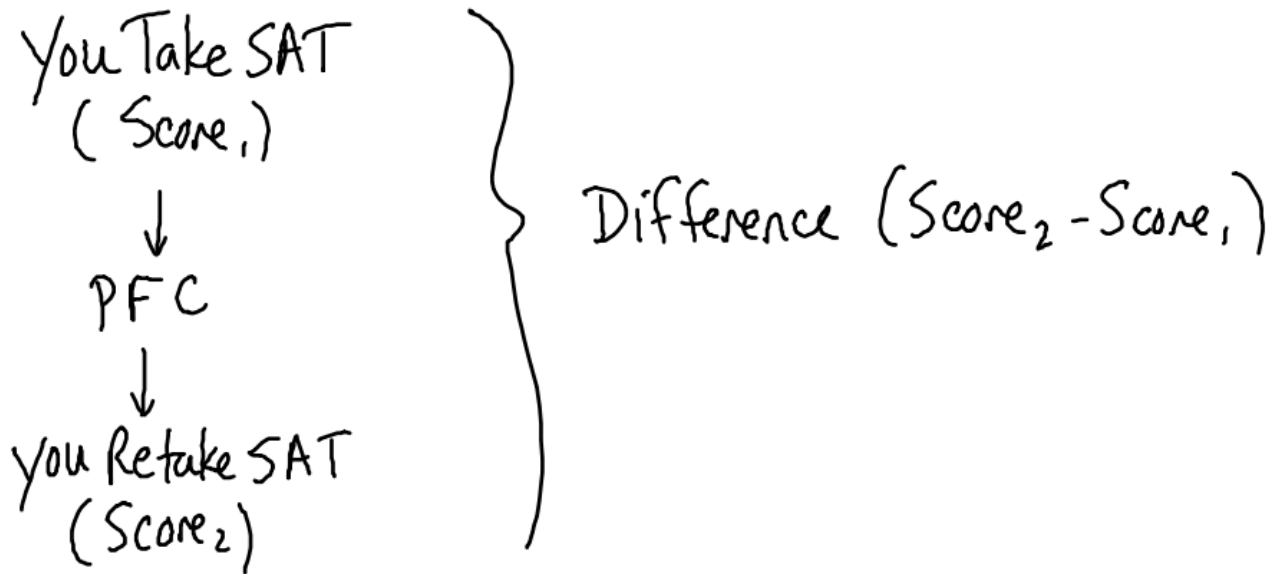
We are 90% confident that the average daily intake of calcium for women between the ages of 18 and 24 years old is between 809 mg and 1043 mg (which reinforces the findings of the test that women receive less than 1200 mg of calcium/day).



t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Within Subjects / Matched Pairs t Test



$$H_0: \mu = 0$$

where μ = mean difference in responses

$$H_a: \mu < 0 \quad H_a: \mu > 0 \quad H_a: \mu \neq 0$$

MATCHED PAIRS T TEST

*This test is used to compare the responses to a treatment in a **within-groups** design (ie, does an SAT prep course improve an individual's SAT scores?).*

A listening test with a maximum score of 36 was administered to Spanish teachers before and after an institute designed to improve Spanish listening skills.

Sub	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Pre	30	28	31	26	20	30	34	15	28	20	30	29	31	29	34	20	26	25	31	29
Post	29	30	32	30	16	25	31	18	33	25	32	28	34	32	32	27	28	29	32	32

Determine if the institute improved listening skills at the 5% significance level.

CALCULATE THE DIFFERENCES BETWEEN THE 2 TREATMENTS:

Sub	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Pre	30	28	31	26	20	30	34	15	28	20	30	29	31	29	34	20	26	25	31	29
Post	29	30	32	30	16	25	31	18	33	25	32	28	34	32	32	27	28	29	32	32
Dif	-1	2	1	4	-4	-5	-3	3	5	5	2	-1	3	3	-2	7	2	4	1	3

P) STATE POPULATION PARAMETER:

μ = mean difference in listening scores (post - pre)
for Spanish teachers attending institute

H) STATE HYPOTHESES:

$H_0: \mu = 0$ $H_a: \mu > 0$


A) VERIFY CONDITIONS REQUIRED FOR TEST:

a) Random

Don't know... results may be invalid!

b) Normal sampling distribution- ~~normal population~~ or large ~~sample size~~ ($n > 30$) or justification for normal distribution ($n < 30$) after omitting outliers

Modified box plot - slightly skewed with no outliers



c) Independence

Number of attendees $> 10 (20) > 200$?

T) PERFORM TEST:

a) USING TABLE B:

- i) Calculate mean (\bar{x}) and standard deviation (s)

$$\bar{x} = 1.45 \quad s = 3.2032$$

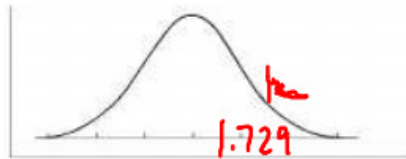
- ii) Calculate t statistic

$$t = \frac{\bar{x} - 0}{\frac{s}{\sqrt{n}}} = \frac{1.45 - 0}{\frac{3.2032}{\sqrt{20}}} = 2.024$$

- iii) Determine degrees of freedom

$$df = 20 - 1 = 19$$

- iv) Determine critical t -value and P -value



$$P\text{-value} < .05$$

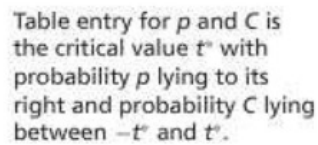
$$t_{cdf}(1.729, 100, 19) \approx .0286$$

b) USING CALCULATOR:

STAT \rightarrow TESTS \rightarrow t Test \rightarrow $t = 2.02$
 $\rightarrow p = .0286$

S) STATE CONCLUSION:

At $\alpha = .05$, there is sufficient ^{evidence} to reject H_0 ($p = .0286$) and conclude the institute improved listening skills for everyone who attended



t distribution critical values

	Upper-tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

CONFIDENCE INTERVAL (Use PAIS):

A 90% confidence interval for the mean increase in listening scores can be found using:

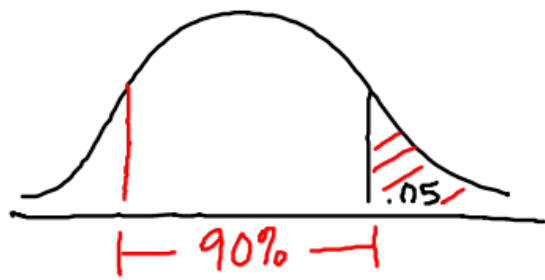
STAT → TESTS → T Interval = (.21, 2.69)

We are 90% confident that the mean increase in the listening scores was between .21 and 2.69 points after teachers participated in the institute.

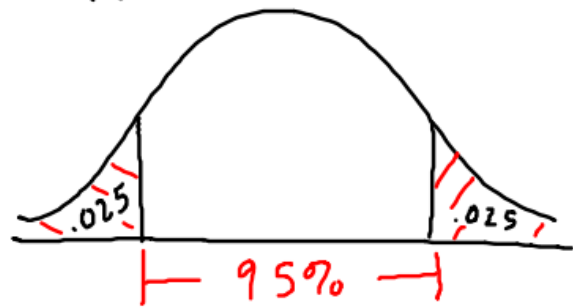
Using Significant Tests Wisely

- 1) Badly designed experiments or surveys produce invalid results

2) Reinforce results with appropriate CIs



One Tailed



Two Tailed

3) Lack of statistical evidence does not mean H_0 is true

* 4) Statistical significance ($p < .05$) is not the same as practical evidence (HSSSE)

2007 HIGH SCHOOL SURVEY OF STUDENT ENGAGEMENT

Students at 100+ schools in over 30 states were asked to complete a survey;
one of the questions was:

Have you ever been bored in high school?

0	1	2	3	4
Never	Once or Twice	Once in a while	Every Day	Every Class

There was overwhelming statistical evidence ($p = .0001$)
that Lawrence North students were **less** bored in school
compared to the other HSSSE schools.

Actual Results:

HSSSE Mean = 2.75

LN Mean = 2.64