If you are one of the first 10 people to class, please write the # of hours of sleep you got last night.

(nearest 0.5 hours)

7	7	3	6	5
5	10	4.5	8	5

Pick up the Warm Up

DATA ANALYSIS 17

Data are obtained from a random sample of adult women with regard to their ages and their monthly expenditures on health products. The resulting regression equation is: $\widehat{\text{Expenditure}} = 43 + 0.23(\text{Age})$ with $r = .27$. What percentage of the variation in expenditures can be explained by looking at ages?

(A) 0.23 percent

(B) 23 percent

(C) 7.29 percent

(D) 27 percent

(E) 52.0 percent

$r^2 =$

Answer: **(C) The coefficient of determination r^2 gives the proportion of the *y*-variance that is predictable from a knowledge of *x*. In this case, $r^2 = (.27)^2 = .0729$ or 7.29 percent.**

DATA ANALYSIS 15

Given that the median is 270 and the interquartile range is 20, which of the following statements is true?

(A) Fifty percent of the data are greater than or equal to 270.

(B) Fifty percent of the data are between 260 and 280.

(C) Seventy-five percent of the data are less than or equal to 280.

(D) The mean is between 250 and 290.

(E) The standard deviation is approximately 13.5.

Answer: **(A) Fifty percent of the data are on either side of the median. The interquartile range gives the distance between Q_1 and Q_3, but doesn't say where the median falls between Q_1 and Q_3. For example, in this example, it is possible that $Q_1 = 265$ and $Q_3 = 285$. Depending upon the shape of the distribution, the mean and standard deviation could be anything.**

HW
LOTTERY

Today:

Testing Claims about Population Proportions

How many hours of sleep did you get last night?

Lesson 9.3 Day 1 **Significant Tests for Means, μ**

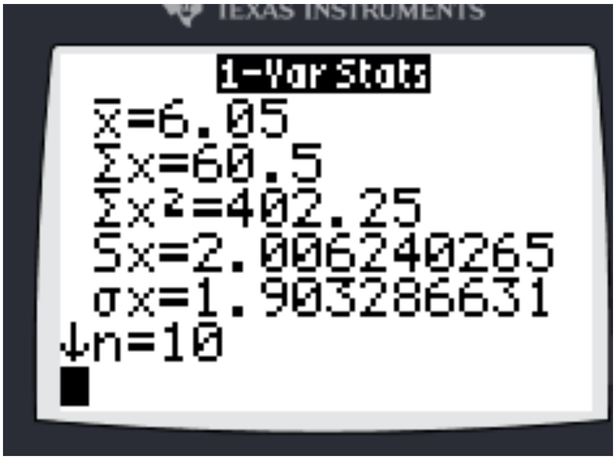
Are you getting enough sleep?

It's recommended that teenagers get 8 hours of sleep a night. Mr. Cedarlund believes his AP Stats students are getting less than the recommended 8 hours of sleep per night. To test this belief, he takes a random sample of 10 students in class and records the number of hours of sleep for each. ***Do these data provide convincing evidence that the AP stats students get less than 8 hours of sleep per night using $\alpha = 0.05$?***

do up through question #3

7 7 3 6 5

5 10 4.5 8 5



Sample of
10 of you

1. Calculate the sample mean and standard deviation from the sample taken.

2. State the appropriate hypotheses for a significance test. Be sure to define the parameter of interest.

3. What conditions must be met? *It turns they are the same conditions as for making a confidence interval for a mean.* Check them.

1. Calculate the sample mean and standard deviation from the sample taken.

$\bar{x} = 6.05 \quad S_x = 2.01$

2. State the appropriate hypotheses for a significance test. Be sure to define the parameter of interest.

$H_0: \mu = 8$

$H_a: \mu < 8$

$\mu \rightarrow$ true mean hours of sleep for AP students at SHS

3. What conditions must be met? *It turns they are the same conditions as for making a confidence interval for a mean.* Check them.

Random

No

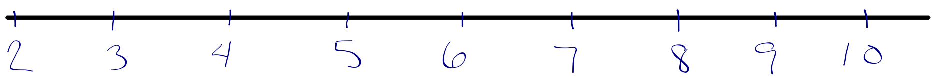
10%

- Not met
- but we'll assume indep.

Normal

$n \geq 30$ CLT ?

Strong skew or outliers?


2
3
4
5
6
7
8
9
10


y
x

What To Do When Conditions Are Not Satisfied
If a condition is not satisfied, go through all of the proper motions for calculating a confidence interval or carrying out a significance test. Then at the very end of your conclusion, write:

We must be cautious of our results, as the ___________ condition was not satisfied.

Now work up to question 5, then wait

4. Give the formulas for the mean and standard deviation of the sampling distribution of $\bar{x}$ and calculate the values.

5. Draw a picture of the appropriate distribution and then calculate the test statistic.

4. Give the formulas for the mean and standard deviation of the sampling distribution of $\bar{x}$ and calculate the values.

$\mu_{\bar{x}} = \mu = 8$

↑ assumed from H_0

5. Draw a picture of the appropriate distribution and then calculate the test statistic.

4. Give the formulas for the mean and standard deviation of the <u>sampling distribution</u> of $\bar{x}$ and calculate the values.

$\mu_{\bar{x}} = \mu = 8$

↑ assumed from H_0

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx$

5. Draw a picture of the appropriate distribution and then calculate the test statistic.

4. Give the formulas for the mean and standard deviation of the <u>sampling distribution</u> of $\bar{x}$ and calculate the values.

$\mu_{\bar{x}} = \mu = 8$

↑ assumed from H_0

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s_x}{\sqrt{n}} = \frac{2.01}{\sqrt{10}} \approx 0.636$

5. Draw a picture of the appropriate distribution and then calculate the test statistic.

4. Give the formulas for the mean and standard deviation of the <u>sampling distribution</u> of $\bar{x}$ and calculate the values.

$\mu_{\bar{x}} = \mu = 8$

assumed from H_0

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s_x}{\sqrt{n}} =$

5. Draw a picture of the appropriate distribution and then calculate the test statistic.

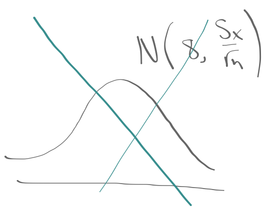
4. Give the formulas for the mean and standard deviation of the <u>sampling distribution</u> of $\bar{x}$ and calculate the values.

$\mu_{\bar{x}} = \mu = 8$

assumed from H_0

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s_x}{\sqrt{n}} =$

5. Draw a picture of the appropriate distribution and then calculate the test statistic.



$z = \frac{\bar{x} - \mu}{\frac{s_x}{\sqrt{n}}}$

In an ideal world, for a test of H_0: $\mu = \mu_0$, our standardized test statistic would be $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

Because the population standard deviation σ is almost always unknown, we use the sample standard deviation s_x in its place. The resulting standardized test statistic has the *standard error* of $\bar{x}$ in the denominator and is denoted by t.

$$t = \frac{\bar{x} - \mu_0}{s_x / \sqrt{n}}$$

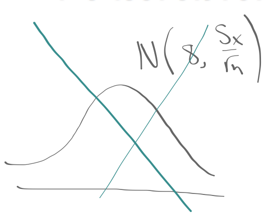
4. Give the formulas for the mean and standard deviation of the <u>sampling distribution</u> of $\bar{x}$ and calculate the values.

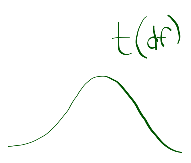
$\mu_{\bar{x}} = \mu = 8$

assumed from H_0

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx \frac{s_x}{\sqrt{n}} = .636$

5. Draw a picture of the appropriate distribution and then calculate the test statistic.

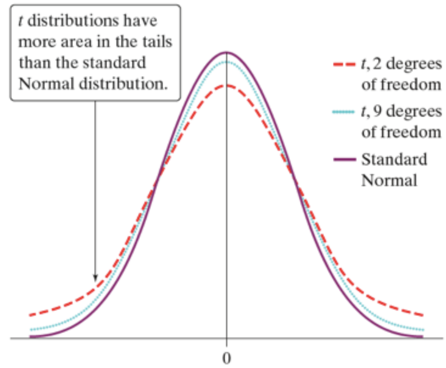




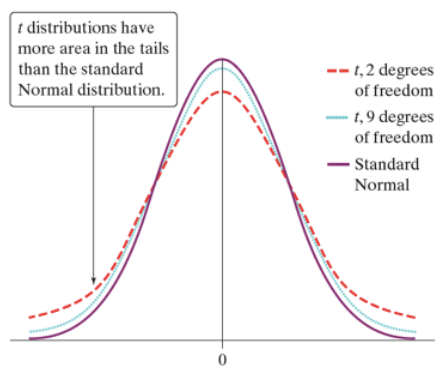
$z = \frac{\bar{x} - \mu}{\frac{s_x}{\sqrt{n}}}$

$t = \frac{\bar{x} - \mu}{\frac{s_x}{\sqrt{n}}} = \frac{6.05 - 8}{.636} \approx -3.07$

The *t* distributions have more variability than the standard Normal distribution, because the *t* distributions have more area in the tails.



As the degrees of freedom increase, the *t* distributions approach the standard Normal distribution.



Now, to question 6 and 7

$t = -3.07$ $df = 10 - 1 = 9$

6. Remember, since we are working with means, the test statistic is a *t* value. Use table B to find the P-value. *Go to row* $df = n - 1$. *Then find the closest t-values and match to the tail probability. Sometimes a range of probabilities are reported.*

P-Value is in between .5% and 1%

7. What conclusion can we make?

reject Ho ?

fail to reject Ho ?

1% .5%

Tail probability p								
.10	.05	.025	.02	.01	.005	.0025	.001	.0005
3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437

Use Table B to find the P-Value

We can use Table B to find a *P*-value from the appropriate *t* distribution when performing a test about a population mean.

In the "Better batteries" example,
H_0: $\mu = 30$
H_a: $\mu > 30$

$$t = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}} = \frac{33.93 - 30}{9.82/\sqrt{15}} = 1.55$$

df = $n - 1 = 15 - 1 = 14$

	Upper-tail probability *p*		
df	0.10	0.05	0.025
13	1.350	1.771	2.160
14	1.345	1.761	2.145
15	1.341	1.753	2.131
	80%	90%	95%
	Confidence level *C*		

$P(t \geq 1.55)$ is between 0.10 and 0.05.

	Upper-tail probability p		
df	0.10	0.05	0.025
13	1.350	1.771	2.160
14	1.345	1.761	2.145
15	1.341	1.753	2.131
	80%	90%	95%
	Confidence level C		

$P(t \geq 1.55)$ is between 0.10 and 0.05.

We'll report it as an interval of probabilities

Table B has limitations for finding *P*-values.

- It gives an interval of possible *P*-values for a significance test.
- The table includes probabilities only for *t* distributions with degrees of freedom from 1 to 30 and then skips to df = 40, 50, 60, 80, 100, 1000.

	Upper-tail probability p		
df	0.10	0.05	0.025
13	1.350	1.771	2.160
14	1.345	1.761	2.145
15	1.341	1.753	2.131
	80%	90%	95%
	Confidence level C		

Table B has limitations for finding *P*-values.

- It gives an interval of possible *P*-values for a signi…
- The table shows probabilities for only *posit*… value for a negative value of *t*, we use … distributions.
- The table includes probabi… with degrees of freedom from 1 to 30 … , 60, 80, 100, 1000.

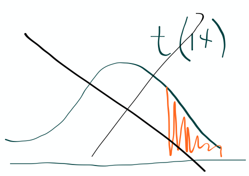
CAUTION:

If the df you need isn't provided in Table B, use the next smaller df that is available.

		0.025
		2.160
	1.761	2.145
	1.753	2.131
…0%	90%	95%
	Confidence level *C*	

Given the limitations of Table B, the advice is to use technology to find P-values when carrying out a significance test about a population mean.

(Just for the P-Value)



$t_{cdf}($ -1000, -3.07, 9)

~~1.55~~ 1000 14

Lower, upper, df

~~= 0.0717~~

~~7.17%~~

$P(t \geq 3.17)$

```
NORMAL FLOAT AUTO REAL RADIAN MP
tcdf(1.55,1000,14)
                    0.0717235647
```

$P(t \leq -3.17 \text{ or } t \geq 3.17)$

```
NORMAL FLOAT AUTO REAL RADIAN MP
tcdf(-1000,-3.17,36)*2
                    0.0031080065
```

Significance Test for a Population Mean, μ.

Important ideas:

CONDITIONS

Random

10%

Normal/~~Large Counts~~ Large Sample

- Pop Normal
- $n \geq 30$ CLT
- No strong skew or outliers

same as for a confid. interval for a mean

Significance Test for a Population Mean, μ.

Important ideas:

CONDITIONS

Random

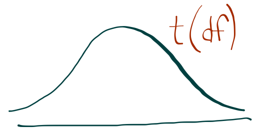
10%

Normal/Large ~~Counts~~ Sample

- Pop. Normal
- $n \geq 30$ CLT
- No strong skew or outliers

t - Statistic

$$t = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$$



t(df)

$df = n - 1$

P-Value

t^* is for confidence intervals only.

Significance Test for a Population Mean, μ.

Important ideas:

CONDITIONS

Random

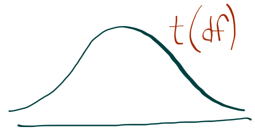
10%

Normal/Large Counts

- Pop. Normal
- $n \geq 30$ CLT
- No strong skew or outliers

t - Statistic

$$t = \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$$



t(df)

P-Value

Use ~~table~~ B with ~~df~~ → Report an interval of prob.

calculator

tcdf(lower, upper, df)

will give you the P-Value

t^* is for confidence intervals only.

Significance Test for a Population Mean, μ.

Important ideas:

CONDITIONS

Random

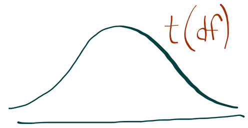
10%

Normal/Large Counts

- Pop Normal
- $n \geq 30$ CLT
- No strong skew or outliers

t - Statistic

$$t = \frac{\bar{x} - \mu_0}{s_x / \sqrt{n}}$$



$t(df)$

P-Value

Use ~~table~~ B with ~~df~~

calculator

t_{cdf}(lower, upper, df)

multiply by 2 for two sided H_a

Putting It All together

Carrying Out a Significance Test for μ

•

Significance Test: **One-Sample t Test for μ**

The level of dissolved oxygen (DO) in a stream or river is an important indicator of the water's ability to support aquatic life. A researcher measures the DO level at 15 randomly chosen locations along a stream. Here are the results in milligrams per liter (mg/l):

4.53	5.04	3.29	5.23	4.13	5.50	4.83	4.40
5.42	6.38	4.01	4.66	2.87	5.73	5.55	

An average dissolved oxygen level below 5 mg/l puts aquatic life at risk.
Do the data provide convincing evidence at the $\alpha = 0.05$ significance level that aquatic life in this stream is at risk?

State

Plan

Do

Conclude

State

[Define parameters, Show evidence for H_a (show the statistic), Hypotheses, and significance level, α]

State

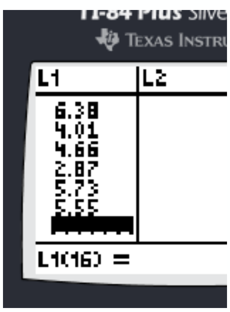
[Define parameters, Show evidence for H_a(show the statistic), Hypotheses, and significance level, α]

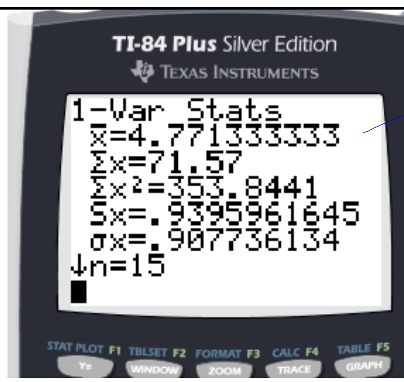
parameter μ = true mean DO level in the stream.

statistic $\bar{x}$ = ← Need to calculate

Hypothesis

α





$\bar{x} = 4.771$

State

[Define parameters, Show evidence for H_a(show the statistic), Hypotheses, and significance level, α]

parameter μ = true mean DO level in the stream (in mg/l)

Statistic $\bar{x} = 4.771$

Hypothesis H_0: $\mu = 5$

H_a: $\mu < 5$

α $\alpha = 0.05$

Plan [Name the procedure and Check all conditions]

Name:

Random:

10%:

Normal/Large Sample:

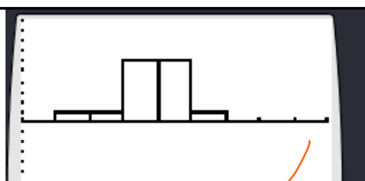
Plan [Name the procedure and Check all conditions]

One sample t Test for μ

Random: Researcher selected 15 random locations in stream

10%: there are an infinite number of locations in the stream so it is not necessary to check 10% condition.

Normal/Large Sample: Need to look at distribution (histogram, dot plot)



looks roughly symmetric with no strong outliers

Do [Give the sample statistic(s), Calculate the standardized test statistic, Find the P-Value]

For test statistic: General Formula, Specific Formula, then with numbers, then final answer

$\bar{x} = 4.771$

$S_x =$

GDC

Do [Give the sample statistic(s), Calculate the standardized test statistic, Find the P-Value]

For test statistic: General Formula, Specific Formula, then with numbers, then final answer

$\bar{x} = 4.771$

$S_x = 0.9396$

$$\text{Test Stat.} = \frac{\text{stat} - \text{param}}{SD}$$

$$t = \frac{\bar{x} - \mu}{S_x / \sqrt{n}}$$

$$t = \frac{4.771 - 5}{.9396 / \sqrt{15}}$$

$$t = -0.94$$

Do [Give the sample statistic(s), Calculate the standardized test statistic, Find the P-Value]

For test statistic: General Formula, Specific Formula, then with numbers, then final answer

$\bar{x} = 4.771$

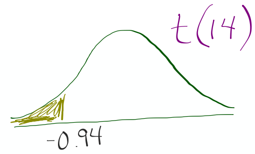
$S_x = 0.9396$

$$\text{Test Stat.} = \frac{\text{Stat} - \text{param}}{SD}$$

$$t = \frac{\bar{x} - \mu}{S_x/\sqrt{n}}$$

$$t = \frac{4.771 - 5}{.9396/\sqrt{15}}$$

$$t = -0.94$$



P-Value

Table B Between 0.15 and .20

or

P-Value = $t_{cdf}(-1000, -.94, 14)$

~~= 0.186~~

0.182

Because the P-Value range of 0.15 to .20 is > .05

We fail to reject H_0.

or...

Alternative

$t_{cdf}(\underset{\text{Lower}}{-1000}, \underset{\text{Upper}}{-.94}, \underset{\text{df}}{14}) = 0.186$

Show

Conclude [Make a conclusion about the hypothesis in the context of the problem, two-sentence structure]

Because the P-Value range of 0.15 to .20 is $> .05$ we fail to reject H_A.

$\therefore$ We don't have convincing evidence that the true mean DO level in the stream is less than 5 mg/l

Given your conclusion in part (a), which kind of mistake—a Type I error or a Type II error—could you have made? Explain what this mistake would mean in context.

Given your conclusion in part (a), which kind of mistake—a Type I error or a Type II error—could you have made? Explain what this mistake would mean in context.

Because we failed to reject H_0, we could have made a Type II error (failing to reject when H_a is true)

Given your conclusion in part (a), which kind of mistake—a Type I error or a Type II error—could you have made? Explain what this mistake would mean in context.

Because we failed to reject H_0, we could have made a Type II error (failing to reject when H_a is true)

If we did, then the true DO level μ is less than 5 mg/l but we didn't find convincing evidence.

Given your conclusion in part (a), which kind of mistake—a Type I error or a Type II error—could you have made? Explain what this mistake would mean in context.

Because we failed to reject H_0, we could have made a Type II error (failing to reject when H_a is true)

If we did, then the true DO level μ is less than 5 mg/l but we didn't find convincing evidence.

That would imply aquatic life in this stream is at risk but we were not able to detect that.

Given your conclusion in part (a), which kind of mistake—a Type I error or a Type II error—could you have made? Explain what this mistake would mean in context.

Because we failed to reject H_0, we could have made a Type II error (failing to reject when H_a is true)

If we did, then the true DO level μ is less than 5 mg/l but we didn't find convincing evidence.

That would imply aqualic life in this stream is at risk but we were not able to detect that.

Context

Computer Output - will do t-test calculations

Session

One-Sample T: DO (mg/L)

Test of mu = 5 vs < 5

Variable	N	Mean	StDev	SE Mean	T	P
DO (mg/L)	15	4.771	0.940	0.243	-0.94	0.181

Read top of p.594

Session

One-Sample T: DO (mg/L)

Test of mu = 5 vs < 5

Variable	N	Mean	StDev	SE Mean	T	P
DO (mg/L)	15	4.771	0.940	0.243	-0.94	0.181

$$SE = \frac{S_x}{\sqrt{n}} = \frac{.940}{\sqrt{15}} = .243$$

$$t = \frac{4.771 - 5}{.243}$$

AP Exam Tip

another risk in using the Calculator for the "DO" step is:

LOSS OF UNDERSTANDING

- UNDERSTANDING THAT IS Required ON M/C QUESTIONS AND CAREFULLY CRAFTED FR QUESTIONS

9.3......65, 67, 69, 73, 77, 79 and <u>109</u>

study pp. 585-594

read about →

$$\underbrace{t}_{2} = \frac{\overbrace{\bar{x}}^{1} - \mu_0}{\underbrace{\frac{s_x}{\sqrt{n}}}_{3}}$$

There are three conditions that must be met for the formula for the standardized test statistic to be valid—one for each of the three components in the formula.

$$\underbrace{t}_{2} = \frac{\overbrace{\bar{x}}^{1} - \mu_0}{\underbrace{\frac{s_x}{\sqrt{n}}}_{3}}$$

There are three conditions that must be met for the formula for the standardized test statistic to be valid—one for each of the three components in the formula.

$$t = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}}$$

1. The Random condition

This condition helps ensure that $\bar{x} - \mu_0$ is a good estimate for the difference between the true value of μ and the null value of μ_0.

There are three conditions that must be met for the formula for the standardized test statistic to be valid—one for each of the three components in the formula.

$$t = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}}$$

2. The Normal/Large Sample condition

This condition allows us to model the distribution of the standardized test statistic *t* using a *t* distribution with n – 1 degrees of freedom.

There are three conditions that must be met for the formula for the standardized test statistic to be valid—one for each of the three components in the formula.

$$t = \frac{\bar{x} - \mu_0}{\frac{s_x}{\sqrt{n}}}$$

3. The 10% condition

This condition allows us to use the familiar formula for the standard deviation of the sampling distribution of $\bar{x}$ (with s_x replacing σ) when we are sampling without replacement from a finite population.