

# SECTION 9.3 Exercises

NAME:

Need 2 Yellow "TEST OF SIGNIFICANCE TEMPLATES"

71. Sweetening colas Cola makers test new recipes for loss of sweetness during storage. Trained tasters rate the sweetness before and after storage. From experience, the population distribution of sweetness losses will be close to Normal. Here are the sweetness losses (sweetness before storage minus sweetness after storage) found by tasters from a random sample of 10 batches of a new cola recipe:

2.0 0.4 0.7 2.0 -0.4 2.2 -1.3 1.2 1.1 2.3

Are these data good evidence that the cola lost sweetness? Carry out a test to help you answer this question.

71

See template

73. Healthy bones The recommended daily allowance (RDA) of calcium for women between the ages of 18 and 24 years is 1200 milligrams (mg). Researchers who were involved in a large-scale study of women's bone health suspected that their participants had significantly lower calcium intakes than the RDA. To test this suspicion, the researchers measured the daily calcium intake of a random sample of 36 women from the study who fell in the desired age range. The Minitab output below displays descriptive statistics for these data, along with the results of a significance test.

## Descriptive Statistics: Calcium intake (mg)

Variable	N	Mean	SE Mean	StDev	Min	Q1	Med	Q3	Maximum
Calcium	36	856.2	51.1	306.7	374.0	632.3	805.0	1090.5	1425.0

## One-Sample T: Calcium intake (mg)

Test of  $\mu = 1200$  vs  $< 1200$

Variable	N	Mean	StDev	SE Mean	T	P
Calcium	36	856.2	306.7	51.1	-6.73	0.000

(a) Determine whether there are any outliers. Show your work.

(b) Interpret the P-value in context.

(c) Do these data give convincing evidence to support the researchers' suspicion? Carry out a test to help you answer this question.

73A Show work here

73B+C: See template

$$IQR = Q3 - Q1 = 1090.5 - 632.3 = 458.2$$

$$Q3 + 1.5 IQR = 1090.5 + 1.5(458.2) = 1777.8 > \max = 1425 \text{ (No outlier)}$$

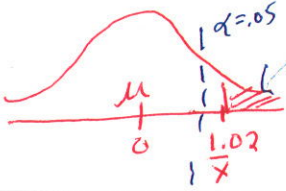
$$Q1 - 1.5 IQR = 632.3 - 1.5(458.2) = -55 < \min = 374.0 \text{ (No outlier)}$$

7.3B The output shows a p-value = 0.000. IF THE MEAN DAILY CALCIUM INTAKE FOR women 18 to 24 IS REALLY 1200 mg, then the likelihood of getting a sample of 36 women with a mean intake of 856.2 mg or smaller is ROUGHLY 0.

OVER →

#71

## Test of Significance Template

Parameter of Interest	$\mu$ = actual mean amount of sweetness <sup>loss</sup> (sweetness before storage minus sweetness after storage)	
Choice of Test	1 SAMPLE TTEST FOR $\mu$	
Level of Significance	$\alpha = .05$ Since $\alpha$ was not given	
Null Hypothesis	English: Symbols: $H_0: \mu = 0$	
Alternative Hypothesis	English: Symbols: $H_a: \mu > 0$	
Conditions of Test	① $\sigma$ IS UNKNOWN (T inference) ② Normal - Previous experience, population distribution is Normal ③ Random sample of 10 batches ④ Independent - there are at least 10(10) = 100 batches of the new Soda available.	
Sampling Distribution	Sketch of the sampling distribution of the sample statistic under the null hypothesis, indicating the mean: 	
Test Statistic	Formula: $t = \frac{\bar{x} - \mu}{s_x / \sqrt{n}}$	Plug-ins & Value: $\mu = 0$ $s_x = 1.196$ $t = \frac{1.02 - 0}{1.196 / \sqrt{10}} = \frac{1.02}{.3782} = 2.70$ $\bar{x} = 1.02$ $n = 10$ $d = 9$
P-value	Use correct probability notation. $P(t \geq 2.70) = \text{tcdf}(2.70, E99, 9) = .0122$	
Meaning of the P-value	Since $p = .0122 < \alpha = .05$ , Reject $H_0$	
Conclusions	<input checked="" type="checkbox"/> Reject null hypothesis <input type="checkbox"/> Significant result <input type="checkbox"/> Fail to reject null hypothesis <input type="checkbox"/> Not Significant result English: Since the p-value is less than the .05 significance level, we Reject $H_0$ . It appears that there is an average loss of sweetness for this cola.	

$t = 2.70$   
 $p = .0122$

Data:  
 $\mu_0 = 0$   
 $L1$   
 $> \mu_0$

STAT  
 Ttest  
 2: Ttest

CALC

ENTER DATA INT L1



#73

PARTS B+C

## Test of Significance Template

Parameter of Interest	$\mu$ = the actual mean daily calcium intake of women 18-24	
Choice of Test	1 SAMPLE T TEST FOR $\mu$	
Level of Significance	$\alpha = .05$ (since not given)	
Null Hypothesis	English:	
Alternative Hypothesis	English:	
Conditions of Test	Symbols: $H_0: \mu = 1200 \text{ mg}$ Symbols: $H_A: \mu < 1200 \text{ mg}$	
Sampling Distribution	① $\sigma$ IS UNKNOWN (T INFERENCE) ② Random sample of 36 women ③ Normal - the sample was large enough $n = 36 > 30$ ④ Independent - there are clearly more than 360 ( $36 \cdot 10$ ) women in the U.S.	
Test Statistic	Formula:	Plug-ins & Value:
P-value	$t = \frac{\bar{x} - \mu}{s_x / \sqrt{n}}$ $\mu = 1200 \quad \bar{x} = 856.2 \quad n = 36 \quad s_x = 306.7 \quad t = \frac{856.2 - 1200}{306.7 / \sqrt{36}} = -6.73$	
Meaning of the P-value	Use correct probability notation. $P(t \leq -6.73) = \text{tcdf}(-1.899, -6.73, 35) \approx 0$	
Conclusions	The p-value is extremely small (about 0) so Reject $H_0$ <input checked="" type="checkbox"/> Reject null hypothesis <input type="checkbox"/> Significant result <input type="checkbox"/> Fail to reject null hypothesis <input type="checkbox"/> Not Significant result English: Since p-value is extremely small, we Reject $H_0$ . It appears that women in this age group are getting less than 1200mg calcium daily, on average.	

**75** **Growing tomatoes** An agricultural field trial compares the yield of two varieties of tomatoes for commercial use. Researchers randomly select 10 Variety A and 10 Variety B tomato plants. Then the researchers divide in half each of 10 small plots of land in different locations. For each plot, a coin toss determines which half of the plot gets a Variety A plant; a Variety B plant goes in the other half. After harvest, they compare the yield in pounds for the plants at each location. The 10 differences (Variety A - Variety B) give  $\bar{x} = 0.34$  and  $s_x = 0.83$ . A graph of the differences looks roughly symmetric and single-peaked with no outliers. Is there convincing evidence that Variety A has the higher mean yield? Perform a significance test using  $\alpha = 0.05$  to answer the question.

**175** → SEE TEMPLATE

**77** **The power of tomatoes** The researchers who carried out the experiment in Exercise 75 suspect that the large P-value (0.114) is due to low power.

(a) Describe a Type I and a Type II error in this setting. Which type of error could you have made in Exercise 75? Why?

(b) Explain two ways that the researchers could have increased the power of the test to detect  $\mu = 0.5$ .

Ⓐ TYPE I ERROR: Experts conclude that Variety A has a higher mean yield when it actually doesn't

TYPE II ERROR: EXPERTS conclude that there is no mean difference in yield when in fact Variety A has a higher mean yield

We could have made a type II error since we failed to reject  $H_0$



Ⓑ 2 ways to increase power

① increase the significance level ( $\alpha$ )

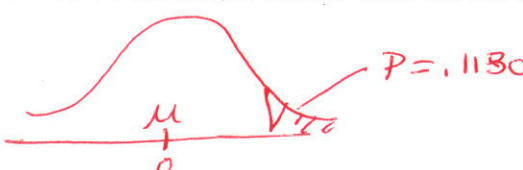
② increase the sample size ( $n$ )



HW

#75

## Test of Significance Template

Parameter of Interest	$\mu$ = the true mean difference in yield between Variety A + B tomato plants	
Choice of Test	one sample t-test for $\mu$	
Level of Significance	$\alpha = .05$	
Null Hypothesis	English: Symbols: $H_0: \mu = 0$	
Alternative Hypothesis	English: Symbols: $H_A: \mu > 0$	
Conditions of Test	① Random - There was random assignment ② $\sigma$ is unknown ( $t$ inference) ③ Independent - There are more than 100 of each variety of plants ④ Normal - Graphs were done and there were no outliers and they were roughly symmetric	
Sampling Distribution	Sketch of the sampling distribution of the sample statistic under the null hypothesis, indicating the mean: 	
Test Statistic	Formula: $t = \frac{\bar{x} - \mu}{s_x / \sqrt{n}}$	Plug-ins & Value: $\mu = 0$ $n = 10$ $\bar{x} = .34$ $df = 9$ $s_x = .83$ $t = \frac{.34 - 0}{.83 / \sqrt{10}} = \frac{.34}{.2625} = 1.30$
P-value	Use correct probability notation. $P = P(t \geq 1.30) = \text{tcdf}(1.3, E99, 9) = .1130$	
Meaning of the P-value	Since the p-value is large and greater than $\alpha$ , FAIL TO REJECT $H_0$ $.1130 > .05$	
Conclusions	English: Since the p-value is larger than $\alpha = .05$ , we FAIL to Reject $H_0$ . We do not have enough evidence to conclude that Variety A has a higher mean yield than Variety B.	

 $t = 1.295$   
 $P = .1137$ 
 $s_x = .83$   
 $n = 10$   
 $\bar{x} > \mu_0$ 
 $\mu_0 = 0$   
 $\bar{x} = .34$ 

T Test

STAT

Tests

# SECTION 9.3 Exercises

NAME

KEY

Need 1 Yellow Test Template

89. Right versus left The design of controls and instruments affects how easily people can use them. A student project investigated this effect by asking 25 right-handed students to turn a knob (with their right hands) that moved an indicator. There were two identical instruments, one with a right-hand thread (the knob turns clockwise) and the other with a left-hand thread (the knob must be turned counterclockwise). Each of the 25 students used both instruments in a random order. The following table gives the times in seconds each subject took to move the indicator a fixed distance:<sup>30</sup>

Subject	Right thread	Left thread
20	89	93
21	78	76
22	100	116
23	89	78
24	85	101
25	88	123

- (a) Explain why it was important to randomly assign the order in which each subject used the two knobs.

(answer here ↓)

IT IS IMPORTANT TO RANDOMLY ASSIGN SO THAT WE AVERAGE OUT ANY EFFECT DUE TO DOING THE ACTIVITY BETTER THE SECOND TIME NO MATTER WHICH KNOB IS USED SECOND.

Subject	<u>L1</u> Right thread	<u>L2</u> Left thread
1	113	137
2	105	105
3	130	133
4	101	108
5	138	115
6	118	170
7	87	103
8	116	145
9	75	78
10	96	107
11	122	84
12	103	148
13	116	147
14	107	87
15	118	166
16	103	146
17	111	123
18	104	135
19	111	112

$L3 = L2 - L1$  (Left - Right)

Calculator Tip

STAT TESTS  
2: T-Test

DATA  
 $\mu_0 = 0$   
L3  
FREQ: 1

CALCULATE

$\mu > 0$   
 $t = 2.904$   
 $p = .0038$

NOTE: You would get the same conclusion if you did Right-Left. You would just change  $H_A$ .

$H_0: \mu_d = 0$   
 $H_A: \mu_d < 0$   
 $t = -2.904$   
 $p = .0039$

TRY THIS:  $H_0: \mu_d = 0$   $H_A: \mu_d \neq 0$

$t = -2.904$   $p = .0078$   $\therefore$  STILL REJECT  $H_0$

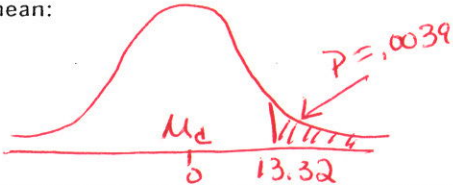
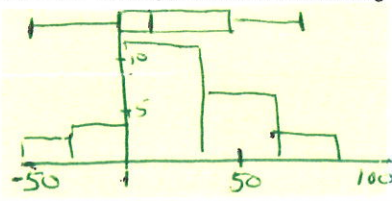
- (b) The project designers hoped to show that right-handed people find right-hand threads easier to use. Carry out a significance test at the 5% significance level to investigate this claim.

$\bar{x} = 13.32$   
 $s_x = 22.93$   
 $n = 25$

Complete Test Template



## Test of Significance Template

Parameter of Interest	$\mu_d$ = actual mean difference (left-right) in the time it takes to turn the knob with left thread and right thread	
Choice of Test	PAIRED T-Test for $\mu$	
Level of Significance	$\alpha = .05$ since no $\alpha$ was given	
Null Hypothesis	English: RIGHT HANDED SAMPLE Symbols: $H_0: \mu_d = 0$ seconds	
Alternative Hypothesis	English: Does it take longer to turn knob <u>left</u> than <u>right</u> . Symbols: $H_A: \mu_d > 0$ seconds	
Conditions of Test	① $\sigma$ is unknown (t-inference) ② THIS IS A RANDOMIZED EXPERIMENT. ③ INDEPENDENT: we aren't sampling (10%). The difference in times for individuals/subjects should be independent if the experiment is conducted properly. ④ Normal - Small sample ( $n = 25 < 30$ ). A HISTOGRAM LOOKS FAIRLY SYMMETRIC AND A BOX PLOT SHOWS NO OUTLIERS. SEE GRAPHS BELOW.	
Sampling Distribution	Sketch of the sampling distribution of the sample statistic under the null hypothesis, indicating the mean:  	
Test Statistic	Formula: $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ Plug-ins & Value: $\mu = 0$ $\bar{x} = 13.32$ $t = \frac{13.32 - 0}{22.936/\sqrt{25}}$ $n = 25$ $s_x = 22.936$ $df = 24$	$t = 2.904$
P-value	Use correct probability notation. $p = P(t \geq 2.904) = \text{tcdf}(2.904, E99, 24) = .0039$	
Meaning of the P-value	$P$ is very small so we Reject $H_0$	
Conclusions	<input checked="" type="checkbox"/> Reject null hypothesis <input type="checkbox"/> Significant result <input type="checkbox"/> Fail to reject null hypothesis <input type="checkbox"/> Not Significant result English: Since our $p$ -value $< .05$ , we reject $H_0$ . We have enough evidence to conclude that it takes longer for right handed students to complete the task when the knob has a left hand thread, on average.	

94. Significance and sample size A study with 5000 subjects reported a result that was statistically significant at the 5% level. Explain why this result might not be particularly large or important.

#'s 94-97 answer here

The study may have rejected  $H_0$ , But with such a large sample size, such a rejection might occur even if the actual differs only slightly from the hypothesized value. For example, the difference between  $\mu=10$  and  $\mu=10.5$  might have no practical importance.

95. Sampling shoppers A marketing consultant observes 50 consecutive shoppers at a supermarket, recording how much each shopper spends in the store. Explain why it would not be wise to use these data to carry out a significance test about the mean amount spent by all shoppers at this supermarket.

ANY NUMBER OF THINGS COULD GO WRONG WITH A CONVENIENCE SAMPLE. DEPENDING ON THE TIME OF DAY OR THE DAY OF THE WEEK, CERTAIN TYPES OF SHOPPERS WOULD OR WOULD NOT BE PRESENT.

Remember! THE ONLY WAY TO SHOW CAUSE AND EFFECT IS WITH A WELL-DESIGNED, WELL-CONTROLLED EXPERIMENT!

3 COMPONENTS ① Randomization ② CONTROL ③ REPLICATION

96. Ages of presidents Joe is writing a report on the backgrounds of American presidents. He looks up the ages of all the presidents when they entered office. Because Joe took a statistics course, he uses these numbers to perform a significance test about the mean age of all U.S. presidents. Explain why this makes no sense.

We have information about the whole population of interest.

97. Do you have ESP? A researcher looking for evidence of extrasensory perception (ESP) tests 500 subjects. Four of these subjects do significantly better ( $P < 0.01$ ) than random guessing.

(a) Is it proper to conclude that these four people have ESP? Explain your answer.

(b) What should the researcher now do to test whether any of these four subjects have ESP?

<sup>would</sup>  
① No we expect about 5 of the 500 subjects who don't have ESP to do better than randomly guessing just by chance.  
 $500(.01) = 5$

② The researcher should repeat the procedure on these 4 to see if they again perform well