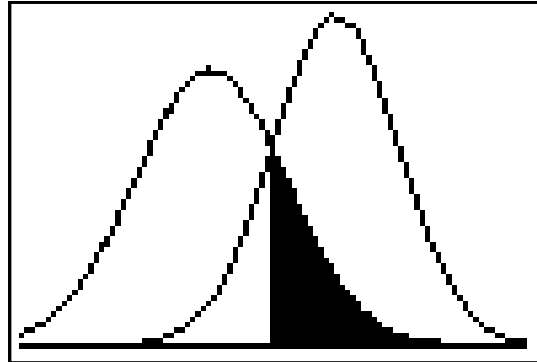


CHAPTER

17



Inference about a Population Mean

17.1	Inference for the Mean of a Population
17.2	Paired Samples
17.3	Selected Exercise Solutions

Introduction

In this chapter, we demonstrate the various t procedures that are used for confidence intervals and significance tests about the mean of a normal population for which the *population* standard deviation is unknown. We also consider comparing means from paired samples, since the inference is based on the differences of the individual observations.

17.1 Inference for the Mean of a Population

We begin with a short program that allows us to find a critical value t^* upon specifying the degrees of freedom and confidence level. This is a built-in function on TI-84 and TI-89 calculators. TI-83s need this program (or you could use Table C). This program asks for degrees of freedom rather than sample size, since t distributions are used in inference for more than just a single sample.

The TSCORE Program

PROGRAM:TSCORE	: "tcdf(0,X,M) "[STO]Y1
:Disp "DEG. OF FREEDOM"	:solve(Y1-R/2,X,2) [STO]Q
:Input M	:Disp "T SCORE"
:Disp "CONF. LEVEL"	:Disp round(Q,3)
:Input R	

Example 17.1 Finding t^* For a t distribution with 9 degrees of freedom (df), what point on this distribution has probability 0.05 to its right?

Solution. With 0.05 to the right of the desired point, there is 0.95 to the left. Program **TSCORE** is designed to find two-sided t^* values for confidence intervals. For that program, we'll ask for 90% confidence since that area in the center puts 5% on each tail. Below are the outputs of the **TSCORE** program and from the built-in functions. The correct t^* is 1.833.

```
DEG OF FREEDOM
?9
CONF LEVEL
?.9
T SCORE
1.833
Done
```

```
invT(.95,9)
1.833112923
```

One-sample t Confidence Interval

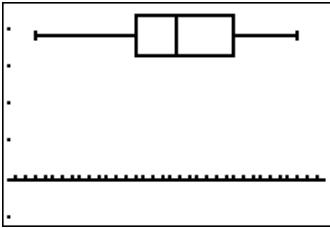
We now examine confidence intervals for one mean for which we will use the **TInterval** feature (item 8) from the **STAT TESTS** menu ([F7] **Ints** on a TI-89). As with the **ZInterval** feature that we used in Chapter 14, we can enter the summary statistics or use data in a list.

Example 17.2 Healing of skin wounds. In Example 14.3, we looked at a study of the healing rate of skin wounds in newts. Repeated below are the healing rates (in

micrometers per hour) after a razor cut was made in the skin of an anesthetized newt. We want a 95% confidence interval for the mean rate μ for all newts of this species.

29 27 34 40 22 28 14 35 26
35 12 30 23 18 11 22 23 33

Solution. In Example 14.3, we assumed the population standard deviation was $\sigma = 8$ micrometers per hour. Here, we'll use the variability of the newts in this study to estimate the variability of the population. Enter the data in a list (say, L1). We first must check the “simple conditions” to determine whether our inference is valid. As stated before, we are willing to view these newts as a random sample from the population of all newts of this species. Since this is a small sample, we need to check for outliers or skewness in our sample. Define a boxplot of the data and examine it. Our boxplot shown below is relatively symmetric around the median and indicates no outliers, so we may proceed.



```
TInterval
Inpt: Data Stats
List: L1
Freq: 1
C-Level: .95
Calculate
```

```
TInterval
(21.527, 29.806)
x̄=25.66666667
Sx=8.324308839
n=18
```

We use `8:TInterval` from the `STAT TESTS` menu to find the interval, using the `Data` option. Using t procedures, we'll say we are 95% confident the mean healing rate for this species of newt is between 21.5 and 29.8 micrometers per hour. Note that the standard deviation of this sample is a bit bigger than the 8 micrometers per hour assumed in Example 14.3, where we obtained a confidence interval of 22.0 and 29.4 micrometers per hour.

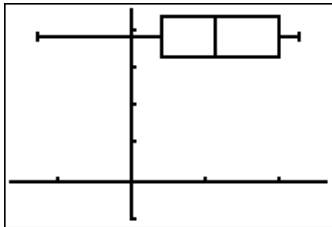
One-sample t test

We now perform significance tests about the mean using the `T-Test` feature (item 2) from the `STAT TESTS` menu.

Example 17.3 Sweetening colas. A more realistic analysis of the cola-sweetening example from Chapter 14 would use the standard deviation of the actual data to test the hypotheses $H_0 : \mu = 0$ (no sweetness loss after storage) against $H_a : \mu > 0$ (there is a loss of sweetness, on average). The data are repeated below for convenience.

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

Solution. With the data in L1, check a boxplot for skewness and outliers. This boxplot is skewed left; with a sample size of only 10, we'll have to use caution in interpreting our results (the p -value may not be accurate). Bring up the T-Test screen from the STAT TESTS menu and adjust Inpt to Data. Enter the value of $\mu_0 = 0$ and the list name. Select the alternative $> \mu_0$, then scroll down to Calculate and press **ENTER**.



```
T-Test
Inpt:DATA Stats
μ₀:0
List:L₁
Freq:1
μ:≠μ₀ <μ₀ >μ₀
Calculate Draw
```

```
T-Test
μ>0
t=2.69668949
p=.0122631561
x̄=1.02
Sx=1.196104789
n=10
```

We obtain a t test statistic of 2.70 and a p -value of 0.0123. Because the p -value is small, we have significant evidence to reject H_0 . If the true mean were 0, a sample mean of 1.02 or higher with a sample of size 10 should only happen about 1 in 100 times; even with the skew in the data, we should be safe in concluding that this cola *does* lose sweetness with storage, on average.

17.2 Paired Samples

Oftentimes, studies are conducted where there is a natural pairing between observations, such as the before- and after-storage ratings by the cola tasters. Each taster tasted the cola in both situations. In using the differences (which were given there) as the data for the hypothesis test, we are eliminating any variability due to individual taster preferences and focusing on how their ratings changed. There are many situations where this is advantageous.

Example 17.4 Do chimpanzees collaborate? Humans often collaborate to solve problems. Will chimpanzees recruit another chimp when solving a problem requires collaboration? Researchers presented chimpanzee subjects with food outside their cage that they could bring within reach by pulling two ropes, one attached to each end of the food tray. If a chimp pulled only one rope, the rope became loose and the food was lost. Another chimp was available as a partner, but only if the subject unlocked a door joining the two cages. The same 8 chimpanzees faced this problem in two versions: the two ropes were close enough together that one chimp could pull both (no collaboration needed) or the two ropes were too far apart for one chimp to pull both (collaboration needed). The data below show how often in 24 trials each chimp opened the door to recruit another chimp as a partner. Is there evidence that chimpanzees recruit partners more often when a problem requires collaboration?

Chimpanzee	Collaboration Needed?	
	Yes	No
Namuiska	16	0
Kalema	16	1
Okech	23	5
Baluku	19	3
Umugenzi	15	4
Indi	20	9
Bili	24	16
Asega	24	20

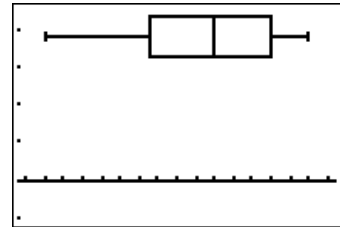
Solution: Since these are paired data (the same chimps under two different circumstances), we need to find the differences to use as data. We have entered the data in L1 (collaboration needed) and L2 (no collaboration needed). Highlight the L3 list name and enter the command shown below to perform the subtraction. The simple conditions here pertain to the differences, not the original data, so define a boxplot using the differences and examine it for skewness or outliers. There are no outliers; it appears that there is some left skew to this distribution, but with only 8 observations, it is not severe.

L1	L2	L3	3
16	0	-----	
16	1		
23	5		
19	3		
15	4		
20	9		
24	16		

L3 = L1 - L2

```

2ND [2] Plot2 Plot3
OFF Off
Type: L1 L2 L3
Freq: 1
Mark: [ ] + .
    
```



We proceed to test hypotheses $H_0 : \mu = 0$ (no difference in asking for help) against $H_a : \mu > 0$ (chimps will ask for help more often when it is needed). *Note that since we found the differences as collaboration needed minus collaboration not needed, this is the correct form of the alternate; if you had done the subtraction in the other order, the alternate would have the opposite sign.* Compute the test statistic and find the p -value using T-Test.

```

T-Test
Inpt: [ ] Stats
μ₀: 0
List: L3
Freq: 1
μ: ≠ μ₀ < μ₀ > μ₀
Calculate Draw
    
```

```

T-Test
μ > 0
t = 7.370255429
P = 7.6628461E-5
x̄ = 12.375
Sx = 4.749060057
n = 8
    
```

Based on this study, with $t = 7.37$ and p -value 7.66×10^{-5} , we reject the null hypothesis. This study indicates strongly the chimpanzees *do* ask for help when it is needed.

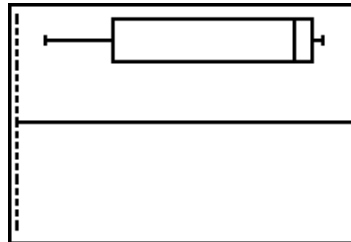
17.3 Selected Exercise Solutions

17.7 The data have been entered in L1. If μ is the mean nitrogen content of Cretaceous era air, we'd like a 90% confidence interval estimate. First, check the conditions: we're assuming our data come from a SRS; can we believe these data came from an (approximately) Normal distribution? With only 9 data values, a histogram will not show the distribution very well. Define and display a boxplot to look for skewness and outliers. There are no outliers, but the distribution is definitely skewed; observe the median far to the right in the box. Use of t procedures might not be valid, we can only proceed with caution.

```

Plot1 Plot2 Plot3
Off Off
Type: L1 L2 L3 L4
Xlist:L1
Freq:1
Mark: + .

```



To create the interval, use option 8:TInterval from the [STAT], TESTS menu. We have data already entered in a list, so the input method is Data. Based on these samples, we estimate that Cretaceous era air had between 55.7% and 63.5% nitrogen, with 90% confidence (assuming the distribution is really approximately Normal).

```

TInterval
Inpt:Data Stats
List:L1
Freq:1
C-Level:.9
Calculate

```

```

TInterval
(55.712,63.466)
x=59.58888889
Sx=6.255286653
n=9

```

17.29 Because each patient had activity measured with no treatment and placebo, this is a matched pairs design. Assuming the differences have a Normal distribution (we don't have the actual data to check), use 1:T-Test from the [STAT], Tests menu to compute the test statistic and P -value for a test of $H_0: \mu_D = 0$ against the two-tailed alternate. With $t = -4.41$ and $P = 0.0069$, we reject the idea that there is no difference between treatment and control. Since the mean difference is negative, this suggests less activity with the treatment.

```

T-Test
Inpt:Data Stats
mu0:0
x: -.326
Sx: .181
n:6
mu:mu0 <mu0 >mu0
Calculate Draw

```

```

T-Test
mu≠0
t=-4.411788156
P=.0069449715
x= -.326
Sx=.181
n=6

```

Continue your practice with these exercises:

- 17.3 Critical values.
- 17.9 Is it significant?
- 17.11 The brain responds to sound.
- 17.13 Diamonds.
- 17.25 Read carefully.
- 17.27 Reading scores in Atlanta.
- 17.31 Learning Blissymbols.
- 17.33 An outlier's effect.
- 17.35 Genetic engineering for cancer treatment.
- 17.37 Growing trees faster.
- 17.39 Weeds among the corn.
- 17.43 How much better does nature heal?
- 17.45 Right versus left.
- 17.47 Practical significance.

Chapter 17 TI Solutions

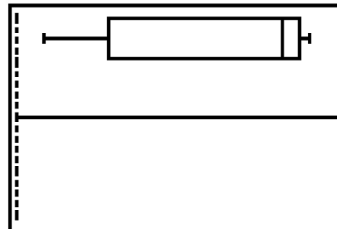
17.3 With a TI-83, use program **TSCORE** to find the critical value. Note that the **TSCORE** program uses the confidence level, not the α level. A one-sided 5% α level corresponds to 90% confidence. With a TI-84 or -89, use **invT** from the **DISTR** menu. Ignoring the negative sign on the **invT** result, we have the same critical value of 2.015. For 98% confidence, there is 1% on either tail. Using **invT**, we find the critical value (again ignoring the negative sign) is 2.518.

```
DEG OF FREEDOM
?5
CONF LEVEL
?.9
T SCORE
2.015
Done
```

```
invT(.05,5)
-2.015048342
invT(.01,21)
-2.517647992
```

17.7 The data have been entered in L1. If μ is the mean nitrogen content of Cretaceous era air, we'd like a 90% confidence interval estimate. First, check the conditions: we're assuming our data come from a SRS; can we believe these data came from an (approximately) Normal distribution? With only 9 data values, a histogram will not show the distribution very well. Define and display a boxplot to look for skewness and outliers. There are no outliers, but the distribution is definitely skewed; observe the median far to the right in the box. Use of t procedures might not be valid, we can only proceed with caution.

```
Plot1 Plot2 Plot3
On Off
Type: L1 L2 L3
Xlist:L1
Freq:1
Mark: [ ] +
```



To create the interval, use option 8: **TInterval** from the **[STAT]**, **TESTS** menu. We have data already entered in a list, so the input method is **Data**. Based on these samples, we estimate that Cretaceous era air had between 55.7% and 63.5% nitrogen, with 90% confidence (assuming the distribution is really approximately Normal).

```
TInterval
Inpt: DATA Stats
List:L1
Freq:1
C-Level: .9
Calculate
```

```
TInterval
(55.712,63.466)
x=59.58888889
Sx=6.255286653
n=9
```

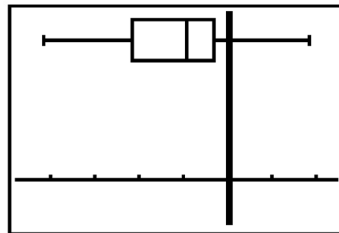
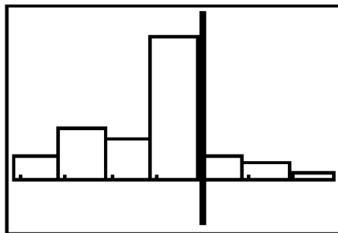

17.9 With $n = 25$, there are $25 - 1 = 24$ degrees of freedom. Using technology, we'll find an exact P -value for this test. Since the test is two-tailed, we'll double the area to the right of $t = 1.12$. The parameters for `tcdf` on the `DISTR` menu are `low,high,df`. With a P -value of 0.2738, this test is not significant at any normal alpha level.

```
tcdf(1.12,99999,
24)
Ans*2
.2737972632
```

17.11 This is a matched pairs situation (we have the response to both tone and call for each neuron), so we'll work with the differences. The tone responses have been entered in `L1` and the call responses in `L2`. We form the differences by highlighting the `L3` name and entering the subtraction command. We should next check to see if the distribution of the differences is approximately Normal (although with a sample of $n = 37$, t procedures will be robust). The `ZOOM` [9] histogram looks reasonably Normal, and our boxplot looks symmetric (with no outliers), so t procedures are justified.

L1	L2	<code>□</code>	3
474	500		-----
256	138		
241	485		
226	338		
185	194		
174	159		
176	341		

`L3 = L1 - L2`



The question of interest is whether the call response is stronger than the tone response, so bearing in mind that our differences are $D = \text{Tone} - \text{Call}$, we will test the hypotheses

$$H_0: \mu_D = 0$$

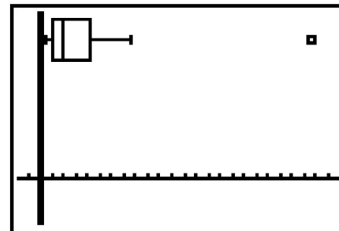
$$H_a: \mu_D < 0$$

(if you subtract the other way, the direction of the alternate hypothesis will change). We use option 2: `T-Test` from the `STAT`, `Tests` menu to perform the mechanics. With a test statistic of $t = -4.84$ and P -value of 0.00001, this is extremely strong evidence that the call response is stronger than the tone response, on average, in macaque monkeys.

```
T-Test
Inpt: DATA Stats
 $\mu_0$ : 0
List: L3
Freq: 1
 $\mu$ :  $\neq \mu_0$  < \mu_0 > \mu_0
Calculate Draw
```

```
T-Test
 $\mu < 0$ 
t = -4.840152879
p = 1.2231168E-5
 $\bar{x}$  = -70.37837838
Sx = 88.44657878
n = 37
```

17.13 The nitrogen data have been entered in `L1`. We'll create a (modified) boxplot to examine the shape of this distribution. This data set has an extreme outlier; further, the main portion seems rather skewed right. We can't trust t procedures for these data.



17.25 We compute the correct t statistic,

$$t = \frac{\bar{x} - 0}{s/\sqrt{n}}$$

for each group. For the student group, we have $t = 0.749$ with P -value 0.470 (the conclusion was correct, however, there is no significant effect here). For the non-student group, we have $t = 3.277$ with P -value 0.0074; there is a significant effect in this group.

```
.08/ (.37/√(12))
      .7489949438
2*tcdf(.749,9999
9,11)
      .4695712459
```

```
.35/ (.37/√(12))
      3.276852879
2*tcdf(3.277,999
999,11)
      .0073729472
```

17.27 With a sample size of $n = 1470$, this is certainly large enough to appeal to the Central Limit Theorem, and call \bar{x} approximately Normal. To find the confidence interval, we'll first solve for the standard deviation (we were given standard error). We use 8:TInterval from the [STAT], Tests menu to compute the interval as shown below. Based on this information, we're 99% confident Atlanta eighth-graders should have a mean TUDA score between 237.2 and 242.8. Since the high end of this interval is below 243 (basic), our indications are that Atlanta eighth-graders, on average, perform below this level.

```
1.1*√(1470)
      42.17463693
```

```
TInterval
Inpt:Data state
x̄:240
Sx:42.175
n:1470
C-Level:.99
Calculate
```

```
TInterval
(237.16,242.84)
x̄=240
Sx=42.175
n=1470
```

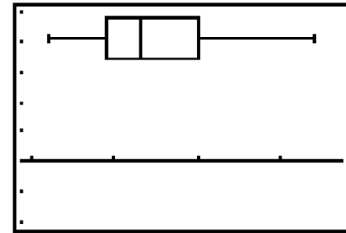
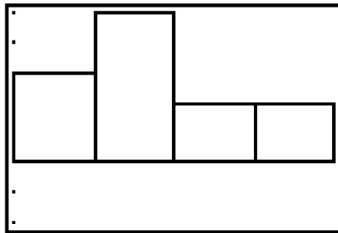
17.29 Because each patient had activity measured with no treatment and placebo, this is a matched pairs design. Assuming the differences have a Normal distribution (we don't have the actual data to check), use 1:T-Test from the [STAT], Tests menu to compute the test statistic and P -value for a test of $H_0: \mu_D = 0$ against the two-tailed alternate. With $t = -4.41$ and $P = 0.0069$, we reject the idea that there is no difference between treatment and control. Since the mean difference is negative, this suggests less activity with the treatment.

```
T-Test
Inpt:Data  Stats
μ₀:0
x̄: -.326
Sx: .181
n:6
μ: μ₀ < μ₀ > μ₀
Calculate Draw
```

```
T-Test
μ ≠ 0
t = -4.411788156
p = .0069449715
x̄ = -.326
Sx = .181
n = 6
```

17.31 TI Calculators can't make stemplots, but we can create a histogram to duplicate the shape we'd see in one. With the data entered in L1, define a histogram and use the **WINDOW** settings shown below. Be sure to press **GRAPH** instead of **ZOOM**[9] to display the plot. The histogram shows no overt skewness nor outliers. Keeping in mind this is a small sample, we can check the (modified) boxplot as well. It shows the right side of the distribution somewhat longer than the left, but again no indication of outliers.

```
WINDOW
Xmin=5
Xmax=25
Xscl=5
Ymin=-2
Ymax=5
Yscl=1
Xres=1
```



Use **TInterval** to compute the desired confidence interval. We're 90% confident the average count of correct Blissymbols among children using this program will be between 10.4 and 15.2.

```
TInterval
Inpt: Data Stats
List: L1
Freq: 1
C-Level: .9
Calculate
```

```
TInterval
(10.424, 15.243)
x̄ = 12.83333333
Sx = 4.648231987
n = 12
```

17.33 The parameter of interest is μ_D , the mean difference between the experimental and control limbs. We want to know whether the electrical field slows healing, so forming differences as Control – Experimental, we have hypotheses $H_0: \mu_D = 0$ and $H_a: \mu_D > 0$. Enter the given values in L1 and L2. We calculate the differences into L3 by highlighting the list name and entering the command L2-L1 and pressing **ENTER**. Examine a boxplot of the differences in L3 (TI calculators do not create stemplots). Note we have a high (extreme) outlier as well as a low one. Now, use **T-Test** from the **STAT TESTS** menu, using the differences in L3 as the data. With a P -value of 0.0310, we

conclude at the 5% significance level that the electrical field does slow healing, on average. Now, delete the high outlier (31) from L3 and recalculate the test. We now have a P -value of 0.0520; this is insufficient evidence at the 0.05 level that the electrical field (or lack thereof) made a difference in average healing rates.

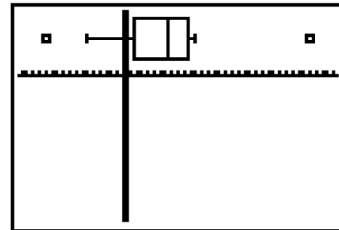
L1	L2	L3	3
28	36		
31	41		
32	39		
33	42		
33	44		
38	38		
45	39		

L3 = L2 - L1

```

10001 Plot2 Plot3
Off Off
Type: L1 L2 L3
Xlist: L3
Freq: 1
Mark: +

```



```

T-Test
Inpt: DATA Stats
μ₀: 0
List: L3
Freq: 1
μ: ≠ μ₀ < μ₀ > μ₀
Calculate Draw

```

```

T-Test
μ > 0
t = 2.076123824
P = .0310563331
x̄ = 6.416666667
Sx = 10.70648345
n = 12

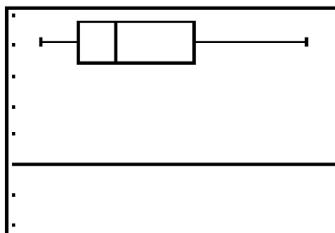
```

```

T-Test
μ > 0
t = 1.788110899
P = .0520231256
x̄ = 4.181818182
Sx = 7.756522182
n = 11

```

17.35 With the data entered in L1, define a boxplot to examine the shape of the distribution. The distribution is rather skewed right, but there are no outliers; further, all the data values are reasonably close to one another. We'll use option 8: TInterval from the [STAT], TESTS menu to do the computations. Based on this sample, the mean doubling time is between 0.92 and 1.42 days; however, based on the shape of the data distribution, we'd hesitate to use this for inference about all possible similar patients.



```

TInterval
Inpt: DATA Stats
List: L1
Freq: 1
C-Level: .9
Calculate

```

```

TInterval
(.921, 1.4245)
x̄ = 1.172727273
Sx = .4606319769
n = 11

```

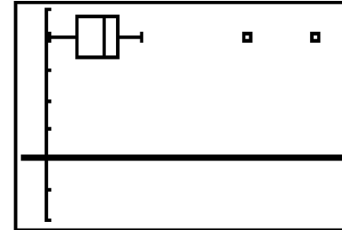
17.37 Because the investigators believed that extra CO_2 would cause the trees to grow faster, the hypotheses are $H_0: \mu_D = 0$ and $H_a: \mu_D > 0$, where μ_D is the mean difference, treatment - control. We enter the Treatment values in L1 and the Control in L2, then form the differences in L3. Use T-Test to perform the mechanics. The test statistic is $t = 3.16$ with P -value 0.0436. While significant at the 5% level, a sample of only $n = 3$ is not very convincing, and risky because we do not have a good idea of the real variation that might occur.

L1	L2	3
10.587	8.752	-----
9.244	7.263	
8.675	5.742	
L3 = L1 - L2		

```
T-Test
Inpt: DATA Stats
μ₀: 0
List: L3
Freq: 1
μ ≠ μ₀ < μ₀ > μ₀
Calculate Draw
```

```
T-Test
μ > 0
t = 3.159644096
p = .0436279547
x̄ = 1.916333333
Sx = 1.050493852
n = 3
```

17.39 Use of *t* procedures with these data is not safe; we have a sample of $n = 28$ plants. The boxplot shown at right clearly identifies two large outliers.



17.43 We entered the experimental values in L1 and the control in L2, then form the differences in L3. Use *TInterval* to perform the mechanics. Based on these data, we are 90% confident the difference in mean healing rate will be between 0.71 and 10.71. This indicates the control limb has the faster healing rate.

L1	L2	3
24	25	-----
23	13	
47	44	
42	45	
26	57	
46	42	
38	50	
L3 = L2 - L1		

```
TInterval
Inpt: DATA Stats
List: L3
Freq: 1
C-Level: .9
Calculate
```

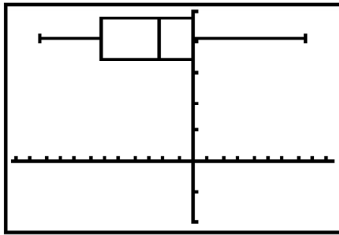
```
TInterval
(.71418, 10.714)
x̄ = 5.714285714
Sx = 10.56429816
n = 14
```

17.45 Since all subjects will use both instruments, we'll flip a coin for each to see which had is used first. We entered the right thread values in L1 and the left thread in L2, then form the differences (right – left) in L3. Since we believe the right hand times should be faster, we will test

$$H_0 : \mu_D = 0$$

$$H_a : \mu_D < 0$$

(if you subtract the other way, the direction of the alternate hypothesis will change). We first examine the shape of the distribution of differences for skewness and outliers by looking at a (modified) boxplot. This distribution is very symmetric, with no outliers; *t* procedures are justified. With a test statistic of $t = -2.90$ and *P*-value 0.0049, we reject the null hypothesis of no difference. This experiment does show that people find right-hand threads easier to use.



```
T-Test
Inpt:  Data  Stats
μ₀: 0
List: L₃
Freq: 1
μ: ≠ μ₀  < μ₀  > μ₀
Calculate Draw
```

```
T-Test
μ < 0
t = -2.903732389
p = .0038957552
x̄ = -13.32
Sx = 22.93599791
n = 25
```

17.47 Our differences are in L3 (from Exercise 17.35 above). We are 90% confident that the right-hand thread will save between 5.5 and 21.2 seconds. This could be of great importance if a task were performed over and over – a minute might be saved for every three repetitions.

```
TInterval
Inpt:  Data  Stats
List: L₃
Freq: 1
C-Level: .9
Calculate
```

```
TInterval
(-21.17, -5.472)
x̄ = -13.32
Sx = 22.93599791
n = 25
```