

Introduction to Statistics Using the TI-83 Graphing Calculator

By

Dr. Robert Knight

This document is a working text that is designed specifically for the course of Introductory Statistics that I teach. In writing this booklet I have presented problems that we will cover in class and methods of solution that I think are the best and most straight forward. This working text is meant to be used in conjunction with the class lecture and required textbook.

I urge students to work in study groups and to meet regularly to go over homework and test questions. If you have a question about the course or need to contact me, my e-mail address is robert.knight@evc.edu

ELEMENTARY STATISTICS

Math 63

Table of Contents

The following table of contents is meant to follow the chapters in Mario Triola's text, Elementary Statistics (ninth edition).

Chapter 1; **Introduction to Statistics**. This chapter is covered in the power point presentation available on my web site. There really is no use made of the TI 83 in this chapter.

Chapter 2; **Describing, Exploring and Comparing Data**.
Pages 3 – 15

Chapters 5 & 6; **Normal Probability Distributions, Estimates and Sample Sizes**.
Pages 16 – 40

Chapters 7 & 8; **Hypothesis Testing, Inference from Two Samples**
Pages 41 – 69

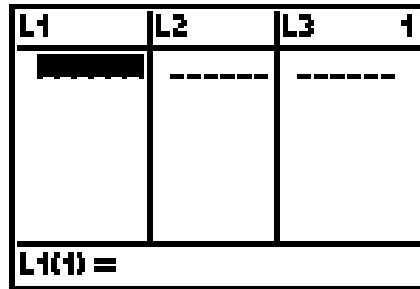
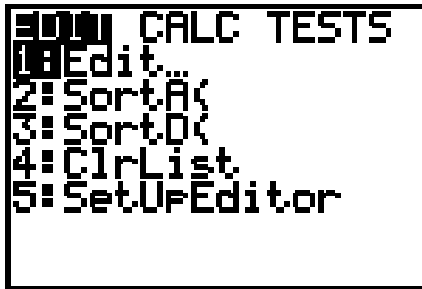
Chapter 9; **Correlation and Regression**
Pages 70 – 75

Chapter 10; **Multinomial Experiments and Contingency Tables**
Pages 76 – 82

Chapter 11; **Analysis of Variance (ANOVA)**
Pages 83 - 88

Data and Lists

When doing homework and taking tests, you will need to put lists of information (DATA) into your calculator. Do this by pressing STAT 1:Edit...

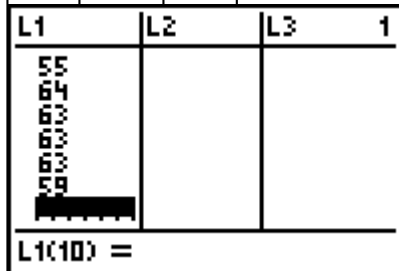


The TI – 83 gives you six lists (L1 – L6). You go from one list to the next using the left or right BLUE arrows.

With the cursor in this position you are ready to enter data into List 1(position 1) by typing the number and pressing ENTER.

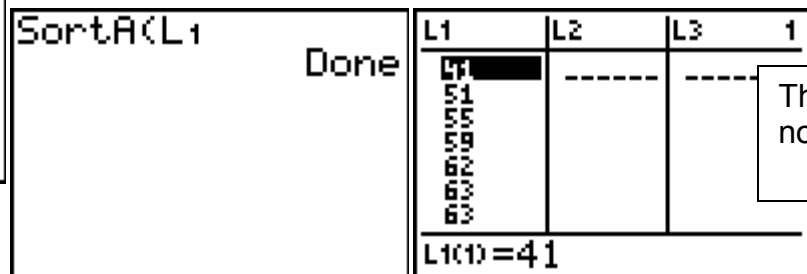
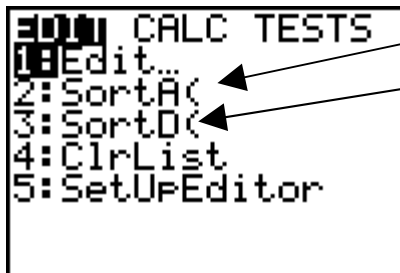
Now enter the numbers that appear in the table below into list 1.

41	55	63
62	64	63
51	63	59



Once you have entered the above data, the calculator screen should look like the one above; (Notice that you cannot see the entire list because it moves up)

The data in any list can be put in 'size order' by using the Sort A (ascending) or Sort D (descending) commands under the Stat button



The List 1 is now sorted

If you ever make a mistake inputting a number into the list you can correct the mistake by simply going to the number that you wish to correct, type the correct number and press the ENTER key.

Try replacing the number 41 with the number 40.

L1	L2	L3	1
41	-----	-----	
51			
55			
59			
62			
63			
L1(1)=40			

Go to the number 41, type in the number 40 and press the ENTER key. 41 has been replaced by 40

L1	L2	L3	1
40	-----	-----	
51			
55			
59			
62			
63			
L1(2)=51			

If you make a mistake you can overwrite the entry as we just did above. Sometimes you might put too many numbers into your list. To get rid of one of them, go to the number you want to delete and press **DEL** (delete).

Now let's say you forgot one of the numbers and you want to put the number into the list in a specific location. Place the cursor **where you want the number to be placed** and press **INS** (insert). Now try inserting the number 52 between 40 and 51. Place the cursor on the number 51, press INS and then 52. Your screen will look like this. When you press ENTER the 52 will be put into the list between 40 and 51. TRY IT!

L1	L2	L3	1
40	-----	-----	
51			
55			
59			
62			
63			
L1(2)=0			

L1	L2	L3	1
40	-----	-----	
51			
55			
59			
62			
63			
L1(2)=52			

L1	L2	L3	1
40	-----	-----	
52			
51			
55			
59			
62			
63			
L1(3)=51			

Now try sorting the data in L1 and putting the data into a stem & leaf plot.

SortA(L1	Done
----------	------

L1	L2	L3	1
40	-----	-----	
51			
52			
55			
59			
62			
63			
L1(1)=40			

The stem leaf plot	
4	0
5	1259
6	2333

CLEARING LISTS

Now let's clear the lists so that you can do more work. There are two ways to do this.

Method 1: Go to the list editor (STAT 1: Edit). Move the cursor to the top of any list you wish to clear so that you highlight the list name. Press CLEAR then ENTER.

Method 2: Press STAT 4: ClrList and enter the list(s) you want cleared with a comma separating them. Clear Lists 1 through 6. Press ENTER and you should see this.

```
ClrList L1,L2,L3
,L4,L5,L6
Done
█
```

All your lists are now cleared ready for your next homework or test question.

Finding out about a list's statistics

Enter the following data into L1 of your List editor

```
5 | 1557
6 | 222
7 | 35
8 | 9
```

Now go to STAT > CALC and select 1:1-Var Stats

```
EDIT 1-VAR TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7:QuartReg
```

This will be pasted to the home screen and follow it by L1 as shown next. Then Press ENTER.

```
1-Var Stats L1█
```

```
1-Var Stats
x̄=64.1
Σx=641
Σx²=42307
Sx=11.6375828
σx=11.04038043
n=10
```

```
1-Var Stats
n=10
minX=51
Q1=55
Med=62
Q3=73
maxX=89
█
```

Five number summary

You now have the summary statistics for list L1. These include the mean, standard deviation, number in the list, median and 5-number summary.

Frequency Tables

The following frequency table will now be input into the TI-83. Clear all lists.

x	f
12-26	2
27-41	5
42-56	8
57-71	4

First enter the lower class limits and upper class limits into lists L1 and L2 respectively. Now move the cursor to L3 and enter the formula for determining the class marks.

L1	L2	L3	3
12	26	---	
27	41	---	
42	56	---	
57	71	---	
---	---	---	
L3=(L1+L2)/2			

Press ENTER and the class marks will now appear in L3. (Notice the () in the formula).

L1	L2	L3	3
12	26	19	
27	41	34	
42	56	49	
57	71	64	
---	---	---	
L3(1)=19			

Now enter the class frequencies in L4 as shown below.

L2	L3	L4	4
26	19	2	
41	34	5	
56	49	8	
71	64	4	
---	---	---	
L4(5) =			

We can find out the summary statistics from our frequency table. First go to the home screen by using 2nd > QUIT. Again go to STAT > CALC and select 1: 1-Var Stats. Put in the lists where the class marks and frequencies are stored remembering to use a comma. In this case 1: 1-Var Stats L3,L4

```
1-Var Stats L3,L4
4
```

Now press ENTER

```
1-Var Stats
x̄=45.05263158
Σx=856
Σx²=42094
Sx=14.00187957
σx=13.62842882
↓n=19
```

```
1-Var Stats
↑n=19
minX=19
Q1=34
Med=49
Q3=49
maxX=64
```

Five number summary

These are the summary statistics for the frequency table. They also include the mean, standard deviation, total number, median and five number summary.

Another statistic we will talk about is the interquartile range. It is gotten by performing the subtraction $Q3 - Q1$ or $49 - 34 = 15$ in this example.

One other statistic is the variance. You can find the variance by pressing LIST > MATH and scroll down or up to 8: variance (enter L3, L4 and press ENTER).

```
NAMES OPS MATH
2: ↑max(
3: mean(
4: median(
5: sum(
6: Prod(
7: stdDev(
8: variance(
```

```
variance(L3,L4)
196.0526316
```

The variance is approximately 196.05.

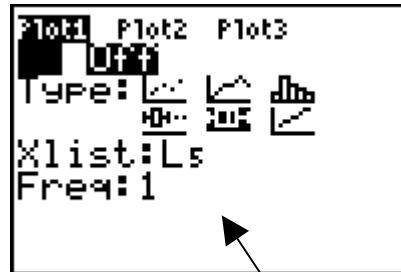
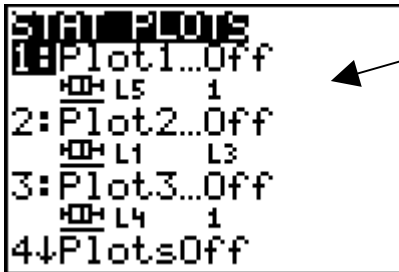
Boxplots

Enter the data found in the first stem plot into L5.

```
5 | 1557
6 | 222
7 | 35
8 | 9
```

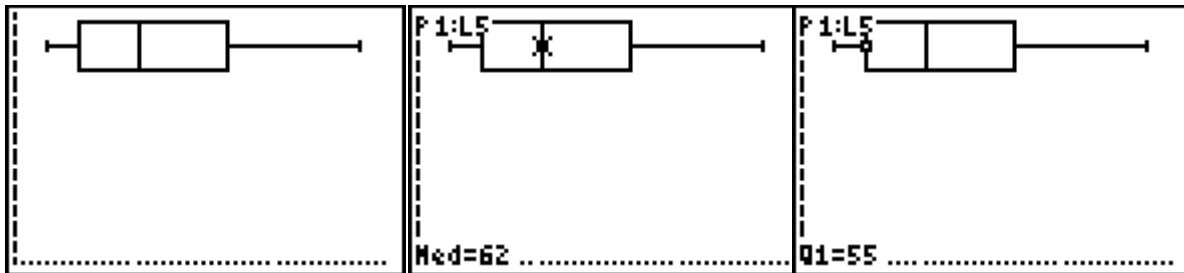
We will now have the calculator draw a boxplot of this data to get an idea of what the distribution of the data looks like.

Press 2nd and STAT PLOT. The screen will look like this.



Now press either 1 or ENTER and use the direction arrows to get this screen.

Turn the plot ON. Now press ZOOM 9 and the screen will look like this. Using the TRACE key and right or left arrows you can find the 5 number summary.



By the way, what is the shape of this distribution? (Ans. Skewed Right)

Histograms

Next we will graph a histogram (what you used to call bar graph) for the data contained in the frequency table which we stored in L1 through L4.

x	f
12-26	2
27-41	5
42-56	8
57-71	4

Press STAT PLOT and select Plot 2, make sure all the other plots are turned off. If they are not select 4:PlotsOff and press ENTER. Now select Plot1 and using the cursor,

```

STAT PLOTS
2:Plot2...Off
  FINAL 1
3:Plot3...Off
  L4 1
4:PlotsOff
  PlotsOn
        
```

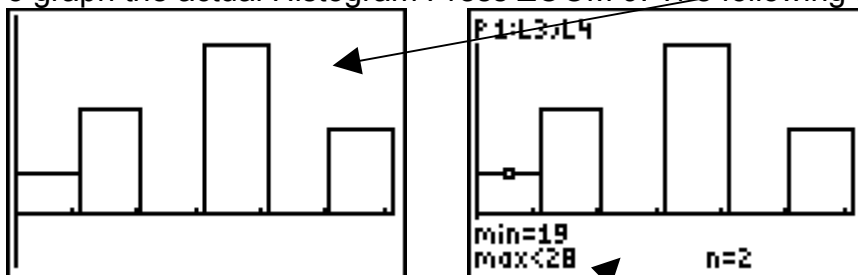
```

Plot2 Plot3
Off
Type: [Histogram]
Xlist:L3
Freq:L4
        
```

1. Turn the plot on.
2. Select Histogram
3. List containing class marks.
4. List containing class frequencies

select the following settings as shown in this screen.

To graph the actual Histogram Press ZOOM 9. The following screen should appear.



Press TRACE and the screen looks like this. Notice that this is not a very good histogram for the frequency table because it has too many classes and some of the classes have no members (frequency = 0). So to make the correct histogram appear we will modify the window by pressing WINDOW.

```

WINDOW
Xmin=19
Xmax=73
Xscl=9
Ymin=-2.40552
Ymax=9.36
Yscl=.1
Xres=1
        
```

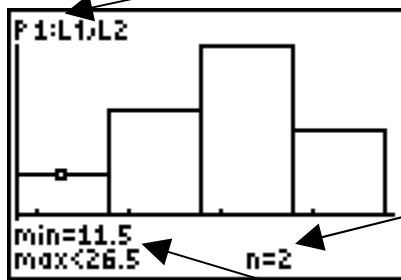
```

WINDOW
Xmin=11.5
Xmax=73
Xscl=15
Ymin=-2.40552
Ymax=9.36
Yscl=.1
Xres=1
        
```

This is the window the calculator sets up. By changing Xmin to the lower class boundary of the first class and Xscl to the class width we will get the histogram you see below.

Make the two changes you see above and then press **GRAPH** (not ZOOM 9). Press the TRACE button and see the screen you see on the next page.

This denotes that the Plot1 uses L3 and L4.



This is the class frequency which corresponds to how height of the bar.

This means that the lower class boundary of the first class is 11.5 the upper class boundary is 26.5.

Using the right arrow we can get the other class's boundaries and frequencies.

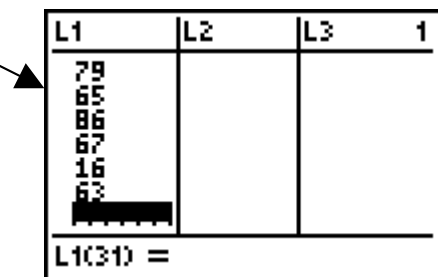
Now let's create a frequency table and histogram from a list of data, Additionally, we will specify the number of classes we want the frequency table to have.

For the data given below, create a frequency table having eight (8) classes.

18	56	25	45	79
32	46	12	54	65
59	29	19	98	86
29	75	49	22	67
92	10	37	43	16
94	19	20	71	63

The first step is to enter all the data into one list of your list editor. It does not matter if you enter going across or up & down, but all the data must be in one list.

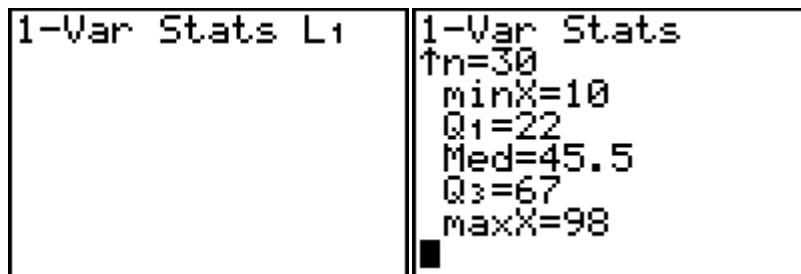
The calculator screen will look something like this.



The next step will be to determine the range of numbers for this set of data.

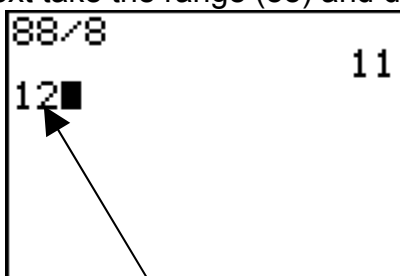
The range = Maximum – Minimum. The easiest way to find the **Max** and **Min** for or data is to have the calculator determine the **1 – Var Stats** for this data set. Go to **Stat > Calc** and select 1: **1 – Var Stats** followed by the list containing the data (L1 in this case). Once the **1 – Var Stats** have been calculated, scroll to the five number summary to get the Max and Min for the data set.

The following screens demonstrate this process.



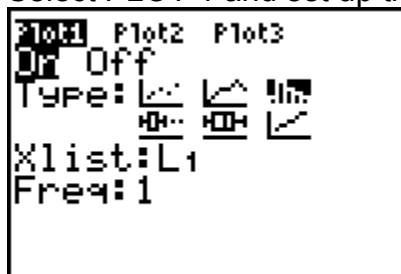
So the range of the data will be $98 - 10 = 88$.

Next take the range (88) and divide by the number of classes required (8 in this case)

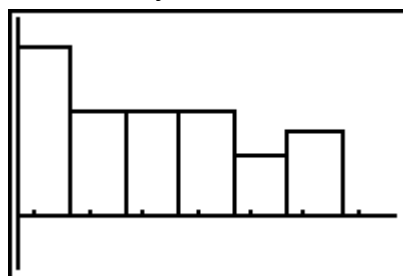


Round this number up to the **next whole number** (**Note**: If the answer was a whole number, go to the next whole number). This gives the class width required. In this case the class width is **12**.

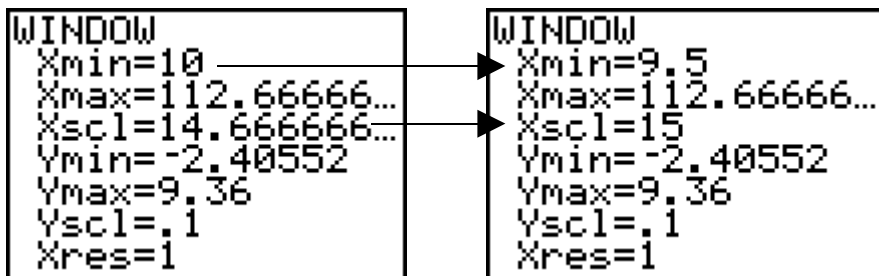
Next, go to the STAT PLOT screen and check to see that all the PLOTS are off. Select PLOT 1 and set up the screen to look like the screen below.



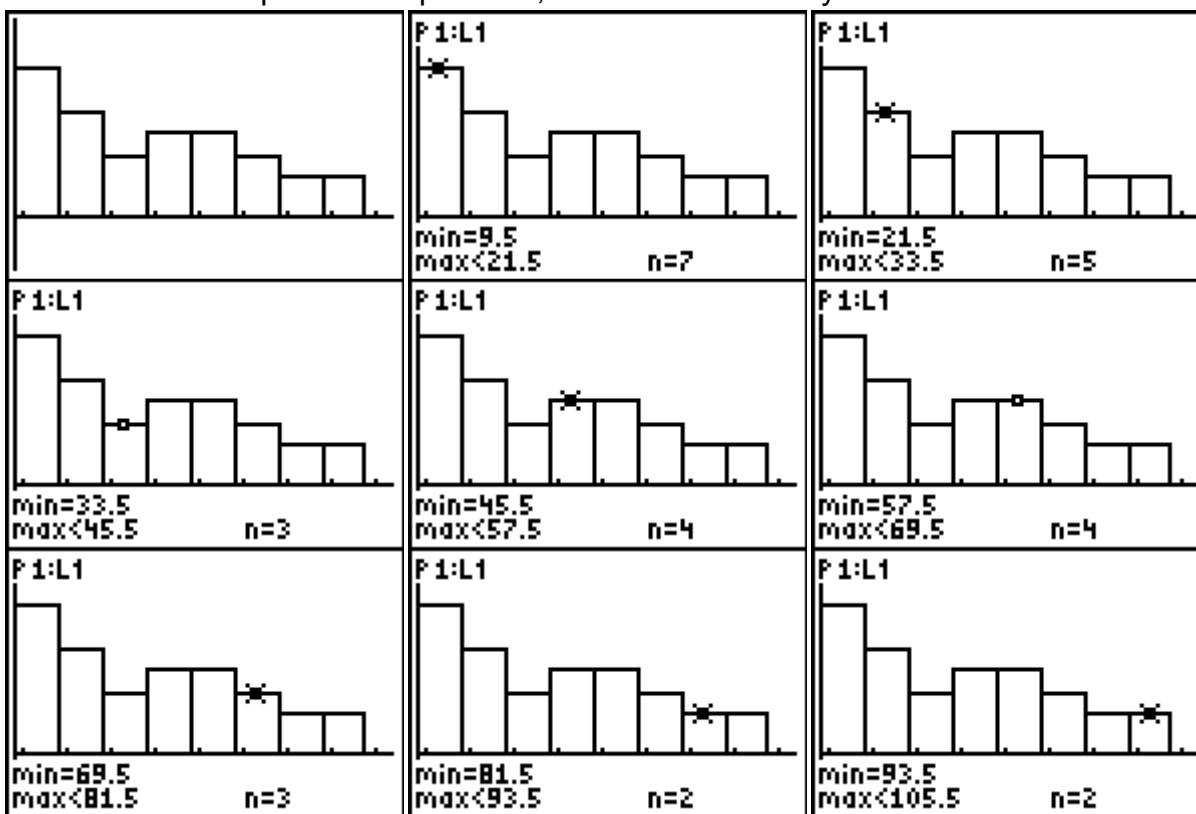
Now Press ZOOM 9 and you should see the following screen. Notice only six classes.



Next, Press the **WINDOW** key and make the following two changes.
 The **Xmax** will be changed to the lower class boundary of the first class.
 T **Xscl** will become the class width that you calculated above.



Now Press the **GRAPH** key and you will see a histogram which has the classes with the proper class width and boundaries for the data. To see the actual boundaries of the classes and their respective frequencies, use the **TRACE** key.



This information can be put into a frequency table like the one below.

X	f
9.5 - 21.5	7
21.5 - 33.5	5
33.5 - 45.5	3
45.5 - 57.5	4
57.5 - 69.5	4
69.5 - 81.5	3
81.5 - 93.5	2
93.5 - 105.5	2

Note that this frequency table uses class boundaries. You can also make a frequency table using class limit like the one below.

X	f
10 - 21	7
22 - 33	5
34 - 45	3
46 - 57	4
58 - 69	4
70 - 81	3
82 - 93	2
94 - 105	2

Methods of determining “Normality”

(How does the shape of the distribution compare to the Bell Curve?)

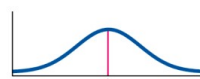
There are three ways of finding out how the graph of a set of data compares to the Bell Curve. This is what is called determining “normality”. All three methods involve using the STAT PLOTS function on the TI 83. Two of the graphs under the STAT PLOTS function have already been covered. These graphs are the box plot and the histogram covered earlier in this chapter.

When looking at the general shape of a histogram, try to fit its general shape to one of the shapes in the diagram below.

Skewness

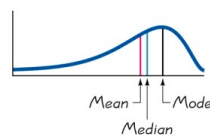


Figure 2-11

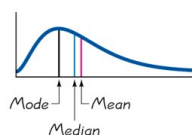


Mode = Mean = Median

(b) Symmetric

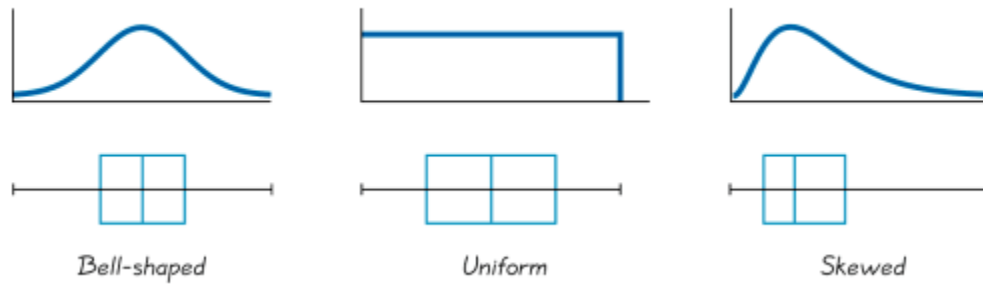


(a) Skewed to the Left
(Negatively)



(c) Skewed to the Right
(Positively)

When considering the case of box plots, the shape of the distribution can be determined by comparing it to shapes found in the diagram below.

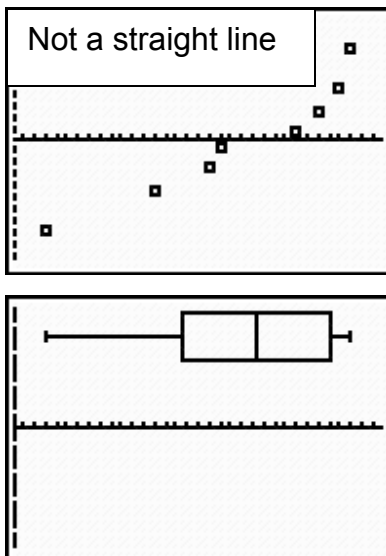


A third method is by using another type of graph found on the TI 83. Under the STAT PLOTS key select a plot (1 – 3) and choose the very last type of graph.

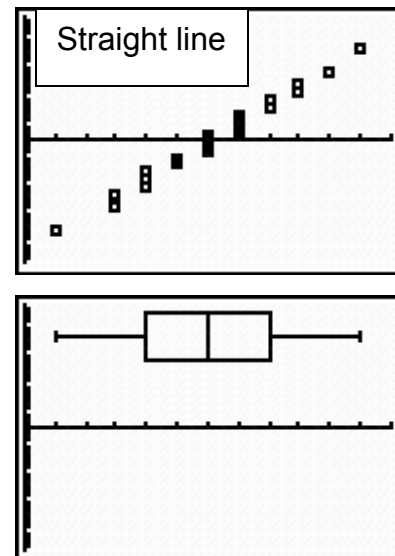
This graph can also be used to test for normality. Below are two different graphs with their respective graphs for testing normality. Notice that the closer the graph for testing normality comes to a **straight line**, the more normal the distribution. Compare the graphs respective box plots to understand this idea.



Screwed Distribution



Normal Distribution



Important formulas for the First Midterm

Z score formula and the definition of statistically unusual (greater than two standard deviations beyond the mean in either the positive or negative direction).

$$Z = \frac{(x - \bar{x})}{s_x}$$

Range = Max - Min

Mid Range =(Min + Max)/2

$$InterquartileRange = \frac{(Q_3 - Q_1)}{2}$$

Measure of relative standing in a finite data set

$$L = \frac{k}{100} * n$$

Refer to flow chart on page 96

Chapters 5 & 6

The Normal Distribution

In drawing the normal distribution we will make use of a saved window to graph the standard normal distribution. Press the WINDOW button and enter the values you see in the following screen very, very carefully.

```
WINDOW
Xmin=-3.2
Xmax=3.2
Xscl=1
Ymin=-.12
Ymax=.45
Yscl=.1
Xres=1
```

Note that these two **negative** values require using the **gray negative** button, **not** the blue subtract button.

Note: this is .1 not 1

We will now store this window to be used in the future so that we do not have to put in these numbers every time we wish to graph a normal distribution. First check by pressing STAT PLOT that all the plots are turned off. Also check by pressing the Y= button that no equations are in the function toolbox.

```
STAT PLOTS
1:Plot1...Off
  ▽ L3  L4
2:Plot2...Off
  ▽ L3  L4
3:Plot3...Off
  ▽ L4  1
4:PlotsOff
```

All plots are turned Off

```
Plot1 Plot2 Plot3
\Y1=
\Y2=
\Y3=
\Y4=
\Y5=
\Y6=
\Y7=
```

No equations in Y= function tool box.

Now press 2nd DRAW and go to the STO menu to see the screen below

```
DRAW POINTS STO
1:StorePic
2:RecallPic
3:StoreGDB
4:RecallGDB
```

Select 3 Store GDB (Graph Data Base) by pressing 3.

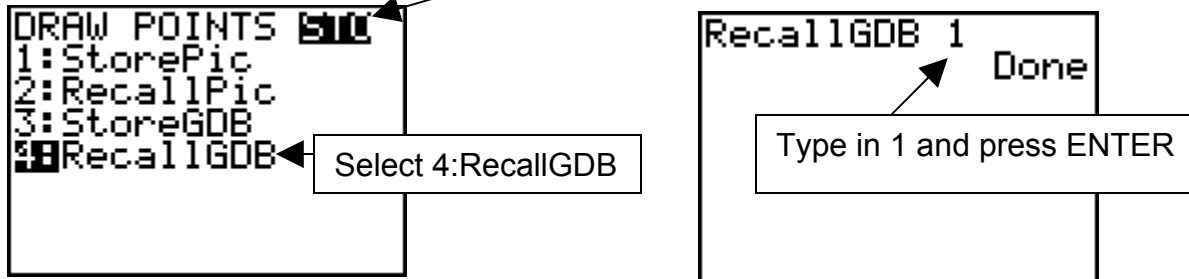
Store GDB will be pasted to the home screen. Now press the number 1.

```
StoreGDB 1
Done
```

Now press ENTER and the screen will tell you it is DONE.

Standard Normal Distribution

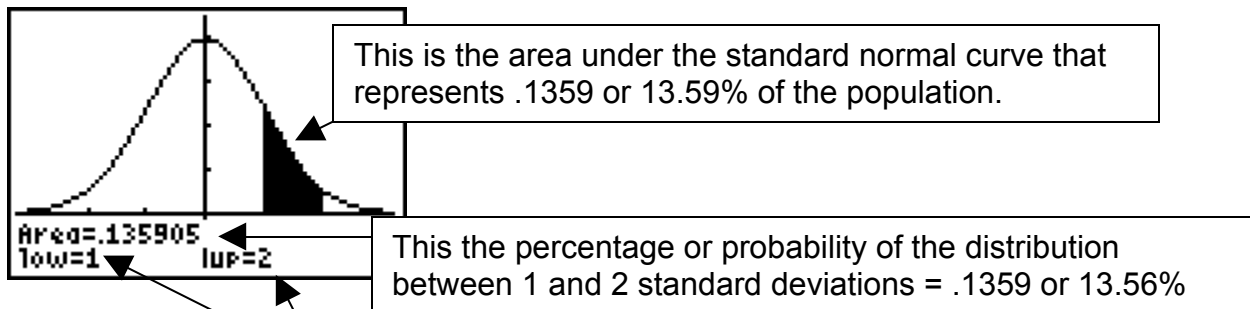
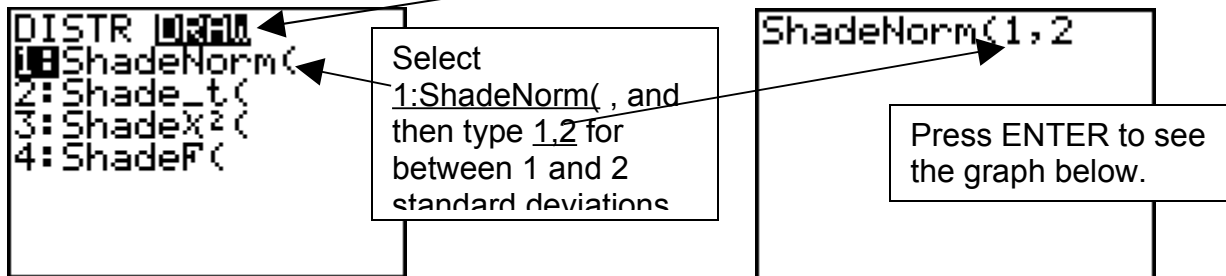
We are now ready to do problems dealing with graphing the standard normal distribution. First we will need to know how to recall the window we stored as graph data base 1 (GDB1). Press 2 nd DRAW and use the arrows to go to the STO menu. Select 4:Recall GDB and press ENTER. This pastes the Recall GDB command to the home screen. Now press the number 1 and then Press ENTER and the calculator tells you it is DONE.



Let say we want to know what proportion or percentage of the population in a standard normal distribution that is between 1 and 2 standard deviations above the mean. Another way this question could be asked is what is the probability that, if a member is randomly selected from the population, that member will be between 1 and 2 standard deviations above the mean?

We will do this problem finding the percentage or probability and also graphing the area under the standard normal curve that represents this percentage or probability.

Press 2 nd DISTR and right arrow to the DRAW menu shown below.



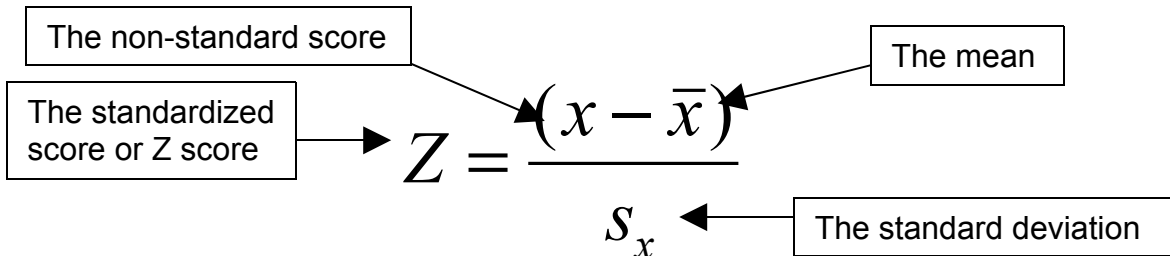
These are the 1 and 2 standard deviations between which the graph is drawn.

When we are finished go to 2 nd DRAW and select 1:ClrDraw to clear your graph.

Non-Standard Normal Distribution

We will now do some problems that deal with other populations which are normal distributions with means and standard deviations other than 0 and 1 respectively. In other words, **non-standard** normal distributions.

How do you deal with a non-standard normal distribution? **Standardize it!** You've heard of standardized tests. How do you standardize? The Z score formula.



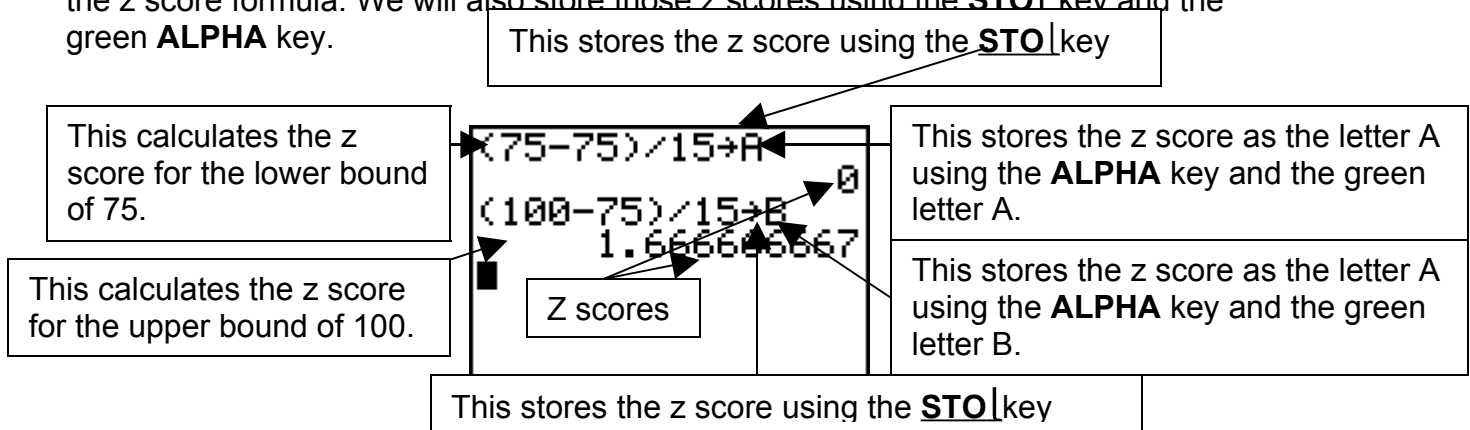
Now we will do a typical non-standard normal deviation problem.

An educational testing corporation has designed a standard test of mechanical aptitude. Scores on this test are normally distributed with a mean of 75 and a standard deviation of 15.

If a subject is randomly selected and tested, find the probability that his score will be between 75 and 100.

Go to the home screen by pressing 2 nd QUIT. Clear the screen by pressing CLEAR.

We will calculate the z scores for the two upper and lower bounds of 75 and 100 using the z score formula. We will also store those z scores using the **STO** key and the green **ALPHA** key.



Now that the z scores have been calculated and stored as A and B, we will graph the
 First press 2 nd DISTR and go to the DRAW menu. Select 1:ShadeNorm(by pressing 1

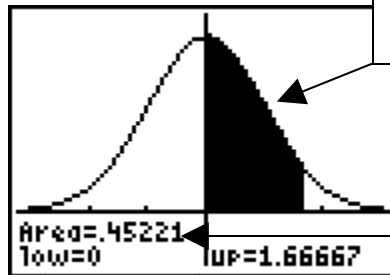
```
DISTR 1:ShadeNorm(
2:Shade_t(
3:ShadeX^2(
4:ShadeF(
```

ShadeNorm is now pasted to the home screen

```
(75-75)/15→A
(100-75)/15→B
1.666666667
ShadeNorm(A,B)
```

Enter A and B by using the ALPHA keys followed by A and B

Now press ENTER to see the graph below.



Here is the graph of the area representing the probability

Here is the actual probability = .4522 or 45.22%

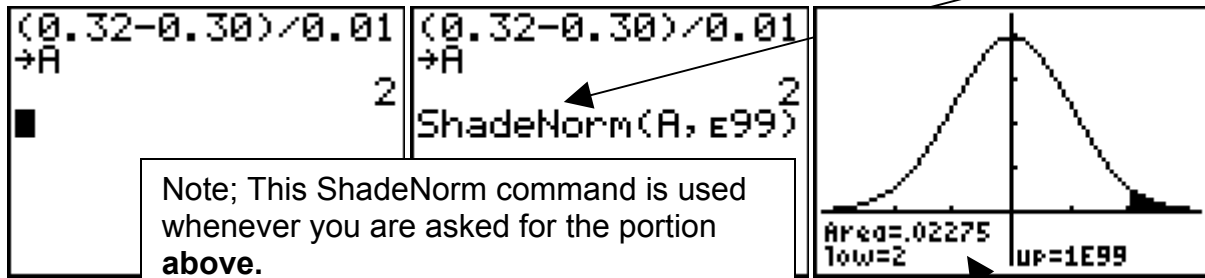
Here are the lower and upper bound z scores.

Do not forget to clear the graph by pressing 2 nd DRAW and select 1:ClrGraph

Now another problem which uses ShadeNorm and the Z score formula

The diameters of bolts produced by a certain machine are normally distributed with a mean of 0.30 inches and a standard deviation of 0.01 inches. What percentage of bolts will have a diameter greater than 0.32 inches?

First, since the question asks for the percentage of bolts with a diameter greater than 0.32 inches, we need to find the Z score for 0.32 in this normal distribution and save it as A. (screen 1) Then use the ShadeNorm command to shade above this Z score value.



The resulting graph should look like this. The area under the curve (.02275) is the decimal equivalent of the percentage we are looking for. Hence, the answer is **2.275%**

Problems which ask for the probability, percentage or portion of the population below a particular value in a normal distribution are done in much the same manner. However the ShadeNorm command will look like this

```
ShadeNorm(-E99, A
```

Percentiles in a Normal Distribution

The amount of rainfall in January in a certain city is normally distributed with a mean of 3.4 inches and a standard deviation of 0.5 inches. Find the value of the quartile Q_1 .

This problem is a little tricky because you need to know that Q_1 is the 25th percentile. So, we are really being asked to find the 25th percentile. The command is under the Dist button. **3 : InvNorm(** . The syntax for this command is to enter the percentile as a decimal, the mean, the standard deviation then press ENTER.

```
0:Stat DRAW
1:normalpdf(
2:normalcdf(
3:invNorm(
4:tpdf(
5:tcdf(
6:χ²pdf(
7:χ²cdf(
```

```
invNorm(.25,3.4,
0.5
3.062755125
█
```

This answer should be rounded **up** to one more decimal place than the data given in the problem. So the answer should be 3.07 or the closest answer to it in a multiple-choice type of question.

The Central Limit Theorem

In problems dealing with samples of a certain size being taken randomly from a population, think about the Central Limit Theorem. The major difference between a Central Limit Theorem problem and problems we have just done is the **recalculation of the standard deviation** for the Central Limit Theorem problems. Other than that the problem can be handled in exactly the same manner as non-standard distribution problems.

For example for the distribution problem that we were doing earlier, a Central Limit Theorem variation would be as follows.

An educational testing corporation has designed a standard test of mechanical aptitude. Scores on this test are normally distributed with a mean of 75 and a standard deviation of 15. **Eighty different subjects** are randomly selected and tested. Find the probability that **this sample of 80** has a mean is below 71.

Again the key here is that a sample of 80 subjects are randomly selected. So first we will recalculate the new standard deviation for this distribution remembering that the original distributions standard deviation was = 15.

```
15/√(80)→S
1.677050983
```

This uses the formula for calculating the new standard deviation, $S_{\bar{x}} = \frac{S}{\sqrt{n}}$ stores in using the STO| button and the ALPHA button followed by the green letter S

The actual value of the standard deviation that will be used in this Central Limit Theorem problem and which is stored as S.

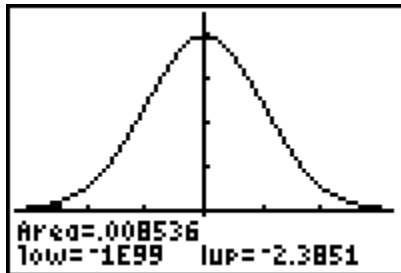
Now we proceed with the problem by calculating the z score for the score of 71 using this newly calculated standard deviation (S) and storing the resulting z score as A.

```
15/√(80)→S
1.677050983
(71-75)/S→A
-2.385139176
```

Finally we go to the 2 nd DISTR button and then to the DRAW menu and select 1:ShadeNorm(as before, thus pasting it to the home screen.

```
15/√(80)→S
1.677050983
(71-75)/S→A
-2.385139176
ShadeNorm(-E99,A
█
```

As before, since this is a problem asking for the probability that a score will be below a particular number (71), we use negative infinity in the form $-E99$. Press ENTER.



Here is the screen giving the probability of **.0085** or **.85%** along with the graph.

Problems;

In one region, the September energy consumption levels for single-family homes are found to be normally distributed with a mean of 1050 kwh and a standard deviation of 218 kwh.

- 1. For a randomly selected home, find the probability that the September energy consumption level is between 1050 kwh and 1250 kwh.
- 2. For a randomly selected home, find the probability that the September energy consumption level is below 1200 kwh.
- 3. For a randomly selected home, find the probability that the September energy consumption level is between 1100 kwh and 1225 kwh.
- 4. For a randomly selected home, find the probability that the September energy consumption level is above 1175 kwh.
- 5. For a randomly selected home, find the probability that the September energy consumption level is between 1000 kwh and 1200 kwh.
- 6. Find the 45th percentile.
- 7. If 50 different homes are randomly selected, find the probability that their mean energy consumption level for September is greater than 1075 kwh.
- 8. The mean serum cholesterol of a large population of overweight adults is 220 milligrams per deciliter (mg/ dl), and the standard deviation is 16.3 mg/dl. If a sample of 30 adults is selected, find the probability that the mean will be between 220 and 222 mg/dl.

9.

Toby's Trucking Company determined that on an annual basis, the distance traveled per truck is normally distributed with a mean of 50.0 thousand miles and a standard deviation of 12.0 thousand miles.

- (a) How many of the 1,000 trucks in the fleet are expected to travel between 30.0 and 60.0 thousand miles in the year?
- (b) How many miles will be traveled by at least 80% of the trucks?

10.

A set of final examination grades in an introductory statistics course was found to be normally distributed with a mean of 73 and a standard deviation of 8.

- (a) What is the probability of getting a grade no higher than 91 on this exam?
- (b) What is the final exam grade if only 5% of the students taking the test scored higher?
- (c) If the professor “curves” (gives A’s to the top 10% of the class regardless of the score), are you better off with a grade of 81 on this exam or a grade of 68 on a different exam where the mean is 62 and the standard deviation is 3?
Show statistically and explain.

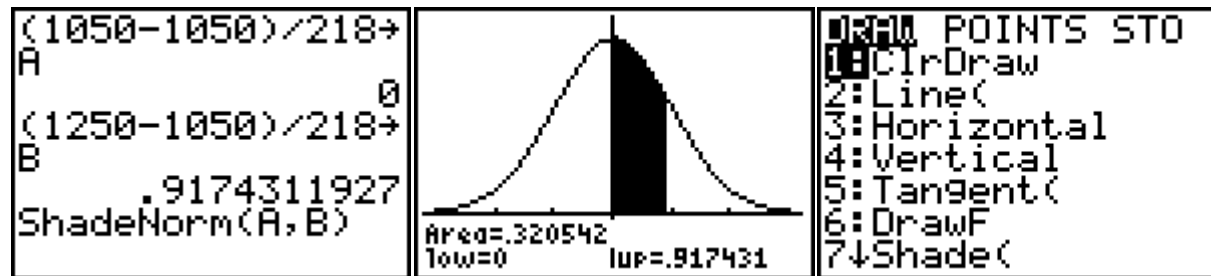
11.

Wages for workers in a particular industry average \$11.90 per hour and the standard deviation is \$0.40. If the wages are assumed to be normally distributed:

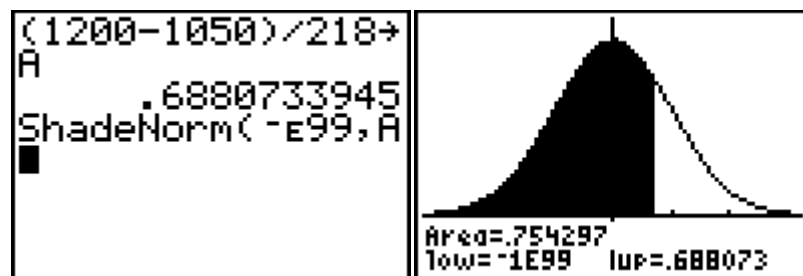
- (a) What percentage of workers receive wages between \$12.20 and \$13.10?
- (b) What percentage of workers receive wages less than \$11.00?
- (c) What percentage of workers receive wages more than \$12.95?
- (d) What must the wage be if only 10% of all workers in this industry earn more?
- (e) What must the wage be if 25% of all workers in this industry earn less?

Answer Key

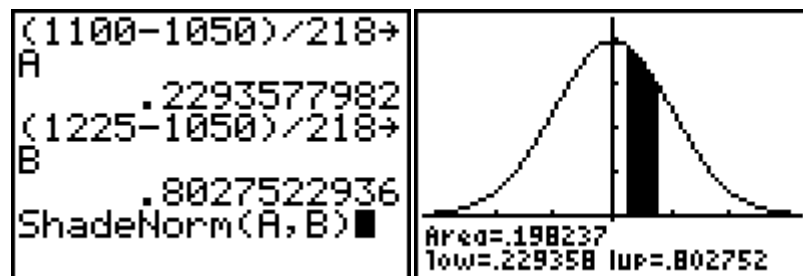
1.



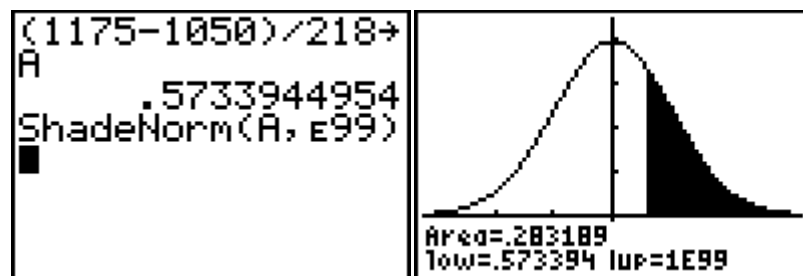
2.



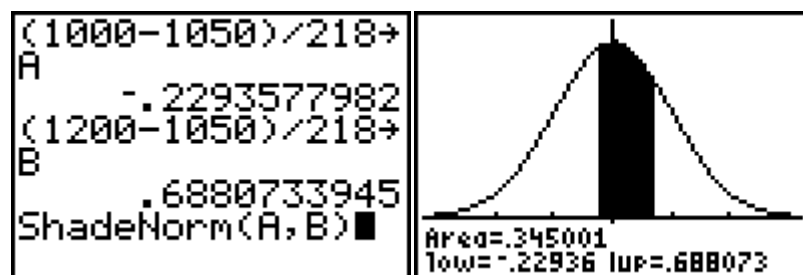
3.



4.



5.



6.

<pre> DISK DRAW 1:normalpdf(2:normalcdf(8invNorm(4:tpdf(5:tcdf(6:X²pdf(7↓X²cdf(</pre>	<pre> invNorm(.45,1050 ,218) 1023 1022.605828 1023 1023 </pre>
--	--

7.

<pre> 218/√(50)→S 30.82985566 (1075-1050)/S→A .810902272 ShadeNorm(A, E99) </pre>	<p>Area=.208711 low=.810902 up=1E99</p>
---	---

8.

<pre> 16.3/√(30)→S 2.975959229 (220-220)/S→A 0 (222-220)/S→B .6720522178 ShadeNorm(A, B)■ </pre>	<p>Area=.249225 low=0 up=.672052</p>
--	--

9.a

<pre> (30-50)/12→A -1.666666667 (60-50)/12→B .8333333333 ShadeNorm(A, B) </pre>	<p>Area=.749881 low=-1.6667 up=.833333</p>	<pre> .7499*1000 749.9 750 750 </pre>
---	--	---------------------------------------

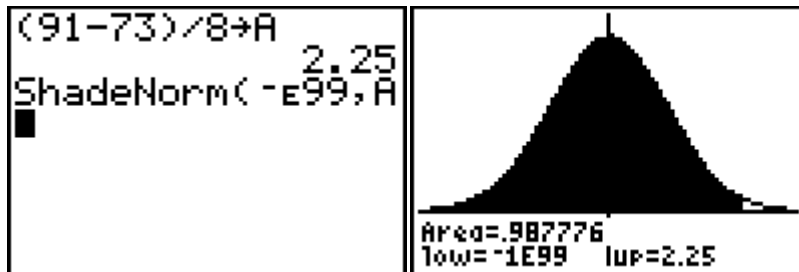
9.b.

<pre> invNorm(.20,50,1 2) 39901 39.9005452 39901 39901 </pre>

10.a.

```
invNorm(.95,73,8
      86.15882901
87
      87
```

10.b.

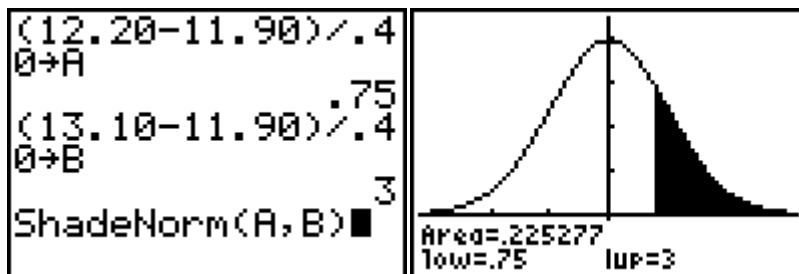


10.c.

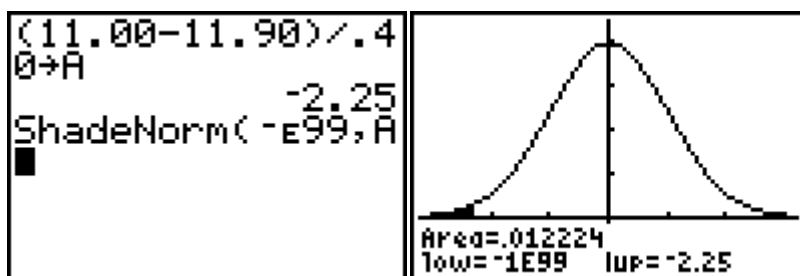
A grade of 68 which is 2 standard deviations above the mean.

```
(81-73)/8
      1
(68-62)/3
      2
```

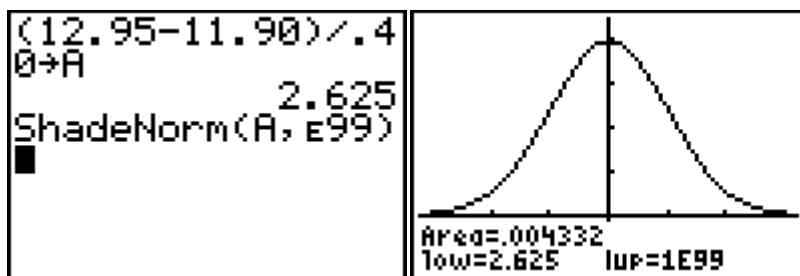
11.a.



11.b.



11.c.



11.d.

```
invNorm(.90,11.9  
0,.40)  
          12.41262063  
12.42  
          12.42
```

11.e.

```
invNorm(.25,11.9  
0,.40)  
          11.6302041  
11.64  
          11.64
```

THE CENTRAL LIMIT THEOREM

1.

For a certain large group of individuals, the mean hemoglobin level in the blood is 21.0 grams per milliliter (g/ml). The standard deviation is 2 g/ml. If a sample of 25 individuals is selected, find the probability that the sample mean will be greater than 21.3 g/ml.

2.

The mean weight of 18-year-old females is 126 pounds, and the standard deviation is 15.7. If a sample of 25 females is selected, find the probability that the mean of the sample will be greater than 128.3 pounds.

3.

The mean grade point average of the engineering majors at a large university is 3.23, with a standard deviation of 0.72. In a class of 48 students, find the probability that the mean grade point average of the 48 students is less than 3.15.

4.

The mean of the number of ounces of coffee a vending machine dispenses is 10. The standard deviation is 0.1. If a sample of 10 cups is selected, find the probability that the mean of the sample will be less than 9.93 ounces.

5.

Assume that the mean systolic blood pressure of normal adults is 120 millimeters of mercury (mmHg), and the standard deviation is 5.6.

a. If an individual is selected, find the probability that the individual's pressure will be between 120 and 121.8 mmHg.

b. If a sample of 30 adults is randomly selected, find the probability that the sample mean will be between 120 and 121.8 mmHg.

c. Why is the answer to part a so much smaller than the answer to part b?

6.

At a large university, the mean age of graduate students who are majoring in psychology is 32.6 years, and the standard deviation is 3 years.

a. If an individual from the department is randomly selected, find the probability that the individual's age will be between 31.0 and 33.2 years.

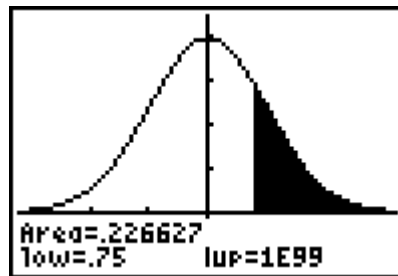
b. If a random sample of 15 individuals is selected, find the probability that the mean age of the students in the sample will be between 31.0 years and 33.2 years.

7.

The average breaking strength of a certain brand of steel cable is 2000 pounds, with a standard deviation of 100 pounds. A sample of 20 cables is selected and tested. Find the sample mean that will cut off the upper 95% of all of the samples of size 20 taken from the population.

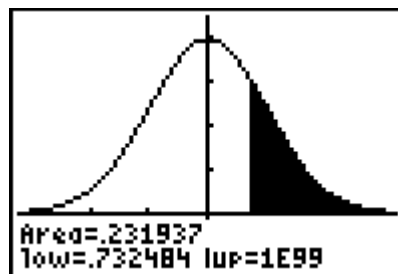
1.

```
2/√(25)→S
(21.3-21.0)/S→A
.75
ShadeNorm(A, 1E99)
█
```



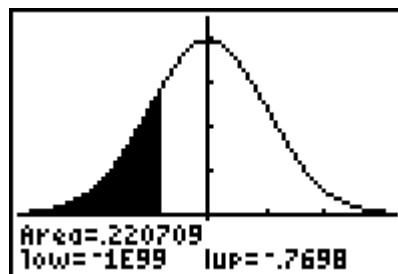
2.

```
15.7/√(25)→S
(128.3-126)/S→A
.7324840764
ShadeNorm(A, 1E99)
```



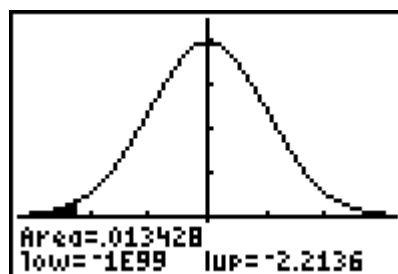
3.

```
.72/√(48)→S
.1039230485
(3.15-3.23)/S→A
-.7698003589
ShadeNorm(-1E99, A)
```



4.

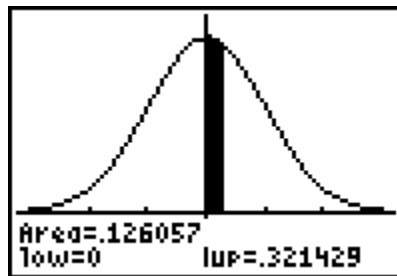
```
.1/√(10)→S
.0316227766
(9.93-10)/S→A
-2.213594362
ShadeNorm(-1E99, A)
█
```



5.

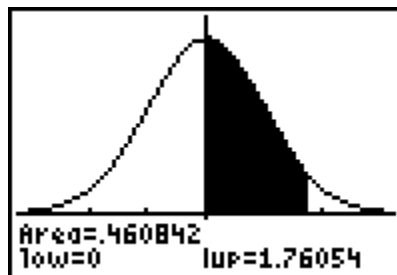
a.

```
(121.8-120)/5.6→A
.3214285714
ShadeNorm(0,A)
```



b.

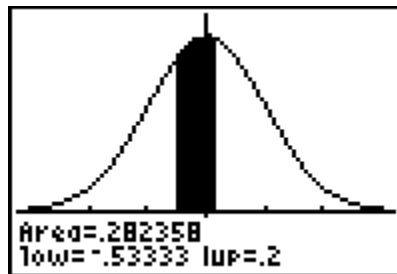
```
5.6/√(30)→S
1.022415441
(121.8-120)/S→A
1.760536792
ShadeNorm(0,A)
```



c. Because the standard deviation (variation) of the sample of 30 is so much smaller (central limit theorem), the area or percentage of the population of sample means in the same interval is much greater.

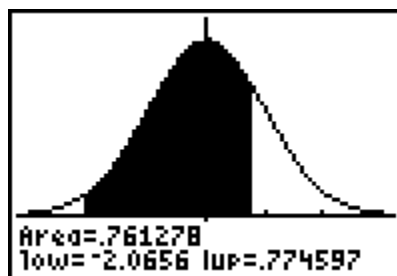
6.a.

```
(31-32.6)/3→A
-.5333333333
(33.2-32.6)/3→B
.2
ShadeNorm(A,B)
```



b.

```
3/√(15)→S
.7745966692
(31-32.6)/S→A
-2.065591118
(33.2-32.6)/S→B
.7745966692
ShadeNorm(A,B)
```



7.

```
100/√(20)→S
22.36067977
invNorm(.05,2000,S)
1963.219955
1964
1964
```


Confidence Intervals

Pressing the STAT key and using the right arrow to go to the TESTS menu starts the calculation of confidence intervals. Then use the down arrow to scroll down to the group of different confidence interval tests.

```
EDIT CALC TESTS
1:Z-Test...
2:T-Test...
3:2-SampZTest...
4:2-SampTTest...
5:1-PropZTest...
6:2-PropZTest...
7:ZInterval...
```

```
EDIT CALC TESTS
7:ZInterval...
8:TInterval...
9:2-SampZInt...
0:2-SampTInt...
A:1-PropZInt...
B:2-PropZInt...
C:χ²-Test...
```

Numbers 7 to B are all of the Interval Tests used in this course.

We will start with numbers 7:Z Interval, 8:T Interval and A:1-PropZInt. (The remaining intervals will be covered in later chapters.)

Estimating a Population Mean

We will start by doing a problem that demonstrates **7:Z Interval**. A sociologist develops a test to measure attitudes about public transportation, and 47 randomly selected subjects are given the test. Their mean is 76.2 and their standard deviation is 21.4. Construct the 95% confidence interval for the mean score of all such subjects. Why is this a problem involving a **Z Interval**? Because the size of the sample is **>30**.

```
ZInterval
Inpt:Data TESTS
σ:21.4
x̄:76.2
n:47
C-Level:95
Calculate
```

Notice that if $n > 30$, then $s = s$ for the sample.

Input the remainder of the summary statistic, go to Calculate and press ENTER

```
ZInterval
(70.082,82.318)
x̄=76.2
n=47
```

This is the estimation interval. The center of the interval is 76.2. The ends of the interval are 70.082 and 82.318.

The Margin of Error is calculated by subtracting $82.318 - 76.2 = 6.118$ or by subtracting $76.2 - 70.082 = 6.118$.

Either way the Margin of Error is 6.118. The interval could be written as **76.2 ± 6.118** .

Now we will do a problem that uses **8:T Interval**.

```
EDIT CALC TESTS
7↑ZInterval...
8TInterval...
9:2-SampZInt...
0:2-SampTInt...
A:1-PropZInt...
B:2-PropZInt...
C↓X2-Test...
```

Twenty cars were randomly selected and weighed. This sample has a mean of 3650 pounds and a standard deviation of 420 pounds. Construct the 99% confidence interval for the population's mean.

Notice that this problem also asks for a confidence interval for the population mean, but that the size of the sample is **≤30**, namely twenty cars.

We will input the above summary statistics into the **T Interval** screen as shown below.

```
TInterval
Inpt:Data STAT
x̄:3650
Sx:420
n:20
C-Level:99
Calculate
```

This screen gives the estimation interval for the population's mean from this sample of 20 cars.

```
TInterval
(3381.3, 3918.7)
x̄=3650
Sx=420
n=20
```

The **Margin of Error** is calculated by subtracting $3918.7 - 3650 = 268.7$ or by subtracting $3381.3 - 3650 = 268.7$. The interval could also be written as **3650 ± 268.7**

Estimating a Population Proportion or Percentage

Next we will consider estimation of a population proportion or percentage. This something you see every time there is an election and the exit polls sampling is used to predict the winning candidate or proposition.

A survey of 300 union members in New York State reveals that 37.3% favor the Republican candidate for governor. Construct the 98% confidence interval for the true population proportion of all New York State union members who favor the Republican candidate.

```
EDIT CALC TESTS
7↑ZInterval...
8:TInterval...
9:2-SampZInt...
0:2-SampTInt...
11-PropZInt...
B:2-PropZInt...
C↓X2-Test...
```

Because this is a problem of estimating a proportion (or percentage) for one population we use A: 1-1PropZInt...

Notice that we are not told how many for the 300 union members favor the Republican candidate, but instead that 37.3% of the 300 favor the Republican candidate for governor. This information will be input into the calculator as follows

Since you only know the percent of workers who support the candidate we input the number as 300 times the percent (in decimal form).

```
1-PropZInt
x:300*.373
n:300
C-Level:98
Calculate
```

When you move the cursor the x: changes to this value.x:111.9

```
1-PropZInt
x:111.9
n:300
C-Level:98
Calculate
```

Since 111.9 is not a whole number (person) we go to the next number namely 112 and input that value for x:112. This **rounding up** not off.

```
1-PropZInt
x:112
n:300
C-Level:98
Calculate
```

Note: In problem where we were given the number of workers who support the candidate this process is unnecessary. Just put in the number as x:

Now go to Calculate and press ENTER.

```
1-PropZInt
(.30837,.4383)
p=.3733333333
n=300
```

Here is the confidence interval estimate for the population percentage (or proportion) who favor the candidate.

The **Margin of Error** in this estimate is $.4383 - .3733 = .065$ or $.3733 - .3083 = .065$

In terms of % the interval could be reported as $37.33\% \pm 6.5\%$

Usually the way this sort of thing is reported in the new media is to say that the candidate is supported by 37.33% of the workers and then somewhere in fine print to tell you that the study carries a margin of error of 6.5%. Look for this when you see projection of the next election returns.

Calculation of Sample Size

When planning a study or project involving estimating a population mean or a population proportion it is a good idea to know in advance how big your study has to be to fulfill certain statistical guidelines. How confident do you want to be in your result? 95%, 98%, 99%, 99.99%? And how large a margin of error are you willing to accept?

These considerations are especially valid for graduate students planning original research projects and dissertations or for professionals planning or critically reviewing a study. Was the study large enough to be statistically valid?

There are basically two formulas used to determine the same size necessary. One formula is for determining the size necessary to estimate a population mean. The other formula is for determining the size necessary to estimate a population proportion or percentage.

I will first give you these two formulas and then show how the calculator can be used to work the formulas through to the correct answer.

The formula above is used for determining the number necessary for a population mean

$$n_{\mu} = \frac{Z_{\alpha/2} \cdot \sigma}{E}^2$$

given certain criteria.

$$n_p = \frac{Z_{\alpha/2}}{E}^2 \times 0.25$$

The formula above is used for determining the number necessary for a population proportion or percentage given certain criteria.

Next we will do two problems, one each using these formulas.

We want to estimate the mean energy consumption level for a home in one region. We want to be 90% confident that our sample mean is within 25 kwh of the true population mean, and past data strongly suggest that the population standard deviation is 137 kwh. **How large must our sample be?**

Notice that this asks the question... How large must our sample be? or How many must be sampled?

In this particular problem we want to estimate the population **mean**. So we will use the first of or two formulas. Namely...

$$n_{\mu} = \frac{z_{\alpha/2} \cdot \sigma}{E}^2$$

In this formula

n_{μ}

stands for the number necessary to estimate the population mean. This is the answer we want!

The next part of the formula we will calculate is

$$z_{\alpha/2}$$

This comes from the part of the problem stating we want to be 90% confident. The way it works is $\alpha = 100\% - 90\% = 10\%$. So that $\alpha/2 = 5\% = .05$. We now use the calculator to figure the rest.

Press 2 nd DISTR and see the following screen.

```

0: normalpdf(
1: normalcdf(
2: normalcdf(
3: invNorm(
4: tpdf(
5: tcdf(
6: x²pdf(
7: x²cdf(
    
```

Select 3: invNorm(and press ENTER. The following screen appears.

```

invNorm(
    
```

Modify this by adding .05) and then using the STO button followed by ALPHA and the green letter Z. Press ENTER and the value you store will be used as $z_{\alpha/2}$.

```

invNorm(.05)→Z
-1.644853626
    
```

Now we will actually do the calculation of

$$n_{\mu} = \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2$$

Again the problem is as stated.

We want to estimate the mean energy consumption level for a home in one region. We want to be 90% confident that our sample mean is within 25 kwh of the true population mean, and past data strongly suggest that the population standard deviation is 137 kwh. How large must our sample be?

Clear the home screen and use the keys to type the following;

```
(2*137/25)^2  
81.24855204
```

This is the stored Z that we just did on the previous page.

Press ENTER and the answer of 81.24855204 appears. However since we are interested in the number to be included in our survey we must round up to 82.

So the answer to the problem is 82.

The next problem involves estimating the population proportion or percentage.

$$n_p = \frac{z_{\alpha/2}^2 k}{E^2} \times 0.25$$

Here is the problem.

We want to estimate, with a maximum error of 3%, the true proportion of adult Americans who are opposed to nuclear power, and we want 95% confidence in our results. How many people should we survey?

Here we use the calculator to determine $z_{\alpha/2}$ in the same way we did in the previous problem. Since we want to be 95% confident, $\alpha = 5\%$ so that $\alpha/2 = 2.5\% = .025$. Now press 2nd DISTR as in the last problem. Select 3:invNorm(

```

0:QUIT DRAW
1:normalpdf(
2:normalcdf(
3:invNorm(
4:tpdf(
5:tcdf(
6:x²pdf(
7:↓x²cdf(
    
```

And modify the screen as below

```

invNorm(.025)→Z
-1.959963986
    
```

This is the value of $Z_{\alpha/2}$ which will be used in the formula.

We will now calculate the answer n as follows

```

(Z/.03)².25
1067.071896
1068
1068
    
```

Here is the answer that we must again **round up** since we are interested in the number of people. So, the answer to the problem is **1068**.

The answer is **1068**.

Statistics Practice Midterm #2

For the following problems, show all work and draw all distributions for full credit.

1.

A physical fitness researcher devises a test of strength and finds that the results are normally distributed with a mean of 110.0 pounds and a standard deviation of 10.4 pounds.

- (a) If a subject is randomly selected and measured, find the probability of a score between 110.0 pounds and 125.0 pounds. a. _____
- (b) If a subject is randomly selected and measured, find the probability of a score between 90.0 pounds and 115.0 pounds. b. _____
- (c) If a subject is randomly selected and measured, find the probability of a score below 95.0 pounds. c. _____
- (d) If a subject is randomly selected and measured, find the probability of a score greater than 85.0 pounds. d. _____
- (e) If a subject is randomly selected and measured, find the probability of a score between 80.0 pounds and 100.0 pounds. e. _____
- (f) Find the 20th percentile. f. _____
- (g) Sixty different subjects are randomly selected and tested. Find the probability that their mean is below 108.0 pounds. g. _____

2.

The Consumer Protection Corporation wants to estimate the proportion of Chevrolet Corvettes that have crashed. In order to obtain a sample proportion that is in error by no more than 0.04 (with 98% confidence), how many cars must be sampled?

3.

Of 346 vending machines tested, 12 are found to be defective. Construct the 98% confidence interval for the proportion of defective vending machines.

4.

A botanist measures the heights of 24 seedlings and obtains a mean and standard deviation of 41.6 cm and 4.8 cm, respectively. Construct the 98% confidence interval for the population mean.

5.

A study of the career paths of hotel general managers sent questionnaires to a simple random sample of 160 hotels belonging to major US chains. There were 114 responses. The average time the 114 managers had spent with their current company was 11.78 years with a standard deviation of 3.2 years. Give the 99% confidence interval for the mean number of years general managers of major-chain hotels have spent with their current company

6.

In order to determine rates, a motel manager wishes to find the 90% confidence interval of the mean rate that competing motels charge. He wishes to be accurate within \$2.00. If the standard deviation of the rates is \$10.00, how large a sample must he selected to get the desired information?

7. A study of 40 English composition professors showed that they spent, on average, 12.6 minutes correcting a student's term paper.

- a. Find the 90% confidence interval of the mean time for all composition papers when $\sigma = 2.5$ minutes.
- b. If a professor stated that he spent, on average, 30 minutes correcting a term paper, what would be your reaction?

8.

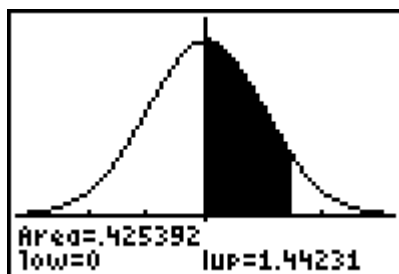
A recent study of 25 students showed that they spent an average of \$18.53 for gasoline per week. The standard deviation of the sample was \$3.00. Find the 95% confidence interval of the true mean of the population.

Answer key for Statistics Mid-term #2

1.

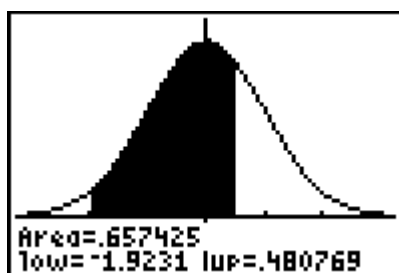
a.

```
(110-110)/10.4→A
0
(125-110)/10.4→B
1.442307692
ShadeNorm(A,B)■
```



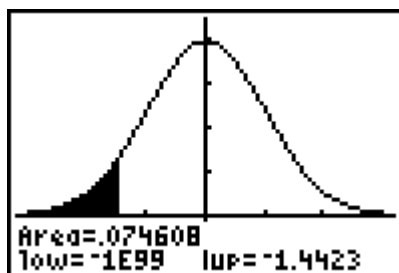
b.

```
(90-110)/10.4→A
-1.923076923
(115-110)/10.4→B
.4807692308
ShadeNorm(A,B)
```



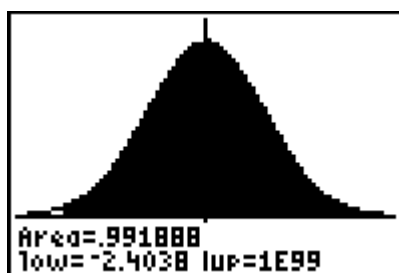
c.

```
(95-110)/10.4→A
-1.442307692
ShadeNorm(-E99,A)
■
```



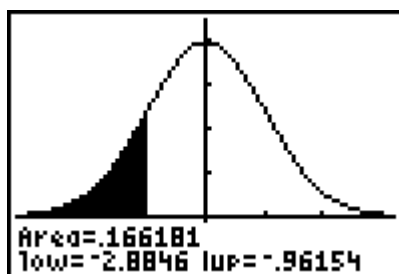
d.

```
(85-110)/10.4→A
-2.403846154
ShadeNorm(A,E99)
```



e.

```
(80-110)/10.4→A
-2.884615385
(100-110)/10.4→B
-.9615384615
ShadeNorm(A,B)■
```

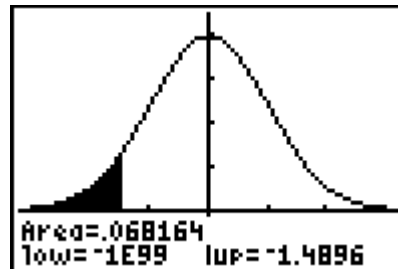


f.

```
invNorm(.20,110,
10.4)
101.2471392
102
102
```

9.

```
10.4/√(60)→S
1.342634227
(108-110)/S→A
-1.489608979
ShadeNorm(-E99,A
```



2.

```
invNorm(.01)→Z
-2.326347877
(Z/.04)²* .25
845.6085071
846
846
```

3.

```
1-PropZInt
x:12
n:346
C-Level:98
Calculate
```

```
1-PropZInt
(.0118,.05757)
P=.0346820809
n=346
```

4.

```
TInterval
Inpt:Data STATES
x̄:41.6
Sx:4.8
n:24
C-Level:98
Calculate
```

```
TInterval
(39.151,44.049)
x̄=41.6
Sx=4.8
n=24
```

5.

```
ZInterval
Inpt:Data STATS
σ:3.2
x̄:11.78
n:114
C-Level:99
Calculate
```

```
ZInterval
(11.008, 12.552)
x̄=11.78
n=114
█
```

6.

```
invNorm(.05)→Z
-1.644853626
(Z*10/2)²
67.63858627
68
68
```

7.

```
ZInterval
Inpt:Data STATS
σ:2.5
x̄:12.6
n:40
C-Level:90
Calculate
```

```
ZInterval
(11.95, 13.25)
x̄=12.6
n=40
```

8.

```
TInterval
Inpt:Data STATS
x̄:18.53
Sx:3
n:25
C-Level:95
Calculate
```

```
TInterval
(17.292, 19.768)
x̄=18.53
Sx=3
n=25
█
```

Yet Another Practice Midterm

1.

A study of 40 bowlers showed that their average score was 186. The standard deviation of the population is 6.

a. Find the 95% confidence interval of the mean score for all bowlers.

b. Find the 95% confidence interval of the mean score if a sample of 100 bowlers is used instead of a sample of 40.

c. Which interval is smaller? Explain why.

2.

An insurance company is trying to estimate the average number of sick days that full-time food service workers use per year. A pilot study found the standard deviation to be 2.5 days. How large of a sample must be selected if the company wants to be 95% sure of getting an interval that contains the true mean with a maximum error of 1 day?

3.

A medical researcher wishes to determine the percentage of females under 20 years of age who take vitamins. He wishes to be 99% confident that the estimate is within 2 percentage points of the true proportion. How large should the sample size be? A recent study of 180 females who took vitamins showed that 25% were under 20 years of age.

4.

The mean score on a dexterity test for 12-year-olds is 30. The standard deviation is 5. If a psychologist administers the test to a class of 22 students, find the probability that the mean of the sample will be between 27 and 31.

5.

A statistical analysis of 1,000 long-distance telephone calls made from the headquarters of Johnson & Shurgot Corporation indicates that the length of these calls is normally distributed with a mean of 240 seconds and a σ of 40 seconds.

(a) What percentage of these calls lasted less than 180 seconds?

(b) What is the probability that a particular call lasted between 180 and 300 seconds?

(c) **How many** calls lasted less than 180 seconds or more than 300 seconds?

(d) What percentage of the calls lasted between 110 and 180 seconds?

(e) What is the length of a particular call if only 1% of all calls are longer?

1a.

```
ZInterval
Inpt:Data Stats
σ:6
x̄:186
n:40
C-Level:95
Calculate
```

```
ZInterval
(184.14, 187.86)
x̄=186
n=40
```

1b.

```
ZInterval
Inpt:Data Stats
σ:6
x̄:186
n:100
C-Level:95
Calculate
```

```
ZInterval
(184.82, 187.18)
x̄=186
n=100
```

1c.

Increasing the sample size results in a decrease in the margin of error.

2.

```
1.00-.95
Ans/2 .05
invNorm(.025)+Z
-1.959963986
```

```
invNorm(.025)+Z
-1.959963986
(Z*2.5/1)2
24.00911766
25 25
```

3a.

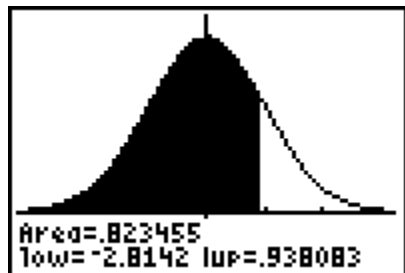
```
1.00-.99
Ans/2 .01
invNorm(.005)+Z
-2.575829303
```

```
invNorm(.005)+Z
-2.575829303
(Z/.02)2*.25
4146.810374
4147 4147
```

4.

```

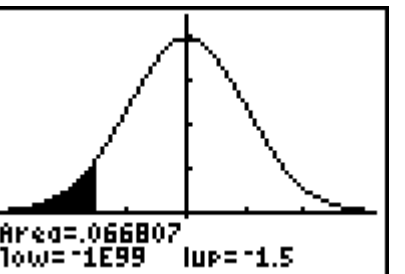
5/√(22)→S
1.066003582
(27-30)/S→A
-2.814249456
(31-30)/S→B
.938083152
ShadeNorm(A,B)■
    
```



5a.

```

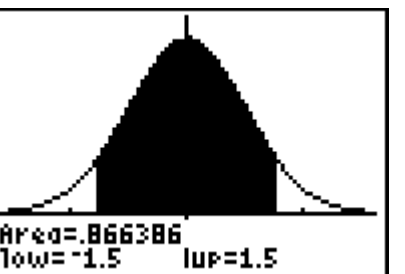
(180-240)/40→A
-1.5
ShadeNorm(-1E99,A)
    
```



5b.

```

(180-240)/40→A
-1.5
(300-240)/40→B
1.5
ShadeNorm(A,B)■
    
```



5c.

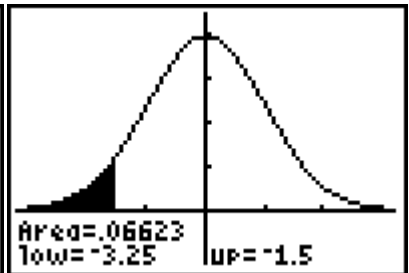
```

1-.8664
.1336
.1336*1000
134
■
    
```

5d.

```

(110-240)/40→A
-3.25
(180-240)/40→B
-1.5
ShadeNorm(A,B)■
    
```



5e.

```

invNorm(.99,240,
40)
333.0539151
334
334
    
```


Chapters 7 & 8

Hypothesis Testing (One Population)

Claims are continually being made by all sorts of people. Politicians, product manufacturers, scientists, lawyers and just every day people make claims involving a large variety of subjects. Testing these claims to see if they are to be supported or not is what is known in statistics as hypothesis testing.

To see the screen that contains these different tests press STAT and go to the TESTS menu. The first six items are all the statistical tests we will use in the course.

```
EDIT CALC TESTS
1: Z-Test...
2: T-Test...
3: 2-SampZTest...
4: 2-SampTTest...
5: 1-PropZTest...
6: 2-PropZTest...
7: ZInterval...
```

The first items 1 – 6 are hypothesis tests.

Now we will do a typical hypothesis (claim) testing problem.

The Chandler Department Store accepts only its own credit card. Among 36 randomly selected cardholders, it was found that the mean amount owed was \$175.37, while the standard deviation was \$84.77. Use a 0.05 level of significance to test the claim that the mean amount charged by all customers is greater than \$150.00.

The first 2 steps are to write the claim in symbolic form, then it's opposite.

Claim: $\mu > 150$ H_1
Opposite: $\mu \leq 150$ H_0

Label the null(w/=) & alternative hypothesis

Since we are testing a claim about a mean with a sample in which $n = 36$ (> 30) we will select 1: Z – Test...

```
EDIT CALC TESTS
1: Z-Test...
2: T-Test...
3: 2-SampZTest...
4: 2-SampTTest...
5: 1-PropZTest...
6: 2-PropZTest...
7: ZInterval...
```

Press ENTER

The next page shows the Z – Test screen with the correct inputs from the problem.

```
Z-Test
Inpt:Data  Stats
μ₀:150
σ:84.77
x̄:175.37
n:36
μ:≠μ₀ <μ₀ >μ₀
Calculate Draw
```

The direction of this symbol is always the same as the symbol in H_1 , the alternative hypothesis

The Chandler Department Store accepts only its own credit card. Among 36 randomly selected cardholders, it was found that the mean amount owed was \$175.37, while the standard deviation was \$84.77. Use a 0.05 level of significance to test the claim that the mean amount charged by all customers is greater than \$150.00.

See that you understand how the other values get filled in on the above screen. Then go to Calculate and press ENTER.

```
Z-Test
μ>150
z=1.795682435
P=.0362724657
x̄=175.37
n=36
█
```

The critical piece of information here is the p-value ($p=.0327$).

Since $p < \alpha$ ($.0327 < .05$) we reject (fail to support) the null hypothesis. Therefore we support the claim that the average amount spent by all customers is greater than \$ 150.

We will now do a problem that uses 2: T – Test under the STAT TESTS menu

```

EDIT CALC TESTS
1:Z-Test...
2:T-Test...
3:2-SampZTest...
4:2-SampTTest...
5:1-PropZTest...
6:2-PropZTest...
7↓ZInterval...
    
```

A test of sobriety involves measuring of the subject's motor skills. Twenty (20) randomly selected sober subjects take the test and produce a mean of 41.0 with a standard deviation of 3.7. At the 0.01 level of significance, test the claim that the mean for this test is equal to 35.0.

Again we symbolically state the claim, it's opposite, then label the null and alternative hypothesis.

Claim: $\mu = 35$ H_0 (null hypothesis)
 Opposite: $\mu \neq 35$ H_1 (alternative hypothesis)

Now select 2: T – Test under the STAT TESTS menu and press ENTER. Set up the screen as seen below. Verify that you know where these numbers come from in the problem.

```

T-Test
Inpt:Data TESTS
μ₀:35
x̄:41
Sx:3.7
n:20
μ: ≠ μ₀ < μ₀ > μ₀
Calculate Draw
    
```

The direction of this symbol is always the same as the symbol in H_1 , the alternative

Go to Calculate and press ENTER. You will see the next screen

```

T-Test
μ≠35
t=7.252112359
P=6.9743223E-7
x̄=41
Sx=3.7
n=20
    
```

The critical piece of information here is again the P – value ($p=6.97E-7$). This really means 6.97×10^{-7} or approximately .0000007.

So since $p < \alpha$ ($.0000007 < .01$), we therefor reject (fail to support) the null hypothesis. Therefore, we reject the claim that the mean for this test is 35.

The last hypothesis test we will do involving one population is testing a claim about a population proportion (or percentage).

A supplier of computer disks claims that no more than 1% of the disks are defective. In a random sample of 600 disks, it is found that 3% are defective, but the supplier claims that this is only a sample fluctuation. At the 0.01 level of significance, test the supplier's claim that no more than 1% are defective.

Again we symbolically state the claim, it's opposite, then label the null and alternative hypothesis.

Claim: $p \leq .01$ H_0 (null hypothesis)
 Opposite $p > .01$ H_1 (alternative hypothesis)

For this problem we will use 5: 1 – PropZTest from the STAT TESTS menu as shown here.

```

EDIT CALC 11:51:16
1:Z-Test...
2:T-Test...
3:2-SampZTest...
4:2-SampTTest...
5:1-PropZTest...
6:2-PropZTest...
7:ZInterval...
    
```

Press ENTER and screen below appears. Fill it in as shown. Verify that you understand where the different numbers come from in the problem.

```

1-PropZTest
P0:.01
x:600*.03
n:600
PROP#P0 <P0 >P0
Calculate Draw
    
```

```

1-PropZTest
P0:.01
x:18
n:600
PROP#P0 <P0 >P0
Calculate Draw
    
```

Notice that in this problem we are not told the number of defective disks, but that 3% were found to be defective. Hence the number is $600 \times .03$. When you move the cursor the calculator figures out what this multiplication problem equals. NOTE: This has to be a whole number like 18.

Press ENTER and the following screen should appear.

```

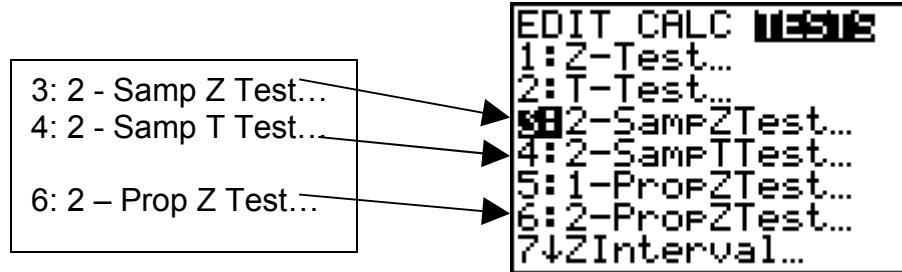
1-PropZTest
PROP>.01
z=4.923659639
P=4.2532555E-7
P̂=.03
n=600
    
```

The important piece of information is $p = 4.25E-7$ which means $p = .000000425$. So since $p < \alpha$ ($.000000425 < .01$), we reject the null hypothesis and therefore reject the supplier's claim

Chapter 6

Hypothesis Testing and Confidence Intervals (Two Populations)

In testing claims made about two different populations we will use the following items from the STAT TESTS menu



First we will do problem which uses 3: 2 – Samp Z Test...

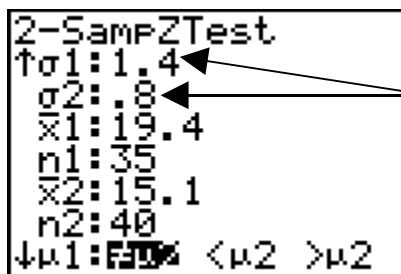
Two types of flares are tested for their burning times (in minutes) and sample results are given at the right. Use a 0.05 significance level to test the claim that the two brands have equal means.

<u>Brand X</u>	<u>Brand Y</u>
$n = 35$	$n = 40$
$\bar{x} = 19.4$	$\bar{x} = 15.1$
$S = 1.4$	$s = 0.8$

Again we symbolically state the claim, it's opposite, then label the null and alternative hypothesis.

Claim: $\mu_1 = \mu_2$ H_0 (null hypothesis)
Opposite $\mu_1 \neq \mu_2$ H_1 (alternative hypothesis)

Because sample sizes are both > 30 we use **3: 2 – Samp Z Test...** from the STAT TESTS menu and enter the information as indicated in the next screen below.



Note: In Z Tests the standard deviations of the samples (s) can be used for the standard deviations of the populations (σ).

Go to Calculate and press ENTER. The following screen should appear.

```
2-SampZTest
μ1≠μ2
z=16.02515384
P=9.00321E-58
x̄1=19.4
x̄2=15.1
↓n1=35
```

The important piece of information is $p = 9E-58$ which means p is very small compared to $\alpha = .05$. So since $p < \alpha$ we reject the null hypothesis and therefore reject the claim that the two populations have equal means.

Next we will answer a related question. What is the 95% confidence interval for the difference between the means of the two populations.

To answer this we go to the STAT TESTS menu and select 9: 2 – Samp Z Int...

```
EDIT CALC TESTS
4: 2-SampTTest...
5: 1-PropZTest...
6: 2-PropZTest...
7: ZInterval...
8: TInterval...
9: 2-SampZInt...
0: 2-SampTInt...
```

```
2-SampZInt
↑σ1: 1.4
σ2: .8
x̄1: 19.4
n1: 35
x̄2: 15.1
n2: 40
↓C-Level: 95
```

Notice that all of the information you input into the screen for 3: 2 – Samp Z Test... is automatically transferred to this screen. You only need to fill in the C – Level: 95

Go to Calculate and press ENTER.

```
2-SampZInt
(3.7741, 4.8259)
x̄1=19.4
x̄2=15.1
n1=35
n2=40
```

This is the 95% Confidence Interval for the difference in the two populations means.

To express this confidence interval using **Margin of Error** notation;

1. Find the 'center' of the interval by averaging the ends $(3.7741+4.8259)/2 = 4.3$
2. Subtract the 'center' or average from the larger end $4.8259 - 4.3 = .5259$
3. The interval can now be written as the 'center' \pm Margin of Error or **$4.3 \pm .5259$**

We will now do a problem similar to the last, but with the major difference that the size of one (or both) of the samples from the two populations we are comparing are less than or equal to 30 ($n \leq 30$).

A statistics student believes that women executives are paid less for the same jobs done by male executives. The student interviews 26 women and finds that their sample mean is \$ 87,736 with a standard deviation of \$ 2,645. A sample of 30 males interviewed resulted in a sample mean of \$ 92,254 and a standard deviation of \$ 3,003. At a .02 level of significance is the students hypothesis correct?

In this problem population 1 will be the female executives and population 2 will be the male executives. It is important to make this assignment for these kinds of problems.

Claim: $\mu_1 < \mu_2$ H_1 (alternative hypothesis)
 Opposite $\mu_1 \geq \mu_2$ H_0 (null hypothesis)

We again go to the STAT TESTS menu and this time select 4: 2 – Samp T Test...

```

EDIT CALC TESTS
1:Z-Test...
2:T-Test...
3:2-SampZTest...
4:2-SampTTest...
5:1-PropZTest...
6:2-PropZTest...
7:ZInterval...
  
```

Press ENTER and input the following summary statistics into the screens you see as you scroll down.

```

2-SampTTest
Inpt:Data TESTS
x1:87736
Sx1:2645
n1:26
x2:92254
Sx2:3003
↓n2:30
  
```

These are the two screens you see when you scroll down

```

2-SampTTest
↑Sx1:2645
n1:26
x2:92254
Sx2:3003
n2:30
μ1:≠μ2 TESTS >μ2
↓Pooled:No WER
  
```

This question Pooled: No Yes is a new. It requires us determining whether the two populations have equal standard deviations. To find that out we need to go to STAT TESTS menu and select D: 2 – Samp F Test...

```

EDIT CALC TESTS
0:2-SampTInt...
A:1-PropZInt...
B:2-PropZInt...
C:X2-Test...
D:2-SampFTest...
E:LinRegTTest...
F:ANOVA(
  
```

When you do you should see the following screen. Notice that all the summary statistics you input into 4: 2 – Samp T Test... has been automatically transferred to this new screen D: 2-Samp F Test.

```

2-SampFTest
Inpt:Data STATS
Sx1:2645
n1:26
Sx2:3003
n2:30
σ1:≠ <σ2 >σ2
Calculate Draw
    
```

We are actually conducting a separate hypothesis test in which;

$\sigma_1 = \sigma_2$ H_0 (null hypothesis)

$\sigma_1 \neq \sigma_2$ H_1 (alternative hypothesis)

Notice the selection of the \neq symbol from the alternative hypothesis.

Go to Calculate and press ENTER. You will have to wait a little while the calculator does its thing and then you should see the following screen.

```

2-SampFTest
σ1≠σ2
F=.7757837678
P=.5226550751
Sx1=2645
Sx2=3003
↓n1=26
    
```

The important piece of information here is $p = .523$. Thus, $p > \alpha$ so we support the null hypothesis that the standard deviations are equal. This means we **POOL YES**.

Now we return to the STAT TESTS menu and select 4: 2 – Samp T Test...

```

2-SampTTest
↑n1:26
x2:92254
Sx2:3003
n2:30
μ1:≠μ2 < >μ2
Pooled:No Yes
Calculate Draw
    
```

Once you have determined from the 2-SampFTest that you will or will not Pool, continue with the remainder of the 2 – SampTTest by going to Calculate and pressing ENTER>

```

2-SampTTest
μ1<μ2
t=-5.931200918
P=1.0990261E-7
df=54
x1=87736
↓x2=92254
    
```

The important piece of information is $p = 1 \times 10^{-7}$ so that $p < \alpha$. So we reject the null hypothesis and support the claim that women executives make less than male executives.

Note: This problem is probably one of the most difficult types of hypothesis testing problems you will encounter. So, if you can follow and eventually do this one, then you are doing well.

As an extension of the last problem we will consider the following. How much different are the female and male executive's salaries?

Construct a 98% Confidence Interval for the difference in the salaries of female and male executives

Go to STAT TESTS and scroll down to 0: 2 – SampTInt...

```

EDIT CALC YES
9↑2-SampZInt...
002-SampTInt...
A: 1-PropZInt...
B: 2-PropZInt...
C: X2-Test...
D: 2-SampFTest...
E↓LinRegTTest...
    
```

You will notice that all of the information you have been putting into the calculator has again been automatically transferred to these new screens. It is a good idea to check it anyway. You will have to input the C – Level 98 for this problem

```

2-SampTInt
Inpt:Data DATA
x1:87736
Sx1:2645
n1:26
x2:92254
Sx2:3003
↓n2:30
    
```

```

2-SampTInt
↑n1:26
x2:92254
Sx2:3003
n2:30
C-Level:95
Pooled:No NO
Calculate
    
```

Go to Calculate and press ENTER. You will have to wait a little bit, then you should see the following screen.

```

2-SampTInt
(-6045, -2991)
df=54
x1=87736
x2=92254
Sx1=2645
↓Sx2=3003
    
```

The above interval shows that women executives make between \$ 6045 and \$ 2991 less than male executives. The two negative signs mean population 1 was less than population 2. To write this using Margin of Error notation;

```

92254-87736      4518
(6045+2991)/2   4518
6045-4518       1527
█
    
```

The middle of the interval can be determined in either of the two methods shown.

The Margin of Error is determined by subtraction of the middle of the interval from either “end”).

The answer; \$ 4518 |\$ 1527

We will next do a problem that tests a claim made involving proportion (or percentage) of two populations.

A radio station manager conducts an audience survey and finds that among 275 men surveyed, 18 were listeners while there were 31 listeners among the 325 women surveyed. At the 0.05 level of significance, test the claim that the proportions of male and female listeners are the same.

Again we symbolically state the claim, it's opposite, then label the null and alternative hypothesis.

Claim: $p_1 = p_2$ H_0 (null hypothesis)
Opposite $p_1 \neq p_2$ H_1 (alternative hypothesis)

Go to STAT TESTS menu and select 6: 2-Prop Z Test

```
EDIT CALC TESTS
1:Z-Test...
2:T-Test...
3:2-SampZTest...
4:2-SampTTest...
5:1-PropZTest...
6:2-PropZTest...
7:ZInterval...
```

The next screen should appear and you should fill it in as shown.

```
2-PropZTest
x1:18
n1:275
x2:31
n2:325
P1:F1 <P2 >P2
Calculate Draw
```

Select Calculate and press ENTER and the next screen should appear.

```
2-PropZTest
P1≠P2
z=-1.333882408
P=.1822425222
P1=.0654545455
P2=.0953846154
↓P=.0816666667
```

The important piece of information is $p = .182$ which means p is greater than $\alpha = .05$. So since $p > \alpha$ we support the null hypothesis and therefore support the claim that the two populations have equal proportions

An extension of the above problem would be to find the 95% confidence interval for estimated difference between the proportions of the two populations. To do this we again go to the STAT TESTS and select B: 2 – Prop Z Int...

```
EDIT CALC TESTS
8↑TInterval...
9: 2-SampZInt...
0: 2-SampTInt...
A: 1-PropZInt...
B: 2-PropZInt...
C:  $\chi^2$ -Test...
D↓2-SampFTest...
```

The next screen shows what will appear. You should see all the inputs from the last problem automatically transferred to this screen.

```
2-PropZInt
x1:18
n1:275
x2:31
n2:325
C-Level:95
Calculate
```

You will have to be sure that the C-Level: is set to 95 as you see here.

Go to Calculate and press ENTER. The next screen should appear.

```
2-PropZInt
(-.0732,.01336)
p1=.0654545455
p2=.0953846154
n1=275
n2=325
```

Here is the 95% confidence interval for estimated difference between the proportions of the two populations. Notice that it includes **zero** since we supported the claim that the two populations had equal proportions.

The final problem we will do is the special case of a claim concerning the means of two dependent populations. This is typically a **before and after** situation with the claim involving the mean difference.

A test of abstract reasoning is given to a random sample of students before and after they completed a formal logic course. The results are given below. At the 0.05 significance level, test the claim that the mean score is improved by the course.

Before	74	83	75	88	84	63	93	84	91	77
After	73	77	70	77	74	67	95	83	84	75

To do this problem we will be using some of our lists L1 – L6. So first clear all the lists in the list editor. Next input the before and after values into L1 and L2 respectively.

L1	L2	L3	2
84	74		
63	67		
93	95		
84	83		
91	84		
77	75		

L2(1) =			

Check to see that all the entries are correct after inputting the values. Also note that the two lists should be the same lengths.

L1	L2	L3	3
74	73		
83	77		
75	70		
88	77		
84	74		
63	67		
93	95		
L3 = L2 - L1			

Next move the cursor to the very top of L3

With the cursor in this position type in the equation for L3 that you see here. $L3 = L2 - L1$

Press ENTER and the next screen should appear.

L1	L2	L3	3
74	73	-1	
83	77	-6	
75	70	-5	
88	77	-11	
84	74	-10	
63	67	4	
93	95	2	
L3(1) = -1			

L3 represents the difference in the scores before and after the course was taken

Before and After Problems: Always do the list subtraction of **BIG – SMALL** based on the story in the claim. For example, in this problem the claim is that the mean score is improved by the course so that the numbers in L2 would be larger than L1 so that $L3 = L2 - L1$. The claim is that the average difference (μ_d) would be positive or > 0 . [If instead the “story” in the problem involved a weight lose program which claimed that the population loses at least 10 pounds with this particular program, then the subtraction would be $L1 - L2$ and the claim is that the average difference (μ_d) would be “at least 10 pounds” or ≥ 10 .]

Again we symbolically state the claim, it’s opposite, then label the null and alternative hypothesis.

Claim: $\mu_d > 0$ H_1 (alternative hypothesis)
 Opposite $\mu_d \leq 0$ H_0 (null hypothesis)

For this problem we go to the STAT TESTS menu and select 2: T – Test...

```

EDIT CALC TESTS
1:Z-Test...
2:T-Test...
3:2-SampZTest...
4:2-SampTTest...
5:1-PropZTest...
6:2-PropZTest...
7↓ZInterval...
  
```

```

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

The list containing the differences is L3
 Data instead of summary Stats

The other inputs into the screen above are the same as for other hypothesis tests. Go to Calculate and press ENTER.

The following screen should appear.

```

T-Test
μ > 0
t = -2.365990288
P = .9789073747
x̄ = -3.7
Sx = 4.945255864
n = 10
  
```

The important piece of information is $p = .989$ which means p is greater than $\alpha = .05$. So since $p > \mu$ we support the null hypothesis and therefore reject the claim that the course improves test scores.

We support the null hypothesis. Therefore, we reject (fail to support) the claim.

Statistics Practice Midterm#3

1.

A large software company gives job applicants a test of programming ability and the mean for that test has been 160. Twenty-five job applicants are randomly selected from one large university and they produce a mean and standard deviation of 183 and 12, respectively. Use a 0.05 level of significance to test the claim that this sample comes from a population with an aptitude level of at least 160.

2.

A poll of 1068 adult Americans reveals that 48% of the voters surveyed prefer the Democratic candidate for the presidency. At the 0.05 level of significance, test the claim that no more than half of all voters prefer the Democrat candidate.

3.

A machine is supposed to dispense a liquid drug into bottles in such a way that the mean of the contents is 150 milliliters. A new machine is tested on a sample of 24 containers and the mean and standard deviation for this sample group is found to be 146 and 26 milliliters respectively. At the 0.05 level of significance, test the claim that the new machine dispenses the liquid drug as it is supposed to.

4. The average income of 16 families who reside in a large metropolitan city is \$54,356, and the standard deviation is \$8256. The average income of 12 families who reside in a suburb of the same city is \$46,512, with a standard deviation of \$1311. At a $\alpha = 0.05$, can one conclude that the standard deviations of the income of the families who reside within the city and those who reside in the suburb are equal?

At $\alpha = 0.05$, can one conclude that the income of the families who reside within the city is greater than that of those who reside in the suburb?

Construct a 95% confidence interval for the difference $\mu_x - \mu_y$ based on the sample data from this problem.

5.

A survey of consumer preferences involves two different market areas. Of 150 consumers surveyed in area A, 20 preferred a certain product. There were 350 consumers surveyed in market area B, and 52 of them preferred the same product. At the 0.01 level of significance, test the claim that the same market shares prefer the product.

Construct a 95% confidence interval for the difference between the two population market share percentages described in this problem.

6.

A test of writing ability is given to a random sample of students before and after they completed a formal writing course. The results are given below. At the 0.05 significance level, test the claim that the mean score is not affected by the course.

Before	70	80	92	99	93	97	76	63	68	71	74
After	69	79	90	96	91	95	75	64	62	64	76

7. Sample data are collected from two brands of a product and the results are given at the right. At the 0.05 significance level, test the claim that these samples come from populations with equal standard deviations.

Brand X

Brand Y

n =	20	n =	27
x =	125.4	x =	119.3
S =	8.2	S =	2.6

Refer to the same sample data given in this problem and test the claim that the two populations have equal means. Use a 0.05 significance level.

Construct a 95% confidence interval for the difference $\mu_x - \mu_y$ based on the sample data from this problem.

8.

An educational testing company has been using a standard test of verbal ability and the mean and standard deviation has been 430 and 130, respectively. In analyzing a new version of that test, it is found that a sample of 100 randomly selected subjects produces a mean and a standard deviation of 424 and 155, respectively. At the 0.05 level of significance, test the claim that the new version has a mean equal to that of the past version.

9.

A manufacturer considers her production process to be out of control when defects exceed 3%. In a random sample of 170 items, the defect rate is 5.9% but the manager claims that this is only a sample fluctuation and production is not really out of control. At the 0.01 level of significance, test the manager's claim.

Answer key to midterm #3

1.

Claim: $\mu \geq 160$ H_0

Opposite: $\mu < 160$ H_1

<pre>T-Test Inpt:Data State μ₀:160 x̄:183 Sx:12 n:25 μ:≠μ₀ μ₀ >μ₀ Calculate Draw</pre>	<pre>T-Test μ<160 t=9.583333333 P=.9999999994 x̄=183 Sx=12 n=25</pre>
---	--

The important piece of information is $p = .999$ which means p is greater than $\alpha = .05$. So since $p > \alpha$ we support the null hypothesis and therefore support the claim that this sample comes from a population with an aptitude level of at least 160.

2.

Claim: $p \leq .50$ H_0

Opposite: $p > .50$ H_1

<pre>1-PropZTest P₀:.5 x:.48*1068 n:1068 PROP≠P₀ <P₀ >P₀ Calculate Draw</pre>	<pre>1-PropZTest P₀:.5 x:512.64 n:1068 PROP≠P₀ <P₀ >P₀ Calculate Draw</pre>	<pre>1-PropZTest P₀:.5 x:513 n:1068 PROP≠P₀ <P₀ >P₀ Calculate Draw</pre>
<pre>1-PropZTest PROP>.5 z=-1.285179129 P=.900635087 P̂=.4803370787 n=1068</pre>		

The important piece of information is $p = .901$ which means p is greater than $\alpha = .05$. So since $p > \alpha$ we support the null hypothesis and therefore support the claim that no more than half of all voters prefer the Democrat candidate.

3.

Claim: $\mu = 150$ H_0

Opposite: $\mu \neq 150$ H_1

<pre>T-Test Inpt:Data STAT2 μ₀:150 x̄:146 Sx:26 n:24 μ:150 <μ₀ >μ₀ Calculate Draw</pre>	<pre>T-Test μ≠150 t=-.7536891516 p=.4586826604 x̄=146 Sx=26 n=24</pre>
---	--

The important piece of information is $p = .459$ which means p is greater than $\alpha = .05$. So since $p > \alpha$ we support the null hypothesis and therefore support the claim that the new machine dispenses the liquid drug as it is supposed to.

4. a.

Claim: $\sigma_1 = \sigma_2$ H_0

Opposite: $\sigma_1 \neq \sigma_2$ H_1

<pre>2-SampFTest Inpt:Data STAT2 Sx1:8256 n1:16 Sx2:1311 n2:12 σ₁:150 <σ₂ >σ₂ Calculate Draw</pre>	<pre>2-SampFTest σ₁≠σ₂ F=39.65829009 p=4.2747726E-7 Sx1=8256 Sx2=1311 ↓n1=16 █</pre>
--	--

The important piece of information is $p = 4.3 \text{ E}-7 = .0000000043$ which means p is much less than $\alpha = .05$. So since $p < \alpha$ we reject the null hypothesis and therefore reject the claim that the two populations have equal standard deviations or variances.

4. b.

Claim: $\mu_1 > \mu_2$ H_1

Opposite: $\mu_1 \leq \mu_2$ H_0

<pre>2-SampTTest Inpt:Data STAT2 x̄₁:54356 Sx₁:8256 n₁:16 x̄₂:46512 Sx₂:1311 ↓n₂:12</pre>	<pre>2-SampTTest ↑n₁:16 x̄₂:46512 Sx₂:1311 n₂:12 μ₁≠μ₂ <μ₂ STAT2 Pooled:STAT2 Yes Calculate Draw</pre>	<pre>2-SampTTest μ₁>μ₂ t=3.738069102 p=8.9609354E-4 df=16.00090808 x̄₁=54356 ↓x̄₂=46512</pre>
--	---	--

The important piece of information is $p = .000896$ which means p is less than $\alpha = .05$. So since $p < \alpha$ we reject the null hypothesis and therefore support the claim that the income of the families who reside within the city is greater than that of those who reside in the suburb.

4. c.

<pre>2-SampTInt Inpt:Data x1:54356 Sx1:8256 n1:16 x2:46512 Sx2:1311 ↓n2:12</pre>	<pre>2-SampTInt ↑n1:16 x2:46512 Sx2:1311 n2:12 C-Level:95 Pooled: Yes Calculate</pre>	<pre>2-SampTInt (3395.6,12292) df=16.00090808 x1=54356 x2=46512 Sx1=8256 ↓Sx2=1311</pre>
--	---	--

<pre>(3395.6+12292)/2 7843.8 12292-7843.8 4448.2</pre>
--

Another way of expressing this same interval uses the center of the interval **7843.8** and the Margin of Error **4448.2**

This same interval could be written;
7843.8 ± 4448.2

5.a.

Claim: $p_1 = p_2$ H_0
 Opposite $p_1 \neq p_2$ H_1

<pre>2-PropZTest x1:20 n1:150 x2:52 n2:350 P1: <P2 >P2 Calculate Draw</pre>	<pre>2-PropZTest P1≠P2 z=-.4447410374 P=.6565069557 P1=.1333333333 P2=.1485714286 ↓P=.144</pre>
---	---

The important piece of information is $p = .657$ which means p is greater than $\alpha = .01$. So since $p > \alpha$ we support the null hypothesis and therefore support the claim that the same market shares prefer the product.

5.b.

<pre>2-PropZInt x1:20 n1:150 x2:52 n2:350 C-Level:95 Calculate</pre>	<pre>2-PropZInt (-.0812,.0507) P1=.1333333333 P2=.1485714286 n1=150 n2=350</pre>
--	--

Since this interval includes 0 it is consistent with the answer to part a.

6.

Claim: $\mu_d = 0$ H_0

Opposite: $\mu_d \neq 0$ H_1

L1	L2	L3	3
70	69		
80	79		
92	90		
99	96		
93	91		
97	95		
76	75		
L3=L2-L1			
T-Test			
Inpt: STAT Stats			
$\mu_0: 0$			
List: L3			
Freq: 1			
$\mu: \mu_0 < \mu_0 > \mu_0$			
Calculate Draw			

L1	L2	L3	3
70	69	-1	
80	79	-1	
92	90	-2	
99	96	-3	
93	91	-2	
97	95	-2	
76	75	-1	
L3(1)=-1			
T-Test			
$\mu \neq 0$			
t=-2.507132682			
P=.0310648639			
$\bar{x} = -2$			
Sx=2.645751311			
n=11			

The important piece of information is $p = .03$ which means p is less than $\alpha = .05$. So since $p < \alpha$ we reject the null hypothesis and therefore reject the claim that the mean score is not affected by the course.

7.a.

Claim: $\sigma_1 = \sigma_2$ H_0

Opposite: $\sigma_1 \neq \sigma_2$ H_1

2-SampFTest	2-SampFTest
Inpt: Data STAT	$\sigma_1 \neq \sigma_2$
Sx1: 8.2	F=9.946745562
n1: 20	P=2.7664961E-7
Sx2: 2.6	Sx1=8.2
n2: 27	Sx2=2.6
$\sigma_1: \sigma_1 < \sigma_1 > \sigma_1$	↓n1=20
Calculate Draw	

The important piece of information is $p = 2.77E-7$ which means p is less than $\alpha = .05$. So since $p < \alpha$ we reject the null hypothesis and therefore reject the claim that that these samples come from populations with equal standard deviations.

7.b.

Claim: $\mu_1 = \mu_2$ H_0

Opposite: $\mu_1 \neq \mu_2$ H_1

<pre>2-SampTTest Inpt:Data STATES x1:125.4 Sx1:8.2 n1:20 x2:119.3 Sx2:2.6 ↓n2:27</pre>	<pre>2-SampTTest ↑n1:20 x2:119.3 Sx2:2.6 n2:27 μ1:6.04 <μ2 >μ2 Pooled:<input checked="" type="checkbox"/> Yes Calculate Draw</pre>	<pre>2-SampTTest μ1≠μ2 t=3.209472799 p=.0040655414 df=21.84671746 x1=125.4 ↓x2=119.3</pre>
---	--	--

From the result of 8.a. Pool No

The important piece of information is $p = .004$ which means p is less than $\alpha = .05$. So since $p < \alpha$ we reject the null hypothesis and therefore reject the claim that that these samples come from populations with equal means.

7.c.

<pre>2-SampTInt Inpt:Data STATES x1:125.4 Sx1:8.2 n1:20 x2:119.3 Sx2:2.6 ↓n2:27</pre>	<pre>2-SampTInt ↑n1:20 x2:119.3 Sx2:2.6 n2:27 C-Level:95 Pooled:<input checked="" type="checkbox"/> Yes Calculate</pre>	<pre>2-SampTInt (2.1567,10.043) df=21.84671746 x1=125.4 x2=119.3 Sx1=8.2 ↓Sx2=2.6</pre>
--	---	---

This answer can be given in terms of the interval difference (2.1567, 10.043) or in terms of the Margin of Error, which is 3.94315.

```
(2.1567+10.043)/
2
6.09985
10.043-6.09985
3.94315
```

Answer:
(2.1567, 10.043) or
6.09985 ± 3.94315

8.

Claim: $\mu = 430$ H_0

Opposite: $\mu \neq 430$ H_1

<pre>Z-Test Inpt:Data [STAT] μ₀:430 σ:155 x̄:424 n:100 μ: [430] <μ₀ >μ₀ Calculate Draw</pre>	<pre>Z-Test μ≠430 z=-.3870967742 P=.698684665 x̄=424 n=100</pre>
---	--

The important piece of information is $p = .699$ which means p is greater than $\alpha = .05$. So since $p > \alpha$ we support the null hypothesis and therefore support the claim that the new version has a mean equal to that of the past version.

9.

Claim: $p \leq .03$ H_0

Opposite: $p > .03$ H_1

<pre>1-PropZTest P₀:.03 x:170*.059 n:170 PROP≠P₀ <P₀ >P₀ Calculate Draw</pre>	<pre>1-PropZTest P₀:.03 x:10.03 n:170 PROP≠P₀ <P₀ >P₀ Calculate Draw</pre>	<pre>1-PropZTest P₀:.03 x:11 n:170 PROP≠P₀ <P₀ >P₀ Calculate Draw</pre>
---	--	---

<pre>1-PropZTest PROP>.03 z=2.65265675 P=.0039930943 P̂=.0647058824 n=170 █</pre>
--

The important piece of information is $p = .004$ which means p is less than $\alpha = .01$. So since $p < \alpha$ we reject the null hypothesis and therefore reject the claim that $p \leq .03$ and reject the claim that this is only a sample fluctuation and production is not really out of control.

Chapter 9

Correlation and Regression Analysis

This chapter will again use the list editor. So, clear all the lists in the list editor and be sure that all PLOTS are OFF and that the Y= function toolbox is clear.

An Algebra test is given to students at the beginning of a calculus course. The Algebra test scores and the final calculus grades are collected for several randomly selected students and the results are given below.

Algebra	36	48	45	43	35	30	20	47
Calculus	77	92	87	81	74	63	52	88

First we will use the list editor and L1 and L2 to input the data for the Algebra and Calculus scores as follows.

L1	L2	L3	Z
45	87		
43	81		
35	74		
30	63		
20	52		
47	88		

L2(9) =			

Double-check the lists to see that the entries are correct and that the lists are of the same length.

- Draw a scatter plot diagram showing the relationship between the Algebra and Calculus test scores.

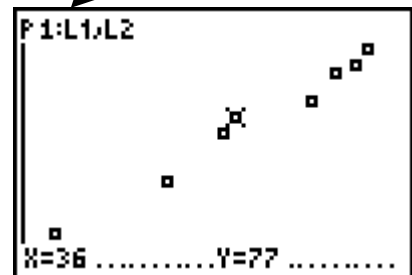
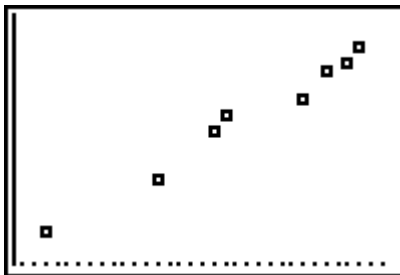
To do this go to STAT PLOT, select 1: Plot 1 and enter the input on the next screen.

Plot1	Plot2	Plot3
Off	Off	Off
Type:		
Xlist:	L1	
Ylist:	L2	
Mark:		

Now use the TRACE button to see the actual values for the points on the lists

Note: This tells you that this is Plot 1: Which uses L1 and L2

Press ZOOM 9 and the following Graph will appear.



Does a significant linear correlation exist between the algebra test scores and the calculus final grades at the 0.05 significance level? Explain your answer.

To answer this question we go to STAT TESTS and scroll to E: LinReg T Test...

```

EDIT CALC TESTS
0: 2-SampTInt...
A: 1-PropZInt...
B: 2-PropZInt...
C: X2-Test...
D: 2-SampFTest...
E: LinRegTTest
F: ANOVA(
    
```

Fill in the input as shown in the screen.

```

LinRegTTest
Xlist: L1
Ylist: L2
Freq: 1
B & P: EQ < 0 > 0
RegEQ: Y1
Calculate
    
```

To paste Y_1 here follow these steps;
 1. Press the VARS button
 2. Use the right arrow to go to Y – VARS
 3. Press 1: Function
 4. Press 1: Y_1

Then go to Calculate and press ENTER. The screens below show what the result is when the screen is scrolled.

This is the general form of the equation for the regression line.

This is the p-value = .000004, which means we reject the null hypothesis that $r=0$ and thus there is a significant correlation

These are the values of a and b to give the regression line's equation

This is the coefficient of determination.

This is the correlation coefficient.

```

LinRegTTest
y=a+bx
B≠0 and P≠0
t=16.11861232
p=3.6249526E-6
df=6
↓a=24.03658537
    
```

```

LinRegTTest
y=a+bx
B≠0 and P≠0
↑b=1.387195122
s=2.204254939
r2=.9774274572
r=.9886493095
    
```

- Draw and find the equation of the regression line. _____

To find the equation of the regression line press the $Y =$ button, the following appears.

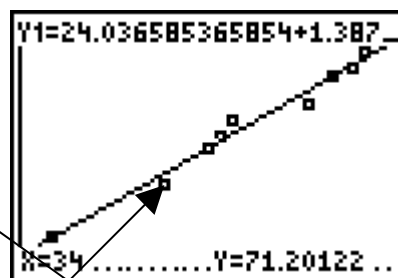
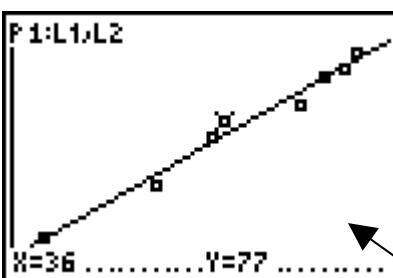
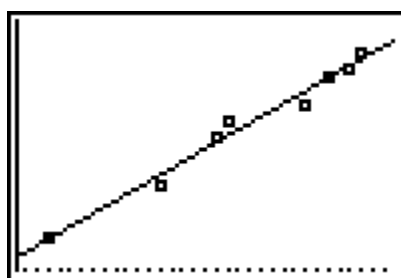
```

P1ot2 P1ot3
\Y1=24.036585365
854+1.3871951219
512X
\Y2=
\Y3=
\Y4=
\Y5=

```

The equation of the regression line can now be rounded of to read;
 $Y = 24 + 1.4 X$

To draw of graph the regression line either press ZOOM 9 or GRAPH.



Press the TRACE button and the up arrow to obtain these second two screens.

- Is there a fairly good predictive value between the score a student obtained on the algebra test and the final grade he/she got in calculus? Explain.

This question should be answered by the hypothesis test involving the $r = 0$ null hypothesis. In other words, if the p – value is very small, the more confident we can be that there is a correlation and therefore the predictive value is good.

- What percent of the data can be explained using the regression line. _____

The answer to this question is gotten using the coefficient of determination (r^2). In this case $r^2 = .9774$ so 97.74% of the data can be explained using the regression model.

- What would be the predicted calculus grade for a person with an algebra test score of 50 according to the regression model?

To answer this question, first go to the home screen and paste Y1.

To paste Y_1 here follow these steps;

1. Press the VARS button
2. Use the right arrow to go to Y – VARS
3. Press 1: Function
4. Press 1: Y_1

➔

$Y_1(50)$
93.39634146

Then type in (50) and press ENTER.

The answer is that the regression model predicts that the person would get about 93.4 as a calculus score.

Problems;

For each of the problems below answer the following questions;

- a. Is there a correlation between the x and y variable? Explain why.
- b. Compute the correlation coefficient. _____.
- c. What is the coefficient of determination? _____.
- d. What percent of these data can be explained using the regression model?
 _____(
- e. Sketch the scatter plot diagram that represents this data on the axes below.
- f. What is the equation of the regression line model for these data?)
- g. Is this a positive or negative correlation? In your own words, what does that mean in this problem?
- h. Predict a y value for any given x value.

1.

The following data were obtained from a survey of the number of years a person smoked and the percentage of lung damage. Predict the percentage of lung damage for a person who has smoked for 30 years.

Years	22	14	31	36	9	41	19
Damage	20	14	54	63	17	71	23

2.

A researcher wishes to determine whether a person's age is related to the number of hours he or she jogs per week. The data for the sample follow,

Age	34	22	40	48	56	62
Hours	7	3	10	15	10	14

3.

A study was conducted to determine the relationship between a person's monthly income and the amount that person spends per month for recreation. The data from the sample follow.

Income	\$800	\$1200	\$1000	\$900	\$850	\$907	\$1100
Amount	\$60	\$200	\$160	\$135	\$45	\$90	\$150

4.

A physician wishes to know whether there is a relationship between a father's weight (in pounds) and his newborn son's weight (in pounds). The data are given below.

Father's Weight	176	160	187	210	196	142	205	215
Son's Weight	6.6	8.2	9.2	7.1	8.8	9.3	7.4	8.6

5.

A study is done to see whether there is a relationship between a student's grade point average and the number of hours the student studies per week. The data are shown below. If there is a significant relationship, predict the grade point average of a student who studies 10 hours per week.

Hours	13	12	9	15	5	7	16
GPA	3.1	3.5	3.0	4.0	1.7	3.2	3.7

ANSWER KEY

1.

LinRegTTest $y=a+bx$ $\beta \neq 0$ and $\rho \neq 0$ $t=7.306674959$ $P=7.5209193E-4$ $df=5$ $\downarrow a=-10.94379712$ ■	LinRegTTest $y=a+bx$ $\beta \neq 0$ and $\rho \neq 0$ $\uparrow b=1.968642906$ $s=7.779573348$ $r^2=.9143652308$ $r=.9562244667$ ■
--	---

2.

LinRegTTest $y=a+bx$ $\beta \neq 0$ and $\rho \neq 0$ $t=3.127464419$ $P=.0352714464$ $df=4$ $\downarrow a=-1.278153846$	LinRegTTest $y=a+bx$ $\beta \neq 0$ and $\rho \neq 0$ $\uparrow b=.2544615385$ $s=2.678001551$ $r^2=.7097460112$ $r=.8424642492$
--	--

3.

LinRegTTest $y=a+bx$ $\beta \neq 0$ and $\rho \neq 0$ $t=4.507613993$ $P=.0063548334$ $df=5$ $\downarrow a=-222.4934051$	LinRegTTest $y=a+bx$ $\beta \neq 0$ and $\rho \neq 0$ $\uparrow b=.3548103945$ $s=27.57373904$ $r^2=.8025166013$ $r=.8958329093$
--	--

4.

LinRegTTest $y=a+bx$ $\beta \neq 0$ and $\rho \neq 0$ $t=-.7862515631$ $P=.4616342049$ $df=6$ $\downarrow a=10.39346387$	LinRegTTest $y=a+bx$ $\beta \neq 0$ and $\rho \neq 0$ $\uparrow b=-.0120373649$ $s=1.037218684$ $r^2=.0934079225$ $r=-.3056270971$ ■
--	---

5.

LinRegTTest $y=a+bx$ $\beta \neq 0$ and $\rho \neq 0$ $t=3.349982476$ $P=.0203292198$ $df=5$ $\downarrow a=1.532212885$ ■	LinRegTTest $y=a+bx$ $\beta \neq 0$ and $\rho \neq 0$ $\uparrow b=.1490196078$ $s=.4492633248$ $r^2=.6917838689$ $r=.8317354561$ ■
--	---

Chapter 10

Contingency Tables and Risk Assessment

Problems that involve categorical data are summarized in contingency tables. Using the following methods we can determine there is indeed a relationship between the two categories of data.

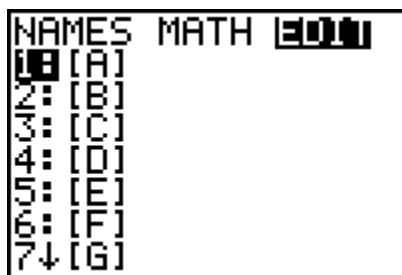
In 1988, the Steering Committee of the Physicians' Health Study Research Group released the results of a 5-year experiment conducted using 22,071 male physicians between the ages of 40 and 84. The physicians had been randomly assigned to two groups. One took an ordinary aspirin tablet every other day while the other group took a "placebo", a pill designed to look just like an aspirin but with no active ingredients. Neither group knew whether or not they were taking the active ingredient. The results of the study are shown in the following table.

	Heart Attack	No Heart Attack	Total
Aspirin	104	10,933	11,037
Placebo	189	10,845	11,034
Total	293	21,778	22,071

1. At a 5% level of significance, is there evidence that a relationship exists between the use of aspirin and heart attacks? Explain your answer fully.
2. Is this a cause and effect relationship? Explain why or why not.
3. What is the increased risk of having a heart attack for men in this age group not taking aspirin? Explain in detail.

We will answer each part of this problem and explain the steps.

To start this problem press the MATRX button and go to the EDIT menu



Press 1: [A] or ENTER.

We first will enter the dimensions of the contingency table in the Matrix window. The dimensions of the matrix are the numbers of ROWS followed by the number of COLUMNS. (THESE DO NOT INCLUDE THE TOTALS).

```
MATRIX[A] 2 x2
[ 0 0 ]
[ 0 0 ]

1, 1=0
```

Next input the actual numbers from the contingency table into the Matrix.

```
MATRIX[A] 2 x2
[ 104 10833 ]
[ 189 10845 ]

2, 2=10845
```

Now go to STAT TESTS and select C: χ^2 - Test

```
EDIT CALC TESTS
0: 2-SampTInt...
A: 1-PropZInt...
B: 2-PropZInt...
C:  $\chi^2$ -Test...
D: 2-SampFTest...
E: LinRegTTest...
F: ANOVA(
```

```
 $\chi^2$ -Test
Observed: [A]
Expected: [B]
Calculate Draw
```

Scroll to Calculate and press ENTER.

```
 $\chi^2$ -Test
 $\chi^2=25.01388397$ 
 $P=5.691897E-7$ 
df=1
```

This screen shows a p – value = 5.7×10^{-7} so that at a level of significance of .05, $p < !$. Therefore we support the claim that a relationship exists between the number of defective components and the day those components were produced.

Is this a cause and effect relationship? Explain why or why not

The answer to this question is YES. Why? Because this was a double blind experiment.

What is the increased risk of having a heart attack for men in this age group not taking aspirin? Explain in detail.

To answer increased or relative risk question, think in terms of percentages. What is the % of heart attacks in the placebo group? What is the % of heart attacks in the treatment group?

```
189/11034→A
.0171288744
104/11037→B
.0094228504
A/B
1.817801794
```

```
.0171288744
104/11037→B
.0094228504
A/B
1.817801794
A/B-1
.8178017945
■
```

The percent of heart attacks in the placebo group was about 1.7%

The percent of heart attacks in the aspirin group was about 0.9%

The relative risk of heart attacks in the placebo group is about 182%

The Increased Risk of heart attack in the placebo group is about 82%.

Problems;

1.

A study is being conducted to determine whether there is a relationship between jogging and blood pressure. A random sample of 210 subjects is selected, and they are classified as shown in the table. At $\alpha = 0.05$, test the claim that jogging and blood pressure are not related.

Jogging Status	Low	Moderate	High
Joggers	34	57	21
Non-joggers	15	63	20

What is the increased risk that a non-jogger will have high blood pressure?

2.

An automobile insurance company wishes to determine whether the age of the insured is related to the amount of liability the driver carries. A sample of 222 drivers shows the following data. At $\alpha = 0.05$, is the amount of coverage independent of the age of the driver?

Age	Under \$50,000	\$50,000-\$100,000	Over \$100,000
21-30	16	25	3
31-40	23	44	15
41-50	15	31	12
51 and over	9	12	11

What is the increased likelihood that a person age 55 will have over \$100,000 liability insurance compared to a person 25 years old?

3.

A study was conducted to determine whether the preference for a two-wheel drive or a four-wheel drive automobile is related to the gender of the purchaser. A sample of 90 buyers was selected, and the data are as shown in the table. At $\alpha = 0.05$, can the researcher conclude that the automobile preference is independent of the gender of the buyer?

Gender	Two-Wheel Drive	Four-Wheel Drive
Male	23	13
Female	43	6

What is the increased likelihood that a man will buy a Four-Wheel Drive compared to a woman?

4.

A study is being conducted to determine whether the age of the customer is related to the type of video movie he or she rents. A sample of renters gives the data shown in the table. At $\alpha = 0.10$, is the type of movie selected related to the age of the customer?

Age	Documentary	Comedy	Mystery
12-20	14	9	8
21-29	15	14	9
30-38	9	21	39
39-47	7	22	17
Over 48	6	38	12

What is the increased likelihood that a person over 48 will watch a comedy compared to a person between 12-20?

5.

A researcher wishes to determine whether the marital status of a student is related to his or her grade in a statistics course. The data below were obtained from a random sample of 142 students. At $\alpha = 0.05$, is the marital status independent of the grade received in the course?

Grade	Married	Single
A	32	27

B	19	14
C	10	16
D or F	8	16

What is the increased probability that a married person will get an A compared to a single person?

6.

A pet store owner wishes to determine whether the type of pet a person selects is related to the individual's gender. The data obtained from a sample are as shown in the table. At $\alpha = 0.10$, is the gender of the purchaser related to the type of pet purchased?

Pet Purchased

Gender	Dog	Cat	Bird
Male	32	4	6
Female	23	24	18

How much more likely is a woman to buy a cat than a man?

ANSWER KEY

1.

χ^2 -Test $\chi^2=6.788575295$ $P=.0335644554$ $df=2$	$20/(15+63+20)\rightarrow A$.2040816327 $21/(34+57+21)\rightarrow B$.1875 $A/B-1$.0884353741
---	--

2.

χ^2 -Test $\chi^2=10.14008662$ $P=.1188752839$ $df=6$ ■	$11/(9+12+11)\rightarrow A$.34375 $3/(16+25+3)\rightarrow B$.0681818182 $A/B-1$ 4.041666667 ■
--	---

3.

χ^2 -Test $\chi^2=6.8106255$ $P=.0090617017$ $df=1$ ■	$13/(23+13)\rightarrow A$.3611111111 $6/(43+6)\rightarrow B$.1224489796 $A/B-1$ 1.949074074
--	--

4.

χ^2 -Test $\chi^2=96.00816624$ $P=2.787319E-17$ $df=8$	$38/(6+38+12)\rightarrow A$.6785714286 $9/(14+9+8)\rightarrow B$.2903225806 $A/B-1$ 1.337301587
--	--

5.

χ^2 -Test $\chi^2=5.123976407$ $P=.1629408463$ $df=3$ ■	$32/(32+19+10+8) \rightarrow A$ $.4637681159$ $27/(27+14+16+16) \rightarrow B$ $.3698630137$ $A/B-1$ ■	A $.4637681159$ $27/(27+14+16+16) \rightarrow B$ $.3698630137$ $A/B-1$ ■ $.2538915727$
--	--	--

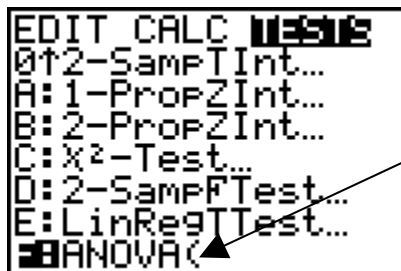
6.

χ^2 -Test $\chi^2=17.6290657$ $P=1.4855834E-4$ $df=2$	$24/(23+24+18) \rightarrow A$ $.3692307692$ $4/(32+4+6) \rightarrow B$ $.0952380952$ $A/B-1$ ■ 2.876923077
---	--

Chapter 11

Analysis of Variance or ANOVA

The program for one-way ANOVA, which comes with the TI-83, is found under the STAT TESTS menu at the very bottom of the list as E: ANOVA.



The limitation of this function is that it can only perform an analysis of variance from lists of data and not from summary statistics. A program called A1ANOVA that performs ANOVA determinations from summary statistics is available and can be linked to your calculator.

Now we will do an example using the ANOVA program that comes with your calculator.

Three car models were studied in a test that involved 4 cars of each model. For each of the 4 cars in each of the five samples, exactly one gallon of gas was placed in the tank and the car was driven until the gas was used up. The distances traveled are given in the table. At the 5% significance level, test the claim that the 5 population means are equal.

A	B	C
16	18	28
22	23	29
17	15	25
17	20	27

Again we symbolically state the claim, its opposite, then label the null and alternative hypothesis.

Claim: $\mu_1 = \mu_2 = \mu_3$

H_0 (null hypothesis)

Opposite $\mu_1 \neq \mu_2 \neq \mu_3$

H_1 (alternative hypothesis)

Notice that $\alpha = .05$

To do this problem we will be using the list editor so clear all lists. Now input the data from each of the columns in the table above using L1 through L3.

L1	L2	L3	1
22	18	28	
17	23	29	
17	15	25	
	20	27	

L1(1)=16			

Go to the STAT TESTS menu and select F: ANOVA

```

EDIT CALC TESTS
0: 2-SampTInt...
A: 1-PropZInt...
B: 2-PropZInt...
C:  $\chi^2$ -Test...
D: 2-SampFTest...
E: LinRegTTest...
F: ANOVA(

```

Enter the names of the lists containing the data into the ANOVA function as shown.

```

ANOVA(L1,L2,L3)

```

Press ENTER and the following screen will appear.

```

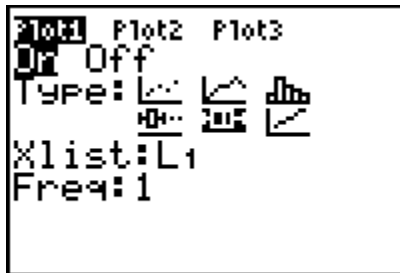
One-way ANOVA
F=14.32818533
P=.0015952113
Factor
df=2
SS=206.166667
↓ MS=103.083333

```

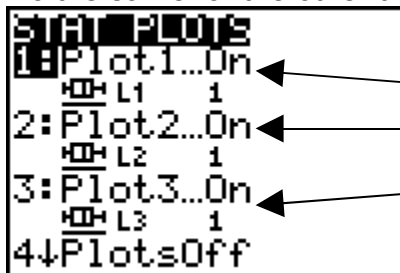
The important piece of information is $p = .002$ which means p is less than $\alpha = .05$. So since $p < \alpha$ we reject the null hypothesis and therefore reject the claim that the three populations have equal means.

Therefore we reject the claim that $\mu_1 = \mu_2 = \mu_3$

Now let us determine which one of the three car models is different. To do this we will use the graphing ability under the STAT PLOT button and draw boxplots of all samples. Go to the STAT PLOT button and select PLOT 1 and set it up in the following way;

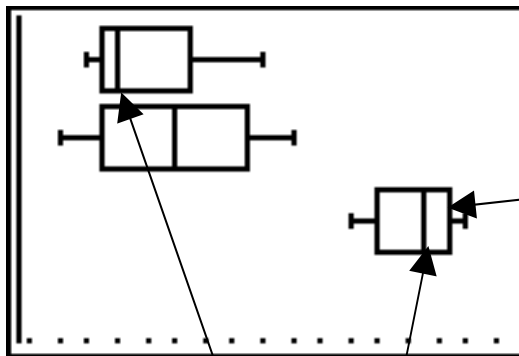


Do the same for the other two plots, PLOT 2 and PLOT 3



Notice that all three plots are turned on with Plot 1 being L1, Plot 2 being L2 and Plot 3 being L3.

When you press ZOOM 9 you should see the following graph with all three boxplots graphed on the same set of axes;



Here we can see that the boxplot for L3 is different (higher values) than the boxplots for L1 or L2.

Well, Brand C is different from Brands A and B. **But how different?**

By examining the medians we can see that the two that are most different are L₁ and L₃.

The two medians, L₁ and L₃ are the furthest apart.

How different is gas mileage between car model A and C? We could determine the 95% confidence interval estimate of the difference between the means for models A and C.

<pre>2-SampTInt Inpt: DATA Stats List1:L1 List2:L3 Freq1:1 Freq2:1 C-Level:95 ↓Pooled:NO Yes</pre>	<pre>2-SampFTest Inpt: DATA Stats List1:L1 List2:L3 Freq1:1 Freq2:1 σ1:EQ <σ2 >σ2 Calculate Draw</pre>	<pre>2-SampFTest σ1≠σ2 F=2.514285714 P=.468853079 Sx1=2.7080128 Sx2=1.70782513 ↓x1=18</pre>
<pre>2-SampTInt ↑List1:L1 List2:L3 Freq1:1 Freq2:1 C-Level:95 Pooled:No Yes Calculate</pre>	<pre>2-SampTInt (-13.17, -5.333) df=6 x1=18 x2=27.25 Sx1=2.7080128 ↓Sx2=1.70782513</pre>	<p>The negative signs here only signify that brand A had the lesser mileage figures compared to brand C</p>

We could also restate this interval using Margin of Error notation.

<pre>(13.17+5.333)/2 9.2515 13.17-9.2515 3.9185 █</pre>	<p>Here is the center of the interval 9.2515 And Here is the Margin of Error 3.9185</p>
---	---

So the interval could also be stated this way; **9.2515 ± 3.9185 mpg.**

Model C gets 9.2515 mpg more than Model A give or take 3.9185 mpg.

Problems

1. A researcher wishes to see whether there is a difference in the average age of nurses, doctors, and X-ray technicians at a local hospital. Employees are randomly selected, and their ages are recorded as shown in the table. At $\alpha = 0.05$, can the researcher conclude that there is a difference in the average ages of each group?

<u>Nurses</u>	<u>Doctors</u>	<u>X-Ray Technicians</u>
23	60	28
25	36	35
26	29	29
35	56	41
42	32	54
22	33	58

2. The grade point averages of students participating in college sports programs are to be compared. The data are shown in the table. At $\alpha = 0.10$, can one conclude that there is a difference in the mean grade point average of the three groups?

<u>Football</u>	<u>Basketball</u>	<u>Hockey</u>
3.2	3.8	3.3
2.6	3.1	2.6
2.4	2.6	1.9
2.4	3.9	1.7
1.9	1.8	2.5

3. A sample of real estate agents located in three different cities is selected, and the weekly commissions of the agents are as recorded below. At $\alpha = 0.10$, is there a difference in the commissions of the agents?

<u>City A</u>	<u>City B</u>	<u>City C</u>
3,500	7,329	8,400
4,000	8,641	7,615
3,271	12,312	3,141
4,268	9,610	1,001
6,788	10,300	
7,286		

Draw boxplots of the three cities commissions from these samples.

Construct a 95% confidence interval estimate of the difference for the cities which appear to vary the most.

ANSWER KEY

1.

```

One-way ANOVA
F=2.166941467
p=.1490360419
Factor
df=2
SS=584.111111
↓ MS=292.055556

```

2.

```

One-way ANOVA
F=1.285610991
p=.3119691358
Factor
df=2
SS=1.18533333
↓ MS=.592666667

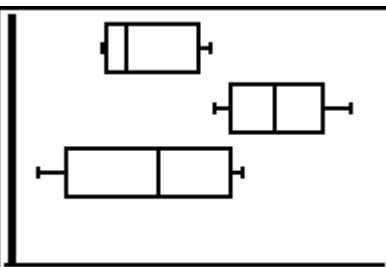
```

3.

```

One-way ANOVA
F=6.651287957
p=.0113788679
Factor
df=2
SS=74074966.9
↓ MS=37037483.5

```



```

2-SampFTest
Inpt: DATA Stats
List1:L1
List2:L2
Freq1:1
Freq2:1
σ1:EQ <σ2 >σ2
Calculate Draw

```

```

2-SampFTest
σ1≠σ2
F=.8662364789
p=.857557813
Sx1=1735.90765
Sx2=1865.12689
↓x1=4852.166667

```

```

2-SampTInt
Inpt: DATA Stats
List1:L1
List2:L2
Freq1:1
Freq2:1
C-Level:95
↓Pooled:No Waa

```

```

2-SampTInt
(-7244, -2328)
df=9
x1=4852.166667
x2=9638.4
Sx1=1735.90765
↓Sx2=1865.12689

```


Practice Final Examination

DO ALL OF THE FOLLOWING QUESTIONS

On all questions, show your work in the space provided. You may use one sheet of notes and your calculator. However, show each step to receive full credit. Show the graph of all distributions that are appropriate for the problem and label the graphs.

In tests of hypothesis, be sure to include the original claim, the null and alternative hypotheses, p values, graphs and conclusions.

Show all work neatly!

If it is illegible...IT IS WRONG !!

1. For the data given in the following stem leaf plot ;

5	345
6	1222
7	1234
8	3344569
9	1679

a. Construct and completely label a Boxplot

b. Find the mean _____

c. Find the standard deviation _____

d. Find the mode _____

e. Find the range _____

f. Find the variance _____

g. Find the mid-range _____

2. Use the frequency table below to find the following.

X	f
0-9	4
10-19	3
20-29	6
30-39	2
40-49	3
50-59	6

a. mean _____

- b. standard deviation _____
- c. class width _____
- d. variance _____
- e. upper class limit of the third class _____
- f. lower class boundary of the third class _____

3. Below neatly draw the histogram corresponding to the frequency table in question 2.

4. The dean of the admissions committee of Harvard Law School notes that the LSAT scores of applicants to the law school are normally distributed with a mean of 750 and a standard deviation of 16.6.

- a. If an applicant is randomly selected, find the probability of the applicant's score being between 760 and 780.
- b. If an applicant is randomly selected, find the probability of a score that is below 740.
- c. What LSAT score must an applicant attain to be places in the top 2%.
- d. If an applicant is randomly selected, find the probability of a score that is above 780.

5. The coffee machine in the faculty lounge dispenses cups of coffee with a mean and standard deviation of 12 oz. and 0.4 oz. respectfully. What is the probability that the dozen cups of coffee that I buy from the machine today will contain an average of between 12.25 and 12.5 ounces? (note: I am not addicted to coffee, I only use it to get my heart started!)

6. The life Trust Insurance Company is suspicious about a cluster of Orange County deaths that resulted from falls. You are asked to estimate the proportion of home accident deaths nationwide that are caused by falls. How many such deaths must you survey to be 94% confident that your sample proportion is within 2% of the true population proportion?

7. A random sample of 288 voters registered in California showed that 141 voted in the last general election. A random sample of 216 voters in Colorado showed that 125 voted in the most recent general election. Do these data indicate that the proportion of voter turnout in Colorado is higher than that in California?

Construct a 95% confidence interval for the difference between the two population proportions referred to in problem 7.

8. A dose of the drug Captopril, designed to lower systolic blood pressure, is administered to 10 randomly selected volunteers, with the following results. At a significance level of 2%, test the claim that Captopril is effective in lowering systolic blood pressure.

Subject	A	B	C	D	E	F	G	H	I	J
Before	120	136	160	98	115	110	180	190	138	128
After	118	122	143	105	98	98	180	175	105	112

Construct a 98% confidence interval for the average difference in systolic blood pressure before and after the use of Captopril. What is the Margin of Error?

9. A random sampling of companies in electric utilities, financial services and food processing gave the following information regarding annual profits per employee (units in thousands of dollars). At a .05 level of significance, is there a difference in the population mean profits per employee in these three types of companies?

Electric utilities	Financial services	Food processing
49.1	55.6	29
43.4	25	27.3
32.9	41.3	10.8
27.8	39.5	32.5
38.3		15.8
36.1		22.6
20.2		

- Construct a boxplot diagram comparing these three industries.
- Construct the 95% confidence interval estimate for the two industries that appear to differ the most in mean profits per employee.

10. Do heavier cars really use more gasoline? Let x be the weight of the car (in hundreds of pounds), and let y be the miles per gallon (mpg). The following is based on data taken from *Consumer Reports*.

X	27	44	32	47	23	40	34	52
Y	30	19	24	13	29	17	21	14

- Is there a correlation between the weight of a car and its gas mileage in mpg? Explain why.
- Compute the correlation coefficient. _____.
- What is the coefficient of determination? _____.
- What percent of these data can be explained using the regression model?
- Sketch the scatter plot diagram that represents this data on the axes below.

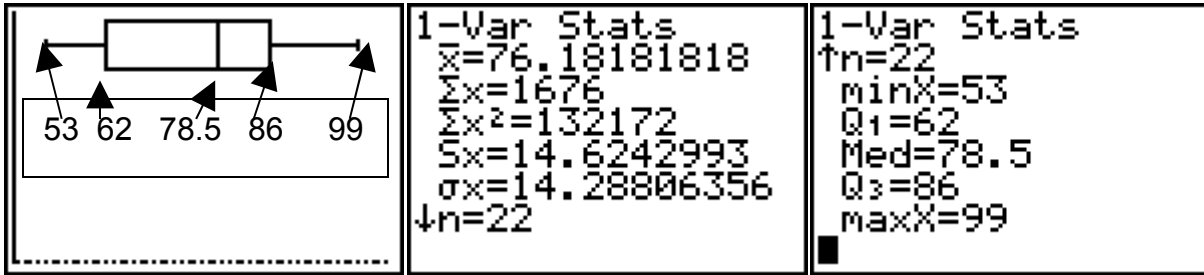


Draw and label the regression line on the scatter plot diagram.

- What is the equation of the regression line model for these data?
- Is this a positive or negative correlation? In your own words, what does that mean in this problem?
- How many miles per gallon would you expect a car 38 hundred pound to get based on this regression model?

ANSWER KEY

1.
a.

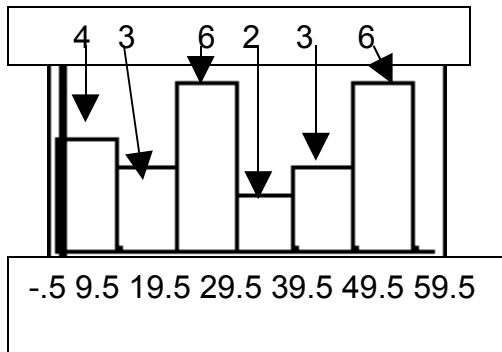
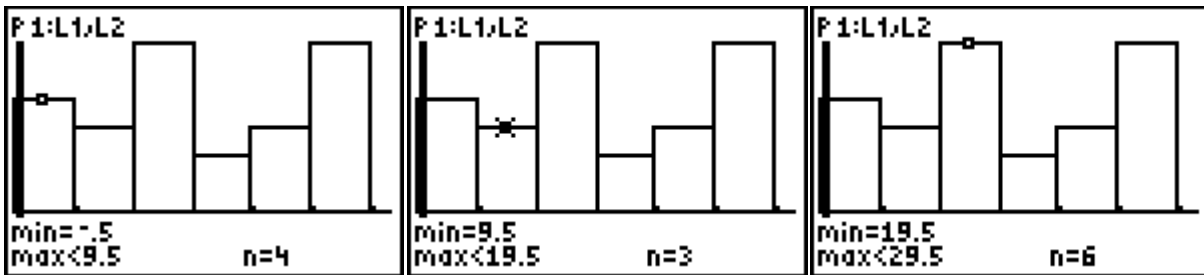


b. 76.18 c. 14.62 d. 62 e. $99-53=46$ f. 213.87 g. $(53+99)/2=76$

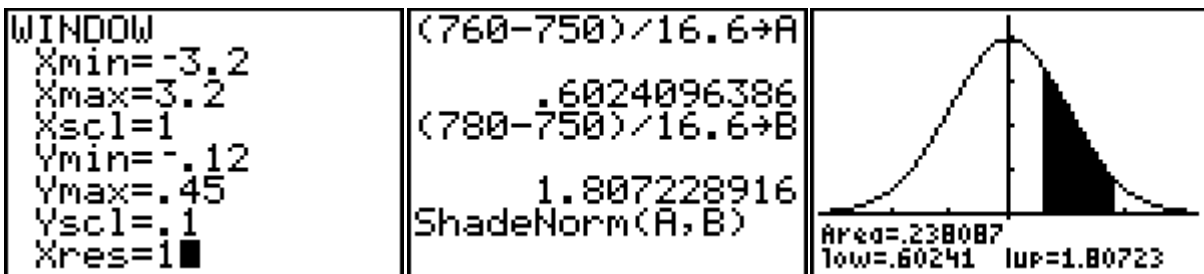
2.

a. 30.75 b. 18.37 c. 10 d. 337.5 e. 29 f. 19.5

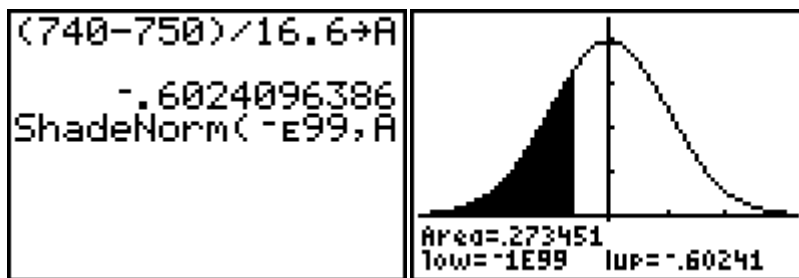
3.



4.
a.



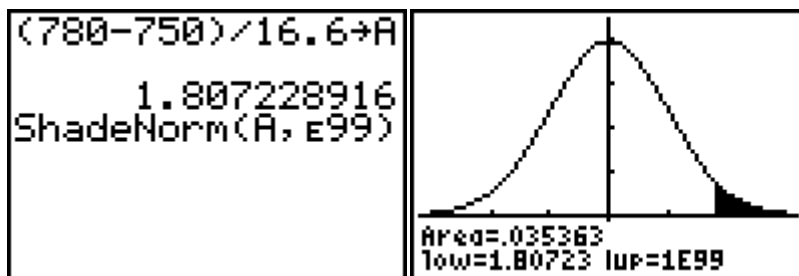
b.



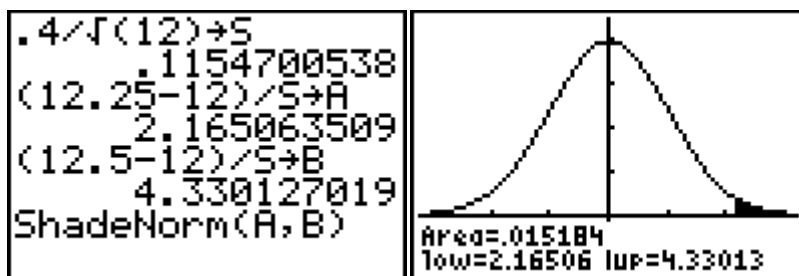
c.

```
invNorm(.98,750,
16.6)
      784.0922319
785
      785
```

d.



5.



6.

<pre>1.00-.94 .06 Ans/2 .03 invNorm(.03)→Z -1.88079361</pre>	<pre>invNorm(.03)→Z .03 -1.88079361 (Z/.02)²*.25 2210.865376 2211 2211</pre>
--	--

7.

Claim: $p_1 < p_2$ H_1

Oppos.: $p_1 \geq p_2$ H_0

```
2-PropZTest
x1:141
n1:288
x2:125
n2:216
P1:≠P2 <P2 >P2
Calculate Draw
```

```
2-PropZTest
P1<P2
z=-1.983289321
P=.0236674906
P1=.4895833333
P2=.5787037037
↓P=.5277777778
```

b.

```
2-PropZInt
x1:141
n1:288
x2:125
n2:216
C-Level:95
Calculate
```

```
2-PropZInt
(-.1767, -.0015)
P1=.4895833333
P2=.5787037037
n1=288
n2=216
```

8.

Claim: $\mu_d > 0$; H_1

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

```
T-Test
μ>0
t=3.365977881
P=.0041549414
x=11.9
Sx=11.17984297
n=10
```

	L3	3
DATA	14	
	17	
	7	
	17	
	12	
	0	

Oppos.: $\mu_d \leq 0$; H_0

8 b.

<pre>TInterval Inpt: DATA Stats List: L3 Freq: 1 C-Level: 98 Calculate</pre>	<pre>TInterval (1.9252, 21.875) x̄=11.9 Sx=11.17984297 n=10</pre>
---	---

21.875 – 11.9 = 9.975 (Margin of Error)

11.9 ± 9.975

9.

<table border="1"> <thead> <tr> <th>L1</th> <th>L2</th> <th>L3</th> <th>3</th> </tr> </thead> <tbody> <tr><td>49.1</td><td>55.6</td><td>29</td><td></td></tr> <tr><td>43.4</td><td>25</td><td>27.3</td><td></td></tr> <tr><td>32.9</td><td>41.3</td><td>10.8</td><td></td></tr> <tr><td>27.8</td><td>39.5</td><td>32.5</td><td></td></tr> <tr><td>38.3</td><td>-----</td><td>15.8</td><td></td></tr> <tr><td>36.1</td><td></td><td>22.6</td><td></td></tr> <tr><td>20.2</td><td></td><td></td><td></td></tr> </tbody> </table>	L1	L2	L3	3	49.1	55.6	29		43.4	25	27.3		32.9	41.3	10.8		27.8	39.5	32.5		38.3	-----	15.8		36.1		22.6		20.2				<pre>ANOVA(L1, L2, L3)</pre>	<pre>One-way ANOVA F=4.315334843 p=.0346748355 Factor df=2 SS=845.207647 ↓ MS=422.603824</pre>
L1	L2	L3	3																															
49.1	55.6	29																																
43.4	25	27.3																																
32.9	41.3	10.8																																
27.8	39.5	32.5																																
38.3	-----	15.8																																
36.1		22.6																																
20.2																																		
L3(?) =																																		

claim: $u_1 = u_2 = u_3 : H_0$

oppos. $u_1 \neq u_2 \neq u_3 : H_1$

$p < \alpha \rightarrow$ Reject H_0 . Therefore, reject the claim that the population means are all equal.

<pre>STAT PLOT 1 Plot1...On L1 1 2 Plot2...On L2 1 3 Plot3...On L3 1 4 PlotsOff</pre>	
---	--

It appears that the two most different populations are those in lists L2 and L3.

To get a 95% confidence interval estimate in the difference in these population means;

<pre>2-SampFTest Inpt: DATA Stats List1: L2 List2: L3 Freq1: 1 Freq2: 1 σ1: EQ <σ2 >σ2 Calculate Draw</pre>	<pre>2-SampFTest σ1≠σ2 F=2.265804348 p=.3970450216 Sx1=12.5141254 Sx2=8.31360331 ↓x1=40.35</pre>	<pre>2-SampTInt Inpt: DATA Stats List1: L2 List2: L3 Freq1: 1 Freq2: 1 C-Level: 95 ↓Pooled: No YES</pre>
---	--	--

<pre> 2-SampTInt (2.3223,32.378) df=8 x1=40.35 x2=23 Sx1=12.5141254 Sx2=8.31360331 </pre>	<pre> (2.3223+32.378)/ 2 17.35015 32.378-17.35015 15.02785 </pre>
---	---

17.35 ± 15.03
10.

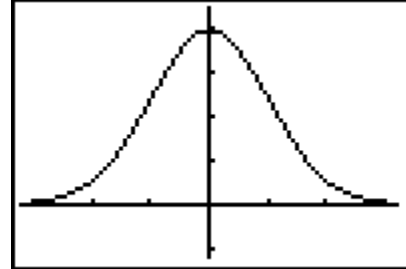
<table border="1"> <thead> <tr> <th>L1</th> <th>L2</th> <th>L3</th> <th>3</th> </tr> </thead> <tbody> <tr><td>27</td><td>30</td><td></td><td></td></tr> <tr><td>44</td><td>19</td><td></td><td></td></tr> <tr><td>32</td><td>24</td><td></td><td></td></tr> <tr><td>47</td><td>13</td><td></td><td></td></tr> <tr><td>23</td><td>29</td><td></td><td></td></tr> <tr><td>40</td><td>17</td><td></td><td></td></tr> <tr><td>34</td><td>21</td><td></td><td></td></tr> </tbody> </table> <pre> L3(1)= </pre>	L1	L2	L3	3	27	30			44	19			32	24			47	13			23	29			40	17			34	21			<pre> LinRegTTest Xlist:L1 Ylist:L2 Freq:1 B & P: $\mu < 0 >$ RegEQ:Y1 Calculate </pre>	<pre> LinRegTTest y=a+bx B≠0 and P≠0 t=-7.167682446 P=3.7233498E-4 df=6 a=43.3262511 </pre>
L1	L2	L3	3																															
27	30																																	
44	19																																	
32	24																																	
47	13																																	
23	29																																	
40	17																																	
34	21																																	
<pre> LinRegTTest y=a+bx B≠0 and P≠0 b=-.6007023705 s=2.23605489 r=.8954260608 r=-.9462695497 </pre>	<pre> Plot2 Plot3 Off Type: <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Xlist:L1 Ylist:L2 Mark: <input checked="" type="checkbox"/> + . </pre>																																	
<pre> Plot2 Plot3 Y1=43.326251097 454+- .6007023705 0044X Y2= Y3= Y4= Y5= </pre>																																		

ANOTHER PRACTICE FINAL EXAMINATION

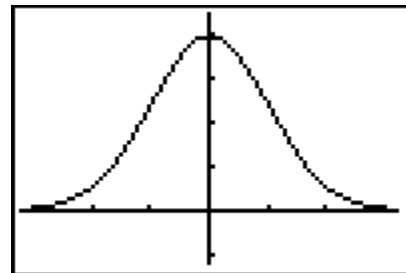
1.

A survey indicates that for each trip to the supermarket, a shopper spends an average of 45 minutes with a standard deviation of 12 minutes. A shopper enters the store. What is the probability that the shopper will be in the store for the lengths of time listed below.

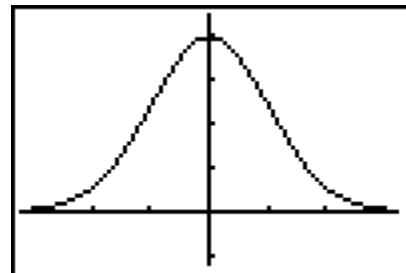
a. Between 24 and 54 minutes.



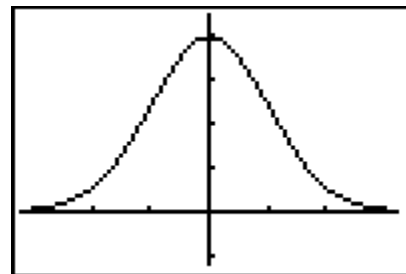
b. More than 40 minutes.



c. Less than 15 minutes.

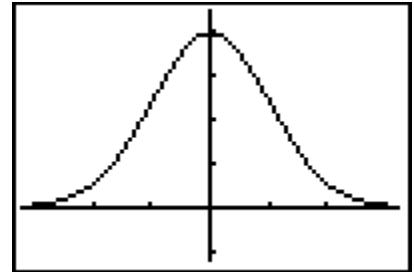


d. 75% of shoppers spend at least how many minutes in the store?



2.

Credit card balances are normally distributed with an average credit card balance of \$2870 and a standard deviation of \$900. If you randomly select 25 credit card holders, what is the probability that their average credit card balance is less than \$2500?



3.

Can you predict a man's height if you only know his shoe size? Below is a table that lists the shoe size and height for 14 randomly selected men

Shoe size	8.5	9.0	9.0	9.5	10	10	10.5	10.5	11	11	11	12	12	12.5
Height	66	68.5	67.5	70	70	72	71.5	69.5	71.5	72	73	73.5	74	74

i. Is there a correlation between a man's shoes size and his height? Explain why.

j. Compute the correlation coefficient. _____

k. What is the coefficient of determination? _____

l. What percent of these data can be explained using the regression model?

m. Sketch the scatter plot diagram that represents this data on the axes below.



Draw and label the regression line on the scatter plot diagram.

n. What is the equation of the regression line model for these data?

o. What height would you predict a man to be if you knew that he wore a size 11.5 shoe?

4.

In crash tests at 5 mph, the mean bumper repair cost for 14 compact cars is \$574 with a standard deviation of \$185. In similar tests of 23 midsize cars, the mean bumper repair cost is \$734 with a standard deviation of \$268. Can you conclude that the mean bumper repair cost for all compact cars is less than for midsize cars?

Construct a 95% confidence interval estimate for the difference in mean bumper repair costs for compact and midsize cars. Express this interval using Margin of Error notation.

5.

The following table shows the price per gallon (in dollars) for a random sample of exterior deck treatments. Can you conclude that the mean prices are the same for the three types of treatments?

Semitransparent treatments	Lightly tinted treatments	Clear treatments
24	21	13
23	14	13
22	21	10
17	16	12
21		22
17		

Use boxplots to analyze which treatments differ most.

Construct a 95% confidence interval estimate of the difference in cost for the two treatments that differ the most.

6.

To test whether a fuel additive improves gas mileage, the gas mileage of nine cars was measured with and without the fuel additive. The results are given below. Can you conclude that the fuel additive improved gas mileage?

Car	1	2	3	4	5	6	7	8	9
With additive	34.5	36.7	34.4	39.8	33.6	35.4	38.4	35.3	37.9
Without additive	36.4	38.8	36.1	40.1	34.7	38.3	40.2	37.2	38.7

Construct a 95% confidence interval estimate for the average increase in gas mileage using the fuel additive. Use Margin of Error Notation.

7.

You want to estimate the mean number of sentences in a magazine advertisement. How many magazine advertisements must be included in the sample if you want to be 95% confident that the sample mean is within one sentence of the true population mean? Assume the standard deviation is 5 sentences.

8.

In a survey of 3420 college students attending private schools, 917 said they had smoked within the last 30 days. In a survey of 5131 college students attending public schools, 1503 said they had smoked within the last 30 days. At a 1% level of significance, can you support the claim that the proportion of college students who said they had smoked in the last 30 day in private schools is less than the proportion in the public schools?

Use Margin of Error notation to represent the 95% confidence interval estimate of the difference in the % of smokers between private and public college students.

9.

The table below shows the results of a random sample of 550 company CEOs classified by age and size of company. At a 1% level of significance, can you conclude that the CEOs ages are related to company size?

Company size	<u>AGE OF CEOs</u>				
	39 & under	40 – 49	50 – 59	60 – 69	70 & over
small/midsize	42	69	108	60	21
large	5	18	85	120	22

What is the increased likelihood that a 35 year old CEO will be working for a small to midsize Company as opposed to a large company?

10.

In an advertisement, a Pizza Company claims that it delivers your pizza in less than 30 minutes. A random sample of 36 delivery times has a sample mean of 28.5 minutes with a standard deviation of 3.5 minutes. At a level of significance of 1%, is there enough evidence to support the Pizza Company's claim?

Construct a 99% confidence interval estimate for the Pizza Company's average delivery time and write this interval using Margin of Error notation.