ELEMENTARY STATISTICS ON THE TI-83

Derek Collis, Harper College

Contents

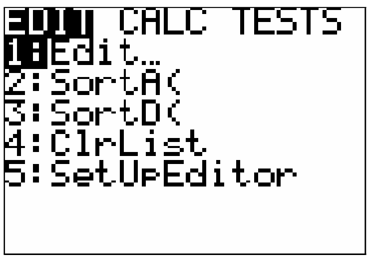
1 Lists

There are six default lists: **L1** through **L6**. Additional lists can be created, named and saved.

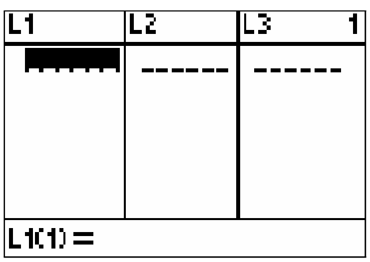
1.1 Entering Data into a List

Example 1 Enter the values 5, 9, 2, 6, 1 into List 1.

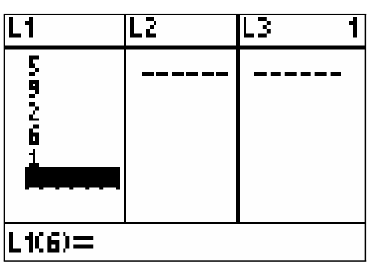
1. Press **STAT** to display the **EDIT** menu.



2. Press **ENTER** or **1** (to select **1:Edit...**)



3. To enter data into list **L1,** type **5**, press **ENTER**; type **9**, press **ENTER**; continue until the entire data set is entered.



4. After all data values are entered, press **STAT** to get back to the **EDIT** menu or **2nd MODE** (for **QUIT**) to return to the Home Screen.

Restriction: At most 999 measurements can be entered into a list.

1.2 Editing Data

Correcting a data value

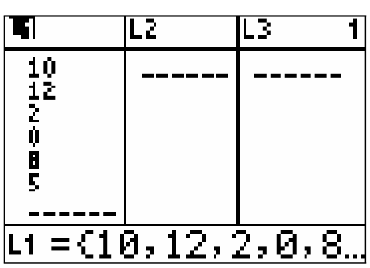
- To correct a data value before pressing **ENTER**, press the left arrow (◀), retype the value and press **ENTER**.
- To correct a data value in a list after pressing **ENTER**, move the cursor to highlight the incorrect value in the list and then type in the correct value and press **ENTER.**
- To delete a data value in a list, move cursor to highlight the value and press **DEL**.
- To change a data value in a list, move cursor to highlight the value, type the correct value and press **ENTER.**

Inserting a data value in a list

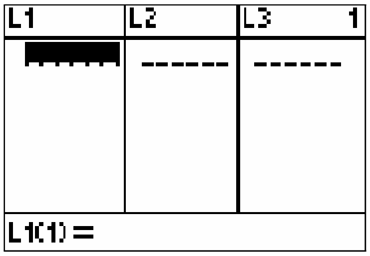
1. Move cursor to position where data value is to be inserted, then press **2nd DEL** (for **INS**).
2. Type data value and press **ENTER**.

1.3 Clearing a List

1. Press the up arrow (▲) to highlight the list name.



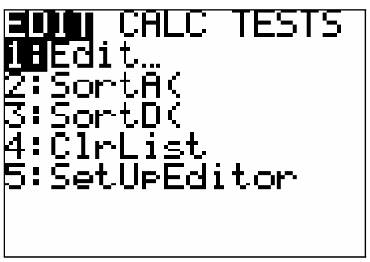
2. Press **CLEAR**, and then press either the down arrow (▼) or **ENTER**.



1.4 Sorting a List

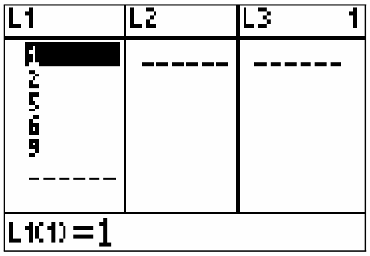
1. Enter the data into list **L1**.

2. Press **STAT 2** to select **SortA(** to sort list in ascending order or **STAT 3** to select **SortD(** to sort list in descending order.



3. Press **2nd 1 ENTER** (for **L1**). The calculator will display **Done**.

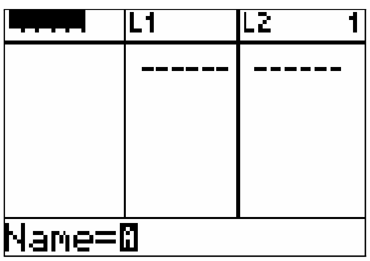
4. Press **STAT ENTER** (or **STAT 1**) to display the sorted list.



1.5 Creating a New List

1. Press **STAT ENTER** (or **STAT 1**).

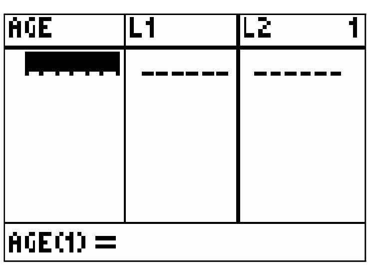
2. Move the cursor to the top of a list to highlight the list name, (the new list will be inserted to the left of highlighted list), then press **2nd DEL** (to select **INS**).



3. Type in a name for the new list. (A maximum of 5 characters is allowed and the first character must be a letter.)

   Note: The calculator is locked in ALPHA mode. To exit from this mode, press **ALPHA**.

4. Press **ENTER** twice.



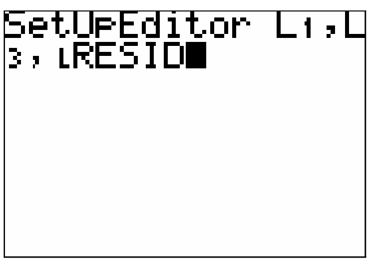
1.6 Storing a List

1. Move cursor to highlight list name.
2. Press **DEL**.

The name and the data are gone but they are stored in the TI-83 memory. To see a display of list names, press **2nd** [**LIST**].

1.7 Retrieving a List

1. Press **STAT 5** to select **SetUpEditor**.
2. To retrieve a default list, press **2nd 1** (to select **L1**) or **2nd 2** (to select **L2**), and so on.
3. To retrieve a user-created list, press **2nd STAT** (to select **LIST**), use ▼ to move the cursor to the list you wish to retrieve. Press **ENTER**. If you want to retrieve more than one list, separate the list names with a comma.



4. Press **ENTER**.
5. Press **STAT ENTER** to view the list data.

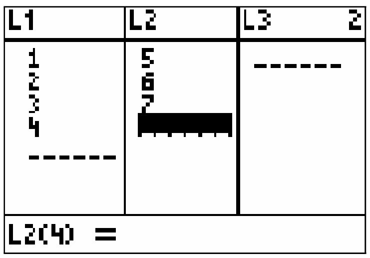
1.8 Restoring the Default Lists

1. Press **STAT 5** to select **SetUpEditor**.
2. Press **ENTER**.

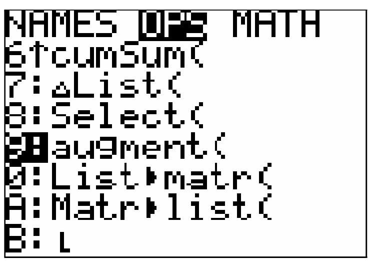
This procedure restores the six default lists, and removes any user-created lists.

1.9 Combining Two or More Lists into a Single List

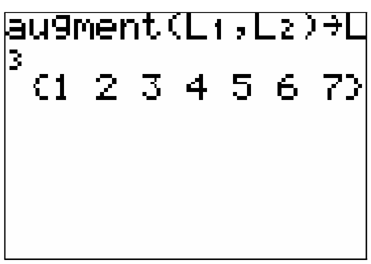
To combine data in List 1 and List 2 and store results into List 3:

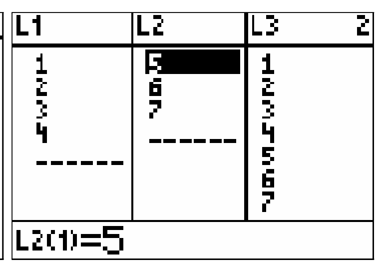
1. 

2. Press **2nd STAT** (to select **LIST**), arrow to **OPS**, and select **9:augment(**.



3. Enter **L1,L2)**, press **STO▶**, and enter **L3**. Press **ENTER**.





Note: If we had entered **L2,L1)** then the entries in List 2 would be listed first in List 3.

This procedure can be useful if you have a large data set and several people need to work with the same data set or a subset of this. Each person can store a part of the data set in a different list in each calculator, and then link and transfer the data and then combine the various lists into one list.

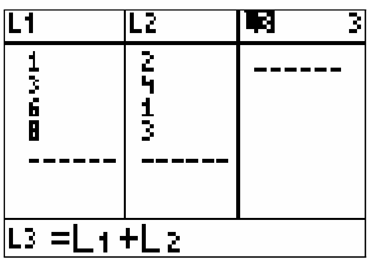
1.10 Applying Arithmetic Operations to Lists

Let's say we have data value in List 1 and List 2 and we wish to add the corresponding entries in these lists and then store these sums in List 3.

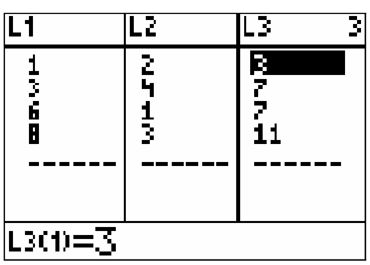
1. Enter the values in **L1** and **L2**.

   Note: the lists must contain the same number of data values, otherwise you will get a dimension mismatch error message in step 3.

2. Move the cursor so that it highlights **L3**.
3. Press **2nd 1** (to select **L1**).
4. Press +.
5. Press **2nd 2** (to select **L2**), then press **ENTER**.



6. The sums appear in **L3**.



2 Graphs

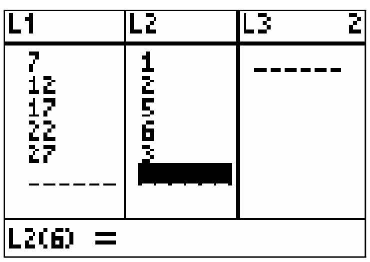
2.1 Frequency Polygon

A frequency polygon is a graph that displays the data by using lines that connect points plotted for the frequencies at the midpoints of the classes. The frequencies are represented by the heights of the points.

Example 1 Generate a frequency polygon for the following frequency distribution.

Class	f	Midpoint
5 – 9	1	7
10 – 14	2	12
15 – 19	5	17
20 – 24	6	22
25 – 29	3	27

1. Enter the class midpoints and frequencies into **L1** and **L2**.



2. Press **2nd Y=** (to select **STAT PLOT**).

3. Press **ENTER** to turn on **Plot1**.

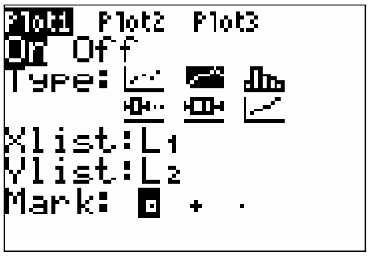
4. Use ▼ to move down to **Type**.

   The cursor flashes on the first type, which is a **Scatterplot**. The other types are: **xyLine**, **Histogram**, **Modified Boxplot**, **Boxplot** and **Normal Probability plot**.

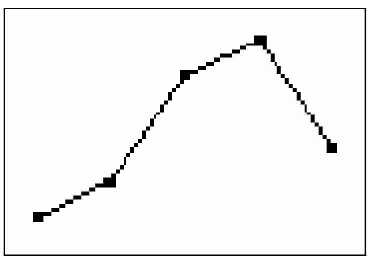
5. Use ▶ to highlight the **xyLine** symbol. Press **ENTER**.

6. Use ▼ to move down to **Xlist**. Set **Xlist** to **L1** and **Ylist** to **L2** (or whatever lists contain your midpoints and frequencies).

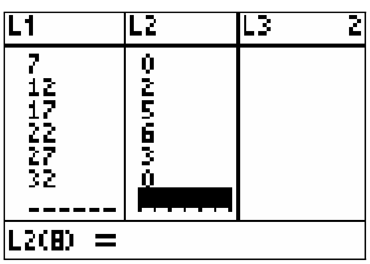
Note the order: Xlist: midpoints **Ylist:** frequencies.



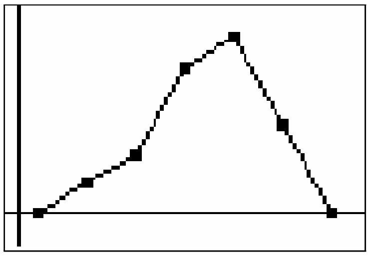
7. Press **ZOOM 9**. (This sets the window appropriately for the values in the lists.)



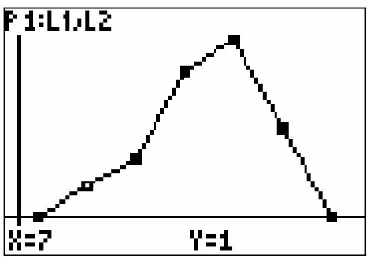
To obtain a graph that begins and finishes at the horizontal axis, subtract the class width from the first midpoint and enter this value before the first midpoint in List 1, and add the class width to the last midpoint and enter this value after the last midpoint in List 1. Since the classes having these midpoints have a frequency of zero, enter a frequency of 0 for both of these in List 2.



Press **GRAPH**.



To obtain the coordinates, press **TRACE**, followed by ◀ or ▶ keys.



2.2 Histogram

A histogram is a graph that displays the data by using contiguous vertical bars (unless the frequency of a class is 0) of various heights to represent the frequencies of the distribution.

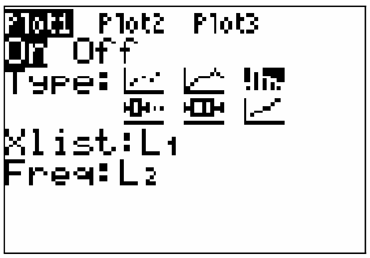
Example 2 Generate a histogram for the following frequency distribution.

Class	f	Midpoint
5 – 9	1	7
10 – 14	2	12
15 – 19	5	17
20 – 24	6	22
25 – 29	3	27

1. Enter the class midpoints and frequencies into **L1** and **L2**.
2. Press **2nd Y=** (to select **STAT PLOT**).
3. Press **ENTER** to turn on **Plot1**
4. Use ▼ to move down to **Type**.
5. Use ▶ to highlight the histogram symbol. Press **ENTER**.

6. Use ▼ to move down to **Xlist**. Set **Xlist** to **L1** and **Ylist** to **L2** (or whatever lists contain your midpoints and frequencies).

   Note the order: Xlist: midpoints **Ylist:** frequencies.



7. Press **WINDOW**. Adjust the values in the window to be:

   Xmin = lower limit of first class

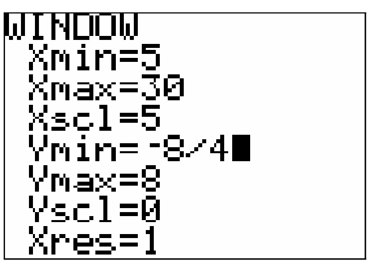
   Xmax = lower limit of last class plus class width (which would be the lower limit of the next class if there were one)

   Xscl = class width

   Ymin = –**Ymax**/4[1]

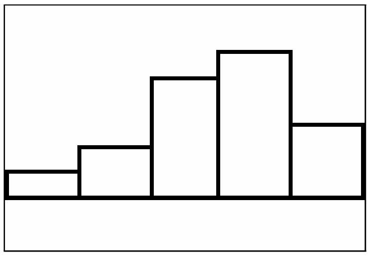
   Ymax = maximum frequency (or a little more) of distribution

   Yscl = 0

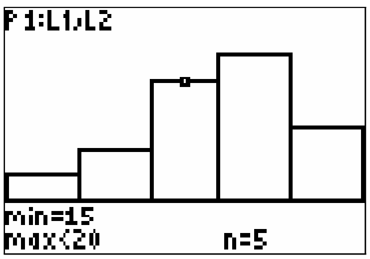


8. Press **GRAPH**.

[1]The purpose of making `Ymin` = –`Ymax`/4 is to allow sufficient space below the histogram so that the screen display is easily read.



9. To obtain various coordinates, press **TRACE**, followed by ◀ or ▶ keys.



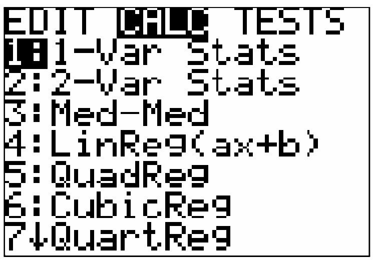
3 Measures of Central Tendency and Variation

3.1 Ungrouped Data

Example 1 The following are the hours per week worked by seven students: 17, 12, 15, 0, 10 and 24. Find the mean, median, standard deviation and variance.

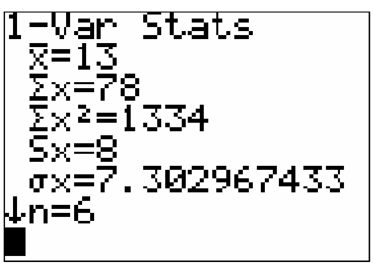
1. Enter data into **L1**.

2. Press **STAT**. Arrow to **CALC**.

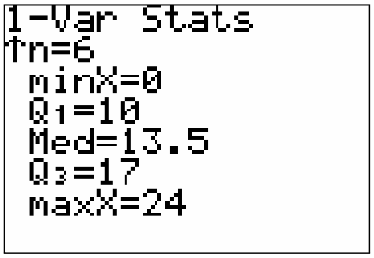


3. Press **1** (or **ENTER**) to select **1-Var Stats**. (If your data is in a list other than List 1, you need to enter the list name; for example, if your data are in List 2, enter **1-Var Stats L2**.)

4. Press **ENTER**.



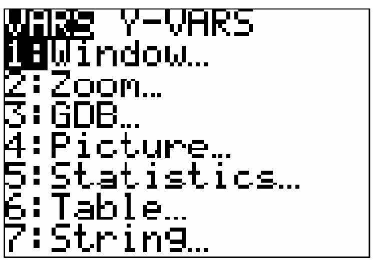
5. Press ▼ to scroll down to see more information.



The mean is 13, the median is 13.5, the sample standard deviation is 8 and the population standard deviation is 7.3.

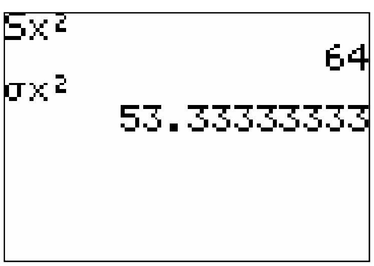
To obtain the variance, perform the following. (This procedure avoids using a rounded standard deviation value to obtain the variance.)

1. Press **VARS,** select **5:Statistics...**



2. Select **3:**S$_x$ (or **4:**σ_x)

3. Press $\mathbf{x}^2$, then press **ENTER**.



The sample variance is 64 and the population variance is 53.3.

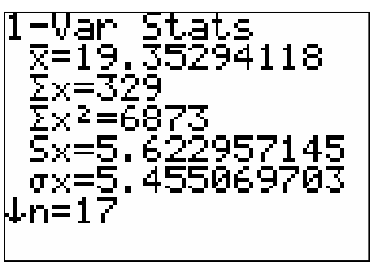
3.2 Grouped Data

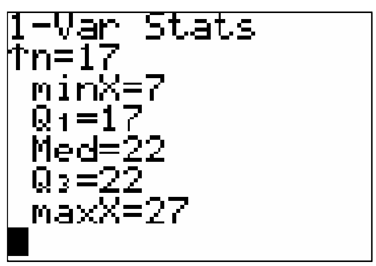
Example 2 The following frequency distribution shows the number of minutes it takes for seventeen students to drive from home to school.

Time to drive from home to school (in minutes)	f	Midpoint
5 – 9	1	7
10 – 14	2	12
15 – 19	5	17
20 – 24	6	22
25 – 29	3	27

1. Enter the class midpoints into **L1** and the frequencies into **L2**.
2. Press **STAT**. Select **CALC**.
3. Press **1** or **ENTER** for **1-Var Stats,** enter the midpoint list and the class frequency list separated by a comma.

4. Press **ENTER**.





The mean time to drive from home to school to 19.4 minutes and the median is 22 minutes. The standard deviation is 5.6 minutes.

4 Boxplots

There are two boxplots on the calculator: the **Modified Boxplot** and the **Standard Boxplot**. The modified boxplot is the fourth symbol in **Type** (located in **STAT PLOT**) and the standard boxplot is the fifth symbol in **Type**.

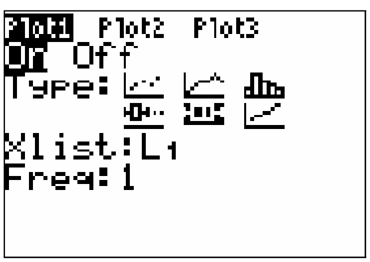
The Standard Boxplot represents the Five-Number Summary: Min, Q_1, Median, Q_2, Max. The Modified Boxplot is more informative as it identifies possible outliers. Instead of extending the whiskers to the minimum and maximum value it extends the whiskers to the smallest value and the largest value in the interval $(Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR)$, where IQR is the interquartile range $(Q_3 - Q_1)$. Generally, values outside this range are considered outliers.

Example 1 Generate a Boxplot and a Modified Boxplot for the values: 1, 2, 3, 3, 4, 5, 5, 5, 6, 7, 8, 9, 25.

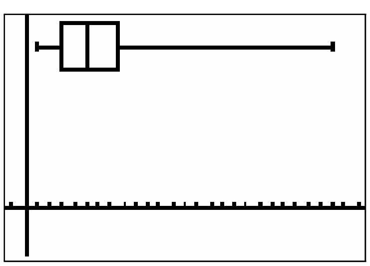
1. Enter the data values into **L1**.

2. Press **2nd Y=** (to select **STAT PLOT**), then press **ENTER**.

3. Set the window as shown.

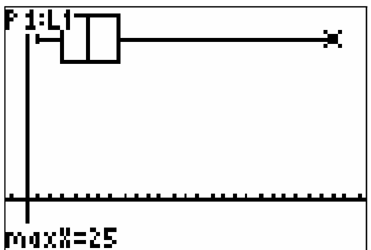


4. Press **ZOOM 9**.

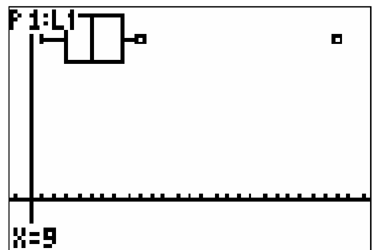


5. Press **TRACE** and press ► twice.

Here the maximum value is shown to be 25. The Modified Boxplot for the same data shows the right whisker now only extends to the value of 9; the value of 25 is shown separate from the boxplot. This is because 25 lies outside the interval $(-3.75, 14.25)$. The data value of 9 is the largest that lies inside this interval. So we have identified 25 as an outlier.



Standard Boxplot



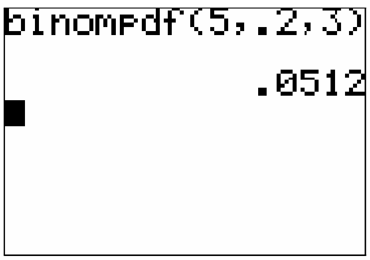
Modified Boxplot

5 Binomial Distribution

5.1 Probability for a Binomial Variable

Example 1 Find $P(X = 3)$ where $n = 5$ and $p = 0.2$.

1. Press **2nd VARS** (to select **DISTR**)
2. Select **0:binompdf(**.
3. Enter **5, .2, 3)**, then press **ENTER.**



5.2 Probabilities for Several Values of a Binomial Variable

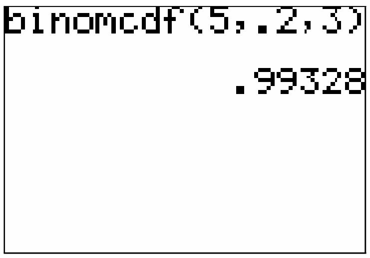
Example 2 Find $P(X = 1, 2)$ where $n = 5$ and $p = 0.2$.

1. Press **2nd VARS** (to select **DISTR**)
2. Select **0:binompdf(**.
3. Enter **5, .2, {1,2})**, then press **ENTER.**

5.3 Cumulative Probability for a Binomial Variable

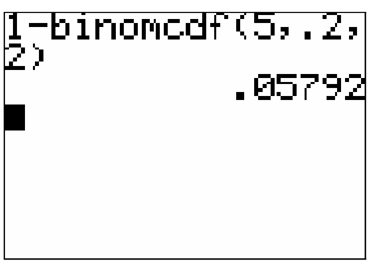
Example 3 Find $P(X \leq 3)$ where $n = 5$ and $p = 0.2$.

1. Press **2nd VARS** (to select **DISTR**)
2. Select **A:binomcdf(**.
3. Enter **5, .2, 3)**, then press **ENTER.**



Example 4 Find $P(X \geq 3)$ where $n = 5$ and $p = 0.2$.

1. Press **1 − 2nd VARS** (to select **DISTR)**
2. Select **A:binomcdf(**.
3. Enter **5, .2, 2)**, then press **ENTER.**

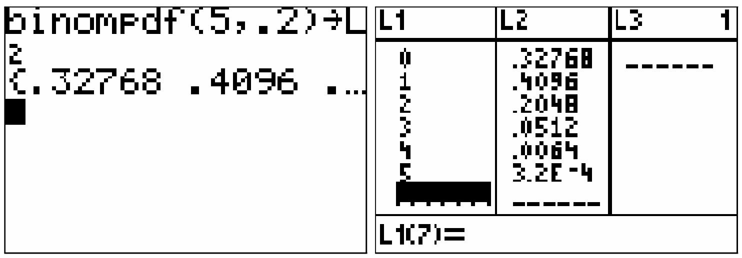


5.4 Constructing a Binomial Probability Distribution

Example 5 Construct a binomial probability distribution for $n = 5$ and $p = 0.2$.

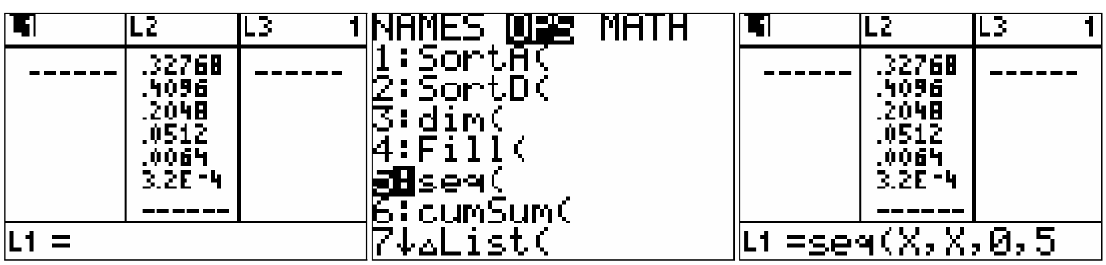
1. Press **2nd VARS** (to select **DISTR)**
2. Select **0:binompdf(**.

3. Enter **5, .2)**
4. Press **STO▶ 2nd 1 (**to select **L2**), then press **ENTER.**



The calculator will store the probabilities in List 2. Type the values 0, 1, 2, 3, 4, 5 into List 1 and you have the binomial probability distribution. Another method for entering a set of integers into a list is the following.

(a) Highlight **L1**.
(b) Press **2nd STAT** (to select **LIST**)
(c) Arrow to **OPS** and select **5:seq(**.
(d) Enter **X**, **X**, **0**, **5**.[2] Press **ENTER**.

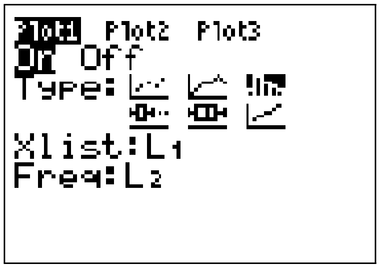


5.5 Constructing a Binomial Probability Histogram

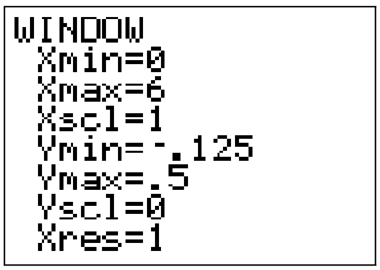
Example 6 Construct a binomial probability distribution for $n = 5$ and $p = 0.2$.

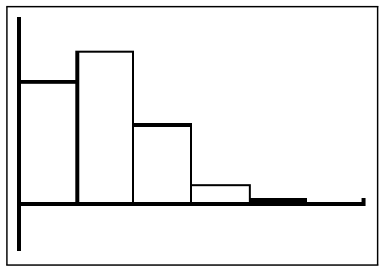
1. Set up **Plot1** for a **Histogram** as shown below.

[2] This instructs the calculator to generate the sequence of X with respect to the variable X starting from 0 and finishing at 5 in increments of 1. The default increment is 1, if some other increment is desired this would be entered as the fifth argument in **seq(**.

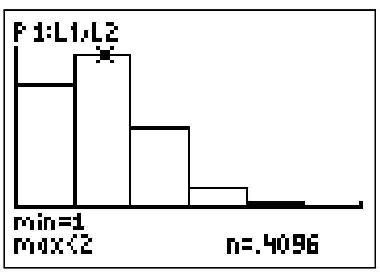


2. Set the window as shown, and press **GRAPH**.

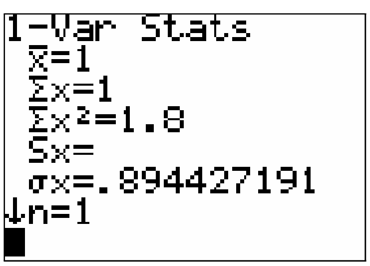




3. Using **TRACE** we can read the probabilities. For example, $P(X = 1) = n = 0.4096$.



4. Using **1-Var Stats L1,L2** we see that $\mu = 1$ and $\sigma = .89$.

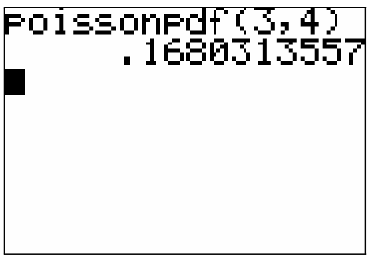


6 Poisson Distribution

6.1 Probability of a Poisson Variable

Example 1 Find $P(x = 4)$ where $\mu = 3$.

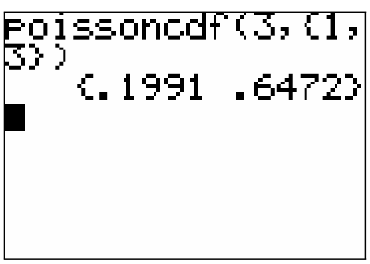
1. Press **2nd VARS** (to select **DISTR**)
2. Select **B:poissonpdf(**.
3. Enter **3,4)**, then press **ENTER.**



6.2 Probability for Several Values of a Poisson Variable

Example 2 Find $P(x = 1, 3)$ where $\mu = 3$.

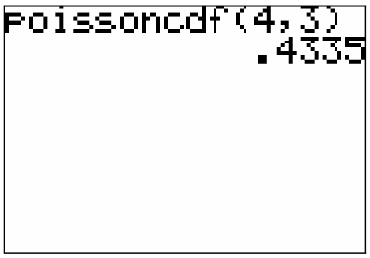
1. Press **2nd VARS (**to select **DISTR**)
2. Select **B:poissonpdf(**.
3. Enter **3, {1,3})**, then press **ENTER.**



6.3 Cumulative Probability for a Poisson Variable

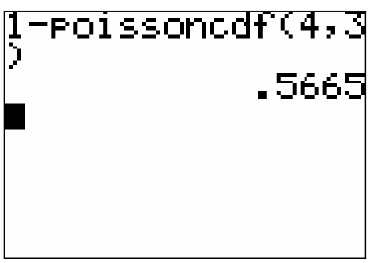
Example 3 Find $P(x \leq 3)$ where $\mu = 4$.

1. Press **2nd VARS (**to select **DISTR**)
2. Select **C:poissoncdf(**.
3. Enter **4, 3)**, then press **ENTER.**



Example 4 Find $P(x \geq 3)$ where $\mu = 4$.

1. Press **1 − 2nd VARS** (to select **DISTR**)
2. Select **C:poissoncdf(**.
3. Enter **4, 2)**, then press **ENTER.**

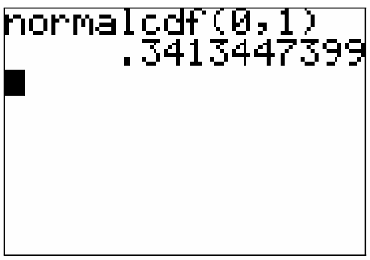


7 Normal Probabilities

7.1 Probability Between Two Z Values

Example 1 Find $P(0 < z < 1)$

1. Press **2nd VARS** (to select **DISTR**)
2. Select 2:**normalcdf(**
3. Enter **0, 1.75)**, then press **ENTER**.

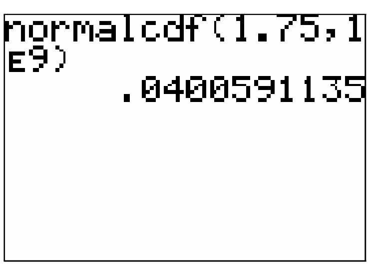


7.2 Probability Greater Than/Less Than a Z Value

Example 2 Find $P\left(z > 1.75\right)$

Method 1.

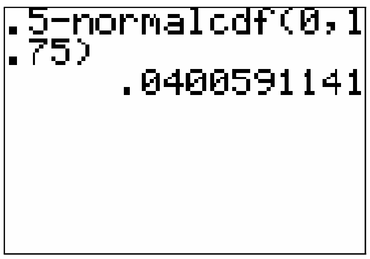
1. Press **2nd VARS** (to select **DISTR**)
2. Select **2**:**normalcdf(**
3. Enter **1.75, 1E9)**, then press **ENTER**.



Note: Since the interval is from 1.75 to ∞, we represent ∞ by a very large number, namely $1,000,000,000$ or, in scientific notation, 1×10^9. This is entered in the calculator as **1 2nd , 9** and is displayed as **1E9**.

Method 2.

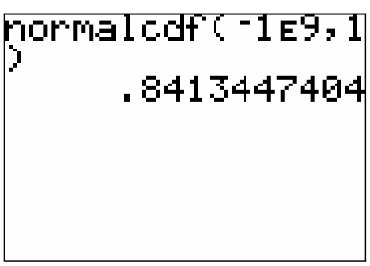
1. Enter **.5** $-$
2. Press **2nd VARS** (to select **DISTR**)
3. Select **2**:**normalcdf(**
4. Enter **0,1.75)**, then press **ENTER**.



Example 3 Find $P(z < 1)$

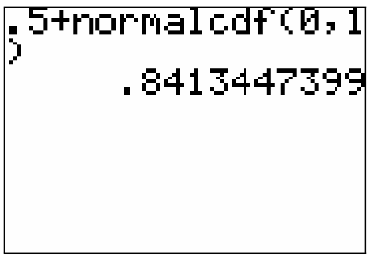
Method 1.

1. Press **2nd VARS** (to select **DISTR**)
2. Select **2**:**normalcdf(**
3. Enter **-1E9,1)**, then press **ENTER**.



Method 2.

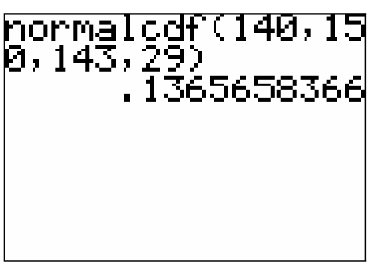
1. Enter **.5** +
2. Press **2nd VARS** (to select **DISTR**)
3. Select **2**:**normalcdf(**
4. Enter **0,1)**, then press **ENTER**.



7.3 Probability Between Two X Values

Example 4 Find $P(140 < x < 150)$ where $\mu = 143$ and $\sigma = 29$

1. Press **2nd VARS** (to select **DISTR**)
2. Select **2**:**normalcdf(**
3. Enter **140,150,143,29)**, then press **ENTER**.

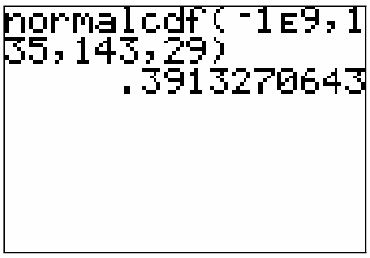


7.4 Probability Less Than an X Value

Example 5 Find $P(x < 135)$ where $\mu = 143$ and $\sigma = 29$.

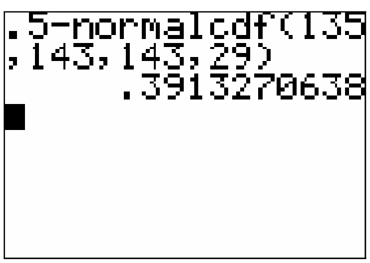
Method 1.

1. Press **2nd VARS** (to select **DISTR**)
2. Select **2**:**normalcdf(**
3. Enter **–1E9, 135,143,29)**, then press **ENTER**.



Method 2.

1. Enter **.5** –
2. Press **2nd VARS** (to select **DISTR**)
3. Select **2**:**normalcdf(**
4. Enter **135,143,143,29)**, then press **ENTER**.

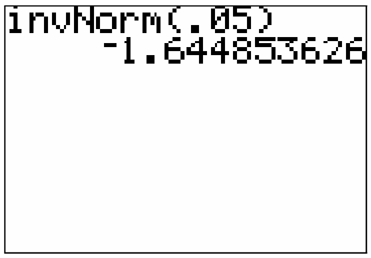


7.5 Finding a Z Value

Example 6 Find z such that 5% of the values are less than z.

1. Press **2nd VARS** (to select **DISTR**)
2. Select **3**:**invNorm(**

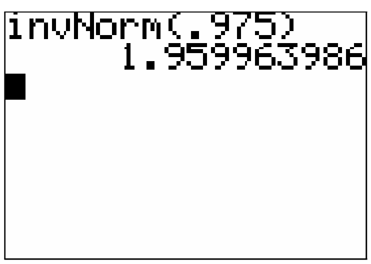
3. Enter **.05)**, then press **ENTER**.



Note: the area to the *left* of the unknown value must be entered.

Example 7 Find z such that 2.5% of the values are greater than z.

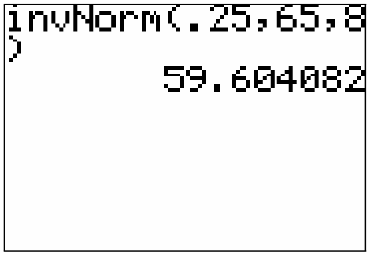
1. Press **2nd VARS** (to select **DISTR**)
2. Select **3**:**invNorm(**
3. Enter **.975)**, then press **ENTER**.



7.6 Finding an X Value

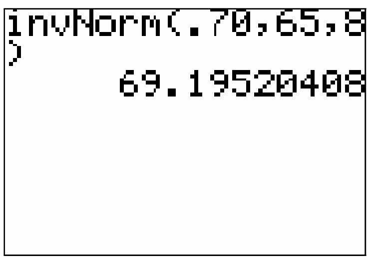
Example 8 Find x such that 25% of the values are less than x, where $\mu = 65$ and $\sigma = 8$.

1. Press **2nd VARS** (to select **DISTR**)
2. Select **3**:**invNorm(**
3. Enter **.25,65,8)**, then press **ENTER**.



Example 9 Find x such that 30% of the values are greater than x, where $\mu = 65$ and $\sigma = 8$.

1. Press **2nd VARS** (to select **DISTR**)
2. Select **3**:**invNorm(**
3. Enter **.70,65,8)**, then press **ENTER**.



8 Confidence Intervals

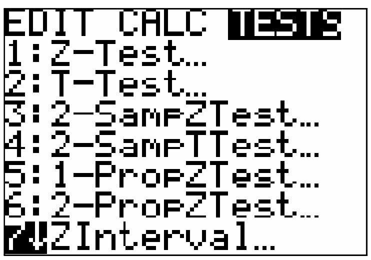
8.1 Confidence Interval for a Population Mean: σ Known

Example 1 Find a 95% confidence interval for the starting salaries of college graduates who have taken a statistics course where $n = 28$, $\bar{x} = \$45,678$, $\sigma = \$9,900$, and the population is normally distributed.

1. Press **STAT**

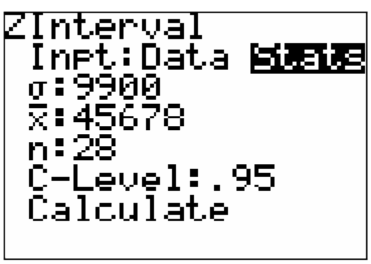
2. Arrow to **TESTS**

3. Select **7:ZInterval...**

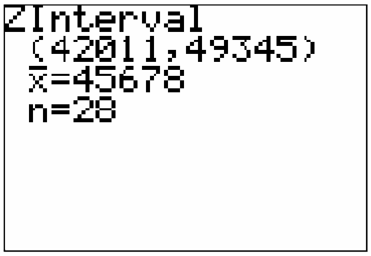


4. Highlight **Stats**

5. Press ▼ and enter the values for $\boldsymbol{\sigma}$, $\bar{\mathbf{x}}$, **n**, and **C-Level**.



6. Highlight **Calculate**, then press **ENTER**.

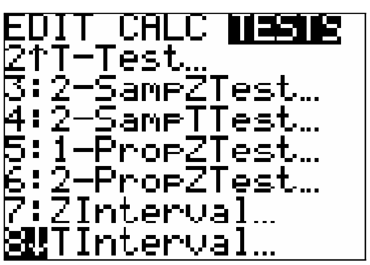


We are 95% confident that the mean starting salary of college graduates that have taken a statistics course is between \$42,011 and \$49,345. This means that if we were to select many different samples of size 28 and construct 95% confidence intervals for each sample, 95% of the constructed confidence intervals would contain μ and 5% would not contain μ. It is **incorrect** to say that "there is a 95% chance that μ will fall between \$42,011 and \$49,345." The population mean, μ, is not a random variable, it is a fixed, but unknown, constant. The probability that this interval contains μ is 0 or 1.

8.2 Confidence Interval for a Population Mean, σ Unknown

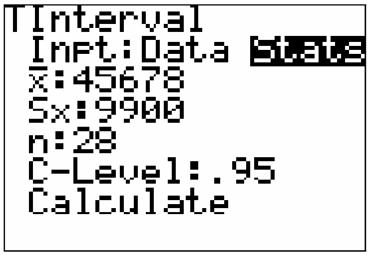
Example 2 Find a 95% confidence interval for the starting salaries of college graduates who have taken a statistics course where $n = 28$, $\bar{x} = \$45,678$, $s = \$9,900$, and the population is normally distributed.

1. Press **STAT**
2. Arrow to **TESTS**
3. Select **8:TInterval...**

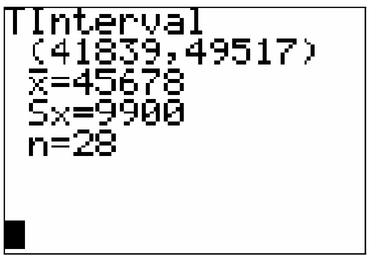


4. Highlight **Stats**

5. Press ▼ and enter the values for $\bar{\mathbf{x}}$, $\mathbf{s}_x$, **n**, and **C-Level**.



6. Highlight **Calculate**, then press **ENTER**.



We are 95% confident that the mean starting salary of college graduates that have taken a statistics course is between \$42,011 and \$49,345.

> **Note:** The confidence interval using the t statistic is wider than the interval using the z statistic, even though the sample sizes are the same and the same value for σ and s is used. The reason for this is that the primary difference between the sampling distribution of t and z is that the t statistic is more variable than the z, which seems obvious when you consider that t contains two random quantities ($\overline{x}$ and s), whereas z contains only one ($\overline{x}$). Thus, the t value will always be larger than a z value for the same sample size.

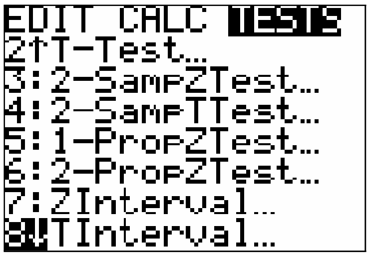
Example 3 The following random sample was selected from a normal distribution: 4, 6, 3, 5, 9, 3. Construct a 95% confidence interval for the population mean, μ.

1. Enter the data into **L1**.

2. Press **STAT**.

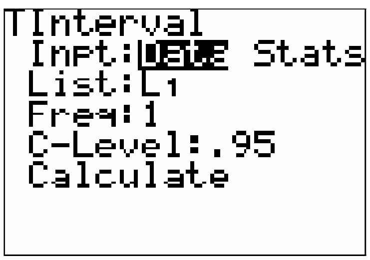
3. Arrow to **TESTS**.

4. Select **8:TInterval...**

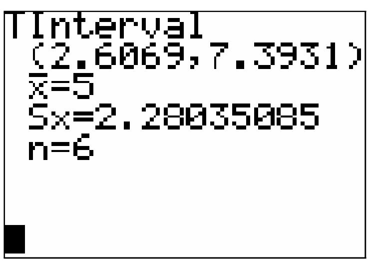


5. Highlight **Data**.

6. Press ▼ and enter the values for **List**, **Freq**, and **C-Level**.



7. Highlight **Calculate**, then press **ENTER**.



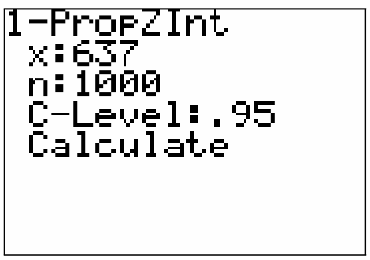
We are 95% confident that the population mean, μ, is between 2.6 and 7.4.

8.3 Confidence Interval for a Population Proportion

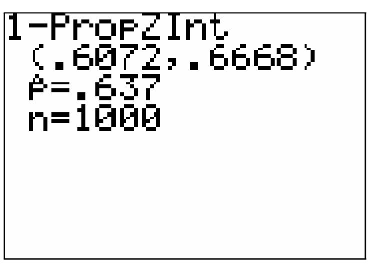
Example 4 Public opinion polls are conducted regularly to estimate the fraction of U.S. citizens who trust the president. Suppose 1,000 people are randomly chosen and 637 answer that they trust the president. Compute a 95% confidence interval for the population proportion of all U.S. citizens who trust the president.

1. Press **STAT**
2. Arrow to **TESTS**
3. Select **A:1-PropZInt...**

4. Enter the values for **x**, **n**, and **C-Level**.



5. Highlight **Calculate**, then press **ENTER**.



We are 95% confident that the true percentage of all U.S. citizens who trust the president is between 60.7% and 66.7%.

9 Hypothesis Tests

9.1 Test for a Mean: Large Sample

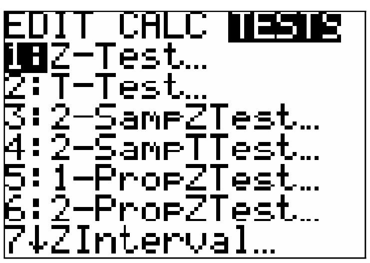
Example 1 A lightbulb manufacturer has established that the life of a bulb has mean 95.2 days with standard deviation 10.4 days. Following a change in the manufacturing process which is intended to increase the life of a bulb, a random sample of 96 bulbs has mean life 96.6 days. Test whether there is sufficient evidence, at the 1% level, of an increase in life.

The hypotheses are:

$$H_0 \;:\; \mu = 95.2$$
$$H_1 \;:\; \mu > 95.2$$

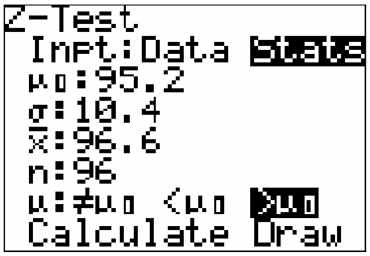
This is a right-tailed test with $\alpha = 0.01$. The critical value is $z = 2.326$. (That is, we will reject H_0 if the test statistic $z \geq 2.326$).

1. Press **STAT**
2. Arrow to **TESTS**
3. Select **1:Z-Test...**

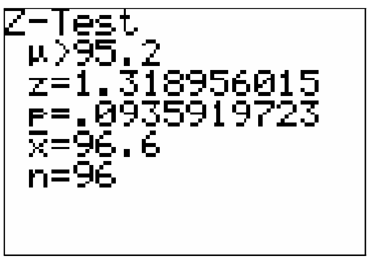


4. Highlight **Stats**.
5. Enter the values
    - μ_0 (value of μ under H_0)
    - σ (use s as estimate for $n \geq 30$)
    - $\bar{x}$ (sample mean)
    - n (sample size)

- μ : $\neq \mu_0$ $< \mu_0$ $> \mu_0$ (form of H_1)



6. Highlight **Calculate**, then press **ENTER**.



Since $z = 1.32$ does not fall in the critical region, we do not reject H_0. Or, since $p = 0.09 > \alpha = 0.01$, we do not reject H_0.
There is not sufficient evidence, at the 1% level, to indicate that the new process has led to an increase in the life of the bulbs. There is a 9.36% chance of observing a sample mean at least as extreme as 96.6 if H_0 is true.

9.2 Test for a Mean: Small Sample

Example 2 An employment information service claims that the mean annual pay for full-time male workers over age 25 and without high school diplomas is less than \$24,600. The annual pay for a random sample of 10 full-time male workers without high-school diplomas is given below. Test the claim at the 5% level of significance. Assume that the income of full-time male workers without high-school diplomas is normally distributed.

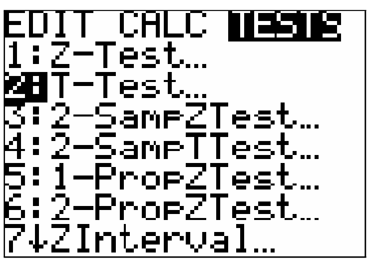
\$22,954	\$23,438	\$24,655	\$23,695	\$25,275
\$19,212	\$21,456	\$25,493	\$26,480	\$28,585

The hypotheses are:

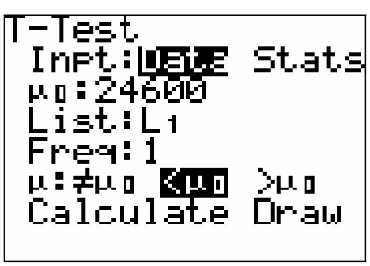
$$\begin{aligned} H_0 &: \mu = 24,600 \\ H_1 &: \mu < 24,600 \end{aligned}$$

This is a left-tailed test with $\alpha = 0.05$. The critical value is $t = -1.833$. (That is, we will reject H_0 if the test statistic $t \leq -1.833$).

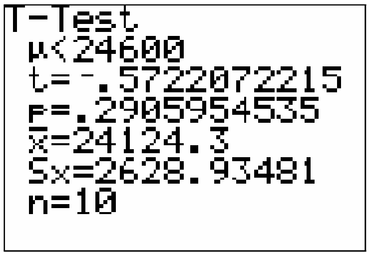
1. Enter the data into **L1**.
2. Press **STAT**
3. Arrow to **TESTS**.
4. Select **2:T-Test...**



5. Highlight **Data**.
6. Enter the values:
   - μ_0 (value of μ under H_0)
   - `List` : (list containing the sample data)
   - `Freq` : (enter 1)
   - μ : $\neq \mu_0$ $< \mu_0$ $> \mu_0$ (form of H_1)



7. Highlight **Calculate**, then press **ENTER**.



Since $t = -.57$ does not fall in the critical region, we do not reject H_0. Or, since $p = 0.29 > \alpha = 0.05$, we do not reject H_0.

There is not sufficient evidence, at the 1% level, to support the claim that the mean annual pay for full-time male workers over age 25 and without high school diplomas is less than $24,600. There is a 29% chance of observing a sample mean at least as extreme as $24,600 if H_0 is true.

9.3 Test for a Proportion

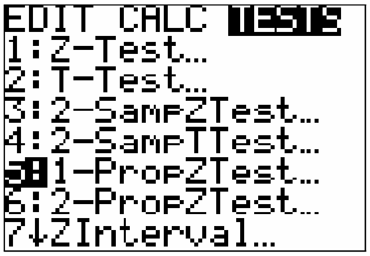
Example 3 A medical researcher claims that less than 20\% of adults in the US are allergic to a medication. In a random sample of 100 adults, 13 say they have such an allergy. Test the researcher's claim at the 5% level of significance.

The hypotheses are:

$$H_0 \ : \ p = 0.2$$
$$H_1 \ : \ p < 0.2$$

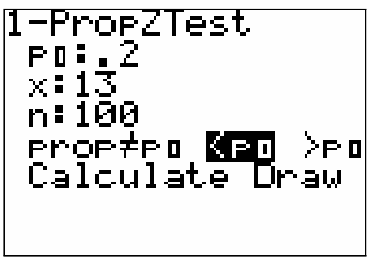
This is a left-tailed test with $\alpha = 0.05$. The critical value is $z = -1.645$. (That is, we will reject H_0 if the test statistic $z \leq -1.645$).

1. Press **STAT**
2. Arrow to **TESTS**
3. Select **5:1-PropZTest...**

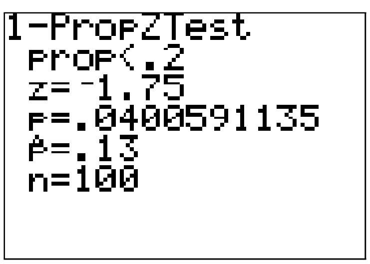


4. Enter the values:

   - p_0 (value of p under H_0)
   - x (number in the sample that have the particular characteristic we are testing)
   - n (sample size)
   - *prop* $\neq p_0 \quad < p_0 \quad > p_0$ (form of H_1)



5. Highlight **Calculate**, then press **ENTER**.



Since $z = 1.75$ falls in the critical region, we reject H_0. Or, alternatively, since $p = 0.04 < \alpha = 0.05$, we reject H_0.

There is sufficient evidence, at the 5% level, to support the claim that less than 20% of adults in the US are allergic to this medication. There is only a 4% chance of observing a sample proportion at least as extreme as .13 if H_0 is true.

10 Two-Sample Hypothesis Tests

10.1 Test for a Difference Between Means (Large Independent Samples)

Example 1 Two machines are used to fill 50-lb bags of dog food. The sample information is given below. Test at the 5% level whether there is a significant difference in the amounts dispensed by the two machines.

	Machine 1	Machine 2
Sample size	81	64
Mean (lb)	51	48
Standard deviation	4	3.5

The hypotheses are:

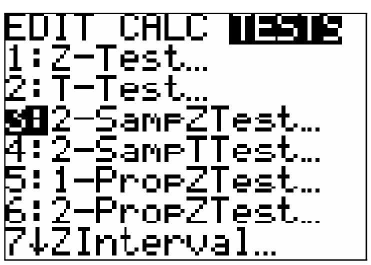
$$H_0 \; : \; \mu_1 = \mu_2$$
$$H_1 \; : \; \mu_1 \neq \mu_2$$

This is a two-tailed test with $\alpha = 0.05$. The critical values are $z = \pm 1.96$. (That is, we will reject H_0 if the test statistic $z \leq -1.96$ or $z \geq 1.96$).

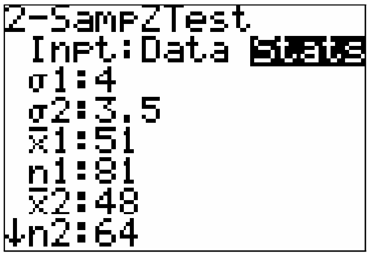
1. Press **STAT**

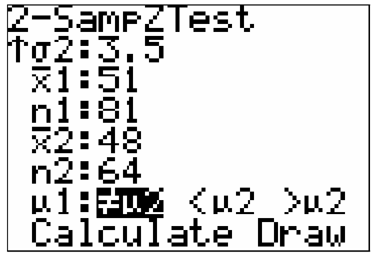
2. Arrow to **TESTS**

3. Select **3:2-SampZTest...**

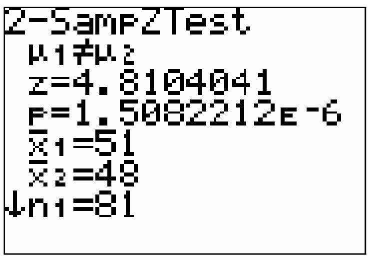


4. Enter the values as shown.





5. Highlight **Calculate**, then press **ENTER**.



Since $z = 4.8$ falls in the critical region, we reject H_0. There is sufficient evidence, at the 5% level, to conclude that there is a difference in the amounts dispensed by the two machines.

10.2 Test for a Difference Between Means (Dependent Samples)

Example 2 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. At a 5% level of significance, is there sufficient evidence to conclude that the fuel additive improved gas mileage?

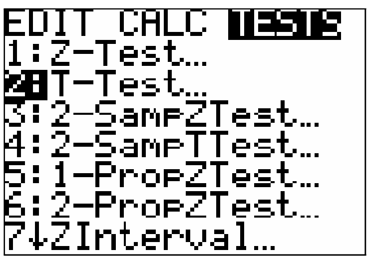
Car	1	2	3	4	5	6	7	8	9
Mileage without additive	34.5	36.7	34.4	39.8	33.6	35.4	38.4	35.3	37.9
Mileage with additive	36.4	38.8	36.1	40.1	34.7	38.3	40.2	37.2	38.7

The hypotheses are:

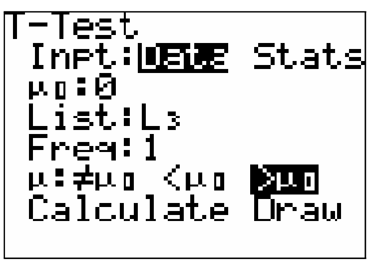
$$H_0 \ : \ \mu_d = 0$$
$$H_1 \ : \ \mu_d > 0$$

This is a right-tailed test with $\alpha = 0.05$. The critical value is $z = 1.645$. (That is, we will reject H_0 if the test statistic $t \geq 1.645$).

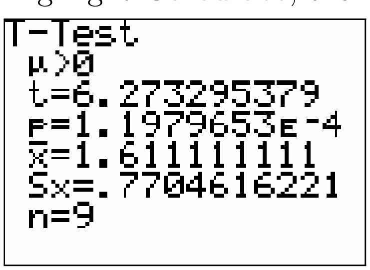
1. Enter data in **L1** and **L2**.
2. Highlight **L3** and enter **L2** − **L1**. Press **ENTER**. The differences are now shown in **L3**.
3. Press **STAT**
4. Arrow to **TESTS**
5. Select **2:TTest...**



6. Highlight **Data**.
7. Enter the values as shown.



8. Highlight **Calculate**, then press **ENTER**.



Since $t = 6.27$ falls in the critical region, we reject H_0. There is sufficient evidence, at the 5% level, to conclude that the fuel additive improved gas mileage. There is only a 0.012% chance of observing these differences if H_0 is true.

10.3 Test for a Difference Between Proportions

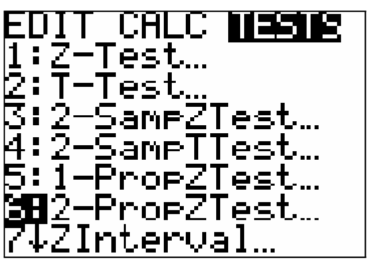
Example 3 A gardener sows two varieties of runner bean seeds. 152 seeds of Variety A yield 82 healthy plants. 100 seeds of Variety B yield 59 healthy plants. At a 5% level of significance, test whether there is a difference between the proportions of healthy plants that can be expected from the two varieties.

The hypotheses are:

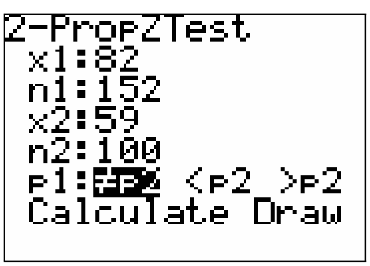
$$\begin{aligned} H_0 &: p_1 = p_2 \\ H_1 &: p_1 \neq p_2 \end{aligned}$$

This is a two-tailed test with $\alpha = 0.05$. The critical values are $z = \pm 1.96$. (That is, we will reject H_0 if the test statistic $z \leq -1.96$ or $z \geq 1.96$).

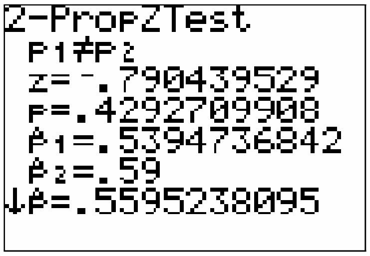
1. Press **STAT**
2. Arrow to **TESTS**
3. Select **6:2-PropZTest...**



4. Enter the values as shown.



5. Highlight **Calculate**, then press **ENTER**.



Since $z = -.79$ does not fall in the critical region, we do not reject H_0. There is not sufficient evidence, at the 5% level, to conclude that there is a difference between the proportions of healthy plants that can be expected from the two varieties. There is a 42.9% chance of observing these differences if H_0 is true.

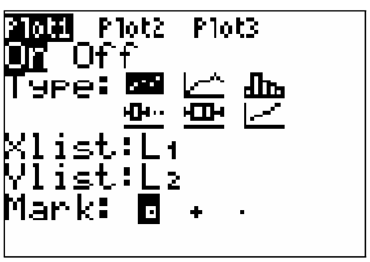
11 Linear Correlation and Regression

11.1 Scatterplot

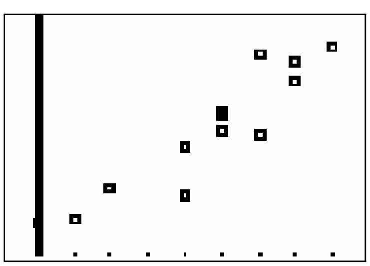
Example 1 The number of hours 13 students spent studying for a test and their scores on that test are given below.

Hours studying, x	0	1	2	4	4	5	5	5	6	6	7	7	8
Test score, y	40	41	51	48	64	69	73	75	68	93	84	90	95

1. Enter the paired data in **L1** and **L2**.
2. Press **2nd Y= 1** (or **2nd Y= ENTER**) to enter **STAT PLOT** menu.
3. Press **ENTER** to turn on `Plot 1`.
4. Press ▼ to move down to **Type** and ▶ to move the cursor over the scatterplot symbol (the first symbol). Press **ENTER** to highlight the symbol.
5. Set **Xlist** to **L1** and **Ylist** to **L2**.



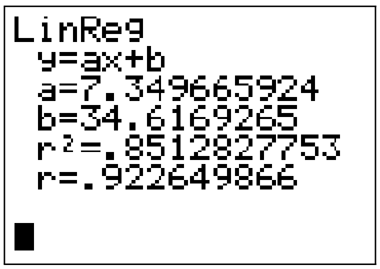
6. Press **ZOOM 9**.



This scatterplot indicates a positive linear association.

11.2 Linear Correlation Coefficient, r

1. Press **STAT**.
2. Arrow to **CALC**.
3. Select **4:LinReg(ax+b)**. Press **ENTER**.



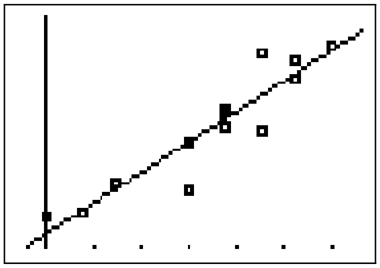
The correlation coefficient is $r = 0.923$.

11.3 Regression Line

The regression line is $\hat{y} = 34.6169 + 7.3497x$ valid for $0 \leq x \leq 8$. The slope of 7.3 indicates that for each additional hour spent studying, the test result will increase by 7.3.

Graph the regression line on the scatterplot.

1. Press **Y=** (Clear Y_1 if necessary)
2. Press **VARS**.
3. Select **5:Statistics...**
4. Arrow to **EQ**
5. Press **ENTER** (or **1**) to select **RegEQ.**
6. Press **GRAPH**



12 Chi-Square Analysis

12.1 Goodness-of-Fit Test

Example 1 Mars, Inc. claims that its M&M plain candies are distributed with the following color percentages: 30% brown, 20% yellow, 20% red, 10% orange, 10% green and 10% blue. A sample of M&Ms were collected with the following observed frequencies. At the 5% level of significance, test the claim that the color distribution is as claimed by Mars, Inc.

Brown	Yellow	Red	Orange	Green	Blue
33	26	21	8	7	5

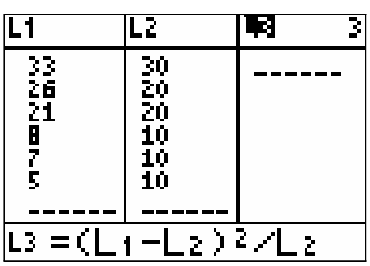
The hypotheses are:

$$H_0 \; : \; \text{The percentages are as claimed by Mars, Inc.}$$
$$H_1 \; : \; \text{At least one percentage is different from the claimed value.}$$

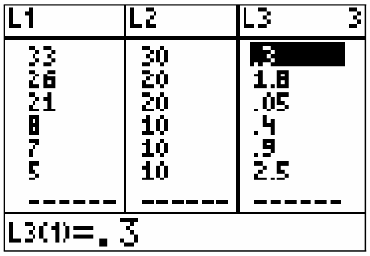
The critical value is $\chi^2 = 11.071$, where the degrees of freedom are $(k-1) = 5$.

1. Enter the observed frequencies into List 1, and the expected frequencies into List 2.

2. Highlight **L3**, and enter $(\mathbf{L_1}\text{-}\mathbf{L_2})^2/\mathbf{L_2}$.



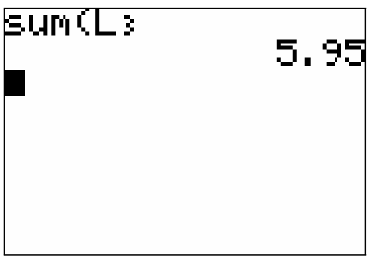
3. Press **ENTER**.



4. Press **2nd MODE** to return to the Home Screen.

5. Press **2nd STAT**, arrow to **MATH**, select **5:sum(**

6. Press **2nd 3** (to select **L3**). Press **ENTER**.



The test statistic is $\chi^2 = 5.95$. Since $\chi^2 = 5.95$ does not fall in the critical region, we do not reject H_0.

There is sufficient evidence, at the 5% level, to support the claim that the distribution of colors is as claimed by Mars, Inc.

12.2 Test for Independence

Example 2 At the 5% level of significance, use the data below to test the claim that when the *Titanic* sank, whether someone survived or died is independent of whether the person was a man, woman, boy or girl.

	Gender/Age			
	Men	Women	Boys	Girls
Survived	332	318	29	27
Died	1360	104	35	18

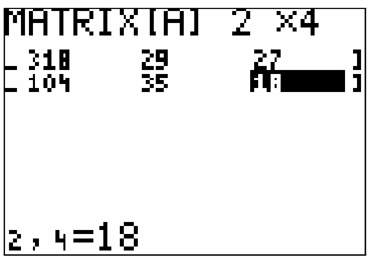
The hypotheses are:

H_0 : Whether a person survived is independent of gender and age.

H_1 : Whether a person survived is not independent of gender and age.

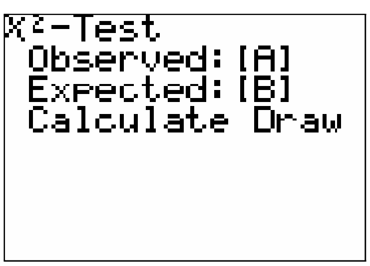
1. The critical value is $\chi^2 = 7.815$, where the degrees of freedom are $(r-1)(c-1) = (2-1)(4-1) = 3$.

2. Enter the data from the contingency table into Matrix A as a 2×4 matrix.

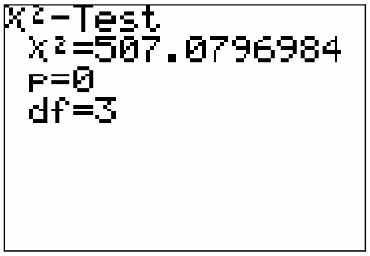


3. Press **STAT,** arrow to **TESTS** and select **C:**χ^2**-Test...**

4. Select **Calculate**.



5. Press **ENTER**.



The test statistic is $\chi^2 = 507.08$. Since $\chi^2 = 507.08$ falls in the critical region, we reject H_0.

There is not sufficient evidence, at the 5% level, to support the claim that whether someone survived or died is independent of whether the person was a man, woman, boy or girl. It appears that whether a person survived the sinking of the Titanic and whether that person was a man, woman, boy or girl are dependent variables.

The expected values are stored in Matrix B. To see the major differences, compare the observed and expected values for each category of the variables.

Category	Observed	Expected	$\frac{(O-E)^2}{E}$
Survived/Man	332	537.4	**78.5**
Survived/Woman	318	134	**252.7**
Survived/Boy	29	20.3	3.7
Survived/Girl	27	14.3	11.3
Died/Man	1360	1154.6	36.5
Died/Woman	104	288	**117.6**
Died/Boy	35	43.7	1.7
Died/Girl	18	30.7	5.3

We see that 318 women actually survived, although we would have expected only 134 if survivability is independent of gender/age. The other major differences are that fewer woman died (104) than expected (288), and fewer men survived (332) than expected (537).

13 Analysis of Variance (ANOVA)

Example 1 Do different age groups have different mean body temperatures? The table below lists the body temperatures of five randomly selected subjects from each of the three different age groups. At the 5% level of significance, test whether the three age-group populations have the same mean body temperature.

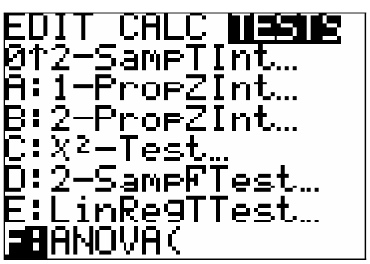
18 − 20	21 − 29	30 and older
98.0	99.6	98.6
98.4	98.2	98.6
97.7	99.0	97.0
98.5	98.2	97.5
97.1	97.9	97.3

The hypotheses are:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$
$$H_1 : \text{Not all the means are equal.}$$

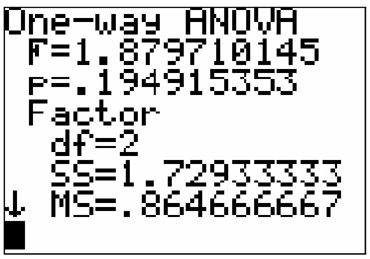
The critical value is $F = 3.8853$, read with $(k-1) = 2$ df in the numerator and $N - k = 12$ df in the denominator.

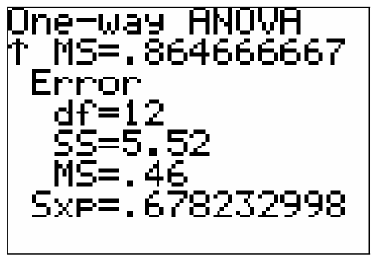
1. Enter the data into List 1, List 2 and List 3.
2. Press **STAT**, arrow to **TESTS** and select **F:ANOVA(**



3. Press **ENTER**, and enter the list labels.

4. Press **ENTER**.





The pooled standard deviation is 0.678, which is the best estimate of the population standard deviation σ. This is the square root of the Mean Square Error. The calculator output corresponds to the following ANOVA table.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Factor	1.729	2	0.865	1.88
Error	5.520	12	0.460	
Total	7.249	14		

Since $F = 1.88$ does not fall in the critical region, we do not reject H_0. There is not sufficient evidence, at the 5% level, to conclude that the three age-group populations have different mean body temperatures.

Example 2 National Computer Products, Inc. (NCP) manufactures printers and fax machines at plants located in Charlotte, Houston and San Diego. To measure how much employees at these plants know about total quality management, a random sample of six employees was selected from each plant and given a quality-awareness examination. The examination scores are shown below. Management would like to test the hypothesis that the mean examination score is the same at each plant.

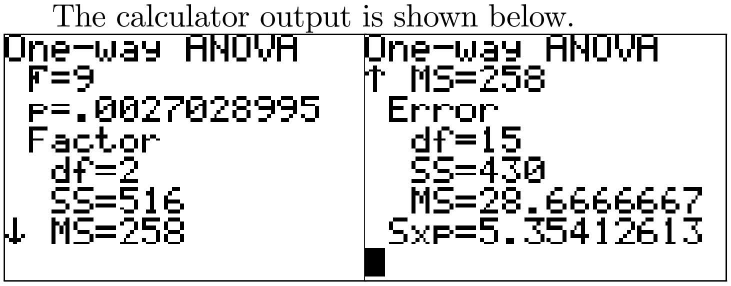
Observation	Plant 1 Charlotte	Plant 2 Houston	Plant 3 San Diego
1	85	71	59
2	75	75	64
3	82	73	62
4	76	74	69
5	71	69	75
6	85	82	67

The hypotheses are:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$
$$H_1 : \text{Not all the means are equal.}$$

The critical value is $F = 3.6823$, read with $(k-1) = 2$ df in the numerator and $N - k = 15$ df in the denominator.

The calculator output is shown below.



Since $F = 9$ falls in the critical region, we reject H_0. There is sufficient evidence, at the 5% level, to conclude that at least one mean is different.

13.1 Scheffé Test

A Scheffé Test can be performed to find which means have a significant difference. The means are compared pair-wise. For each comparison, calculate

$$\frac{(\bar{x}_a - \bar{x}_b)^2}{\dfrac{SS_W}{\sum(n_i - 1)}\left[\dfrac{1}{n_a} + \dfrac{1}{n_b}\right]}$$

where $\bar{x}_a$ and $\bar{x}_b$ are the means being compared and n_a and n_b are the corresponding sample sizes. Then compare the value to the critical value obtained in the one-way

ANOVA multiplied by $k-1$,which gives $CV_{Scheffé} = 7.3646$. To compare Plant 1 and Plant 2:

$$\frac{(\bar{x}_a - \bar{x}_b)^2}{\frac{SS_W}{N-k}\left[\frac{1}{n_a}+\frac{1}{n_b}\right]} = \frac{(79-74)^2}{\frac{430}{15}\left[\frac{1}{6}+\frac{1}{6}\right]} = 2.6163$$

That is,

$(1,2) \rightarrow 2.6163 \rightarrow$ No difference

$(1,3) \rightarrow 17.686 \rightarrow$ Significance difference

$(2,3) \rightarrow 6.698 \rightarrow$ No difference

Thus, there is a significance difference between the mean examination scores at the Charlotte and San Diego plants.