Elementary Statistics on the TI-83 and TI-84

Derek Collis
Harper College

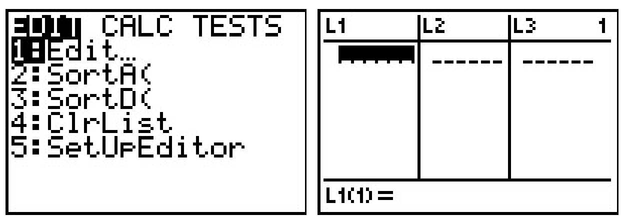
Contents

1 Lists

Data are stored in lists, which can be created and edited using the stat list editor. You can view up to 20 lists in the stat list editor; however, only three lists can be displayed at the same time. There are six default lists: L1 through L6; however, up to 99 lists can be created and named.

1.1 Displaying the stat list editor

1. Press **STAT**. (Figure 1.1(a))
2. Press **ENTER** to select `Edit`. (Figure 1.1(b))



(a) Edit menu (b) Stat list editor

Figure 1.1: Stat list editor

1.2 Entering data

1. Display the stat list editor.
2. Enter the data in `L1` and press **ENTER** after each value. (Figure 1.2)
3. After all the data values are entered, press **STAT** to get back to the `Edit` menu or **2nd** [**QUIT**] to return to the Home Screen.

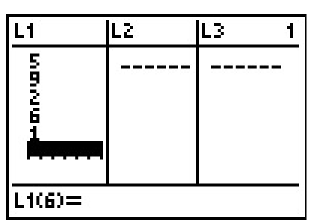


Figure 1.2: Entering data

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

1.3 Editing data

1.3.1 Correcting a data value

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

1.3.2 Inserting a data value into a list

1. Move cursor to position where data value is to be inserted, then press **2nd** [**INS**].
2. Type data value; then press **ENTER**.

1.3.3 Clearing a list

1. Move the cursor onto the list name.
2. Press **CLEAR**; then press either **ENTER** or ▼ (down arrow). (Figure 1.3)

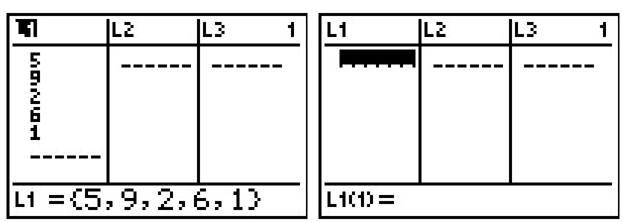


Figure 1.3: Clearing a list

1.4 Sorting data

1. Enter the data into `L1`.
2. Press **STAT 2** to select `SortA` (ascending order) or press **STAT 3** to select `SortD` (descending order).
3. Press **2nd** [**L1**] **ENTER**. The calculator will display `Done`.

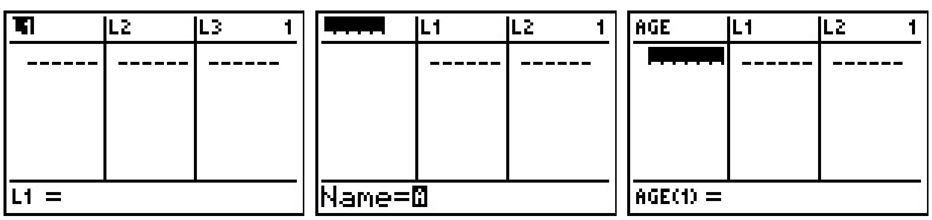
4. Press **STAT ENTER** to display the sorted list. (Figure 1.4)

Figure 1.4: Sorting data

1.5 Creating and naming a list

Create a list and name it `AGE`.

1. Display the stat list editor.
2. Move the cursor onto a list name (the new list will be inserted to the left of highlighted list), then press **2nd** [**INS**]. (Figures 1.5(a) and 1.5(b))
   The `Name=` prompt is displayed and alpha-lock is on. To exit from alpha-lock, press **ALPHA**.
3. Type in a name for the new list. (A maximum of 5 characters is allowed and the first character must be a letter.)
4. Press **ENTER** twice. (Figure 1.5(c))



(a) Stat list editor (b) New list (c) Named list

Figure 1.5: Creating and naming a list

1.6 Removing a list from the stat list editor

1. Move the cursor onto the list name.
2. Press **DEL**.

Note: The list is not deleted from memory; it is only removed from the stat list editor.

1.7 Displaying all list names

Press **2nd** [**LIST**].

1.8 Displaying selected lists in the stat list editor

To display `L1`,`L2` and `L5`.

1. Press **STAT 5** to select `SetUpEditor`.
2. Press **2nd** [**L1**], **2nd** [**L2**] and **2nd** [**L5**].
3. Press **ENTER**.
4. Press **STAT ENTER** to view the Stat List Editor. (Figure 1.6)

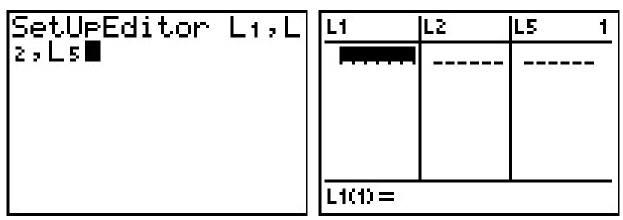


Figure 1.6: Displaying selected lists in stat list editor

1.9 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 from the Stat List Editor.

1.10 Copying one list to another list

To copy the data in `L1` to `L2`.

1. Move the cursor onto `L2`.
2. Press **2nd** [**L1**].
3. Press **ENTER**. The data values from `L1` now appear in `L2`. (Figure 1.7)

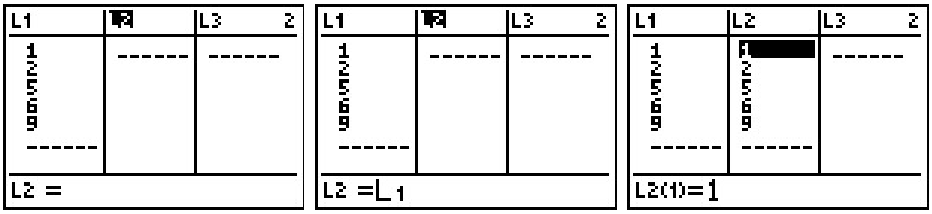


Figure 1.7: Copying data

1.11 Combining two or more lists into a single list

To combine data in `L1` and `L2` and store into `L3`.

1. Enter the data in `L1` and `L2`.
2. Press **2nd** [**LIST**], arrow to `OPS`, then press **9** (to select `augment`).
3. Enter `L1,L2`, press **STO▶**, and enter `L3`; then press **ENTER**. (Figure 1.8)

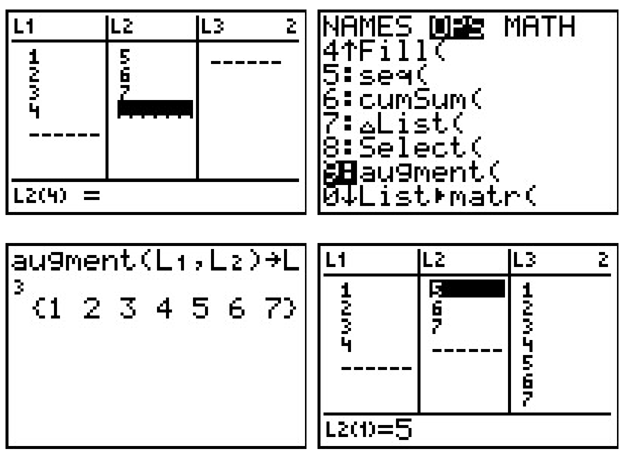


Figure 1.8: Combining lists

If you had entered `L2,L1` then the entries from `L2` would be listed first in `L3`.

1.12 Applying arithmetic operations to lists

To multiply the corresponding entries in `L1` and `L2` and then store these products in `L3`.

1. Enter the data in `L1` and `L2`.

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

2. Move the cursor onto `L3`.
3. Enter `L1*L2`; then press **ENTER**. The sums appear in `L3`. (Figure 1.9)

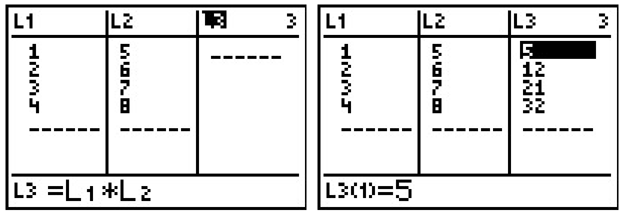


Figure 1.9: Multiplying two lists

All list elements remain, but the formula is detached and the lock symbol disappears.

1.13 Deleting a list from memory

1. Press **2nd** [**MEM**].
2. Press **2** to select `Delete`.
3. Press **4** to select `List`.
4. Arrow to list that you wish to delete:
    - for a TI-83, press **ENTER**
    - for a TI-83 Plus, press **DEL**
    - for a TI-84, press **DEL**

2 Graphs

2.1 Histogram

Example 2.1
Generate a histogram for the frequency distribution in Table 2.1.

Table 2.1: Frequency distribution

Class	f	Class Mark
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`. (Figure 2.1)

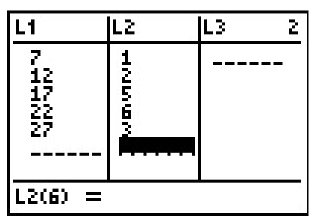


Figure 2.1: Data entered in L1 and L2

2. Press **2nd** [**Y=**] (to select **STAT PLOT**).
3. Press **ENTER** to turn on `Plot1`
4. Arrow down to `Type`. Arrow right to highlight the histogram symbol, then press **ENTER**.
5. Arrow down to `Xlist`. Set `Xlist` to `L1`
6. Arrow down to `Freq`. Set `Freq` to `L2`. (Figure 2.2 on the following page)
7. Press **WINDOW** and make the settings as shown in Figure 2.3 on the next page.

   Table 2.2 on the following page explains the **WINDOW** settings.
8. Press **GRAPH**.
9. To obtain coordinates, press **TRACE**, followed by left or right arrow keys.

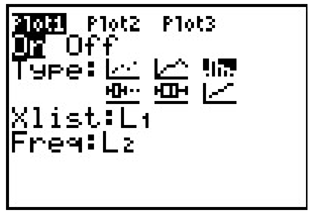


Figure 2.2: Stat plot menu settings

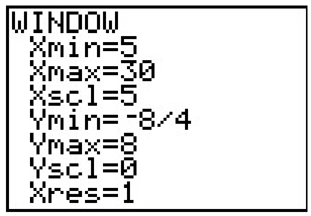
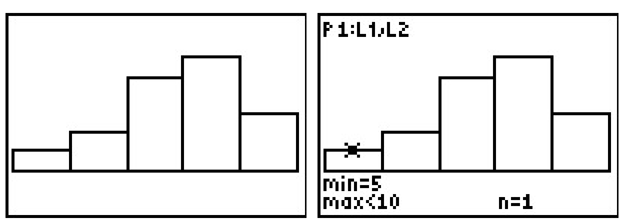


Figure 2.3: Window settings

Table 2.2: Window settings

`Xmin`	=	lower limit of first class
`Xmax`	=	lower limit of last class plus class width (This would be the lower limit of the next class if there were one.)
`Xscl`	=	class width
`Ymin`	=	-Ymax/4 (The purpose of making Ymin = $-$Ymax/4 is to allow sufficient space below the histogram so that the screen display is easily read.)
`Ymax`	=	maximum frequency (or a little more) of distribution
`Yscl`	=	0



(a) Histogram (b) Histogram values

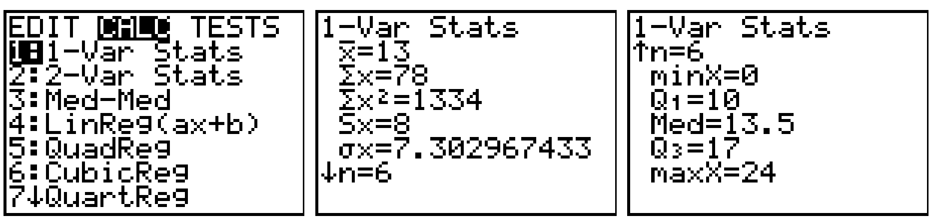
Figure 2.4: Histogram

3 Measures of Center and Variation

3.1 Ungrouped data

Example 3.1
The following are the hours per week worked by a sample of 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`. (Figure 3.1(a))
3. Press **1** or **ENTER** to select `1-Var Stats`. (If your data are in a list other than `L1`, you need to enter the list name; for example, if your data are in `L2`, enter `1-Var Stats L2`.)
4. Press **ENTER**. (Figure 3.1(b))
5. Press ▼ to scroll down to see the median and more information. (Figure 3.1(c))



(a) CALC screen (b) 1-Var screen (c) 1-Var screen

Figure 3.1: Summary statistics for Example 3.1

The mean is 13 hours and the median is 13.5 hours. The standard deviation is 8 hours.

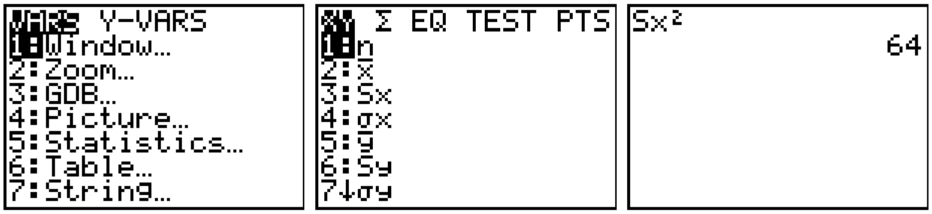
3.1.1 The variance

To obtain the variance for Example 3.1, perform the following.[1]

1. Press **VARS 5** to select `Statistics`. (Figure 3.2(a) on the following page)
2. Press **3** to select `Sx` (or press **4** for σx). (Figure 3.2(b) on the next page)
3. Press $\mathbf{x^2}$; then press **ENTER**. (Figure 3.2(c) on the following page)

This yields a variance of 64.

[1] This procedure avoids using a rounded standard deviation value to obtain the variance.



(a) VARS screen (b) Statistics screen (c) Variance

Figure 3.2: Obtaining the variance for Example 3.1 on the previous page

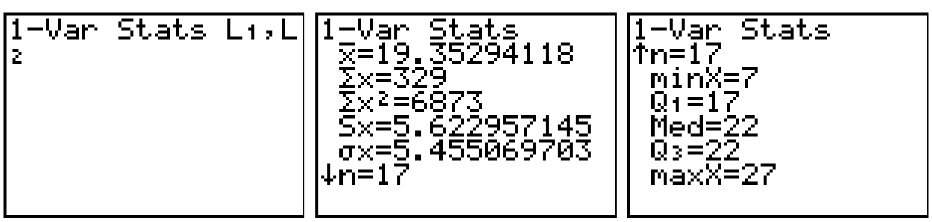
3.2 Grouped data

Example 3.2
The frequency distribution shown in Table 3.1 shows the number of minutes it takes for a sample of seventeen students to drive from home to school.

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

Table 3.1: Frequency distribution for Example 3.2

1. Enter the class marks (midpoints) into `L1` and the frequencies into `L2`.
2. Press **STAT**, arrow to `CALC`, then press **ENTER** to select `1-Var Stats`.
3. Enter `L1,L2` (Figure 3.3(a)), then press **ENTER**. (Figures 3.3(b) and 3.3(c))



(a) 1-Var screen (b) Output (c) Output

Figure 3.3: Obtaining the variance for Example 3.2

The mean drive time is 19.4 minutes and the median is 22 minutes. The standard deviation is 5.6 minutes.

4 Boxplots

There are two types of boxplots: the **standard boxplot** and the **modified boxplot**. The standard boxplot is the fifth symbol in `Type` (located in **STAT PLOT**) and the modified boxplot is the fourth symbol.

The standard boxplot represents the five-number summary: `min`, $\mathtt{Q}_1$, `med`, $\mathtt{Q}_3$, `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 data value and the largest data value in the interval

$$(\textit{lower fence, upper fence}) = (Q_1 - 1.5 \times \text{IQR}, Q_3 + 1.5 \times \text{IQR})$$

where IQR is the interquartile range ($Q_3 - Q_1$). Generally, values outside of this range are considered outliers.

Example 4.1
Generate a standard boxplot and modified boxplot for the values:

$$1 \quad 2 \quad 3 \quad 3 \quad 4 \quad 5 \quad 5 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 25$$

1. Enter the data values into `L1`.
2. Press **2nd** [**STAT PLOT**], then press **ENTER**.
3. Turn `Plot1` on, and set the window as shown in Figure 4.1.

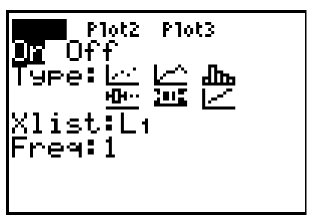
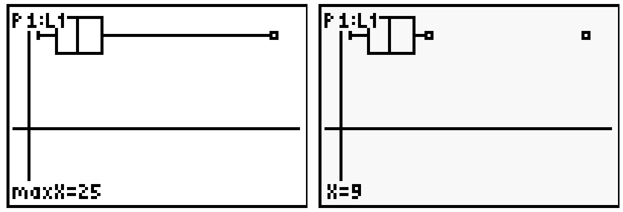


Figure 4.1: Plot1 screen

4. Press **ZOOM 9** (for `ZoomStat`).
5. Press **TRACE** and press ▶ to locate the end of the right whisker.

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 (See Figure 4.2(b) on the next page). This is because 25 lies outside the interval $(-3.75, 14.25)$. The value of 9 is the largest that lies inside this interval. So we have identified 25 as an outlier.



(a) Standard boxplot (b) Modified boxplot

Figure 4.2: Standard and modified boxplots with respect to an outlier

4.1 Comparing two or three boxplots

Boxplots make it easy to compare samples from the same or different populations. Multiple boxplots may be put on the same axes and thus make comparisons easier than multiple histograms, each of which require a separate graph. The TI-83 can compare up to three boxplots.

Example 4.2
The following data represent the number of cold cranking amps of group size 24 and group size 35 batteries. The cold cranking amps number measure the amps produced by the battery at 0^o Fahrenheit. Which type of battery would you prefer?

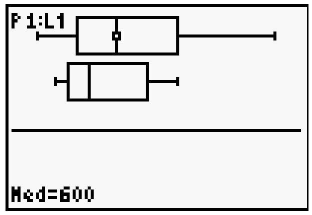
Group Size 24 Batteries			**Group Size 35 Batteries**		
800	600	675	525	620	550
600	525	700	560	675	550
500	660	550	530	570	640
585	675		525	640	640

1. Enter the data values into `L1` and `L2`.
2. Set `Plot1` for a standard or modified boxplot, and set `XList` to `L1`.
3. Set `Plot2` for the same type of boxplot, and set `XList` to `L2`.
4. Press **ZOOM 9**. (See Figure 4.3 on the following page).

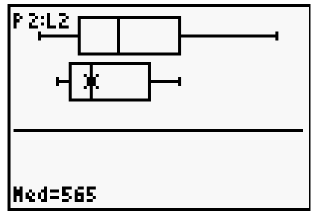
Press the up and down arrows to move between the two boxplots.

Figure 4.3 on the next page shows modified boxplots for the two samples of batteries. The five-number summaries are:

Type of battery	Minimum	Q_1	Median	Q_3	Maximum
Group Size 24	500	550	600	675	800
Group Size 35	525	540	565	640	675



(a) Group size 24 with median of 600 shown



(b) Group size 35 with median of 565 shown

Figure 4.3: Boxplots for group size 24 and 35 batteries

There are no outliers for either type of battery. We see that the group size 24 batteries are higher, on average, than the group size 35 batteries. The display reveals the difference in median cold cranking amps between the two types of batteries: size 24 battery was 600 compared to a median of 565 for the size 35 battery. The upper 25% of the size 24 batteries have greater cold cranking amps than the maximum of the size 35 batteries. Both distributions are right-skewed, with group size 24 battery having more variability. Based on this simple graphical analysis, a group size 24 battery would be preferable.

5 Linear Correlation and Regression

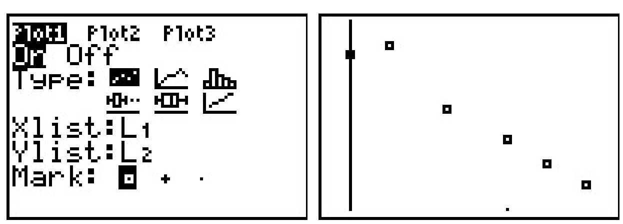
Example 5.1
An educator wants to see how the number of absences a student in his class has affects the student's final grade. The data obtained from a sample are as follows:

No. of absences, x	Final grade, y
10	70
12	65
2	96
0	94
8	75
5	82

The number of absences is the predictor variable and the final grade is the response variable.

5.1 Scatterplot

1. Enter the bivariate data in `L1` and `L2`: the predictor variable values in `L1` and the response variable values in `L2`.
2. Press **2nd** [**STAT PLOT**].
3. Turn on `Plot1` and set `Type` for scatterplot (first symbol in the first row).
4. Set `Xlist` to `L1` and `Ylist` to `L2`. (Figure 5.1(a))
5. Press **ZOOM 9**. (Figure 5.1(b))



(a) Plot screen for scatterplot

(b) Scatterplot

Figure 5.1: Scatterplot input and output screen

The scatterplot indicates a negative linear correlation. That is, as the number of absences increases, the final grade decreases.

5.2 Linear correlation coefficient

1. Press **STAT**.
2. Arrow to `CALC`.
3. Press **4** (to select `LinReg(ax+b)`); then press **ENTER**. (Figure 5.2)

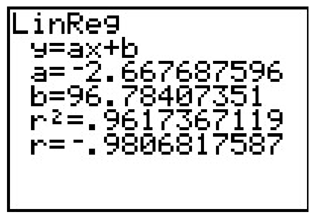


Figure 5.2: LinReg screen

> **r and r^2 not showing?**
> If r and r^2 do not appear on the screen, press **2nd** [**CATALOG**], arrow to `DiagnosticOn` and press **ENTER** twice.

The correlation coefficient, $r = -0.98$, indicates a very strong negative correlation between number of absences and final grade. The coefficient of determination, $r^2 = 0.96$, indicates that about 96% of the variation in final grade is explained by the number of absences. The unexplained variation of 4% is attributable to other factors. *What do you think these could be?*

5.3 Regression line

The regression equation is $\hat{y} = -2.6677x + 96.784$, valid for $0 \leq x \leq 12$. The slope of -2.6677 indicates that for each additional day's absence the final grade decreases, on average, by 2.6677 points. The y-intercept of 96.784 is the predicted score for a student who has no absences.

> **Practical interpretation of y-intercept**
> In linear regression, the estimated y-intercept will often not have a practical interpretation. It will, however, be practical if the value $x = 0$ is meaningful and within the scope of the model.

5.3.1 Graph the regression line on the scatterplot

There are two ways to graph the regression on the scatterplot as shown in below.

Method 1

1. Press **Y=**. (Clear Y_1 if necessary.)
2. Press **VARS 5** (to select `Statistics`).
3. Arrow to `EQ`.
4. Press **ENTER** to select `RegEQ`.
5. Press **GRAPH**. (Figure 5.3)

Method 2

1. Press **STAT**.
2. Arrow to `CALC`.
3. Press **4** (to select `LinReg(ax+b)`).
4. Enter `L1,L2,Y1`, then press **ENTER**.

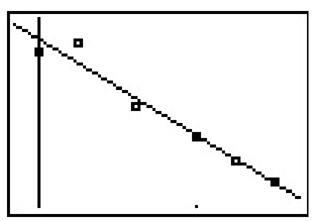


Figure 5.3: Scatterplot with regression line

6 The Binomial Distribution

6.1 Probability for a single value

Example 6.1
Find $P(X = 3)$ where $n = 5$ and $p = 0.2$.

1. Press **2nd** [**DISTR**].
2. Press **0**[1] (for `binompdf`).
3. Enter `5,.2,3)`; then press **ENTER**. (See Figure 6.1).

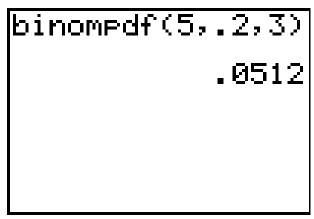


Figure 6.1: $P(X = 3)$ where $n = 5$ and $p = 0.2$

6.2 Probabilities for more than one value

Example 6.2
Find $P(X = 1, 2)$ where $n = 5$ and $p = 0.2$.

1. Press **2nd** [**DISTR**].
2. Press **ALPHA 0** (for `binompdf`.
3. Enter `5,.2,{1,2}`; then press **ENTER**. (See Figure 6.2).

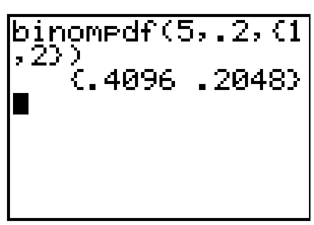


Figure 6.2: $P(X = 1, 2)$ where $n = 5$ and $p = 0.2$

[1] For a TI-84, press **ALPHA** [**A**], to select `binompdf`

6.3 Cumulative probability

Example 6.3
Find $P(X \leq 3)$ where $n = 5$ and $p = 0.2$.

1. Press **2nd** [**DISTR**].
2. Press **ALPHA** [**A**][2] (for `binomcdf`).
3. Enter `5,.2,3`); then press **ENTER** (See Figure 6.3).

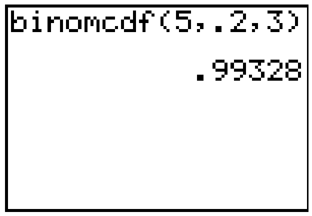


Figure 6.3: $P(X \leq 3)$ where $n = 5$ and $p = 0.2$

Example 6.4
Find $P(X \geq 3)$ where $n = 5$ and $p = 0.2$.

$P(X \geq 3) = 1 - P(X \leq 2)$

1. Press **1 − 2nd** [**DISTR**].
2. Press **ALPHA** [**A**] (for `binomcdf`).
3. Enter `5,.2,2)`; then press **ENTER** (See Figure 6.4).

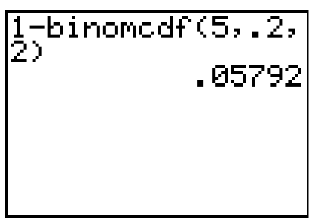


Figure 6.4: $P(X \geq 3)$ where $n = 5$ and $p = 0.2$

Example 6.5
Find $P(3 \leq X \leq 7)$ where $n = 10$ and $p = 0.2$.

$P(3 \leq X \leq 7) = P(X \leq 7) - P(X \leq 2)$

1. Press **2nd** [**DISTR**].
2. Press **ALPHA** [**A**] (for `binomcdf`).

[2]For a TI-84, press **ALPHA** [**B**], to select `binomcdf`

3. Enter `10,.2,7) −`.
4. Press **2nd** [**DISTR**]
5. Press **ALPHA** [**A**] (for `binomcdf`)
6. Enter `10,.2,2`; then press **ENTER** (See Figure 6.5).

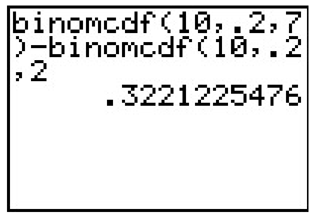


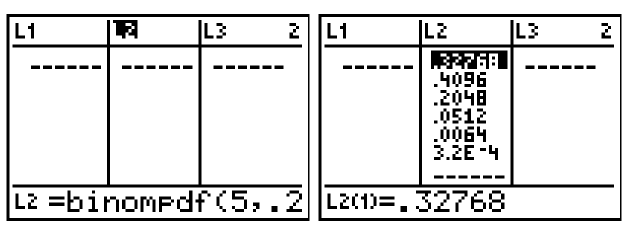
Figure 6.5: $P(3 \leq X \leq 7)$ where $n = 10$ and $p = 0.2$

6.4 Constructing a binomial probability distribution

Example 6.6
Construct a binomial probability distribution for $n = 5$ and $p = 0.2$.

6.4.1 Method 1

1. Display the stat list editor.
2. Move the cursor onto `L2`.
3. Press **2nd** [**DISTR**].
4. Press **0** (fir `binompdf`) and enter `5,.2` as shown in Figure 6.6(a).
5. Press **ENTER**. The probabilities are displayed in `L2` as shown in Figure 6.6(b).



(a) Entering formula in **L2** (b) Probabilities in **L2**

Figure 6.6: Entering probabilities in **L2**

6. Enter the values `0, 1, 2, 3, 4, 5` into `L1`.

6.4.2 Method 2

1. In the home screen press **2nd** [**DISTR**].
2. Press **0** (for `binompdf`) and enter `5,.2`).
3. Press **STO▶ 2nd** [**L2**]; then press **ENTER** (See Figure 6.7).

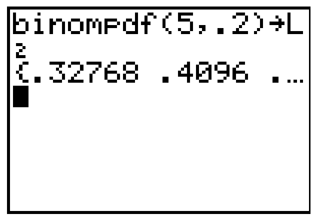


Figure 6.7: Entering formula from the home screen

6.5 Constructing a binomial probability histogram

Example 6.7
Construct a binomial probability histogram for $n = 5$ and $p = 0.2$.

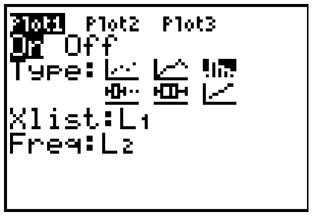
1. Construct the binomial probability distribution in `L1` and `L2`.
2. Set up `Plot1` for a `Histogram` as shown in Figure 6.8(a) on the following page.
3. Set the window as shown in Figure 6.8(b) on the next page[3], and press **GRAPH** to display the histogram as shown in Figure 6.8(c) on the following page.
4. Using **TRACE** we can read the probabilities of the distribution; for example, $P(X = 1) = n = 0.4096$ as shown in Figure 6.8(d) on the next page.

6.6 Quick method for entering a set of integers into a list

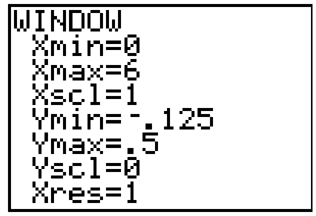
Example 6.8
Enter the integers 0 to 20 into L1.

1. Highlight `L1`.
2. Press **2nd** [**LIST**].
3. Arrow to `OPS` and press **5** (for `seq`).

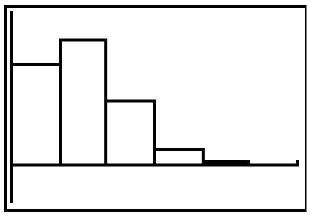
[3]The `Ymax` is chosen as to be larger than the largest probability and the `Ymin` is chosen to be `-Ymax/4` to give the appropriate space below the histogram for reading the **TRACE** values.



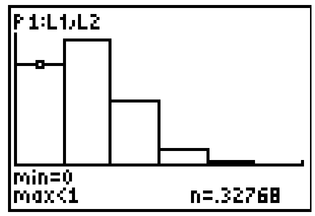
(a) Plot1 settings



(b) Window settings



(c) Histogram



(d) Using **TRACE** to display the probabilities

Figure 6.8: Settings and Histogram

4. Enter `X,X,0,20` as shown in Figure 6.9, then press **ENTER**.

   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`.

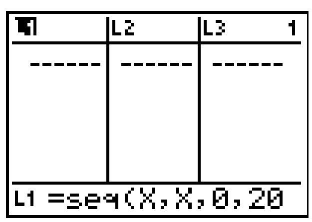


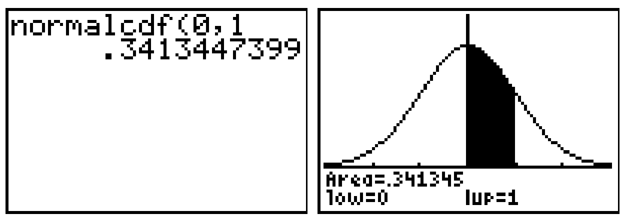
Figure 6.9: Generating a sequence of integers

7 The Normal Distribution

7.1 Probability between two z values

Example 7.1
Find $P(0 < z < 1)$.

1. Press **2nd** [**DISTR**].
2. Press **2** (to select `normalcdf`).
3. Enter `0,1`, then press **ENTER**. (Figure 7.1)



(a) Screen output (b) Area under curve

Figure 7.1: $P(0 < z < 1)$

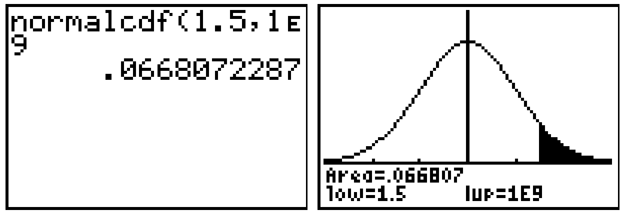
7.2 Probability greater than a z value

Example 7.2
Find $P(z > 1.5)$.

7.2.1 Method 1

1. Press **2nd** [**DISTR**].
2. Press **2** (to select `normalcdf`).
3. Enter `1.5,1E9`; then press **ENTER**. (Figure 7.2 on the following page)

Representing infinity
$P(z > 1.5)$ implies the interval $1.5 < z < \infty$. We represent ∞ by a large number, such as 1,000,000,000 or, in scientific notation, 1×10^9. This is entered into the calculator as `1 2nd` [EE] 9 and is displayed as 1E9.



(a) Screen output (b) Area under curve

Figure 7.2: $P(z > 1.5)$: Method 1

7.2.2 Method 2

1. Enter `.5` −.
2. Press **2nd** [**DISTR**].
3. Press **2** (to select `normalcdf`).
4. Enter `0,1.5`; then press **ENTER**. (Figure 7.3)

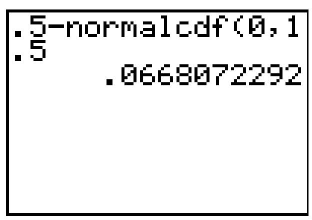


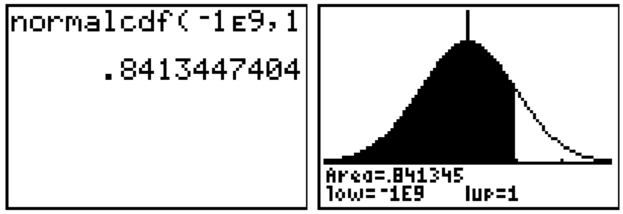
Figure 7.3: $P(z > 1.5)$: Method 2

7.3 Probability less than a z value

Example 7.3
Find $P(z < 1)$.

7.3.1 Method 1

1. Press **2nd** [**DISTR**].
2. Press **2** (to select `normalcdf`).
3. Enter −`1E9,1`; then press **ENTER**. (Figure 7.4 on the following page)



(a) Screen output (b) Area under curve

Figure 7.4: $P(z < 1)$: Method 1

7.3.2 Method 2

1. Enter `.5` +.
2. Press **2nd** [**DISTR**].
3. Press **2** (to select `normalcdf`).
4. Enter `0,1`, then press **ENTER**. (Figure 7.5)

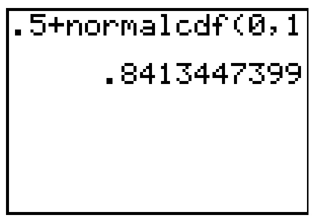


Figure 7.5: $P(z < 1)$: Method 2

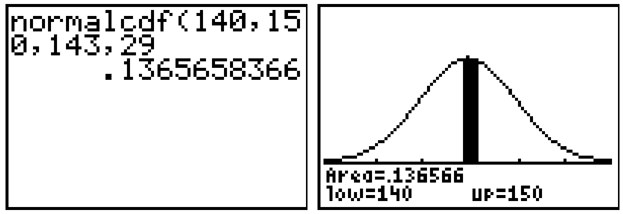
7.4 Probability between two x values

Example 7.4
Find Find $P(140 < x < 150)$ where $\mu = 143$ and $\sigma = 29$.

1. Press **2nd** [**DISTR**].
2. Press **2** (to select `normalcdf`).
3. Enter `140,150,143,29`, then press **ENTER**. (Figure 7.6 on the following page)

7.5 Probability less than an x value

Example 7.5
Find $P(x < 135)$ where $\mu = 143$ and $\sigma = 29$.

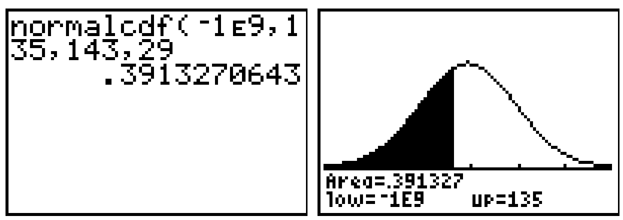


(a) Screen output (b) Area under curve

Figure 7.6: $P(140 < x < 150)$

7.5.1 Method 1

1. Press **2nd** [**DISTR**].
2. Press **2** (to select `normalcdf`).
3. Enter -1`E9,135,143,29`; then press **ENTER**. (Figure 7.7)



(a) Screen output (b) Area under curve

Figure 7.7: $P(140 < x < 150)$: Method 1

7.5.2 Method 2

1. Enter `.5` $-$
2. Press **2nd** [**DISTR**].
3. Press **2** (to select `normalcdf`).
4. Enter `135,143,143,29`; then press **ENTER**. (Figure 7.8 on the following page)

7.6 Finding a z value

Example 7.6
Find z such that 5% of the values are less than z.

1. Press **2nd** [**DISTR**].

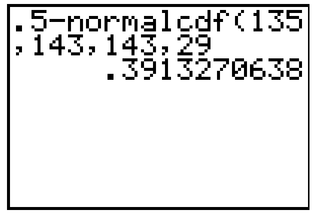
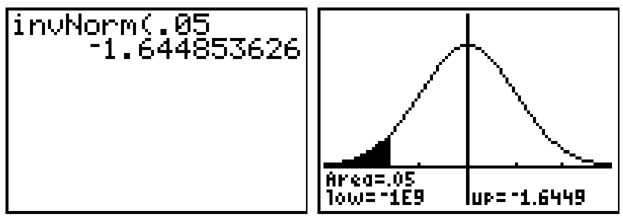


Figure 7.8: $P(140 < x < 150)$: Method 2

2. Press **3** (to select `invNorm`).
3. Enter `.05`, then press **ENTER**. (Figure 7.9)

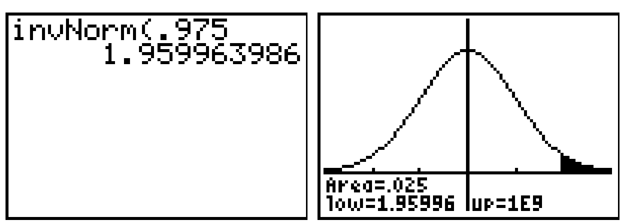


(a) Screen output (b) Area under curve

Figure 7.9: z such that 5% of the values are less than z.

Example 7.7
Find z such that 2.5% of the values are greater than z.

1. Press **2nd VARS** (to select **DISTR**).
2. Select **invNorm**.
3. Enter `.975`, then press **ENTER**. (Figure 7.10)



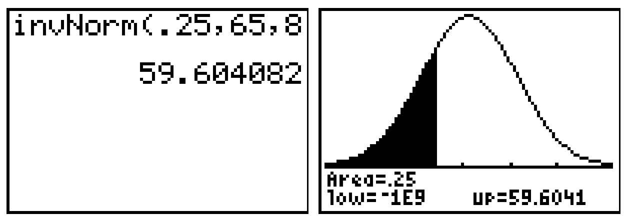
(a) Screen output (b) Area under curve

Figure 7.10: z such that 2.5% of the values are greater than z.

7.7 Finding an x value

Example 7.8
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 **invNorm**.
3. Enter `.25,65,8`, then press **ENTER**. (Figure 7.11)

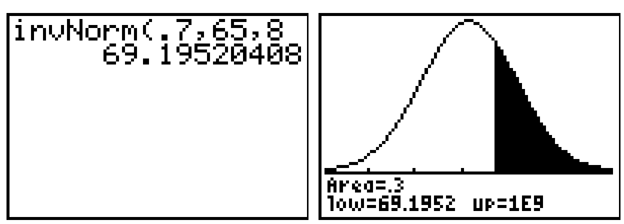


(a) Screen output (b) Area under curve

Figure 7.11: x such that 25% of the values are less than x

Example 7.9
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 **invNorm**.
3. Enter `.7,65 ,8`, then press **ENTER**. (Figure 7.12)



(a) Screen output (b) Area under curve

Figure 7.12: x such that 30% of the values are greater than x

8 Assessing Normality

To assess the likelihood that a sample came from a population that is normally distributed, we use a **normal probability plot**.

8.1 Normal probability plots

Example 8.1
The following data represent the number of miles on a four-year-old Chevy Camaro. Determine whether the data could have come from a population that is normally distributed.

42,544	27,274	34,258	59,177	44,091
35,631	42,371	48,018	58,795	44,832

1. Enter the data into **L1**.
2. Press **2nd** [**STAT PLOT**].
3. Press **ENTER** (to select `Plot1`).
4. Turn `Plot1` on.
5. Arrow down to `Type` and highlight the `Normal Probability Plot` icon; press **ENTER**.
6. Set the `Data List` to `L1` and the `Data Axis` to `X`.
7. Press **ZOOM 9** (to select `ZoomStat`). (Figure 8.1)

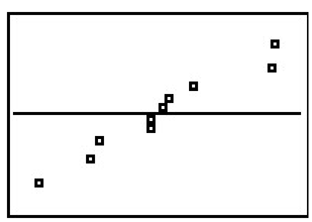


Figure 8.1: Normal probability plot for Example 8.1

The normal probability plot is fairly linear, therefore we can conclude that the sample data came from a population that is approximately normally distributed.

Example 8.2
Determine whether the data could have come from a population that is normally distributed.

43	63	53	50	58	44
53	53	52	14	50	43

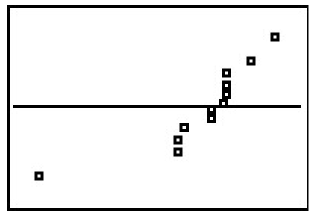


Figure 8.2: Normal probability plot for Example 8.2 on the previous page

The normal probability plot looks fairly linear except for the value of 14, which falls well outside the overall linear pattern, and is a potential outlier. The modified boxplot confirms this.

It would be important to determine whether this value of 14 is an incorrect entry or a correct, but exceptional, observation. This outlier will affect both the mean and the standard deviation, because neither is resistant. The normal probability plot below shows what would result if we removed the value of 14.

Example 8.3
The normal probability plot for a random sample of 15 observations is shown. Determine whether the data could have come from a population that is normally distributed.

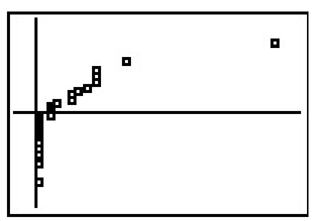


Figure 8.3: Normal probability plot for Example 8.3

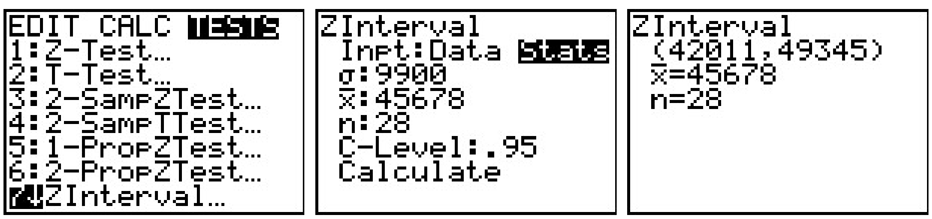
The non-linearity of the normal probability plot suggests that it is unlikely that this sample came from a population that is normally distributed.

9 Confidence Intervals

9.1 Confidence interval for a population mean: σ known

Example 9.1
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`; then press **7** (to select `ZInterval`). (Figure 9.1(a))
3. Highlight `Stats` and press **ENTER**.
4. Enter the values for σ, $\bar{x}$, `n` and `C-Level`[1]. (Figure 9.1(b))
5. Highlight `Calculate`; then press **ENTER**. (Figure 9.1(c))



(a) Tests: ZInterval (b) ZInterval screen (c) ZInterval output

Figure 9.1: Confidence interval for a population mean, σ known, using the summary statistics

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.

Interpretation of the confidence interval
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. We don't know if this particular interval contains μ or not: our confidence is in the procedure, not this particular interval. 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: there is no chance or probability associated with it. The probability that this interval contains μ is 0 or 1.

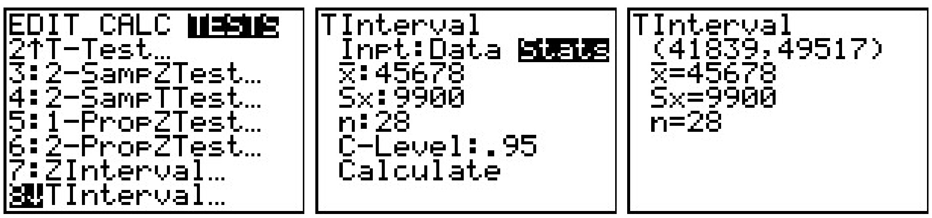
[1] The confidence level can be entered either as a decimal (.95) or as the percentage value (95).

9.2 Confidence interval for a population mean, σ unknown

Example 9.2

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`, then press **8** (to select `TInterval`). (Figure 9.2(a))
3. Highlight `Stats` and press **ENTER**.
4. Enter the values for $\bar{x}$, s_x, n and `C-Level`. (Figure 9.2(b))
5. Highlight `Calculate`; then press **ENTER**. (Figure 9.2(c))



(a) Tests: ZInterval (b) ZInterval screen (c) ZInterval output

Figure 9.2: Confidence interval for a population mean, σ unknown, using the summary statistics

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.

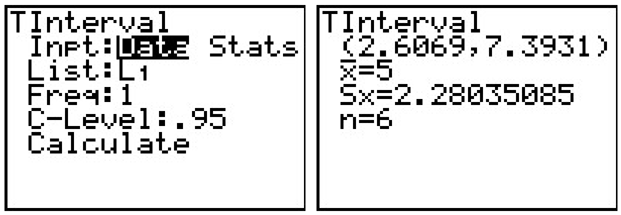
> **Difference between Z interval and T interval**
>
> The confidence interval using the t statistic is wider than the interval using the z statistic, even though the sample sizes are the same and the same value for and s is used. The reason for this is that the primary difference between the sampling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

Example 9.3

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**, then press **8** (to select `TInterval`).
4. Highlight `Data`. (Figure 9.3(a))
5. Enter the values for `List`, `Freq` and `C-Level`.
6. Highlight `Calculate`; then press **ENTER**. (Figure 9.3(b))



(a) TInterval *Data* screen (b) TInterval output

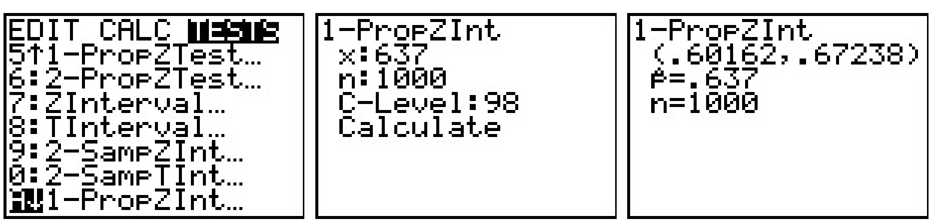
Figure 9.3: Confidence interval for a population mean, σ unknown, using the raw data

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

9.3 Confidence interval for a population proportion

Example 9.4
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 98% confidence interval for the population proportion of all U.S. citizens who trust the president.

1. Press **STAT**.
2. Arrow to **TESTS**, then press **ALPHA** [**A**] (to select `1-PropZInt`). (Figure 9.4(a))
3. Enter the values for x, n, and `C-Level`. (Figure 9.4(b))
4. Highlight `Calculate`, then press **ENTER**. (Figure 9.4(c))



(a) Tests: 1-PropZInt (b) 1-PropZInt screen (c) 1-PropZInt output

Figure 9.4: Confidence interval for a population proportion

We are 98% confident that the true percentage of all U.S. citizens who trust the president is between 60.2% and 67.2%.

10 Hypothesis Tests

10.1 Test for a mean: σ known

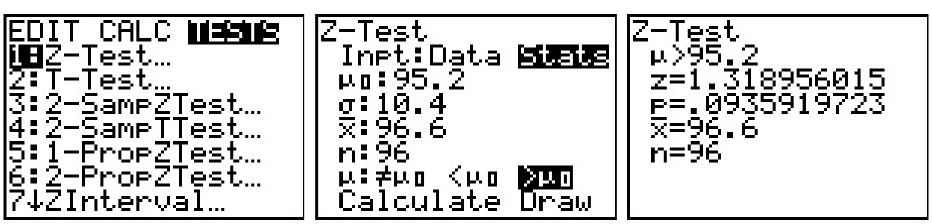
Example 10.1
A lightbulb manufacturer has established that the life of a bulb has mean 95.2 days with standard deviation 10.4 days. Following a change in the manufacturing process which is intended to increase the life of a bulb, a random sample of 96 bulbs has mean life 96.6 days. Test whether there 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`, then press **1** or **ENTER** (to select `Z-Test`). (Figure 10.1(a))
3. Highlight `Stats`, the press **ENTER**.
4. Enter the values:
   - μ_0: (value of μ under H_0)
   - σ: (population standard deviation)
   - $\bar{x}$: (sample mean)
   - n: (sample size)
   - $\mu :\neq \mu_0 \quad < \mu_0 \quad > \mu_0$ (form of H_1) (Figure 10.1(b))
5. Highlight `Calculate`, then press **ENTER**. (Figure 10.1(c))



(a) Tests: Z-Test (b) Z-Test screen (c) Z-Test output

Figure 10.1: Z-Test

Since $z = 1.32$ does not fall in the critical region, we do not reject H_0.

There is not sufficient evidence to indicate that the new process has led to an increase in the life of the bulbs.

10.2 Test for a mean: σ unknown

Example 10.2
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:

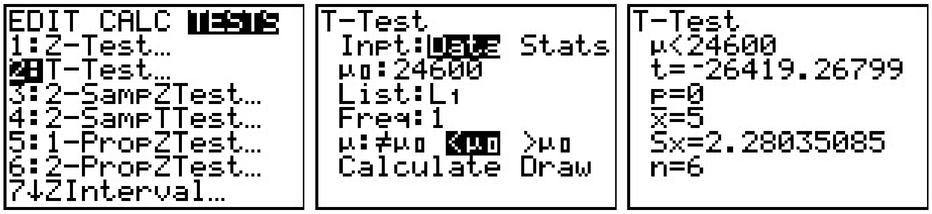
$$H_0 : \mu = 24,600$$
$$H_1 : \mu < 24,600$$

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**, then press **2** (to select `T-Test`). (Figure 10.2(a) on the next page)
4. Highlight `Data`, then press **ENTER**.
5. Enter the values:
   - μ_0 (value of μ under H_0)
   - List: (list containing the sample data)
   - Freq: (enter 1)
   - $\mu :\neq \mu_0 \quad < \mu_0 \quad > \mu_0$ (form of H_1) (Figure 10.2(b) on the following page)
6. Highlight `Calculate`, then press **ENTER**. (Figure 10.2(c) on the next page)

Since $t = -.57$ does not fall in the critical region, we do not reject H_0.

There is not sufficient evidence 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.



(a) Tests: T-Test (b) T-Test screen (c) T-Test output

Figure 10.2: T-Test

10.3 Test for a proportion

Example 10.3
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.

We can assume that the sample size of 100 is less than 5% of the population size (of all adults in the US) and $np(1-p) = 100)(0.13)(0.87) = 11.31 > 10$. So $\hat{p}$ is approximately normally distributed.

The requirements are satisfied, we can proceed with the hypothesis test. The hypotheses are:

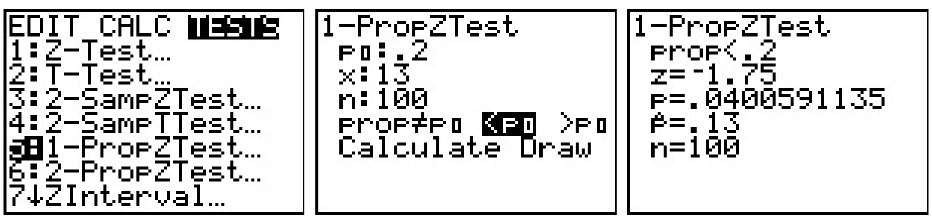
$$H_0 : p = 0.2$$
$$H_1 : p < 0.2$$

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`, then press **5** (to select `1-PropZTest`). (Figure 10.3(a))
3. Enter the values:
   - p_0: (value of p under H_0)
   - x: (number in the sample that have the particular characteristic)
   - n: (sample size)
   - prop: $\neq p_0$ $< p_0$ $> p_0$ (form of H_1) (Figure 10.3(b))
4. Highlight `Calculate`, then press **ENTER**.

Since $z = 1.75$ falls in the critical region, we reject H_0.

There is sufficient evidence to support the claim that less than 20% of adults in the US are allergic to this medication.



(a) Tests: 1-PropZTest (b) 1-PropZTest screen (c) 1-PropZTest output

Figure 10.3: 1-PropZTest

11 Chi-Square Analysis

11.1 Goodness-of-Fit Test

Example 11.1
Mars, Inc. claims that its M&M plain candies are distributed with the following colour 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 colour distribution is as claimed by Mars, Inc.

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

The hypotheses are:

H_0 : The percentages are as claimed by Mars, Inc.
H_1 : 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 `L1`, and the expected frequencies into `L2`.
2. Highlight `L3`, and enter `(L1-L2)`^{2}`/L2`.
3. Press **ENTER**.
4. Press **2nd** [**QUIT**] to return to the Home Screen.
5. Press **2nd** [**LIST**], arrow to `MATH`, then press **5** (to select `sum`).
6. Press **2nd** [**L3**]; then 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 colours is as claimed by Mars, Inc.

11.2 Test for Independence

Example 11.2
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.

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

1. Enter the data from the contingency table into Matrix A as a 2×4 matrix.
2. Press **STAT**, arrow to `TESTS`, then press **3** (to select χ^2−Test).
3. Arrow to `Calculate`, then 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	$\frac{(0-E)^2}{E}$
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).