Elementary Statistics
on the TI-83 and TI-84

Derek Collis
Harper College
# Contents

1 Lists .......................... 1
   1.1 Displaying the stat list editor .............................................. 1
   1.2 Entering data ................................................................. 1
   1.3 Editing data ................................................................. 2
      1.3.1 Correcting a data value .............................................. 2
      1.3.2 Inserting a data value into a list ................................. 2
      1.3.3 Clearing a list ....................................................... 2
   1.4 Sorting data ................................................................. 2
   1.5 Creating and naming a list ................................................. 3
   1.6 Removing a list from the stat list editor ............................... 3
   1.7 Displaying all list names ................................................... 4
   1.8 Displaying selected lists in the stat list editor ...................... 4
   1.9 Restoring the default lists ............................................... 4
   1.10 Copying one list to another list ....................................... 4
   1.11 Combining two or more lists into a single list ....................... 5
   1.12 Applying arithmetic operations to lists ................................ 5
   1.13 Deleting a list from memory ............................................. 6

2 Graphs ...................................... 7
   2.1 Histogram ................................................................. 7

3 Measures of Center and Variation ............................................. 9
   3.1 Ungrouped data ............................................................. 9
      3.1.1 The variance ......................................................... 9
   3.2 Grouped data .............................................................. 10

4 Boxplots ..................................... 11
   4.1 Comparing two or three boxplots ...................................... 12

5 Linear Correlation and Regression .......................................... 14
   5.1 Scatterplot ............................................................... 14
   5.2 Linear correlation coefficient ......................................... 15
   5.3 Regression line .......................................................... 15
      5.3.1 Graph the regression line on the scatterplot .................. 16
         Method 1 ............................................................... 16
         Method 2 ............................................................... 16

6 The Binomial Distribution .................................................. 17
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Probability for a single value</td>
<td>17</td>
</tr>
<tr>
<td>6.2</td>
<td>Probabilities for more than one value</td>
<td>17</td>
</tr>
<tr>
<td>6.3</td>
<td>Cumulative probability</td>
<td>18</td>
</tr>
<tr>
<td>6.4</td>
<td>Constructing a binomial probability distribution</td>
<td>19</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Method 1</td>
<td>19</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Method 2</td>
<td>20</td>
</tr>
<tr>
<td>6.5</td>
<td>Constructing a binomial probability histogram</td>
<td>20</td>
</tr>
<tr>
<td>6.6</td>
<td>Quick method for entering a set of integers into a list</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>The Normal Distribution</td>
<td>22</td>
</tr>
<tr>
<td>7.1</td>
<td>Probability between two z values</td>
<td>22</td>
</tr>
<tr>
<td>7.2</td>
<td>Probability greater than a z value</td>
<td>22</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Method 1</td>
<td>22</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Method 2</td>
<td>23</td>
</tr>
<tr>
<td>7.3</td>
<td>Probability less than a z value</td>
<td>23</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Method 1</td>
<td>23</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Method 2</td>
<td>24</td>
</tr>
<tr>
<td>7.4</td>
<td>Probability between two x values</td>
<td>24</td>
</tr>
<tr>
<td>7.5</td>
<td>Probability less than an x value</td>
<td>24</td>
</tr>
<tr>
<td>7.5.1</td>
<td>Method 1</td>
<td>25</td>
</tr>
<tr>
<td>7.5.2</td>
<td>Method 2</td>
<td>25</td>
</tr>
<tr>
<td>7.6</td>
<td>Finding a z value</td>
<td>25</td>
</tr>
<tr>
<td>7.7</td>
<td>Finding an x value</td>
<td>27</td>
</tr>
<tr>
<td>8</td>
<td>Assessing Normality</td>
<td>28</td>
</tr>
<tr>
<td>8.1</td>
<td>Normal probability plots</td>
<td>28</td>
</tr>
<tr>
<td>9</td>
<td>Confidence Intervals</td>
<td>30</td>
</tr>
<tr>
<td>9.1</td>
<td>Confidence interval for a population mean: $\sigma$ known</td>
<td>30</td>
</tr>
<tr>
<td>9.2</td>
<td>Confidence interval for a population mean, $\sigma$ unknown</td>
<td>31</td>
</tr>
<tr>
<td>9.3</td>
<td>Confidence interval for a population proportion</td>
<td>32</td>
</tr>
<tr>
<td>10</td>
<td>Hypothesis Tests</td>
<td>34</td>
</tr>
<tr>
<td>10.1</td>
<td>Test for a mean: $\sigma$ known</td>
<td>34</td>
</tr>
<tr>
<td>10.2</td>
<td>Test for a mean: $\sigma$ unknown</td>
<td>35</td>
</tr>
<tr>
<td>10.3</td>
<td>Test for a proportion</td>
<td>36</td>
</tr>
<tr>
<td>11</td>
<td>Chi-Square Analysis</td>
<td>38</td>
</tr>
<tr>
<td>11.1</td>
<td>Goodness-of-Fit Test</td>
<td>38</td>
</tr>
<tr>
<td>11.2</td>
<td>Test for Independence</td>
<td>38</td>
</tr>
</tbody>
</table>
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))

![Edit menu and 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.

![Entering data](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](image)

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)

![Sorting data](image)

**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))

![Creating and naming a list](image)

**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)
1.11 Combining two or more lists into a single list

To combine data in \( L_1 \) and \( L_2 \) and store into \( L_3 \).

1. Enter the data in \( L_1 \) and \( L_2 \).
2. Press 2nd [LIST], arrow to OPS, then press 9 (to select augment).
3. Enter \( L_1, L_2 \), press STO\( \downarrow \), and enter \( L_3 \); then press ENTER. (Figure 1.8)

If you had entered \( L_2, L_1 \) then the entries from \( L_2 \) would be listed first in \( L_3 \).

1.12 Applying arithmetic operations to lists

To multiply the corresponding entries in \( L_1 \) and \( L_2 \) and then store these products in \( L_3 \).

1. Enter the data in \( L_1 \) and \( L_2 \).

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

drcollis/harpercollege
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](image)

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**

drcollis/harpercollege
2 Graphs

2.1 Histogram

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

Table 2.1: Frequency distribution

<table>
<thead>
<tr>
<th>Class</th>
<th>f</th>
<th>Class Mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>5–9</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>10–14</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>15–19</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>20–24</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>25–29</td>
<td>3</td>
<td>27</td>
</tr>
</tbody>
</table>

1. Enter the class midpoints and frequencies into L1 and L2. (Figure 2.1)

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

<table>
<thead>
<tr>
<th>Setting</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xmin</td>
<td>lower limit of first class</td>
</tr>
<tr>
<td>Xmax</td>
<td>lower limit of last class plus class width (This would be the lower limit of the next class if there were one.)</td>
</tr>
<tr>
<td>Xscl</td>
<td>class width</td>
</tr>
<tr>
<td>Ymin</td>
<td>-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.)</td>
</tr>
<tr>
<td>Ymax</td>
<td>maximum frequency (or a little more) of distribution</td>
</tr>
<tr>
<td>Yscl</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 2.4: Histogram

(a) Histogram

(b) Histogram values

drcollis/harpercollege
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))

\[
\begin{array}{|c|c|c|}
\hline
\text{EDIT CALC TESTS} & \text{1-Var Stats} & \text{1-Var Stats} \\
\hline
1:1-Var Stats & \bar{x}=13 & \bar{x}=13 \\
2:2-Var Stats & \sum x=78 & \sum x=78 \\
3:Med-Med & \sum x^2=1334 & \sum x^2=1334 \\
4:LinReg(ax+b) & \bar{x}=8 & \bar{x}=8 \\
5:QuadReg & \sigma x=7.302967433 & \sigma x=7.302967433 \\
6:CubicReg & \bar{n}=6 & \bar{n}=6 \\
7:QuartReg & \max x=24 & \max x=24 \\
\hline
\end{array}
\]

(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 \(S_x\) (or press 4 for \(\sigma x\)). (Figure 3.2(b) on the next page)
3. Press \(x^2\); 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.
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.

<table>
<thead>
<tr>
<th>Driving time</th>
<th>f</th>
<th>Class mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>5–9</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>10–14</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>15–19</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>20–24</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>25–29</td>
<td>3</td>
<td>27</td>
</tr>
</tbody>
</table>

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))

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, Q1, med, Q3, 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

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

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 2 3 3 4 5 5 5 6 7 8 9 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.
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.
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° Fahrenheit. Which type of battery would you prefer?

<table>
<thead>
<tr>
<th>Group Size</th>
<th>Group Size</th>
<th>Group Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 Batteries</td>
<td>25 Batteries</td>
<td></td>
</tr>
<tr>
<td>800 600 675</td>
<td>525 620 550</td>
<td></td>
</tr>
<tr>
<td>600 525 700</td>
<td>560 675 550</td>
<td></td>
</tr>
<tr>
<td>500 660 550</td>
<td>530 570 640</td>
<td></td>
</tr>
<tr>
<td>585 675</td>
<td>525 640 640</td>
<td></td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Type of battery</th>
<th>Minimum</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Size 24</td>
<td>500</td>
<td>550</td>
<td>600</td>
<td>675</td>
<td>800</td>
</tr>
<tr>
<td>Group Size 35</td>
<td>525</td>
<td>540</td>
<td>565</td>
<td>640</td>
<td>675</td>
</tr>
</tbody>
</table>
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:

<table>
<thead>
<tr>
<th>No. of absences, $x$</th>
<th>Final grade, $y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td>12</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>96</td>
</tr>
<tr>
<td>0</td>
<td>94</td>
</tr>
<tr>
<td>8</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>82</td>
</tr>
</tbody>
</table>

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))

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)

![LinReg screen](image)

Figure 5.2: LinReg screen

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 Y₁ if necessary.)
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 (for binompdf).
3. Enter 5,.2,3); then press ENTER. (See Figure 6.1).

\[
\text{binompdf}(5,0.2,3) = 0.0512
\]

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).

\[
\text{binompdf}(5,0.2,{1,2}) = 0.4096, 0.2048
\]

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].
3. Enter 5, .2, 3; then press ENTER (See Figure 6.3).

\[
\text{binomcdf}(5, .2, 3) = 0.99328
\]

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].
3. Enter 5, .2, 2; then press ENTER (See Figure 6.4).

\[
1 - \text{binomcdf}(5, .2, 2) = 0.05792
\]

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].

\[\text{binomcdf}(10, .2, 2)\]

\[\text{binomcdf}(10, .2, 7)\]

\[0.05792\]

\[0.0041\]

\[0.0538\]

$^2$For a TI-84, press ALPHA [B], to select binomcdf

drcollis/harpercollege
3. Enter $10, \cdot 2, 7$.
4. Press 2nd [DISTR].
5. Press ALPHA [A] (for binomcdf).
6. Enter $10, \cdot 2, 2$; 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 (for 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).

```
binomPdf(5,.2)→L
.32768 .4096 ...
```

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.

2. Press 2nd [LIST].
3. Arrow to OPS and press 5 (for seq).

\(^3\)The \( Y_{\text{max}} \) is chosen as to be larger than the largest probability and the \( Y_{\text{min}} \) is chosen to be \(-Y_{\text{max}}/4\) to give the appropriate space below the histogram for reading the TRACE values.

drcollis/harpercollege
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.
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)

![Screen output and area under curve](image)

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 \( \infty \) by a large number, such as 1,000,000,000 or, in scientific notation, \( 1 \times 10^9 \). This is entered into the calculator as 1 2nd [EE] 9 and is displayed as 1E9.
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 7.2: P(z > 1.5): Method 1](a) Screen output  (b) Area under curve

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)
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)

7.4 Probability between two $x$ values

Example 7.4
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 $\sigma = 29$. 

drcollis/harpercollege
The Normal Distribution

7.5.1 Method 1

1. Press 2nd [DISTR].
2. Press 2 (to select normalcdf).
3. Enter \(-1e9, 135, 143, 29\); then press ENTER. (Figure 7.7)

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

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)
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)

![Screen output](image1.png)  ![Area under curve](image2.png)

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)

![Screen output](image3.png)  ![Area under curve](image4.png)

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.

\[
\begin{array}{ccccccc}
42,544 & 27,274 & 34,258 & 59,177 & 44,091 \\
35,631 & 42,371 & 48,018 & 58,795 & 44,832 \\
\end{array}
\]

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](image)

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.
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: $\sigma$ 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 $\text{STAT}$.
2. Arrow to TESTS; then press 7 (to select $\text{ZInterval}$). (Figure 9.1(a))
3. Highlight Stats and press ENTER.
4. Enter the values for $\sigma$, $\bar{x}$, n and C-Level$^1$. (Figure 9.1(b))
5. Highlight Calculate; then press ENTER. (Figure 9.1(c))

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 $\mu$ and 5% would not contain $\mu$. We don’t know if this particular interval contains $\mu$ or not: our confidence is in the procedure, not this particular interval. It is incorrect to say that “there is a 95% chance that $\mu$ will fall between $42,011 and $49,345”.

The population mean, $\mu$, 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 $\mu$ 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, \( \sigma \) 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 \( \text{STAT} \).
2. Arrow to \( \text{TESTS} \), then press \( 8 \) (to select \( \text{TInterval} \)). (Figure 9.2(a))
3. Highlight \( \text{Stats} \) and press \( \text{ENTER} \).
4. Enter the values for \( \bar{x}, s, n \) and \( \text{C-Level} \). (Figure 9.2(b))
5. Highlight \( \text{Calculate} \); then press \( \text{ENTER} \). (Figure 9.2(c))

![Figure 9.2: Confidence interval for a population mean, \( \sigma \) unknown, using the summary statistics](image)

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 \( \sigma \) 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, \( \mu \).

1. Enter the data into \( L1 \).
2. Press \( \text{STAT} \).

\[ \text{Difference between Z interval and T interval} \]

\text{drcollis/harpercollege}
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))

Figure 9.3: Confidence interval for a population mean, $\sigma$ unknown, using the raw data

We are 95% confident that the population mean, $\mu$, 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))

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: $\sigma$ 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:
   - $\mu_0$: (value of $\mu$ under $H_0$)
   - $\sigma$: (population standard deviation)
   - $\bar{x}$: (sample mean)
   - $n$: (sample size)
   - $\mu \neq \mu_0$, $\mu > \mu_0$ (form of $H_1$) (Figure 10.1(b))
5. Highlight Calculate, then press ENTER. (Figure 10.1(c))

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: $\sigma$ 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.

$22,954 \quad 23,438 \quad 24,655 \quad 23,695 \quad 25,275$

$19,212 \quad 21,456 \quad 25,493 \quad 26,480 \quad 28,585$

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:
   - $\mu_0$ (value of $\mu$ 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 $t = -0.57$ 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 $24,600.$

drcollis/harpercollege
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)
   - \( \text{prop} \neq p_0, p_0 < p_0 > p_0 \) (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.
### Figure 10.3: 1-PropZTest

<table>
<thead>
<tr>
<th>(a) Tests: 1-PropZTest</th>
<th>(b) 1-PropZTest screen</th>
<th>(c) 1-PropZTest output</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EDIT CALC TESTS</strong></td>
<td><strong>1-PropZTest</strong></td>
<td><strong>1-PropZTest</strong></td>
</tr>
<tr>
<td>1: 2-Test...</td>
<td>( \hat{p} = 0.2 )</td>
<td>( \hat{p} = 0.2 )</td>
</tr>
<tr>
<td>2: T-test...</td>
<td>( x = 13 )</td>
<td>( z = -1.76 )</td>
</tr>
<tr>
<td>3: 2-SampZTest...</td>
<td>( n = 100 )</td>
<td>( p = 0.0400591135 )</td>
</tr>
<tr>
<td>4: 2-SampTTest...</td>
<td>( \text{prop}\neq \hat{p} ) ( \text{prop} &gt; \hat{p} )</td>
<td>( \hat{p} = 0.15 )</td>
</tr>
<tr>
<td>5: 2-PropZTest...</td>
<td>Calculate Draw</td>
<td>( n = 100 )</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Brown</th>
<th>Yellow</th>
<th>Red</th>
<th>Orange</th>
<th>Green</th>
<th>Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>26</td>
<td>21</td>
<td>8</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

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 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.
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 \( 2 \times 4 \) matrix.
2. Press \text{STAT}, arrow to \text{TESTS}, then press 3 (to select \( \chi^2 \)-Test).
3. Arrow to \text{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.

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

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).