Chapter 3
Relationships Between Quantitative Variables

Introduction

In this chapter, we will learn how to describe the relationship between two quantitative variables. Remember (from Chapter 2) that the terms quantitative variable and measurement variable are synonyms for data that can be recorded as numerical values and then ordered according to those values. The relationship between weight and height is an example of a relationship between two quantitative variables.

The questions we ask about the relationship between two variables often concern specific numerical features of the association. For example, we may want to know how much weight will increase on average for each 1-inch increase in height. Or, we may want to estimate what the college grade point average will be for a student whose high school grade point average was 3.5. In this chapter, you will learn how to create simple summaries and pictures from various kinds of raw data.

After reading this chapter you should be able to:
1. Display a scatterplot of two quantitative variables.
2. Display subgroups of two quantitative variables on a scatterplot.
3. Display a scatterplot with the regression equation superimposed upon the scatterplot.
4. Make predictions using a regression equation.
5. Obtain the residuals.
6. Find the correlation coefficient and the coefficient of determination for two quantitative variables.
7. Obtain the regression output, identifying the slope, intercept, \( r^2 \), SSTO, and SSE for two quantitative variables.

Keystrokes Introduced
1. \text{2nd} [STAT PLOT] > scatterplot displays a scatterplot of two quantitative variables.
2. [STAT] CALC> 8: LinReg (a + bx) calculates a regression equation for two quantitative variables.
3. \text{2nd} [CATALOG] >DiagnosticOn displays \( r \), the correlation coefficient, and \( r^2 \), the coefficient of determination when a linear regression equation is ob-
3.1 Looking for Patterns With Scatterplots

A scatterplot is a two-dimensional graph of the measurements for two numerical variables. A point on the graph represents the combination of measurements for an individual observation. The vertical axis, which is called the y axis, is used to locate the value of one of the variables. The horizontal axis, called the x axis, is used to locate the value of the other variable.

Questions to Ask About a Scatterplot
What is the average pattern? Does it look like a straight line or is it curved?
What is the direction of the pattern?
How much do individual points vary from the average pattern?
Are there any unusual data points?

Example 3.1 Height and Handspan

Tables 3.1a and 3.1b display the observations of a dataset that includes the heights (in inches) and fully stretched hand spans (in centimeters) of 167 college students. The data values for all 167 students are the raw data for studying the connection between height and handspan. Imagine how difficult it is to see the pattern in the data from all 167 observations were shown in Table 3.1. Even when we just look at the data for the first 12 students, it takes a while to confirm that there does seem to be a tendency for taller people to have larger handspans.

Follow these steps to display a scatterplot of handspan and height measurements for all 167 students.

1. Preparations:
   a. Turn off all "Y=" functions.
      
      Press \[ Y= \] and press \[ \text{CLEAR} \] to remove all functions. For each line that is not blank, place the cursor on the function and press \[ \text{CLEAR} \]. Press \[ \text{2nd} \] \[ \text{QUIT} \].
   b. Clear all lists in the Stat editor:
      
      Press \[ \text{STAT} \], selecting 4: \ClrList. Enter each list name: L1, L2, L3, L4,
Chapter 3  Relationships Between Quantitative Variables

L5, L6, as shown in Figure 3.1. Press [ENTER] to execute the command.

<table>
<thead>
<tr>
<th>Ss</th>
<th>Height</th>
<th>Hand</th>
<th>Ss</th>
<th>Height</th>
<th>Hand</th>
<th>Ss</th>
<th>Height</th>
<th>Hand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>68</td>
<td>21.5</td>
<td>31</td>
<td>67</td>
<td>20.0</td>
<td>61</td>
<td>64</td>
<td>20.0</td>
</tr>
<tr>
<td>2</td>
<td>71</td>
<td>23.5</td>
<td>32</td>
<td>67</td>
<td>20.0</td>
<td>62</td>
<td>65</td>
<td>20.0</td>
</tr>
<tr>
<td>3</td>
<td>73</td>
<td>22.5</td>
<td>33</td>
<td>66</td>
<td>19.0</td>
<td>63</td>
<td>74</td>
<td>24.0</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>18.0</td>
<td>34</td>
<td>62</td>
<td>17.0</td>
<td>64</td>
<td>68</td>
<td>21.0</td>
</tr>
<tr>
<td>5</td>
<td>68</td>
<td>23.5</td>
<td>35</td>
<td>72</td>
<td>22.0</td>
<td>65</td>
<td>68</td>
<td>21.5</td>
</tr>
<tr>
<td>6</td>
<td>59</td>
<td>20.0</td>
<td>36</td>
<td>71</td>
<td>22.0</td>
<td>66</td>
<td>69</td>
<td>18.5</td>
</tr>
<tr>
<td>7</td>
<td>73</td>
<td>23.0</td>
<td>37</td>
<td>61</td>
<td>17.5</td>
<td>67</td>
<td>68</td>
<td>23.0</td>
</tr>
<tr>
<td>8</td>
<td>75</td>
<td>24.5</td>
<td>38</td>
<td>63</td>
<td>19.0</td>
<td>68</td>
<td>67</td>
<td>23.0</td>
</tr>
<tr>
<td>9</td>
<td>65</td>
<td>21.0</td>
<td>39</td>
<td>66</td>
<td>19.0</td>
<td>69</td>
<td>61.5</td>
<td>20.5</td>
</tr>
<tr>
<td>10</td>
<td>69</td>
<td>20.5</td>
<td>40</td>
<td>71</td>
<td>22.0</td>
<td>70</td>
<td>63</td>
<td>16.5</td>
</tr>
<tr>
<td>11</td>
<td>69</td>
<td>20.5</td>
<td>41</td>
<td>71</td>
<td>22.0</td>
<td>71</td>
<td>67</td>
<td>19.5</td>
</tr>
<tr>
<td>12</td>
<td>64</td>
<td>18.5</td>
<td>42</td>
<td>66</td>
<td>18.5</td>
<td>72</td>
<td>71</td>
<td>23.0</td>
</tr>
<tr>
<td>13</td>
<td>67</td>
<td>21.0</td>
<td>43</td>
<td>70</td>
<td>20.0</td>
<td>73</td>
<td>73</td>
<td>22.5</td>
</tr>
<tr>
<td>14</td>
<td>67</td>
<td>19.5</td>
<td>44</td>
<td>67</td>
<td>20.5</td>
<td>74</td>
<td>63</td>
<td>18.5</td>
</tr>
<tr>
<td>15</td>
<td>69</td>
<td>22.0</td>
<td>45</td>
<td>69</td>
<td>21.0</td>
<td>75</td>
<td>61</td>
<td>18.5</td>
</tr>
<tr>
<td>16</td>
<td>73</td>
<td>22.0</td>
<td>46</td>
<td>67</td>
<td>19.5</td>
<td>76</td>
<td>67</td>
<td>21.5</td>
</tr>
<tr>
<td>17</td>
<td>62</td>
<td>20.0</td>
<td>47</td>
<td>68</td>
<td>20.0</td>
<td>77</td>
<td>72</td>
<td>20.5</td>
</tr>
<tr>
<td>18</td>
<td>69</td>
<td>22.5</td>
<td>48</td>
<td>67</td>
<td>21.5</td>
<td>78</td>
<td>72</td>
<td>20.5</td>
</tr>
<tr>
<td>19</td>
<td>64</td>
<td>18.5</td>
<td>49</td>
<td>68</td>
<td>22.5</td>
<td>79</td>
<td>68</td>
<td>20.0</td>
</tr>
<tr>
<td>20</td>
<td>74</td>
<td>21.5</td>
<td>50</td>
<td>71</td>
<td>20.0</td>
<td>80</td>
<td>66</td>
<td>21</td>
</tr>
<tr>
<td>21</td>
<td>73</td>
<td>24.5</td>
<td>51</td>
<td>70</td>
<td>22.5</td>
<td>81</td>
<td>67</td>
<td>21.5</td>
</tr>
<tr>
<td>22</td>
<td>66</td>
<td>20.5</td>
<td>52</td>
<td>74</td>
<td>24.5</td>
<td>82</td>
<td>67</td>
<td>20.5</td>
</tr>
<tr>
<td>23</td>
<td>74</td>
<td>24.5</td>
<td>53</td>
<td>60</td>
<td>18.5</td>
<td>83</td>
<td>72</td>
<td>20.5</td>
</tr>
<tr>
<td>24</td>
<td>73</td>
<td>21.0</td>
<td>54</td>
<td>65</td>
<td>20.0</td>
<td>84</td>
<td>67.5</td>
<td>21.0</td>
</tr>
<tr>
<td>25</td>
<td>69</td>
<td>21.0</td>
<td>55</td>
<td>72</td>
<td>24.0</td>
<td>85</td>
<td>63.75</td>
<td>21.5</td>
</tr>
<tr>
<td>26</td>
<td>64</td>
<td>18.5</td>
<td>56</td>
<td>76</td>
<td>23.5</td>
<td>86</td>
<td>72</td>
<td>21.5</td>
</tr>
<tr>
<td>27</td>
<td>67</td>
<td>18.0</td>
<td>57</td>
<td>66</td>
<td>21.0</td>
<td>87</td>
<td>69</td>
<td>22.5</td>
</tr>
<tr>
<td>28</td>
<td>60</td>
<td>19.5</td>
<td>58</td>
<td>64.5</td>
<td>19.5</td>
<td>88</td>
<td>68</td>
<td>21.0</td>
</tr>
<tr>
<td>29</td>
<td>75</td>
<td>20.5</td>
<td>59</td>
<td>71</td>
<td>20.0</td>
<td>89</td>
<td>71</td>
<td>21.0</td>
</tr>
<tr>
<td>30</td>
<td>64</td>
<td>21.0</td>
<td>60</td>
<td>69</td>
<td>22.5</td>
<td>90</td>
<td>71</td>
<td>22.0</td>
</tr>
</tbody>
</table>

Table 3.1a

2. Enter data using the [STAT] list editor.

Press [STAT] [ENTER] to select the [STAT] list editor.

a. Enter the data for the quantitative variables "height" and "handspan" in lists L1 and L2.

Place the cursor on list L1 row 1 to make L1(1) the active list row. Enter the height data: 68, 71, 73, ... pressing [ENTER] after each entry.
Place the cursor on list L2 row 1 to make L2(1) the active list row. Enter the
3.1 Looking for Patterns With Scatterplots

hand data: 21.5, 23.5, 22.5, ... in L2 pressing ENTER after each entry,
as shown in Figure 5.2.

![Figure 3.1]

![Figure 3.2]

<table>
<thead>
<tr>
<th>Ss</th>
<th>Height</th>
<th>Hand</th>
<th>Ss</th>
<th>Height</th>
<th>Hand</th>
</tr>
</thead>
<tbody>
<tr>
<td>91</td>
<td>63</td>
<td>19.0</td>
<td>92</td>
<td>70</td>
<td>23.0</td>
</tr>
<tr>
<td>93</td>
<td>68</td>
<td>20.5</td>
<td>94</td>
<td>67.5</td>
<td>20.5</td>
</tr>
<tr>
<td>95</td>
<td>75</td>
<td>21.0</td>
<td>96</td>
<td>75</td>
<td>24.0</td>
</tr>
<tr>
<td>97</td>
<td>71</td>
<td>22.0</td>
<td>98</td>
<td>71</td>
<td>21.0</td>
</tr>
<tr>
<td>99</td>
<td>64</td>
<td>19.5</td>
<td>100</td>
<td>71</td>
<td>21.0</td>
</tr>
<tr>
<td>101</td>
<td>69</td>
<td>19.5</td>
<td>102</td>
<td>65</td>
<td>19.0</td>
</tr>
<tr>
<td>103</td>
<td>69</td>
<td>23.0</td>
<td>104</td>
<td>63</td>
<td>20.5</td>
</tr>
<tr>
<td>105</td>
<td>70</td>
<td>24.0</td>
<td>106</td>
<td>71</td>
<td>22.0</td>
</tr>
<tr>
<td>107</td>
<td>64</td>
<td>20.0</td>
<td>108</td>
<td>63</td>
<td>21.5</td>
</tr>
<tr>
<td>109</td>
<td>65</td>
<td>19.0</td>
<td>110</td>
<td>66</td>
<td>19.0</td>
</tr>
<tr>
<td>111</td>
<td>66</td>
<td>20.0</td>
<td>112</td>
<td>65</td>
<td>19.5</td>
</tr>
<tr>
<td>113</td>
<td>67.5</td>
<td>20.0</td>
<td>114</td>
<td>57</td>
<td>16.0</td>
</tr>
<tr>
<td>115</td>
<td>72</td>
<td>22.5</td>
<td>116</td>
<td>64</td>
<td>17.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L1</th>
<th>L2</th>
<th>L3</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.5</td>
<td>23.5</td>
<td>22.5</td>
</tr>
</tbody>
</table>

Table 3.1b

3. Plot the statistical data by creating a scatterplot of hands and height measurements for all 167 students.

Press 2nd STAT PLOT accessing the stat plot menu.
Press ENTER selecting Plot 1. Place the cursor on ON and press ENTER.
Use the down arrow key and the right arrow key to select the first icon in the first row, the scatterplot. Press ENTER. Use the down arrow key to select list
Chapter 3  Relationships Between Quantitative Variables

L1 as the list, 2nd [L1]. Use the down arrow key to enter list L2 as the Ylist:
2nd [L2]. Use the down arrow key to select the second icon for the mark. The
settings for Plot 1 are shown in Figure 3.3.

4. View the graph.

Press ZOOM/9: ZoomStat to view the graph, as shown in Figure 3.4.

![Figure 3.3](image1.png)  ![Figure 3.4](image2.png)

5. Save list L1 as HIGHT and list L2 as HAND.

Press 2nd [L1] STO→ 2nd [A-LOCK] and type HIGHT; press ENTER.
Press 2nd [L2] STO→ 2nd [A-LOCK] and type HAND; press ENTER.

Figure 3.4 is a scatterplot that displays the handspan and height measurements
for all 167 students. The handspan measurements are plotted along the vertical
axis (y), and the height measurements are plotted along the horizontal axis (x).
Each point represents the two measurements for an individual.
We see that taller people tend to have greater handspan measurements than
shorter people do. When two variables tend to increase together, as they do
in Figure 3.4, we say that they have a positive association. Another noteworthy
characteristic of the graph is that we can describe the general pattern of
this relationship with a straight line. In other words, the handspan and height
measurements may have a linear relationship.

**Indicating Groups Within the Data on Scatterplots**

When we examined the connection between height and handspan in Example 3.1,
you may have wondered whether we should be concerned about student gender.
Both height and handspan tend to be greater for men than for women, so we should
consider the possibility that gender differences might be completely responsible for
the observed relationship.
It's easy to indicate subgroups on a scatterplot. We just use different symbols or
different colors to represent the different groups.
Example 3.1 Height and Handspan Continued

The data for females is displayed in Table 3.2. The data for males is displayed in Table 3.3.

<table>
<thead>
<tr>
<th>Ss</th>
<th>Height</th>
<th>Hand</th>
<th>Ss</th>
<th>Height</th>
<th>Hand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>68</td>
<td>21.5</td>
<td>31</td>
<td>64</td>
<td>20.0</td>
</tr>
<tr>
<td>2</td>
<td>64</td>
<td>18.0</td>
<td>32</td>
<td>65</td>
<td>20.0</td>
</tr>
<tr>
<td>3</td>
<td>59</td>
<td>20.0</td>
<td>33</td>
<td>68</td>
<td>21.5</td>
</tr>
<tr>
<td>4</td>
<td>65</td>
<td>21.0</td>
<td>34</td>
<td>69</td>
<td>18.5</td>
</tr>
<tr>
<td>5</td>
<td>69</td>
<td>20.5</td>
<td>35</td>
<td>68</td>
<td>23.0</td>
</tr>
<tr>
<td>6</td>
<td>64</td>
<td>18.5</td>
<td>36</td>
<td>61.5</td>
<td>20.5</td>
</tr>
<tr>
<td>7</td>
<td>67</td>
<td>21.0</td>
<td>37</td>
<td>63</td>
<td>16.5</td>
</tr>
<tr>
<td>8</td>
<td>67</td>
<td>19.5</td>
<td>38</td>
<td>67</td>
<td>19.5</td>
</tr>
<tr>
<td>9</td>
<td>62</td>
<td>20.0</td>
<td>39</td>
<td>63</td>
<td>18.5</td>
</tr>
<tr>
<td>10</td>
<td>64</td>
<td>18.5</td>
<td>40</td>
<td>61</td>
<td>18.5</td>
</tr>
<tr>
<td>11</td>
<td>66</td>
<td>20.5</td>
<td>41</td>
<td>68</td>
<td>20.0</td>
</tr>
<tr>
<td>12</td>
<td>64</td>
<td>18.5</td>
<td>42</td>
<td>66</td>
<td>21.0</td>
</tr>
<tr>
<td>13</td>
<td>67</td>
<td>18.0</td>
<td>43</td>
<td>67</td>
<td>20.5</td>
</tr>
<tr>
<td>14</td>
<td>60</td>
<td>19.5</td>
<td>44</td>
<td>63.75</td>
<td>21.5</td>
</tr>
<tr>
<td>15</td>
<td>64</td>
<td>21.0</td>
<td>45</td>
<td>72</td>
<td>21.5</td>
</tr>
<tr>
<td>16</td>
<td>67</td>
<td>20.0</td>
<td>46</td>
<td>68</td>
<td>21.0</td>
</tr>
<tr>
<td>17</td>
<td>67</td>
<td>20.0</td>
<td>47</td>
<td>63</td>
<td>19.0</td>
</tr>
<tr>
<td>18</td>
<td>66</td>
<td>19.0</td>
<td>48</td>
<td>68</td>
<td>20.5</td>
</tr>
<tr>
<td>19</td>
<td>62</td>
<td>17.0</td>
<td>49</td>
<td>67.5</td>
<td>20.5</td>
</tr>
<tr>
<td>20</td>
<td>61</td>
<td>17.5</td>
<td>50</td>
<td>64</td>
<td>19.5</td>
</tr>
<tr>
<td>21</td>
<td>63</td>
<td>19.0</td>
<td>51</td>
<td>69</td>
<td>19.5</td>
</tr>
<tr>
<td>22</td>
<td>66</td>
<td>18.5</td>
<td>52</td>
<td>65</td>
<td>19.0</td>
</tr>
<tr>
<td>23</td>
<td>67</td>
<td>20.5</td>
<td>53</td>
<td>63</td>
<td>20.5</td>
</tr>
<tr>
<td>24</td>
<td>67</td>
<td>19.5</td>
<td>54</td>
<td>64</td>
<td>20.0</td>
</tr>
<tr>
<td>25</td>
<td>68</td>
<td>20.0</td>
<td>55</td>
<td>63</td>
<td>21.5</td>
</tr>
<tr>
<td>26</td>
<td>71</td>
<td>20.0</td>
<td>56</td>
<td>65</td>
<td>19.0</td>
</tr>
<tr>
<td>27</td>
<td>60</td>
<td>18.5</td>
<td>57</td>
<td>66</td>
<td>19.0</td>
</tr>
<tr>
<td>28</td>
<td>65</td>
<td>20.0</td>
<td>58</td>
<td>66</td>
<td>20.0</td>
</tr>
<tr>
<td>29</td>
<td>66</td>
<td>21.0</td>
<td>59</td>
<td>63</td>
<td>19.5</td>
</tr>
<tr>
<td>30</td>
<td>64.5</td>
<td>19.5</td>
<td>60</td>
<td>67</td>
<td>20.0</td>
</tr>
</tbody>
</table>

Table 3.2

Follow these steps to display a scatterplot of handsap and height measurements for the 89 female students and the 78 male students.

1. Preparations:
   a. Turn off all "Y=" functions.
      Press \[ \text{Y-} \] and press \[ \text{CLEAR} \] to remove all functions. For each line that is not blank, place the cursor on the function and press \[ \text{CLEAR} \].

2nd QUIT
Chapter 3  Relationships Between Quantitative Variables

b. Clear all lists in the Stat editor.

Press STAT selecting: ClrList. Enter each list name: L1, L2, L3, L4, L5, L6, as shown in Figure 3.5. Press ENTER to execute the command.

Height and HandSpans for 78 Males

<table>
<thead>
<tr>
<th>Ss</th>
<th>Height</th>
<th>Hand</th>
<th>Ss</th>
<th>Height</th>
<th>Hand</th>
<th>Ss</th>
<th>Height</th>
<th>Hand</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71</td>
<td>23.5</td>
<td>27</td>
<td>72</td>
<td>24.0</td>
<td>53</td>
<td>71</td>
<td>22.0</td>
</tr>
<tr>
<td>2</td>
<td>73</td>
<td>22.5</td>
<td>28</td>
<td>76</td>
<td>23.5</td>
<td>54</td>
<td>72</td>
<td>22.5</td>
</tr>
<tr>
<td>3</td>
<td>68</td>
<td>23.5</td>
<td>29</td>
<td>71</td>
<td>20.0</td>
<td>55</td>
<td>67</td>
<td>21.5</td>
</tr>
<tr>
<td>4</td>
<td>73</td>
<td>23.0</td>
<td>30</td>
<td>69</td>
<td>22.5</td>
<td>56</td>
<td>72</td>
<td>23.5</td>
</tr>
<tr>
<td>5</td>
<td>75</td>
<td>24.5</td>
<td>31</td>
<td>74</td>
<td>24.0</td>
<td>57</td>
<td>74</td>
<td>22.0</td>
</tr>
<tr>
<td>6</td>
<td>69</td>
<td>20.5</td>
<td>32</td>
<td>68</td>
<td>21.0</td>
<td>58</td>
<td>69</td>
<td>18.0</td>
</tr>
<tr>
<td>7</td>
<td>69</td>
<td>22.0</td>
<td>33</td>
<td>67</td>
<td>23.0</td>
<td>59</td>
<td>74</td>
<td>23.5</td>
</tr>
<tr>
<td>8</td>
<td>73</td>
<td>22.0</td>
<td>34</td>
<td>71</td>
<td>23.0</td>
<td>60</td>
<td>73</td>
<td>24.0</td>
</tr>
<tr>
<td>9</td>
<td>69</td>
<td>22.5</td>
<td>35</td>
<td>73</td>
<td>22.5</td>
<td>61</td>
<td>76</td>
<td>24.5</td>
</tr>
<tr>
<td>10</td>
<td>74</td>
<td>21.5</td>
<td>36</td>
<td>67</td>
<td>21.5</td>
<td>62</td>
<td>76</td>
<td>23.0</td>
</tr>
<tr>
<td>11</td>
<td>73</td>
<td>24.5</td>
<td>37</td>
<td>72</td>
<td>20.5</td>
<td>63</td>
<td>69</td>
<td>22.5</td>
</tr>
<tr>
<td>12</td>
<td>74</td>
<td>24.5</td>
<td>38</td>
<td>72</td>
<td>20.5</td>
<td>64</td>
<td>75</td>
<td>24.5</td>
</tr>
<tr>
<td>13</td>
<td>73</td>
<td>21.0</td>
<td>39</td>
<td>67</td>
<td>21.5</td>
<td>65</td>
<td>69</td>
<td>22.0</td>
</tr>
<tr>
<td>14</td>
<td>69</td>
<td>21.0</td>
<td>40</td>
<td>72</td>
<td>20.5</td>
<td>66</td>
<td>67</td>
<td>22.0</td>
</tr>
<tr>
<td>15</td>
<td>75</td>
<td>20.5</td>
<td>41</td>
<td>67.5</td>
<td>21.0</td>
<td>67</td>
<td>74</td>
<td>24.5</td>
</tr>
<tr>
<td>16</td>
<td>72</td>
<td>22.0</td>
<td>42</td>
<td>69</td>
<td>22.5</td>
<td>68</td>
<td>74</td>
<td>24.0</td>
</tr>
<tr>
<td>17</td>
<td>71</td>
<td>22.0</td>
<td>43</td>
<td>71</td>
<td>21.0</td>
<td>69</td>
<td>71</td>
<td>21.5</td>
</tr>
<tr>
<td>18</td>
<td>66</td>
<td>19.0</td>
<td>44</td>
<td>71</td>
<td>22.0</td>
<td>70</td>
<td>67</td>
<td>22.0</td>
</tr>
<tr>
<td>19</td>
<td>71</td>
<td>22.0</td>
<td>45</td>
<td>70</td>
<td>23.0</td>
<td>71</td>
<td>73</td>
<td>23.0</td>
</tr>
<tr>
<td>20</td>
<td>71</td>
<td>22.0</td>
<td>46</td>
<td>75</td>
<td>21.0</td>
<td>72</td>
<td>78</td>
<td>25.5</td>
</tr>
<tr>
<td>21</td>
<td>70</td>
<td>20.0</td>
<td>47</td>
<td>74</td>
<td>24.0</td>
<td>73</td>
<td>72</td>
<td>20.5</td>
</tr>
<tr>
<td>22</td>
<td>69</td>
<td>21.0</td>
<td>48</td>
<td>71</td>
<td>22.0</td>
<td>74</td>
<td>74</td>
<td>24.0</td>
</tr>
<tr>
<td>23</td>
<td>67</td>
<td>21.5</td>
<td>49</td>
<td>71</td>
<td>21.0</td>
<td>75</td>
<td>70</td>
<td>22.0</td>
</tr>
<tr>
<td>24</td>
<td>68</td>
<td>22.5</td>
<td>50</td>
<td>71</td>
<td>21.0</td>
<td>76</td>
<td>70</td>
<td>23.5</td>
</tr>
<tr>
<td>25</td>
<td>70</td>
<td>22.5</td>
<td>51</td>
<td>69</td>
<td>23.0</td>
<td>77</td>
<td>73</td>
<td>23.0</td>
</tr>
<tr>
<td>26</td>
<td>74</td>
<td>24.5</td>
<td>52</td>
<td>70</td>
<td>24.0</td>
<td>78</td>
<td>73</td>
<td>21.0</td>
</tr>
</tbody>
</table>

Table 3.3

2. Enter data using the STAT list editor.

Press STAT ENTER to select the STAT list editor.

a. Enter the data for the quantitative variables "height" and "handspan" for females in lists L1 and L2.

Place the cursor on list L1 row 1 to make L1(1) the active list row. Enter the height data for females: 68, 64, 59, ... pressing ENTER after each entry.
3.1 Indicating Groups Within the Data on Scatterplots

Place the cursor on list L2 row 1 to make L2(1) the active list row. Enter the hand data for females: 21.5, 18.0, 20.0, ... in L2 pressing [ENTER] after each entry, as shown in Figure 3.6.

b. Enter the data for the quantitative variables "height" and "handspan" for males in lists L3 and L4.

Place the cursor on list L3 row 1 to make L3(1) the active list row. Enter the height data for males: 71, 73, 68, ... pressing [ENTER] after each entry. Place the cursor on list L4 row 1 to make L4(1) the active list row. Enter the hand data for males: 23.5, 22.5, 23.5, ... in L2 pressing [ENTER] after each entry, as shown in Figure 5.7.

Figure 3.5

Figure 3.6

Figure 3.7

3. Plot the statistical data by creating a scatterplot indicating groups within the data.


Create a scatterplot of female heights and handspans with heights on the horizontal axis and handspan on the vertical axis.

Press [ENTER], selecting Plot 1. Place the cursor on ON and press [ENTER]. Use the down arrow key and the right arrow key to select the first icon in the first row, the scatterplot. Press [ENTER]. Use the down arrow key to select list L1 as the Xlist, [2nd] L1. Use the down arrow key to enter list L2 as the Ylist: [2nd] L2. Use the down arrow key to select the second icon for the mark. The settings for Plot 1 are shown in Figure 3.8.

Create a scatterplot of male heights and handspans with heights on the horizontal axis and handspan on the vertical axis.

Use the up arrow key, to select Plot 2. Place the cursor on ON and press [ENTER]. Use the down arrow key and the right arrow key to select the first icon in the first row, the scatterplot. Press [ENTER]. Use the down arrow key to select list L3 as the Xlist, [2nd] L3. Use the down arrow key to enter list L4 as the Ylist: [2nd] L2. Use the down arrow key to select the third icon for the mark. The settings for Plot 2 are shown in Figure 3.9.

4. Set the Window viewing variables in order to view the graph.
Press \textit{WINDOW}, row 1, column 2. Set \textit{Xmin} to 55, \textit{Xmax} to 80; \textit{Xscl} to 1; \textit{Ymin} to 15. Set \textit{Ymax} to 26; \textit{Yscl} to 10; \textit{Xres} to 1. These settings are illustrated in Figure 3.10.

5. View the graph.

Press \textit{GRAPH} to view the graph, as shown in Figure 3.11.

6. Save list L1 as HGHTF and list L2 as HANDF.

Press \textit{2nd L1} \textit{STO} \押\textit{2nd A-LOCK} and type HGHTF; press \textit{ENTER}.

Press \textit{2nd L2} \textit{STO} \押\textit{2nd A-LOCK} and type HANDF; press \textit{ENTER}.

7. Save list L3 as HGHTM and list L4 as HANDM.

Press \textit{2nd L1} \textit{STO} \押\textit{2nd A-LOCK} and type HGHTM; press \textit{ENTER}.

Press \textit{2nd L2} \textit{STO} \押\textit{2nd A-LOCK} and type HANDM; press \textit{ENTER}.

Notice that the positive association between hands pan and height appears to hold within each sex. For both men and women, hands pan tends to increase as height increases.

3.2 Describing Linear Patterns With a Regression Line

Scatter plots show us a lot about a relationship, but we often want more specific numerical descriptions of how the response and explanatory variables are related. Imagine, for example, that we are examining the weights and heights of a sample of college women. We might want to know what the increase in average weight is for each 1-inch increase in height. Or, we might want to estimate the average weight for women with a specific height, like 5'10".

Regression analysis is the area of statistics used to examine the relationship between a quantitative response variable and one or more explanatory variables. A key element of regression analysis is the estimation of a regression equation that describes how, on average, the response variable is related to the explanatory vari-
3.2 Describing Linear Patterns With a Regression Line

ables. This regression equation can be used to answer the types of questions that we just asked about the weights and heights of college women.

A regression equation can also be used to predict values of a response variable using known values of an explanatory variable. For instance, it might be useful for colleges to have an equation for the connection between verbal SAT score and college grade point average (GPA). They could use that equation to predict the potential GPAs of future students, based on their verbal SAT scores. Some colleges actually do this kind of prediction to decide whom to admit, but they use a collection of variables to predict GPA.

There are many types of relationships and many types of regression equations. The simplest kind of relationship between two variables is a straight line, and that's the only type we will discuss here. Straight-line relationships occur frequently in practice, so this is a useful and important type of regression equation. Before we use a straight-line regression model, however, we should always examine a scatterplot to verify that the pattern actually is linear.

Example 3.2 Driver Age and the Maximum Legibility Distance of Highway Signs In a study of the legibility and visibility of highway signs, a Pennsylvania research firm determined the maximum distance at which each of 30 drivers could read a newly designed sign. The 30 participants in the study ranged in age from 18 to 82 years old. The government agency that funded the research hoped to improve highway safety for older drivers and wanted to examine the relationship between age and the sign legibility distance.

Table 3.4 lists the data. We will use the TI calculator to display a scatterplot to show that the relationship between "maximum distance" and "age" has a straight line pattern and to find the "best" line for this set of measurements. We will display a line that describes the average relationship between the two variables.

<table>
<thead>
<tr>
<th>Age</th>
<th>Distance</th>
<th>Age</th>
<th>Distance</th>
<th>Age</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>510</td>
<td>37</td>
<td>420</td>
<td>68</td>
<td>300</td>
</tr>
<tr>
<td>20</td>
<td>590</td>
<td>41</td>
<td>460</td>
<td>70</td>
<td>390</td>
</tr>
<tr>
<td>22</td>
<td>560</td>
<td>46</td>
<td>450</td>
<td>71</td>
<td>320</td>
</tr>
<tr>
<td>23</td>
<td>510</td>
<td>49</td>
<td>380</td>
<td>72</td>
<td>370</td>
</tr>
<tr>
<td>23</td>
<td>460</td>
<td>53</td>
<td>460</td>
<td>73</td>
<td>280</td>
</tr>
<tr>
<td>25</td>
<td>490</td>
<td>55</td>
<td>420</td>
<td>74</td>
<td>420</td>
</tr>
<tr>
<td>27</td>
<td>560</td>
<td>63</td>
<td>350</td>
<td>75</td>
<td>460</td>
</tr>
<tr>
<td>28</td>
<td>510</td>
<td>65</td>
<td>420</td>
<td>77</td>
<td>360</td>
</tr>
<tr>
<td>29</td>
<td>460</td>
<td>66</td>
<td>300</td>
<td>79</td>
<td>310</td>
</tr>
<tr>
<td>32</td>
<td>410</td>
<td>67</td>
<td>410</td>
<td>82</td>
<td>360</td>
</tr>
</tbody>
</table>

Table 3.4

Follow these steps to display a scatterplot with the regression equation superimposed upon the scatterplot.

1. Preparations:
Chapter 3 Relationships Between Quantitative Variables

a. Turn off all "Y=" functions.

Press \( Y= \) and press \( \text{CLEAR} \) to remove all functions. For each line that is not blank, place the cursor on the function and press \( \text{CLEAR} \). Press \( 2^\text{nd} \) \( \text{QUIT} \).

b. Clear all lists in the Stat editor.

Press \( \text{STAT} \), selecting 4: ClrList. Enter each list name: L1, L2, L3, L4, L5, L6. Press \( \text{ENTER} \) to execute the command.

c. Turn Diagnostics On to display \( r \), the correlation coefficient, and \( r^2 \), the coefficient of determination.

Press \( 2^\text{nd} \) \( \text{CATALOG} \), located on the bottom row, 2nd column from the left above the \( \mathbf{X} \). Press \( \text{ALPHA} \) and use the down arrow key to locate \( \text{DiagnosticOn} \), as shown in Figure 5.12. Press \( \text{ENTER} \) to select the command and press \( \text{ENTER} \) once again to execute the command.

![CATALOG](image)

Figure 3.12

2. Enter data using the \( \text{STAT} \) list editor.

Press \( \text{STAT} \) \( \text{ENTER} \) to select the \( \text{STAT} \) list editor.

a. Place the cursor on list L1 row 1 to make L1(1) the active list row. Enter the "Age" data: 18, 20, 22, ... pressing \( \text{ENTER} \) after each entry.

Place the cursor on list L2 row 1 to make L2(1) the active list row. Enter the "Distance" data: 510, 590, 560, ... in L2 pressing \( \text{ENTER} \) after each entry, as shown in Figure 5.13.

![Data Table](image)

Figure 3.13

3. Obtain the regression equation.

Press \( \text{STAT} \) \( \rightarrow \) to obtain the \( \text{STAT} \) CALC menu.
3.2 Describing Linear Patterns With a Regression Line

a. Use the down arrow key \( \downarrow \) seven times and press \( \text{ENTER} \) or just press \( \times \) to select 8: LinReg \((a+bx)\), as shown in Figure 3.14. Press \( \text{2nd} \) \( \downarrow \) to select the "Age" data. Press \( \text{2nd} \) \( \downarrow \) \( \downarrow \) to select the "Distance" data, as shown in Figure 5.15. Press \( \text{ENTER} \) to execute the command. The output from the TI calculator is displayed in Figure 5.16.

![Figure 3.14](image1)
![Figure 3.15](image2)
![Figure 3.16](image3)

The regression line \( y = 577 - 3x \) describes how the maximum sign legibility distance (the y variable) is related to driver age (the x variable).

4. Obtain data points to plot the regression equation.

Press \( \text{VARS} \) row 4, column 4. Select 5: Statistics. Use the right arrow, \( \downarrow \), twice, selecting 2: a. Press \( \text{VARS} \) 5: Statistics, and the right arrow, \( \downarrow \), twice, selecting 3: b. Press \( \boxed{x} \) \( \text{2nd} \) \( \downarrow \) \( \downarrow \). Press \( \text{STO} \rightarrow \text{2nd} \) \( \downarrow \) \( \downarrow \) \( \downarrow \) \( \downarrow \). Your screen should look like Figure 5.17. These data points represent the predicted values of "Distance" from the "Age" variable stored in list L1. These predicted values of "Distance" are stored in L3.

![Figure 3.17](image4)

5. Display a scatterplot with the regression equation superimposed upon the scatterplot.

Press \( \text{2nd} \) \( \text{STAT PLOT} \) accessing the StatPlot menu.

Press \( \text{ENTER} \) selecting Plot 1. Place the cursor on ON and press \( \text{ENTER} \). Use the down arrow key and the right arrow key to select the first icon in the first row, the scatterplot. Press \( \text{ENTER} \). Use the down arrow key to select list L1 as the Xlist, \( \text{2nd} \) L1. Use the down arrow key to enter list L2 as the Ylist: \( \text{2nd} \) L2. Use the down arrow key to select the second icon for the mark. The settings for Plot 1 are shown in Figure 3.18.

Use the up arrow key to place the cursor on Plot2. Place the cursor on ON and press \( \text{ENTER} \). Use the down arrow key and the right arrow key to select the second icon in the first row, the xyLine. Press \( \text{ENTER} \). Use the down arrow key to select L1 as the Xlist \( \text{2nd} \) L1. Use the down arrow key to select L3 as the Ylist \( \text{2nd} \) L3. The settings for Plot 2 are shown in Figure 3.19.

47
Chapter 3  Relationships Between Quantitative Variables

6. View the graph.

Press ZOOM 9: ZoomStat to view the graph, as shown in Figure 3.20.

Earlier, we asked these two questions about distance and age:

(i) How much does the distance decrease when age is increased?
(ii) For drivers of any specific age, what is the average distance at which the sign can be read?

The slope of the equation can be used to answer the first question. Remember that the slope is the number that multiples the variable and the sign of the slope indicates the direction of the association. Here, the slope tells us that, on average, the legibility distance decreases 3 feet when age increases by one year. This information can be used to estimate the average change in distance for any difference in ages. For an age increase of 30 years, the estimated decrease in legibility distance is 90 feet because the slope is -3 feet per year.

The question about estimating the average legibility distances for a specific age is answered by using the specific age as the x value in the regression equation. To emphasize this use of the regression line, we write it as

\[ \text{Average distance} = 577 - 3 \text{ Age} \]

1. Make predictions for specific ages, 20, 50, and 80, finding the average distance at which the sign can be read.


3.3 Measuring Strength and Direction with Correlation

calculations are shown in Figure 3.21.

\[
\begin{align*}
\text{a} + \text{b} \times \text{c} &= 516.5452298 \\
\text{a} + \text{b} \times \text{d} &= 426.3401687 \\
\text{a} + \text{b} \times \text{e} &= 336.1351076
\end{align*}
\]

Figure 3.21

For any given line, we can calculate the predicted value \( \hat{y} \) for each point in the observed data. To do this for any particular point, we use the observed \( x \) value in the regression equation. The prediction error for an observation is the difference between the observed \( y \) value and the predicted value \( \hat{y} \); the formula is error = \( (y - \hat{y}) \). The terminology "error" is somewhat misleading, since the amount by which an individual differs from the line is usually due to natural variation rather than "errors" in the measurements. A more neutral term for the difference \( (y - \hat{y}) \) is that it is the residual for that individual.

2. Obtain the residuals.

Recall that the predicted values of "Distance", based upon the "Age" variable are stored in L3 and the observed values of "Distance" are in L2.

Press \[ \text{STAT} \text{ ENTER} \] to select the \[ \text{STAT} \] list editor.

a. Place the cursor at the top of list L4. Press \[ \text{2nd} \text{ L2 \ - \ 2nd \ L3} \], as shown in Figure 3.22, pressing \[ \text{ENTER} \] to obtain the residuals.

b. Place the cursor at the top of list L5. Press \[ \text{2nd} \text{ LSTAT} \], selecting the list RESID, as shown in Figure 3.23. Press \[ \text{ENTER} \] to drive the residuals, that are automatically generated on the TI calculator with that list name, into L5. Observe that the residuals displayed in lists L4 and L5 are identical, as shown in Figure 3.24.

\[
\begin{align*}
\text{L2} & | \text{L3} | \text{L5} | \text{e} \\
52.56 & | 54.55 & | -2.96 \\
54.55 & | 54.55 & | 0.00 \\
50.16 & | 48.32 & | -1.84 \\
48.32 & | 50.16 & | -1.84 \\
\text{L4} = \text{L2} - \text{L3} \\
48.50 & | 48.50 & | 0.00 \\
48.50 & | 48.50 & | 0.00 \\
48.50 & | 48.50 & | 0.00 \\
48.50 & | 48.50 & | 0.00
\end{align*}
\]

Figure 3.22

\[
\begin{align*}
\text{L2} & | \text{L3} | \text{L4} | \text{e} \\
52.56 & | 54.55 & | -2.96 \\
54.55 & | 54.55 & | 0.00 \\
50.16 & | 50.16 & | 0.00 \\
48.32 & | 48.32 & | 0.00 \\
\text{L5} = \text{LRESID} \\
48.50 & | 48.50 & | 0.00 \\
48.50 & | 48.50 & | 0.00 \\
48.50 & | 48.50 & | 0.00
\end{align*}
\]

Figure 3.23

\[
\begin{align*}
\text{L2} & | \text{L3} | \text{L4} | \text{e} \\
52.56 & | 54.55 & | -2.96 \\
54.55 & | 54.55 & | 0.00 \\
50.16 & | 50.16 & | 0.00 \\
48.32 & | 48.32 & | 0.00 \\
\text{L5} = \text{LRESID} \\
48.50 & | 48.50 & | 0.00 \\
48.50 & | 48.50 & | 0.00 \\
48.50 & | 48.50 & | 0.00
\end{align*}
\]

Figure 3.24

3. Turn off all plots and return the graph window to standard viewing.

Press \[ \text{2nd} \text{ STAT PLOT} \], selecting PlotsOff and press \[ \text{ENTER} \]. Press \[ \text{ZOOM} \] and select 6: ZStandard to restore the default graph window settings.
3.3 Measuring Strength and Direction with Correlation

The linear pattern is so common that a statistic was created to characterize this type of relationship. The statistical correlation between two quantitative variables is a number that indicates the strength and the direction of a straight-line relationship.

(i) The strength of the relationship is determined by the closeness of the points to a straight line.

(ii) The direction is determined by whether one variable generally increases or generally decreases when the other variable increases.

As used in statistics, the meaning of the word correlation is much more specific than it is in everyday life. A statistical correlation only describes linear relationships. Whenever a correlation is calculated, a straight line is used as the frame of reference for evaluating the relationship. When

Example 3.2 Driver Age and the Maximum Legibility Distance of Highway Signs Revised In a study of the legibility and visibility of highway signs, a Pennsylvania research firm determined the maximum distance at which each of 30 drivers could read a newly designed sign. The 30 participants in the study ranged in age from 18 to 82 years old. The government agency that funded the research hoped to improve highway safety for older drivers and wanted to examine the relationship between age and the sign legibility distance.

Table 3.3 lists the data. We will use the TI calculator to determine the correlation coefficient between “maximum distance” and “age”.

Follow these steps to find the correlation coefficient and the coefficient of determination.

1. Preparations:
   a. Turn off all "Y=" functions.

      Press [Y=] and press [CLEAR] to remove all functions. For each line that is not blank, place the cursor on the function and press [CLEAR]. Press [2nd] [QUIT].

   b. Clear all lists in the Stat editor: Caution: If the "Age" and "Distance" data are within L1 and L2, Do NOT execute this step.

      Press [STAT], selecting 4: ClrList. Enter each list name: L1, L2, L3, L4, L5, L6. Press [ENTER] to execute the command.

   c. Turn Diagnostics On to display r, the correlation coefficient, and r², the coefficient of determination.

      Press [2nd] [CATALOG], located on the bottom row, 2nd column from the left above the Ï. Press [ALPHA] [D], and use the down arrow key to
locate DiagnosticOn, as shown in Figure 3.12. Press [ENTER] to select the command and press [ENTER] once again to execute the command.

2. Enter data using the STAT list editor.

Press STAT [ENTER] to select the STAT list editor.

a. Place the cursor on list L1 row 1 to make L1(1) the active list row. Enter the "Age" data: 18, 20, 22, ..., pressing [ENTER] after each entry. Place the cursor on list L2 row 1 to make L2(1) the active list row. Enter the "Distance" data: 510, 590, 560, ... in L2 pressing [ENTER] after each entry, as shown in Figure 3.13.

3. Obtain the regression equation, correlation coefficient and the coefficient of determination.

Press STAT [►] to obtain the STAT CALC menu.

a. Use the down arrow key [▼], seven times and press [ENTER], or just press [8] to select 8: LinReg (a+bx), as shown in Figure 3.14. Press 2nd [L] to select the "Age" data. Press 2nd [L2] to select the "Distance" data, as shown in Figure 3.15. Press [ENTER] to execute the command. The output from the TI calculator is displayed in Figure 3.16 and Figure 3.25.

For the data shown in Figure 3.20 relating driver age and sign legibility distance, the correlation is $r = -0.80$. This value indicates a somewhat strong negative association between the variables.

### Calculating the Sum of Squared Errors

A least squares line has the property that the sum of squared differences between the observed values of $y$ and the predicted values is smaller for that line than it is for any other line. Put more simply, the least squares line minimizes the sum of squared prediction errors for the observed data set. The notation SSE, which stands for sum of squared errors, is used to represent the sum of squared prediction errors. The least squares line (the regression line) has a smaller SSE than any other regression line that might be used to predict the response variable.
Chapter 3  Relationships Between Quantitative Variables

**Example - Exam Scores** Suppose that $x =$ score on exam 1 in a course and $y =$ score on exam 2, and that the first two rows in Table 5.5 (shown below) give $x$ values and $y$ values for $n = 6$ students. We will use the TI calculator to obtain the regression output, identifying the slope, intercept, $r^2$, SSTO, and SSE for this set of measurements.

\[
\begin{align*}
  x &= \text{Exam 1 score} & 70 & 75 & 80 & 80 & 85 & 90 \\
  y &= \text{Exam 2 score} & 75 & 82 & 80 & 86 & 90 & 91 \\
\end{align*}
\]

*Table 3.5*

Follow these steps to obtain the regression output, identifying the slope, intercept, $r^2$, SSTO, and SSE for this set of measurements.

1. Preparations:
   a. Turn off all "Y=" functions.

   Press $\text{[Y=]}$ and press $\text{CLEAR}$ to remove all functions. For each line that is not blank, place the cursor on the function and press $\text{CLEAR}$ Press $\text{[2nd]}$ $\text{[QUIT]}$.

   b. Clear all lists in the Stat editor.

   Press $\text{[STAT]}$, selecting 4: ClrList. Enter each list name: L1, L2, L3, L4, L5, L6. Press $\text{[ENTER]}$ to execute the command.

   c. Turn Diagnostics On to display $r$, the correlation coefficient, and $r^2$, the coefficient of determination.

   Press $\text{[2nd]}$ $\text{[CATALOG]}$, located on the bottom row, 2nd column from the left above the $\text{[0]}$. Press $\text{[ALPHA]}$ $\text{[D]}$, and use the down arrow key to locate $\text{DiagnosticOn}$, as shown in Figure 5.12. Press $\text{[ENTER]}$ to select the command and press $\text{[ENTER]}$ once again to execute the command.

2. Enter data using the $\text{[STAT]}$ list editor.

   Press $\text{[STAT]}$ $\text{[ENTER]}$ to select the $\text{STAT}$ list editor.

   a. Place the cursor on list L1 row 1 to make L1(1) the active list row. Enter the "$x = \text{Exam 1 score}"$ data: 70, 75, 80, ... pressing $\text{[ENTER]}$ after each entry.

   Place the cursor on list L2 row 1 to make L2(1) the active list row. Enter the "$y = \text{Exam 2 score}"$ data: 75, 82, 80, ... in L2 pressing $\text{[ENTER]}$ after...
3.3 Calculating the Sum of Squared Errors

each entry, as shown in Figure 3.13.

<table>
<thead>
<tr>
<th>L1</th>
<th>L2</th>
<th>L3</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>86</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>L1-L3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.26

3. Obtain the regression equation.

Press [STAT] ▶ to obtain the [STAT] CALC menu.

a. Use the down arrow key, ▼, seven times and press [ENTER] or just press 8 to select 8: LinReg (a+bx), as shown in Figure 3.27. Press 2nd L1 to select the "x = Exam 1 score" data. Press 2nd L2 to select the "y = Exam 2 score" data, as shown in Figure 3.28. Press [ENTER] to execute the command. The output from the TI calculator is displayed in Figure 3.29.

Figure 3.27

The regression equation is \( y = 20 + 0.8x \); the y-intercept is 20 and the slope is 0.8. The correlation coefficient, \( r = 0.918 \), describes a moderately strong positive association. The squared correlation is \( r^2 = 0.842 \). The "x = Exam 1 score" explains 84.2% of the variation among the "y = Exam 2 score" data.

4. Obtain the sum of square errors.

Press [VARS], row 4, column 4. Select 5: Statistics. Use the right arrow, ▶, twice, selecting 2: a. Press [VARS] 5: Statistics, and the right arrow, ▶, twice, selecting 3: b. Press \( \frac{x}{2nd L1} \). Press \( STO \cdot \cdot \cdot 2nd L3 \). Your screen should look like Figure 3.30. These data points represent the predicted values, \( \hat{y} \), from "x = Exam 1 score" variable stored in list L1. These predicted values, \( \hat{y} \), of "Exam 2 score" are stored in L3.

Press [STAT] [ENTER] to select the [STAT] list editor.

a. Place the cursor at the top of list L4. Press 2nd L2 - 2nd L3, as shown in Figure 3.31, pressing [ENTER] to obtain the residuals. The resid-

53
Chapter 3  Relationships Between Quantitative Variables

The predicted y values are shown in Figure 3.32.

b. Press \[ \text{2nd QUIT} \] Press \[ \text{2nd LIST} \rightarrow \rightarrow \] selecting 5: \text{sum}. Press \[ \text{2nd L4} \ \text{x^2} \] \] Press \[ \text{ENTER} \] Your screen should look like Figure 3.33.

The sum of squared errors is SSE = 30.

5. Obtain the total sum of squares, \( SSE = \sum (y - \bar{y})^2 \).

To obtain the mean of the predicted y values, press \[ \text{STAT} \rightarrow \text{CALC}, \] selecting 1: 1Var Stats. Press \[ \text{ENTER} \] Press \[ \text{2nd L3} \] and \[ \text{ENTER} \] as shown in Figure 5.34. The results are shown in Figure 5.35.

The mean of the predicted y values is 84.

Press \[ \text{STAT} \ \text{ENTER} \] to select the \[ \text{STAT} \] list editor.

a. Place the cursor at the top of list L5. Press \[ \text{2nd L2} \ \text{[8 4]} \] as shown in Figure 3.36, pressing \[ \text{ENTER} \] The results are shown in Figure 3.37.

b. Press \[ \text{2nd QUIT} \] Press \[ \text{2nd LIST} \rightarrow \rightarrow \] selecting 5: \text{sum}. Press \[ \text{2nd L5} \ \text{x^2} \] \] Press \[ \text{ENTER} \] Your screen should look like Figure
3.3 Calculating the Sum of Squared Errors

The total sum of squares, $SSTO = \sum (y - \bar{y})^2$ is 190.

The coefficient of determination, $r^2 = \frac{SSTO - SSE}{SSTO} = \frac{190 - 30}{190} = 0.84211$. 