12-1 Suppose the COMMERCE DATA displayed in Class Exercises for Unit #9 are to be used to investigate the prediction of shipments of dishwashers from investments; the necessary data should be stored in lists named DISH and INVES on your TI-Calculator. (The 9 steps in Class Exercises for Unit #12 will be very helpful in this exercise.)

(a) Identify the response (dependent) variable $Y$ and the explanatory (predictor or independent) variable $X$.

$Y = \text{shipments of dishwashers, and } X = \text{investments}$

(b) Write the equation of the least squares line, followed by a one sentence interpretation of the slope.

The equation of the least squares line can be written as

$$\text{DISH} = -2067.1433 + 92.0582(\text{INVES})$$

With each increase of one billion dollars in investments, shipments of dishwashers increases by about 92.0582 thousand.

(c) Complete the display titled **Scatter Plot of Investments and Shipments of Dishwashers** by (i) adding appropriate titles to the horizontal and vertical axes, and (ii) adding the two missing points to the scatter plot, corresponding to investments of 47.1 and 48.2 billion dollars.
(d) Use the least squares line to predict shipments of dishwashers when investments are 55 billion dollars.

The predicted shipments of dishwashers is $-2067.1433 + 92.0582(55) = 2996.060$ thousand.
(e) Use the least squares line to estimate the mean for shipments of dishwashers when investments are 35 billion dollars.

The estimated mean for shipments of dishwashers is $-2067.1433 + 92.0582(35) = 1154.895$ thousand

(f) Why would it not be appropriate to use the least squares line to estimate the shipments of dishwashers when investments are 25 billion dollars?

The least squares line may not be valid for prediction outside the range of the data.

(g) On average, what is the change in shipments of dishwashers with an increase of 3 billion dollars in investments?

With each increase of 3 billion dollars in investments, shipments of dishwashers increases by about $(3)(92.0582) = 276.1747$ thousand

(h) Add a graph of the least squares line to the display titled **Scatter Plot of Investments and Shipments of Dishwashers**.

$-2067.1433 + 92.0582(30) = 694.604$  
$-2067.1433 + 92.0582(70) = 4376.934$
(i) Use the least squares line to approximate the investments in a year when 2,000,000 dishwashers have been shipped.

Solve \(2000 = -2067.1433 + 92.0582(\text{INVES})\) to get \(\text{INVES} = 44.1801\) billion dollars
(j) Obtain the residuals, and verify that the five-number summary for these residuals is \(-965.3323, -470.4546, 46.6736, 408.7551, 949.9357\); then, do each of the following:

Find the range.

\[ 949.9357 - (-965.3323) = 1915.2680 \]

Find the interquartile range.

\[ 408.7551 - (-470.4546) = 879.2097 \]

Decide whether or not there are any outliers, and explain why or why not.

\[ Q_1 - Min = -470.4546 - (-965.3323) = 494.8777 < 1318.81455 = 1.5(IQR) \]

\[ Max - Q_3 = 949.9357 - 408.7551 = 541.1806 < 1318.81455 = 1.5(IQR) \]

Therefore, there is no potential outlier.

Construct a box plot of the residuals in the display titled \textit{Modified Box Plots of Residuals in the Simple Linear Regression to Predict Shipments of Dishwashers from Investment}. 
(k) Complete the display titled **Residual Plot in the Simple Linear Regression to Predict Shipments of Dishwashers from Investment** by (i) adding appropriate titles to the horizontal and vertical axes, and (ii) adding the missing point corresponding to the largest investment.

\[-2067.1433 + 92.0582(63.8) = 3806.173, \quad 3199 - 3806.173 = -607.173\]
(l) After examination of the residual plot in the display titled **Residual Plot in the Simple Linear Regression to Predict Shipments of Dishwashers from Investment** together with the scatter plot in the display titled **Scatter Plot of Investments and Shipments of Dishwashers**, complete the following statement concerning the relationship and the linearity assumption.

The residual plot appears to suggest that the linearity assumption **is**/ **is not** satisfied, since the points do not appear to be completely randomly scattered; instead, it seems a curve would be better to describe the relationship than a straight line.

The scatter plot appears to suggest that there is **no relationship** / **a positive linear relationship** / **a negative linear relationship** / **a positive nonlinear relationship** / **a negative nonlinear relationship** / **a relationship which is neither positive or negative**.
Suppose the COMMERCE DATA displayed in Class Exercises for Unit #9 are to be used to investigate the prediction of shipments of refrigerators from investments; the necessary data should be stored in lists named REFRG and INVES on your TI-Calculator. (The 9 steps in Class Exercises for Unit #12 will be very helpful in this exercise.)

(a) Identify the response (dependent) variable $Y$ and the explanatory (predictor or independent) variable $X$.

$Y =$ shipments of refrigerators, and $X =$ investments

(b) Write the equation of the least squares line, followed by a one sentence interpretation of the slope.

The equation of the least squares line can be written as

$$REFRIG = 1516.5873 + 75.0685(INVES)$$

With each increase of one billion dollars in investments, shipments of refrigerators increases by about 75.0685 thousand.

(c) Complete the display titled Scatter Plot of Investments and Shipments of Refrigerators by (i) adding appropriate titles to the horizontal and vertical axes, and (ii) adding the two missing points to the scatter plot, corresponding to the two largest investments of 62.4 and 63.8 billion dollars.
(d) Use the least squares line to predict shipments of refrigerators when investments are 55 billion dollars.

The predicted shipments of refrigerators is $1516.5873 + 75.0685(55) = 5645.352$ thousand
Use the least squares line to estimate the mean for shipments of refrigerators when investments are 35 billion dollars.

The estimated mean for shipments of refrigerators is $1516.5873 + 75.0685(35) = 4143.983$ thousand.

Indicate all investment values for which it would not be appropriate to use the least squares line to estimate the shipments of refrigerators.

The least squares line may not be valid for prediction outside the range of the data, which would be investments too much less than 35 billion dollars or too much more than 60 billion dollars.

On average, what is the change in shipments of refrigerators with an increase of 15 billion dollars in investments?

With each increase of 15 billion dollars in investments, shipments of refrigerators increases by about $(15)(75.0685) = 1126.0268$ thousand.

Add a graph of the least squares line to the display titled Scatter Plot of Investments and Shipments of Refrigerators.

$1516.5873 + 75.0685(30) = 3768.641$  $1516.5873 + 75.0685(70) = 6771.379$
(i) Use the least squares line to approximate the investments in a year when five million refrigerators have been shipped.

Solve $5000 = 1516.5873 + 75.0685(\text{INVES})$ to get $\text{INVES} = 46.40315$ billion dollars

---

**Modified Box Plot of Residuals in the Simple Linear Regression to Predict Shipments of Refrigerators from Investment**

<table>
<thead>
<tr>
<th>Thousands of Refrigerators</th>
</tr>
</thead>
<tbody>
<tr>
<td>-600</td>
</tr>
</tbody>
</table>

---
Obtain the residuals, and verify that the five-number summary for these residuals is \(-611.4352, -310.8587, -20.0848, 397.9624, 847.1133\); then, do each of the following:

Find the range.

\[
847.1133 - (-611.4352) = 1458.5485
\]

Find the interquartile range.

\[
397.9624 - (-310.8587) = 708.8211
\]

Decide whether or not there are any outliers, and explain why or why not.

\[
Q_1 - \text{Min} = -310.8587 - (-611.4352) = 300.5765 < 1063.23165 = 1.5(IQR)
\]

\[
\text{Max} - Q_3 = 847.1133 - 397.9624 = 449.1509 < 1063.23165 = 1.5(IQR)
\]

Therefore, there is no potential outlier.

Construct a box plot of the residuals in the display titled Modified Box Plots of Residuals in the Simple Linear Regression to Predict Shipments of Refrigerators from Investment.
Residual Plot in the Simple Linear Regression to Predict Shipments of Refrigerators from Investment

(k) Complete the display titled **Residual Plot in the Simple Linear Regression to Predict Shipments of Refrigerators from Investment** by (i) adding appropriate titles to the horizontal and vertical axes, and (ii) adding the missing point corresponding to the investment of 47.1 billion dollars.

\[1516.5873 + 75.0685(47.1) = 5052.311, \quad 5124 - 5052.311 = 71.689\]
After examination of the residual plot in the display titled *Residual Plot in the Simple Linear Regression to Predict Shipments of Dishwashers from Investment* together with the scatter plot in the display titled *Scatter Plot of Investments and Shipments of Dishwashers*, complete the following statement concerning the relationship and the linearity assumption.

The residual plot appears to suggest that the linearity assumption is / is not satisfied, since the points appear to be randomly scattered

The scatter plot appears to suggest that there is no relationship / a positive linear relationship / a negative linear relationship / a positive nonlinear relationship / a negative nonlinear relationship / a relationship which is neither positive or negative.
The Temperature and Bacteria Data, displayed as Table 12-4, is used in a study of the prediction of the number of thousands of bacteria that will develop in a particular environment from temperature.

(a) Identify the response variable $Y$ and the explanatory variable $X$.

The response variable is $Y = \text{“number of thousands of bacteria”}$, and the explanatory variable is $X = \text{“temperature”}$.

(b) Verify that the least squares line to predict number of bacteria ($bct$) from temperature ($tmp$) is

$$bct = -139.7321 + 4.8821(tmp)$$

(A programmable calculator or appropriate statistical software will be helpful.)

(c) Find the five-number summary and interquartile range for the residuals.

five-number summary for the residuals:

$$-44.4286, -14.9018, -2.1964, +19.8036, +42.8036$$

interquartile range for the residuals

$$= 19.8036 - (-14.9018) = 34.7054$$
(d) Decide whether there appear to be any candidates for outliers, and complete the construction of the modified box plot of the residuals. There are no outliers, since 
\[-14.9018 - (-44.4286) = 29.5268\]
and
\[42.8036 - 19.8036 = 23.0000\]
are both less than \(1.5(34.7054) = 52.0581\).

(e) Complete the construction of the scatter plot of the data, and graph the least squares line on the scatter plot.

(f) From the scatter plot constructed in part (e), decide whether or not you think the residual plot will look random.
The scatter plot constructed in part (e) appears to display a nonlinear relationship, which suggests that the residual plot will not look random.

<table>
<thead>
<tr>
<th>Temperature(°F)</th>
<th>Bacteria(1000s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>71</td>
</tr>
<tr>
<td>50</td>
<td>93</td>
</tr>
<tr>
<td>55</td>
<td>127</td>
</tr>
<tr>
<td>55</td>
<td>149</td>
</tr>
<tr>
<td>60</td>
<td>180</td>
</tr>
<tr>
<td>60</td>
<td>196</td>
</tr>
<tr>
<td>65</td>
<td>197</td>
</tr>
<tr>
<td>65</td>
<td>175</td>
</tr>
<tr>
<td>70</td>
<td>176</td>
</tr>
<tr>
<td>70</td>
<td>199</td>
</tr>
<tr>
<td>75</td>
<td>208</td>
</tr>
<tr>
<td>75</td>
<td>182</td>
</tr>
<tr>
<td>80</td>
<td>243</td>
</tr>
<tr>
<td>80</td>
<td>259</td>
</tr>
<tr>
<td>85</td>
<td>298</td>
</tr>
<tr>
<td>85</td>
<td>284</td>
</tr>
</tbody>
</table>
12-5 - continued

(g) Complete the construction of the residual plot.

(h) Decide whether or not the linearity assumption appears to be reasonable, and state why or why not.

The linearity assumption does not appear to be reasonable, since the residual plot does not look random.
In Exercises 10-5 and 11-1, the Biotin and Weight Data, displayed in Table 10-5, is used in a study of the relationship between the amount of biotin added per kilogram of feed and the weight in grams of newly hatched turkeys after three weeks, with interest in the prediction of weight from amount of biotin. The least squares line was found to be

\[ wt = 191 + 1.2(btn) \]

where \( wt \) represents weight in grams and \( btn \) represents amount of biotin per kilogram of feed.

(a) Find the residuals.

\[ +5, -9, -36, +15, +5, +42, -8, +17, -14, -17 \]

(b) Find the five-number summary and interquartile range for the residuals.

five-number summary for the residuals: \(-36, -14, -1.5, +15, +42\)
interquartile range for the residuals = \( 15 - (-14) = 29 \)
(c) Decide whether there appear to be any candidates for outliers, and complete the construction of the modified box plot of the residuals.

There are no outliers, since $-14 - (-36) = 22$ and $42 - 15 = 27$ are both less than $1.5(29) = 43.5$.

(d) From the scatter plot constructed in Exercise 11-1(h), decide whether or not you think the residual plot will look random; then, complete construction of the residual plot.

The scatter plot constructed in Exercise 11-1(h) appears to display a linear relationship, which suggests that the residual plot will look random.
(e) Decide whether or not the linearity assumption appears to be reasonable, and state why or why not.

The linearity assumption appears to be reasonable, since the residual plot looks random.

12-9 In Exercise 11-3, the Task Training Data, displayed in Table 11-1, is used to obtain the least squares line for the prediction of the number of minutes necessary to perform a job from the number of hours of training to perform the job.

(a) Why would it be difficult to make a decision about whether the linearity assumption is reasonable from a residual plot of this data?

It is difficult to make a decision about whether the linearity assumption is reasonable from a residual plot of this data, because there are so few data points.

(b) The number of hours of training to perform the job in the data ranged from 15 to 45 hours. Do you think the linearity assumption would still be reasonable if the training hours in the data had ranged from 5 to 100 hours? Why or why not?

The linearity assumption would most likely not be reasonable, if the training hours in the data had ranged from 5 to 100 hours, since we would expect the time to perform the job to stop decreasing after a certain number of training hours.
12-11 Each of 1200 people in a survey is asked to name which brand of toothpaste he or she uses most often. The results are displayed in the bar chart on the right. Explain why this bar chart is a misleading display of the results of the survey. Since the vertical axis does not begin at zero, the heights of the bars are distorted when compared with each other. For instance, the bar for “Brand X” is three times as high as the bar for “Other,” but the frequency for “Brand X” is less than two times as high as the frequency for “Other”.

12-13 A line graph is constructed with five different companies represented on the horizontal axis and number of employees represented on the vertical axis. Explain why this graph is an inappropriate and misleading representation of the data.

In a line graph, the variable on the horizontal axis should be time, but the five different companies have been represented on the horizontal axis.
Suppose the COMMERCE DATA displayed in Class Exercises for Unit #9 are to be used to investigate the prediction of expenditures from investments; the necessary data should be stored in lists named EXPEN and INVES on your TI-Calculator. (The 9 steps in Class Exercises for Unit #12 will be very helpful in this exercise.)

(a) Identify the response (dependent) variable $Y$ and the explanatory (predictor or independent) variable $X$. $Y = \text{expenditures}$, and $X = \text{investments}$

(b) Write the equation of the least squares line, followed by a one sentence interpretation of the slope.

The equation of the least squares line can be written as

$$\text{EXPEN} = -33.6990 + 2.9839(\text{INVES})$$

With each increase of one billion dollars in investments, expenditures increases by about 2.9839 billion dollars.

(c) Complete the display titled *Scatter Plot of Investments and Expenditures* by (i) adding appropriate titles to the horizontal and vertical axes, and (ii) adding the three missing points to the scatter plot, corresponding to the investments of 51.2, 52.7, and 53.7 billion dollars.

(d) Use the least squares line to predict expenditures when investments are 57 billion dollars.
The predicted expenditures is $-33.6990 + 2.9839(57) = 136.384$ billion dollars.
(e) Use the least squares line to estimate the mean expenditures when investments are 38 billion dollars.

The estimated mean expenditures is \(-33.6990 + 2.9839(38) = 79.690\) billion dollars.
(f) Indicate all investment values for which it would not be appropriate to use the least squares line to estimate the expenditures.

The least squares line may not be valid for prediction outside the range of the data, which would be investments too much less than 35 billion dollars or too much more than 60 billion dollars.

(g) On average, what is the change in expenditures with an increase of 30 billion dollars in investments?

With each increase of 30 billion dollars in investments, expenditures increase by about $(30)(2.9839) = 89.5175$ billion dollars.

(h) Add a graph of the least squares line to the display titled Scatter Plot of Investments and Expenditures.

$\begin{align*}
-33.6990 + 2.9839(30) &= 55.818 \\
-33.6990 + 2.9839(70) &= 175.175
\end{align*}$
(i) Use the least squares line to approximate the investments in a year when expenditures are 100 billion dollars.

Solve \(100 = -33.6990 + 2.9839(\text{INVES})\) to get \(\text{INVES} = 44.8066\) billion dollars

**Modified Box Plot of Residuals in the Simple Linear Regression to Predict Expenditures from Investment**

![Box Plot](image)
(j) Obtain the residuals, and verify that the five-number summary for these residuals is \(-45.5749, -21.1142, -3.2432, 20.4777, 60.7070\); then, do each of the following:

Find the range.

\[
60.7070 - (-45.5749) = 106.2819
\]

Find the interquartile range.

\[
20.4777 - (-21.1142) = 41.5919
\]

Decide whether or not there are any outliers, and explain why or why not.

\[
Q_1 - Min = -21.1142 - (-45.5749) = 24.4607 < 62.38785 = 1.5(IQR)
\]

\[
Max - Q_3 = 60.7070 - 20.4777 = 40.2293 < 62.38785 = 1.5(IQR)
\]

Therefore, there is no potential outlier.

Construct a box plot of the residuals in the display titled "Modified Box Plots of Residuals in the Simple Linear Regression to Predict Expenditures from Investment."
(k) Complete the display titled **Residual Plot in the Simple Linear Regression to Predict Expenditures from Investment** by (i) adding appropriate titles to the horizontal and vertical axes, and (ii) adding the missing point corresponding to the largest investment.
After examination of the residual plot in the display titled Residual Plot in the Simple Linear Regression to Predict Expenditures from Investment together with the scatter plot in the display titled Scatter Plot of Investments and Expenditures, complete the following statement concerning the relationship and the linearity assumption.

The residual plot appears to suggest that the linearity assumption is satisfied, since the points appear to be randomly scattered.

The scatter plot appears to suggest that there is

no relationship / a positive linear relationship / a negative linear relationship / a positive nonlinear relationship / a negative nonlinear relationship / a relationship which is neither positive or negative.