If the Shoe Fits . . .

How can you tell when two quantities—like age and heart rate—are related? In this module, you explore methods for finding and evaluating mathematical models of some important relationships.

Byron Anderson • Ruth Brocklebank • Pete Stabio

If the Shoe Fits . . .

Introduction

In one sense, moviegoers, doctors, economists, and sports fans all study relationships. You can tell when two characters on the screen might be more than friends from their first conversation. And you can guess from the crack of the bat when a fly ball might become a home run.

In many situations, however, it's not enough just to know that two things are related. It also may be important to know how they are related. For example, physiologists have found that as the amount of alcohol consumed increases, human reaction time also increases. In other words, the more people drink, the more time it takes for them to react to a sharp curve in the road—or a child's ball rolling in front of the car.

Discussion

Describe any relationship that may exist between the following quantities:

a. the number of hours worked by an employee and the employee's gross pay

b. the heights of basketball players and the numbers of points they score

c. the distance a student lives from school and the student's arrival time at school in the morning

d. the ages of drivers and the numbers of automobile accidents they cause

e. the number of hours spent studying and a student's grade-point average

Activity 1

Medical scientists have studied the relationship between exercise and the human body for decades. Much of that research has focused on the body's need for additional oxygen during physical activity.

One way to determine if there is a relationship between two quantities is to gather and analyze data. After the data has been collected, it can be organized into ordered pairs and graphed. If there appears to be a trend or pattern, an equation can often be found to summarize the relationship mathematically. If the equation is a good model, it can then be used to make predictions for one quantity based on specific values of the other quantity.

Data that appears to have a linear pattern can be modeled with an equation of the form $y = mx + b$. Finding an equation that models a set of data is not an exact science. In most cases, the model will not fit every data point. In many cases, however, it is possible to find a line that models the data reasonably well. In this activity, you examine some data for oxygen consumption and heart rate, determine a linear model for the data, then decide how well the line models the data by calculating residuals.

Exploration

During a treadmill test, physiologists monitor heart rate and rate of oxygen consumption while the subject walks on a treadmill. Table **1** shows the results of a hypothetical treadmill test in which the treadmill's elevation was increased 2.5% every 3 min.

Table 1: Heart rate and oxygen consumption during a treadmill test

Heart Rate (beats/min)	Oxygen Consumption (L/min)
60	0.41
80	0.71
100	0.90
120	1.27
140	1.62
150	1.76
160	1.92
180	2.24
200	2.42

a. Make a scatterplot of the data in Table **1**. Let x represent heart rate and y represent oxygen consumption.

b.
1. Draw a line that you think fits the data reasonably well.
2. Determine the equation of your line in the form $y = mx + b$. **Note:** Round the values of m and b to the nearest thousandth.

Mathematics Note

In real-world applications, a **linear model** seldom yields an equation that fits all the data—even when a line is a good model for the situation. The difference between the y-coordinate of a data point and the corresponding y-value of the model is a **residual**. Since data points may be located above or below the model, the values of the residuals may be positive or negative.

The **absolute value of a residual** is a measure of the distance from the data point to the model. The sum of the absolute values of the residuals provides information about how well a model fits the data. In general, the smaller the sum of the absolute values of the residuals, the more closely a model approximates the data.

For example, consider the coordinates of the data points shown in the two left-hand columns of Table **2** below. This data can be modeled by the linear equation $y = -1.46x + 230$. The table also lists the corresponding values of $y = -1.46x + 230$ for each x-value in the data, the absolute values of the residuals, and the sum of the absolute values of the residuals for this model.

Table 2: Data points and residuals

x-value	y-value	predicted y-value	$\lvert$residual$\rvert$
20	200	200.8	0.8
25	193	193.5	0.5
30	188	186.2	1.8
35	180	178.9	1.1
40	170	171.6	1.6
		Sum	5.8

c. **1.** Determine the absolute value of the residual for each data point using your equation from Part **b**.

2. Determine the sum of the absolute values of the residuals.

d. Try to reduce the sum of the absolute values of the residuals by adjusting the slope and the y-intercept of your equation from Part **b**. Record the equation that results in the lowest sum.

Discussion

a. What does your scatterplot indicate about the relationship between heart rate and oxygen consumption?

b. What kinds of information could indicate that a line is a good model of a data set?

c. Is the line you found in Part **b** of the exploration a good model of the data? Explain your response.

d.
1. In Part **d** of the exploration, what was the equation of the linear model that generated the smallest sum of the absolute values of the residuals?
2. Describe how you adjusted the equation of your line from Part **b** to find this model.

e. Describe how you would use a linear model to predict each of the following:
1. the rate of oxygen consumption when the heart rate is 168 beats/min
2. the heart rate when oxygen consumption is 1.5 L/min

f. What are some of the dangers of using models to make predictions?

Assignment

1.1 The following table shows the suggested maximum heart rate (by age) for persons entering aerobic training programs.

Age (years)	Heart Rate (beats/min)
20	200
25	195
30	190
35	185
40	180
45	173
50	166
55	160
60	155
65	150
70	145

Source: Adapted from McArdle, Katch, and Katch, *Exercise Physiology*.

a. Draw a scatterplot of the data.

b. Find a linear equation that models the data. Graph this equation on your scatterplot.

c. Determine the sum of the absolute values of the residuals for the equation from Part **b**.

d. Use the equation from Part **b** to predict the suggested maximum heart rate for a person of age 15.

e. Many physiologists consider a rate of about 220 beats/min to be the upper limit for the human heart. Given this information, does the equation from Part **b** appear to be valid for all ages? Explain your response.

1.2 In recent years, physical fitness has focused on aerobic conditioning. Regular aerobic exercise increases your ability to supply oxygen to muscles. When your muscles receive more oxygen, they work more efficiently.

The table below shows data for heart rate and oxygen consumption for a 20-year-old female, both before and after a 10-week aerobic conditioning program. Since the volume of oxygen consumed depends on a person's mass, oxygen consumption is measured here in milliliters per kilogram per minute.

	Before Aerobic Program	**After Aerobic Program**
Heart Rate (beats/min)	**Oxygen Consumption (mL/kg/min)**	**Oxygen Consumption (mL/kg/min)**
100	21.3	22.5
110	22.6	24.4
120	23.8	26.6
130	25.0	28.4
140	25.9	30.3
150	27.4	32.1
160	28.6	34.0
170	29.7	35.8
180	30.9	37.2

Source: Adapted from McArdle, Katch, and Katch, *Exercise Physiology*.

a. On a single set of axes, create scatterplots of the data for both before and after the aerobic conditioning program. Let x represent heart rate and y represent oxygen consumption.

b. Find a linear equation to model each set of data. Graph both equations on the scatterplot from Part **a**.

c. Which line is the better model of its data set: the line that models the "before" data or the line that models the "after" data? Defend your choice using the sum of the absolute values of the residuals.

d. Use your models to predict the heart rate—both before and after the conditioning program—when the rate of oxygen consumption is 40 mL/kg/min.

* * * * *

1.3 The table below shows the mean amount spent on health care in the United States, per person, for each year from 1984 to 1991.

Year	Amount Spent ($) per Person
1984	1049
1985	1108
1986	1135
1987	1135
1988	1298
1989	1407
1990	1480
1991	1554

Source: U.S. Bureau of the Census, 1995.

a. Make a scatterplot of the data.

b. Find the equation of a line that models the data. Graph this equation on the scatterplot from Part **a**.

c. Determine the sum of the absolute values of the residuals for this model.

d. Using your model, predict the mean amount per person that Americans will spend on health care in the year 2000.

1.4 The table below shows the percentage of the U.S. population who participated in the work force for selected years from 1970 to 1992.

Year	Percentage in Work Force
1970	60.4
1980	63.8
1985	64.8
1990	66.4
1992	66.3

Source: U.S. Bureau of the Census, 1995.

a. Make a scatterplot of this data.

b. Find the equation of a line that models the data.

c. Use your model to estimate the percentage of the population who will participate in the work force during the year 2000.

d. The U.S. Bureau of the Census predicts that 68.7% of the population will participate in the work force in the year 2000. How does this prediction compare with the one you made in Part **c**?

* * * * * * * * * *

Activity 2

Scientists and engineers often make decisions or predictions based on observed data. These decisions can have significant, even life-saving, consequences. For example, the amount of radiation administered to cancer patients is determined using experimental models. Airplane designs depend on data collected using wind tunnels. And the decision to place a new drug on the market is made only after analyzing the results of clinical trials in both humans and animals. In all of these applications, models that accurately represent observed data are of vital importance.

When a set of data appears to have a linear relationship, there are several ways to obtain lines that model the data reasonably well. In the previous activity, you used trial and error to adjust the equation of the linear model until the sum of the absolute values of the residuals was reasonably small.

Another method for determining a linear model is based on a measure of central tendency: the median. The line generated by this method is the **median-median line**. In this activity, you examine the procedure for finding a median-median line.

Mathematics Note

When a set of data appears to have a linear relationship, one reasonable linear model is the **median-median line**.

To find this model, the data points are sorted according to the values of their x-coordinates. They are then divided into three groups of equal size, if possible. If not possible and there is a single extra data point, the middle group is increased by one. If there are two extra data points, the end groups are each increased by one. The coordinates of three **summary points** are found using the median x-coordinate and median y-coordinate of each group.

The median-median line is located parallel to the line through the outer summary points and $1/3$ the vertical distance from that line to the middle summary point. For example, Figure **1** shows a scatterplot of the data points (1,2), (2,2), (3,4), (4,5), (5,4), (6,5), and (7,7).

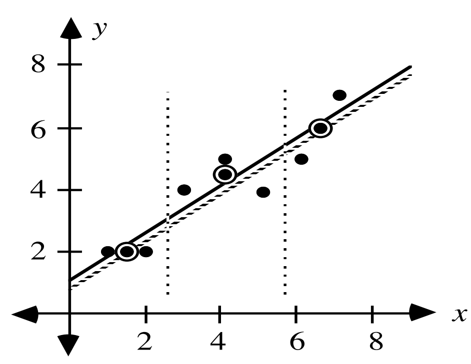


Figure 1: Median-median line through sample data

The dotted vertical lines divide the data into two groups with two points each and a middle group with three data points each. The circled points—(1.5,2), (4,4.5), and (6.5,6), respectively—are the summary points. The solid line is the median-median line. It is located parallel to the line that connects the first and last summary points (the dotted line), and $1/3$ of the vertical distance from the dotted line to the middle summary point. In this case, the equation of the median-median line is $y = 0.8x + 0.97$.

Exploration

In this exploration, you collect data on arm span and shoe size, then use a median-median line to model the relationship between these measurements.

a.
1. Measure and record your arm span in centimeters. (Arm span is the distance across the back from fingertip to fingertip, with arms extended away from the body at shoulder level.)
2. Record your shoe size to the nearest half-size.
3. Compile the class data for arm span and shoe size.

b. Make a scatterplot of the data. Let x represent shoe size and y represent arm span.

c. Use the following steps to divide the data points on your graph into three groups.

1. Create three groups of equal size, if possible. Any data points with the same x-coordinate should be placed in the same group.

   The first group should contain the points that have the smallest x-coordinates; while the third group should contain the points that have the largest x-coordinates. The middle group should contain the remaining data points.

   If the data cannot be divided into three equal groups, divide them as described in Step **2** or Step **3**.
2. If there is one extra point, place it in the middle group.
3. If there are two extra points, place one in the first group and one in the third group.

d. Use the following process to graph the summary point for each group.

1. Find the median of the x-coordinates of the group.
2. Find the median of the y-coordinates of the group.
3. Use the medians of the x-coordinates and the y-coordinates to form an ordered pair.
4. Plot and circle this ordered pair on your scatterplot from Part **b**.

e. Find the median-median line and determine its equation by following the steps below.

1. Draw a line that passes through the summary points of the first and third groups and find the equation of this line.
2. Find the point on the line drawn in Step **1** that has the same x-coordinate as the summary point of the middle group. Label this point on the graph.
3. Find the coordinates of the point that is $1/3$ of the vertical distance between the point labeled in Step **2** and the summary point of the middle group. Graph this point.
4. Draw the line that is parallel to the line in Step **1** and contains the point graphed in Step **3**. This is the median-median line.
5. Find the equation of the median-median line. **Note:** Save this equation, along with the data from this exploration, for use in the assignment and in the next activity.

Discussion

a. Why do you think medians are used to determine summary points?

b. Does the median-median line appear to be a good model of the class data? Defend your response.

c. How does the y-intercept of the line connecting the first and third summary points compare with the y-intercept of the median-median line?

d.
1. Use the median-median line to predict the arm span of a person with a shoe size of 15.
2. Based on your classroom data, does this prediction seem reasonable? Explain your response.

Assignment

2.1 **a.** According to the 1990 *Guinness Book of World Records*, the arm span of the tallest human on record, Robert Wadlow, was 289 cm. Use the median-median line from Part **e** of the exploration to estimate Wadlow's shoe size.

b. Robert Wadlow's actual shoe size was 37. How does your estimate compare with this value?

2.2 **a.** Use technology to find the median-median line of the shoe-size data from the exploration.

b. How does this line compare to the median-median line you calculated by hand in the exploration?

2.3 Since 1985, hospitals have recorded the number of HIV-positive patients discharged after an illness. The table below shows the number of discharged patients who were HIV-positive from 1985 to 1991.

Year	HIV-Positive Patients (in thousands)
1985	23
1986	44
1987	67
1988	95
1989	140
1990	146
1991	165

Source: U.S. Bureau of the Census, 1995.

a. Make a scatterplot of this data.

b. Find the median-median line that models the data.

c. Does the number of discharged patients who are HIV-positive appear to be increasing linearly? Defend your answer.

* * * * *

2.4 The table below shows the number of self-employed workers in the United States for selected years from 1970 to 1992.

Year	Self-Employed Workers (in thousands)
1970	7031
1975	7427
1980	8642
1985	9269
1989	10,008
1990	10,160
1991	10,341
1992	10,017

Source: U.S. Bureau of the Census, 1995.

a. Make a scatterplot of the data.

b. Find the median-median line that models the data.

c. **1.** Use the median-median line to estimate the number of people who were self-employed in 1988.

2. According to the U.S. Bureau of the Census, the actual number of self-employed workers in 1988 was 9917. How does your prediction compare with this value?

2.5 The table below shows the number of births performed by cesarean section in the United States for selected years from 1970 to 1991.

Year	**Number of Cesarean Sections (in thousands)**
1970	195
1980	619
1983	808
1984	813
1985	877
1986	906
1987	953
1988	933
1989	938
1990	945
1991	933

Source: U.S. Bureau of the Census, 1995.

a. Make a scatterplot of this data.

b. Find the median-median line that models the data.

c. Does the median-median line appear to be a reasonable model?

d. Do you think that the number of cesarean births will increase linearly as time passes? Explain your response.

* * * * * * * * * *

Activity 3

As you saw in Activity **2**, the median-median line provides one possible linear model for a set of data. However, the method used to determine the line tends to disregard data points that have large residuals. In this activity, you use residuals to investigate another method for finding the equation of a linear model.

Exploration 1

a. Create a scatterplot of the data for arm span and shoe size from Activity **2**.

b.
1. Draw a line that you think fits the data reasonably well.

2. Determine the equation of your line in the form $y = mx + b$.
Note: In this exploration, round all values to the nearest thousandth.

Mathematics Note

A **linear regression equation** for a set of data is the equation of a linear model that minimizes the sum of the squares of the residuals.

For example, Table **3** lists the coordinates of some data points, the corresponding *y*-values of a linear model, the squares of the residuals, and the sum of the squares of the residuals.

Table 3: Data points and squares of residuals

x-value	y-value	predicted y-value	$(\text{residuals})^2$
20	200	200.8	0.64
25	193	193.5	0.25
30	188	186.2	3.24
35	180	178.9	1.21
40	170	171.6	2.56
		Sum	7.90

Adjusting the equation of a linear model to obtain the smallest sum possible results in the linear regression equation $y = -1.46x + 230$. A method for determining such an equation is called the **least-squares method**. Most forms of technology use the least-squares method to determine linear regression equations.

c. **1.** Determine the square of the residual for each data point using your equation from Part **b**.

2. Determine the sum of the squares of the residuals.

d. Try to reduce the sum of the squares of the residuals by adjusting the slope and the *y*-intercept of your equation from Part **b**. Record the equation that results in the lowest sum.

e. Use technology to find the linear regression equation for the data.

Discussion 1

a. How did the model you found in Part **d** of the exploration compare with each of the following:

1. the median-median line found in Activity **2**?

2. the linear regression equation found using technology?

b. **1.** When would you expect the regression line for a set of data to differ greatly from the median-median line?

2. When would you expect the regression line to be approximately the same as the median-median line?

c. In what types of situations would a regression line provide a better model than a median-median line?

Exploration 2

One way to analyze a linear model is to create a **residual plot**. In the following exploration, you use residual plots to evaluate models of two different data sets. **Note:** In this exploration, round all values to the nearest thousandth.

a. Table **4** shows some data on oxygen consumption and heart rate during a treadmill test.

Table 4: Heart rate, oxygen consumption, and residuals

Heart Rate (beats/min)	Oxygen Consumption (L/min)	Predicted Oxygen Consumption	Residual
60	0.41		
80	0.71		
100	0.90		
120	1.27		
140	1.62		
150	1.76		
160	1.92		
180	2.24		
200	2.42		

One possible model for this data is the linear regression equation $y = 0.015x - 0.509$, where y represents oxygen consumption and x represents heart rate. Create a scatterplot of the data and graph the equation $y = 0.015x - 0.509$ on the same coordinate system.

b.
1. Use the equation $y = 0.015x - 0.509$ to determine the predicted oxygen consumption for each heart rate. Record these values in Table **4**.
2. Determine the value of the residual for each data point. Record these values in Table **4**.

c. Does the line $y = 0.015x - 0.509$ appear to fit the data well? Explain your response.

Mathematics Note

In a **residual plot**, the *x*-values of the data are represented on the *x*-axis and the residuals are represented on the *y*-axis. A scatterplot is created using the ordered pairs (*x*-value, residual). The plot is then examined for possible patterns. A graph in which the points are randomly scattered above and below the *x*-axis typically indicates that a reasonable model has been selected.

For example, consider the coordinates of the data points shown in the two left-hand columns of Table **5** below. This data can be modeled by the linear equation $y = 0.75x + 1.14$. The table also lists the corresponding values of $y = 0.75x + 1.14$ for each *x*-value in the data and the value of the residual for each data point.

Table 5: Data points, predicted *y*-values, and residuals

x-value	y-value	predicted y-value	residual
1	2	1.89	0.11
2	2	2.64	–0.64
3	4	3.39	0.61
4	5	4.14	0.86
5	4	4.89	–0.89
6	5	5.64	–0.64
7	7	6.39	0.61

Figure **2** shows the corresponding residual plot. Since the points appear to be randomly scattered above and below the *x*-axis, the equation appears to be a reasonable model for the data.

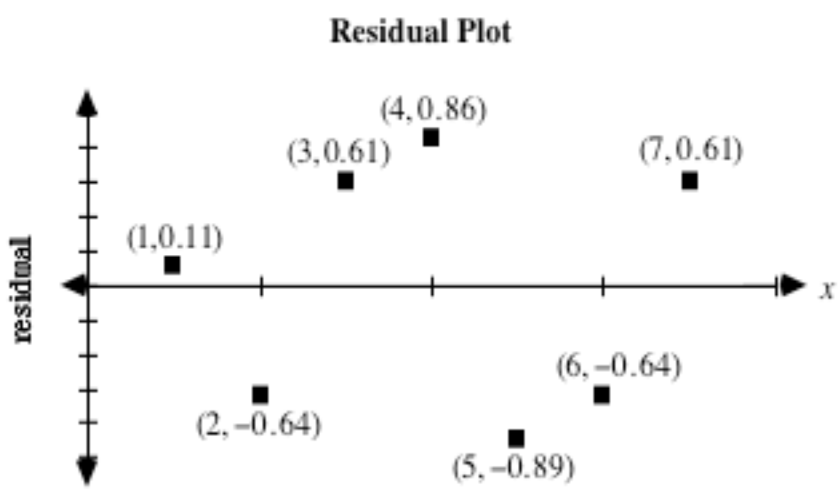


Figure 2: A residual plot

d. Create a residual plot for the equation $y = 0.015x - 0.509$ and the data in Table **4**. Record any patterns you observe.

e. Repeat Parts **a–d** using the equation $y = 0.02x - 1$ as a model of the data.

f. Compare the two residual plots from Parts **d** and **e**.

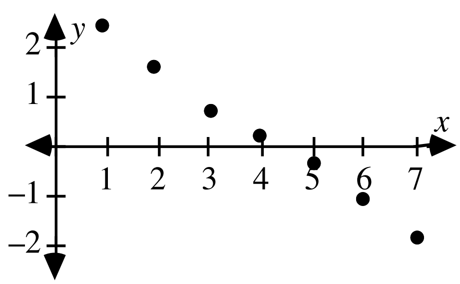
Discussion 2

a. Which equation in the exploration is the better model of the data in Table **4**? Justify your response.

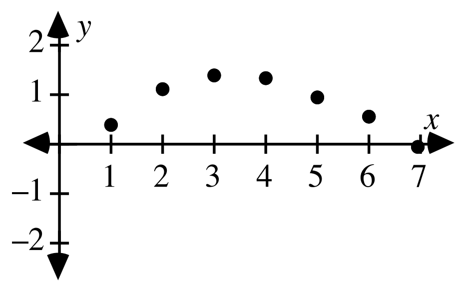
b. Describe how you would use a residual plot to determine whether a linear model is appropriate for a set of data.

c. Each of the four graphs below is a residual plot for a linear model of a different data set. Which plots indicate that the linear model is appropriate for the data? Justify your responses.

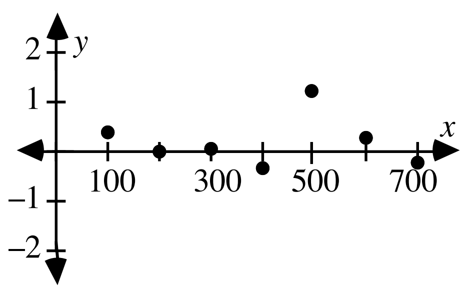
1.



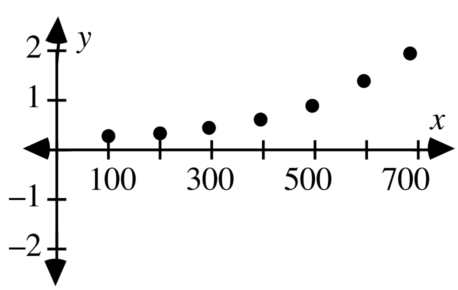
2.



3.



4.



Assignment

3.1 The table below shows the life expectancy for people born in selected years from 1970 to 1991.

Birth Year	Life Expectancy (years)	Birth Year	Life Expectancy (years)
1970	70.8	1983	74.6
1975	72.6	1984	74.7
1976	72.9	1985	74.7
1977	73.3	1986	74.7
1978	73.5	1987	74.9
1979	73.9	1988	74.9
1980	73.7	1989	75.1
1981	74.1	1990	75.4
1982	74.5	1991	75.7

Source: U.S. Bureau of the Census, 1995.

a. Make a scatterplot of the data.

b. Find the equation of the regression line for the data. Graph this equation on the scatterplot from Part **a**.

c. Using the equation from Part **b**, predict the life expectancy of people born in 1973 and people born in 2010.

d. The U.S. Bureau of the Census predicts that people born in 2010 will have a life expectancy of 77.6 years. How does this prediction compare with the one you made in Part **c**?

3.2 The table below shows the energy used, in kilocalories per minute, by a person walking at speeds from 1.45 km/hr to 9.20 km/hr.

Walking Speed (km/hr)	Energy Used (kcal/min)
1.45	2.15
2.75	2.75
3.90	3.40
4.60	4.00
5.20	4.40
6.30	5.60
7.20	6.80
7.70	8.00
8.30	9.25
9.20	11.00

Source: Adapted from McArdle, Katch, and Katch, *Exercise Physiology*.

a. Find a linear regression equation that models this data.

b. Make a residual plot for the equation from Part **a**.

c. What does the residual plot indicate about an appropriate model for the data? Explain your response.

3.3 The table below shows the U.S. birth rate, per 1000 women, for selected years.

Year	Birth Rate per 1000
1970	87.9
1980	68.4
1985	66.3
1989	69.2
1990	70.9

Source: U.S. Bureau of the Census, 1995.

a. Find the linear regression equation that models the data.

b. Make a residual plot for the model from Part **a**.

c. Does the model appear to be appropriate for the data? Explain your response.

* * * * *

3.4 The table below shows the average daily cost of a stay in a U.S. community hospital, for the years from 1984 to 1991.

Year	Daily Cost ($)
1984	411
1985	460
1986	501
1987	539
1988	586
1989	637
1990	687
1991	752

Source: U.S. Bureau of the Census, 1995.

a. Find the linear regression equation for this data.

b. Using the equation from Part **a**, predict the average daily cost of a hospital stay in the year 2000.

c. Make a residual plot for the model from Part **a**.

d. What does the residual plot indicate about an appropriate model for the data? Explain your response.

3.5 The table below shows the number of men and women employed in the United States for the years from 1986 to 1992.

Year	Males (in thousands)	Females (in thousands)
1986	60,892	48,706
1987	62,107	50,334
1988	63,273	51,696
1989	64,315	53,027
1990	64,435	53,479
1991	63,593	53,284
1992	63,805	53,793

Source: U.S. Bureau of the Census, 1995.

a. Find the linear regression equation that models the total employment of men and women for the years shown in the table.

b. According to this model, what is the average yearly increase in the number of American workers?

c. Make a residual plot for the model from Part **a**.

d. What does the residual plot indicate about an appropriate model for the data? Explain your response.

e. Repeat Parts **a–d** using the data for female workers only.

f. Repeat Parts **a–d** using the data for male workers only.

g. Compare the average yearly increases for number of female workers, number of male workers, and total number of workers.

* * * * * * * * * *

Summary Assessment

1. The table below shows some data on body mass and oxygen consumption collected during treadmill tests.

Body Mass (kg)	Oxygen Consumption (L/min)
60	1.50
65	1.66
70	1.78
75	1.93
80	2.07
85	2.17
90	2.30
95	2.46
100	2.56

Source: Adapted from McArdle, Katch, and Katch, *Exercise Physiology*.

a. Make a scatterplot of the data.

b. **1.** Find a linear regression equation that models the data.

2. Find a median-median line that models the data.

c. Find the sum of the squares of the residuals for each model in Part **b**.

d. Make a residual plot for each model in Part **b**.

e. Which model better represents the data? Justify your response.

2. The amount of energy necessary to maintain normal body functions—such as heart beat and respiration—is the **basal metabolic rate**. Basal metabolic rate is typically measured in energy output per unit surface area per hour. It is relatively constant for people of the same age and sex.

The table below shows some data for males from 5 to 70 years of age. Determine whether a linear model is appropriate for this set of data. Justify your answer using graphs, equations, and residuals.

Age (years)	Metabolic Rate ($\mathbf{kcal/m^2/hr}$)
5	50
10	44
20	40
30	38
35	37
55	34
70	33

Source: Adapted from McArdle, Katch, and Katch, *Exercise Physiology*.

Module Summary

- The difference between the y-coordinate of a data point and the corresponding y-value of the model is a **residual**. Since data points may be located above or below the model, the values of the residuals may be positive or negative.
- The **absolute value of a residual** is a measure of the distance from the data point to the model. The sum of the absolute values of the residuals provides information about how well a model fits the data. In general, the smaller the sum of the absolute values of the residuals, the more closely a model approximates the data.
- When a set of data appears to have a linear relationship, one reasonable linear model is the **median-median line**.

  To find this model, the data points are sorted according to the values of their x-coordinates. They are then divided into three groups of equal size, if possible. If not possible and there is a single extra data point, the middle group is increased by one. If there are two extra data points, the end groups are each increased by one. The coordinates of three **summary points** are found using the median x-coordinate and median y-coordinate of each group.

  The median-median line is located parallel to the line through the outer summary points and $1/3$ the vertical distance from that line to the middle summary point.
- A **linear regression equation** for a set of data is the equation of a linear model that minimizes the sum of the squares of the residuals. A method for determining such an equation is called the **least-squares method**. Most forms of technology use the least-squares method to determine linear regression equations.
- In a **residual plot**, the x-values of the data are represented on the x-axis and the residuals are represented on the y-axis. A scatterplot is created using the ordered pairs (x-value, residual). The plot is then examined for possible patterns. A graph in which the points are randomly scattered above and below the x-axis typically indicates that a reasonable model has been selected.

Selected References

McArdle, W. D., F. I. Katch, and V. L. Katch. *Exercise Physiology*. Philadelphia: Lea & Febiger, 1991.

North Carolina School of Science and Mathematics, Department of Mathematics and Computer Science. *Contemporary Precalculus through Applications*. Dedham, MA: Janson Publications, 1992.

Rodahl, K. *The Physiology of Work*. Bristol, PA: Taylor & Francis, 1989.

Thomas, T. R., and B. R. Londeree. "Calories Expended, Walking vs. Jogging." *The Physician and Sports Medicine* 17 (May 1989): 98.

U.S. Bureau of the Census. *Statistical Abstract of the United States: 1995*. Washington, DC: U.S. Government Printing Office, 1995.