Conceptually the complicated computer algorithms simply find the location of the plane that minimizes the sum of squared errors, where an error is defined as the vertical distance from each observation to the model plane. For more than two predictors the model is equivalent to a hyperplane in spaces with four or more dimensions. Although it is impossible to draw a picture of such hyperplanes, the concept is exactly the same. The best location for the hyperplane minimizes the sum of squared errors, and the hyperplane's intercept and slopes define the parameter estimates.

INTERPRETING PARAMETERS IN MULTIPLE REGRESSION

We will use a detailed examination of the data in Figure 6.2 to develop our understanding of the meaning of the partial regression coefficients in a multiple regression model. This dataset, which appears in many SAS examples (SAS Institute, 2008) and in other examples found on the web, contains the weight, height, age, and sex of 19 middle school or junior high students. It is particularly useful for developing our understanding of partial regression coefficients because we have good intuitions about the relationships among these variables. Our goal will be to develop a model of the variable weight. We expect boys to weigh more than girls unless the girls are a lot older and taller than the boys. We expect older students to be heavier, except if the younger students happen to be exceptionally tall. We will see that such intuitions underlie the interpretation of multiple regression models.

We will begin by considering age and height as predictors of weight. That is, our model is:

$$Wt_i = \beta_0 + \beta_1 Age_i + \beta_2 H_t + \varepsilon_i$$

FIGURE 6.2 Sex, age, height, and weight of 19 middle school students

Name	*Sex*	*Age*	*Height (in.)*	*Weight (lb)*
Alfred	M	14	69.0	112.5
Antonia	F	13	56.5	84.0
Barbara	F	13	65.3	98.0
Camella	F	14	62.8	102.5
Henry	M	14	63.5	102.5
Jamal	M	12	57.3	83.0
Jane	F	12	59.8	84.5
Janet	F	15	62.5	112.5
Jayden	M	13	62.5	84.0
John	M	12	59.0	99.5
Joyce	F	11	51.3	50.5
Judy	F	14	64.3	90.0
Louise	F	12	56.3	77.0
Mary	F	15	66.5	112.0
Philip	M	16	72.0	150.0
Robert	M	12	64.8	128.0
Sequan	M	15	67.0	133.0
Thomas	M	11	57.5	85.0
William	M	15	66.5	112.0

where β_1 is the amount we adjust each person's weight estimate up or down depending on his or her age and β_2 is the amount we adjust each person's weight estimate up or down depending on his or her height. Figure 6.3 depicts data for the weights, heights, and ages for these 19 students as a three-dimensional scatterplot. As landmarks, the graph denotes the points for Philip (72 inches tall, age 16, and 150 pounds) and Joyce (51.3 inches tall, age 11, and 50.5 pounds). It is easy to see that the points in Figure 6.3 are not randomly scattered but instead tend to fall along a plane. More importantly, it

FIGURE 6.3 Three-dimensional scatterplot of the relationship of age and height to weight

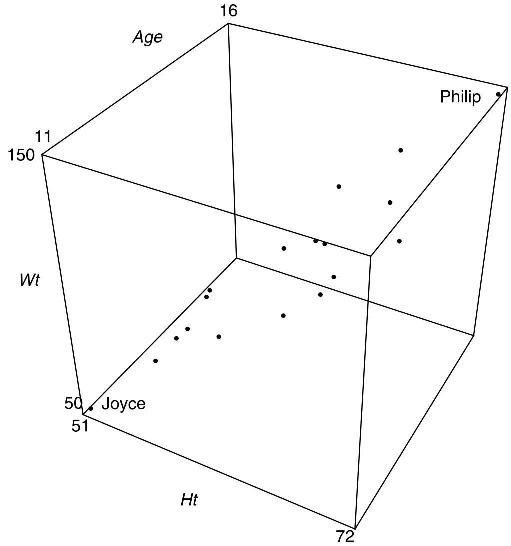
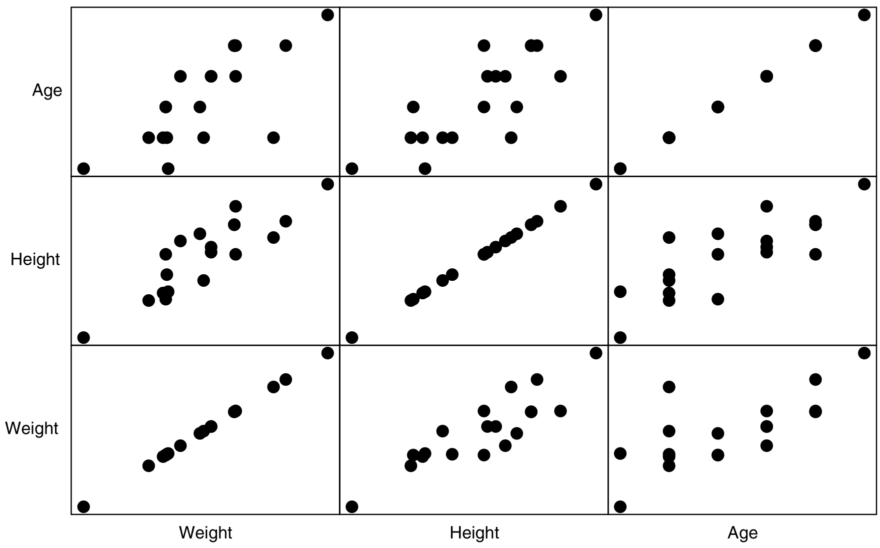


FIGURE 6.4 Matrix of two-dimensional scatterplot of the relationships among age, height, and weight



is easy to see that older, taller students tend to be heavier. Another common way to display data for multiple predictors, especially when more than two predictors precludes viewing the n-dimensional scatterplot, is a matrix of two-way scatterplots as depicted in Figure 6.4. In the last row we can see that taller students tend to be heavier and also that older students tend to be heavier. In the right-most scatterplot in the second row we can see that older students also tend to be taller; that is, height and age are redundant predictors. Because of the redundancy we cannot use the simple formulas. Any multiple regression program provides these estimates of the model parameters:

$$\hat{Wt}_i = -141.22 + 1.28Age_i + 3.60Ht_i$$

But what exactly do these partial regression coefficients mean? We will use a series of simple models and regressions to help us understand the meaning of these coefficients.

A Model of Weight

We begin by considering a simple model for the weights of these students. That is:

$$Wt_i = \beta_0 + \varepsilon_i$$

Of course, the best estimate for this simple model is the mean, so:

$$\hat{Wt}_i = \overline{Wt} = 100.03$$

The weights of the 19 students, arrayed along the x-axis in the same order as they are listed in Figure 6.2, are displayed in the left panel of Figure 6.5. The horizontal line at 100.03 represents the simple model for these data. The points for Philip, the heaviest student, and Joyce, the lightest student, are again labeled as landmarks. The errors or residuals are easily computed as:

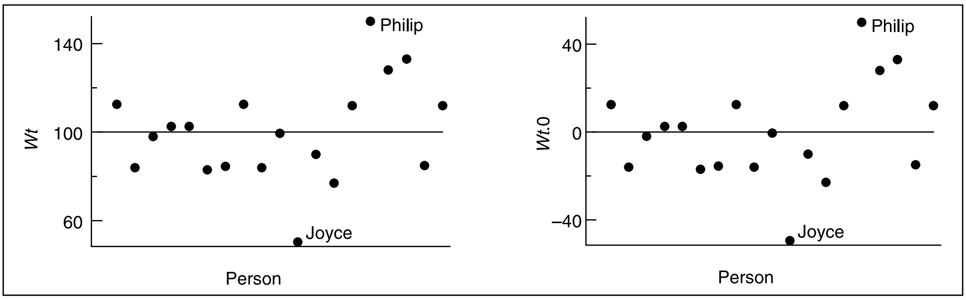
$$e_i = Wt_i - 100.03$$

For example, the errors for Philip and Joyce are, respectively, 151 – 100.03 = +50.97 and 50.5 – 100.03 = –49.53. We will be examining the errors for a number of different models for weight in this and the following sections so we introduce the notation $Wt.0$ to represent the errors from the simple model for weight. You can read "$Wt.0$" as "the part of weight that remains after using a model with β_0." The right panel of Figure 6.5 displays these values of $Wt.0$. A positive value (i.e., those points above the horizontal line at zero) of $Wt.0$ indicates that the student is heavier than the average student, and a negative value (i.e., those points below the line) indicates that the student is lighter than the average student.

A Model of *Wt.*0 using Age

The values of $Wt.0$ in the right panel of Figure 6.5 represent when a student's weight is unexpectedly or surprisingly heavy relative to our simple model of weight. If we are to improve on the simple model, then we need a variable that will help us predict when a student's weight is unusually heavy or light amongst this group of students. The variable Age is an obvious candidate. For reasons of consistent and subsequent interpretation we will construct $Age.0$ in the same way we constructed $Wt.0$. That is, a simple model of age is:

FIGURE 6.5 A simple model for weight



$$Age_i = \beta_0 + \varepsilon_i$$

and the best estimate of the model is the mean, or:

$$\hat{Age}_i = \bar{Age} = 13.32$$

Then we can compute the errors or residuals as:

$$e_i = Age.0_i = Age_i - 13.32$$

For example, *Age*.0 for Philip and Joyce, respectively, are 16 – 13.32 = 2.68 and 11 – 13.32 = –2.32. The value of *Age*.0 for Philip means that he is 2.68 years older than the typical student in this group and the negative value for Joyce means that she is 2.32 years younger than the typical student.

It is now natural to ask whether a student who is older than average also tends to be heavier than average. That is equivalent to a model using *Age*.0 to predict *Wt*.0, or:

$$Wt.0_i = \beta_1 Age.0_i + \varepsilon_i$$

Note that we have omitted the intercept β_0 in this model. We know that the means of both *Wt*.0 and *Age*.0 must be zero. Given that the regression line must go through the point defined by the two means, the intercept is necessarily zero. Hence, there is no reason to include the intercept in the model. Figure 6.6 presents all the values of *Wt*, *Wt*.0, *Age*, and *Age*.0 for the 19 students. Our question is then whether a student being older than average predicts being heavier than average? In other words, does *Age*.0 predict *Wt*.0? Figure 6.7 provides the scatterplot for these two variables. Clearly, as *Age*.0 increases, so does *Wt*.0. We can use the formula for the slope from Chapter 5 or a simple regression program to estimate the model (we urge the reader to do the computations for this and subsequent simple regressions in this section using a hand calculator or using a computer program). The resulting model is:

$$\hat{Wt}.0 = 11.31Age.0$$

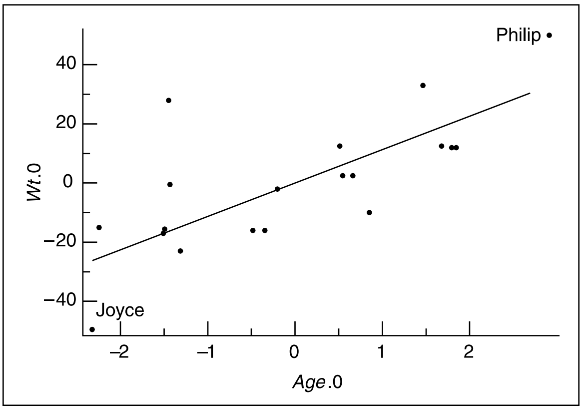
In other words, for each year older (or younger) than average, we expect the student to weigh 11.31 pounds more (or less).

For someone like Philip who is 16, or 2.68 years older than the average student, we expect him to weigh, according to the model, 11.31(2.68) = 30.31 pounds more than the average student. He in fact weighs 49.97 pounds more than the average student. Hence, the error after using both an intercept in the model and age as a predictor is $Wt.0,Age$ = 49.97 – 30.31 = 19.66 pounds. Another way to say this is that Philip, "adjusting for"

FIGURE 6.6 Values of *Wt*, *Wt*.0, *Age*, and *Age*.0

Name	*Wt*	*Wt.0*	*Age*	*Age.0*
Alfred	112.5	12.47	14	0.68
Antonia	84.0	−16.03	13	−0.32
Barbara	98.0	−2.03	13	−0.32
Camella	102.5	2.47	14	0.68
Henry	102.5	2.47	14	0.68
Jamal	83.0	−17.03	12	−1.32
Jane	84.5	−15.53	12	−1.32
Janet	112.5	12.47	15	1.68
Jayden	84.0	−16.03	13	−0.32
John	99.5	−0.53	12	−1.32
Joyce	50.5	−49.53	11	−2.32
Judy	90.0	−10.03	14	0.68
Louise	77.0	−23.03	12	−1.32
Mary	112.0	11.97	15	1.68
Philip	150.0	49.97	16	2.68
Robert	128.0	27.97	12	−1.32
Sequan	133.0	32.97	15	1.68
Thomas	85.0	−15.03	11	−2.32
William	112.0	11.97	15	1.68
Mean	100.03	0	13.32	0

FIGURE 6.7 Relationship between *Age*.0 and *Wt*.0



or “controlling for” his age, is 19.66 pounds heavier than expected. Similarly, Joyce, who is 2.32 years younger than the average student, is expected to weigh 11.31(–2.32) = –26.24 pounds lighter than the average student. However, she is in fact 49.53 pounds lighter so she weighs *Wt*.0,*Age* = –49.53 – (–26.24) = –23.29 pounds less than expected for her age. The values of *Wt*.0,*Age* for all 19 students are displayed in Figure 6.8 (due to rounding errors in the hand calculations, the values are slightly different from those computed above). Note that although most of the errors become smaller (i.e., *Wt*.0,*Age* < *Wt*.0), the errors for a few students become larger. For example, although John’s weight is close to the mean weight (i.e., his *Wt*.0 = –0.53), he is 14.35 pounds heavier than we would expect based on his relatively young age of 12 (i.e., his *Wt*.0,*Age* = 14.35).

FIGURE 6.8 SSE for simple and conditional models

Name	*Wt*	*Wt.0*	*Wt.0,Age*	*Ht.0,Age*
Alfred	112.5	12.47	4.74	4.76
Antonia	84.0	−16.03	−12.46	−4.96
Barbara	98.0	−2.03	1.54	3.84
Camella	102.5	2.47	−5.26	−1.44
Henry	102.5	2.47	−5.26	−0.74
Jamal	83.0	−17.03	−2.15	−1.37
Jane	84.5	−15.53	−0.65	−1.13
Janet	112.5	12.47	−6.56	−4.53
Jayden	84.0	−16.03	−12.46	1.04
John	99.5	−0.53	14.35	0.33
Joyce	50.5	−49.53	−23.35	−4.58
Judy	90.0	−10.03	−17.76	0.06
Louise	77.0	−23.03	−8.15	−2.37
Mary	112.0	11.97	−7.06	−0.53
Philip	150.0	49.97	19.63	2.18
Robert	128.0	27.97	42.85	6.13
Sequan	133.0	32.97	13.94	−0.03
Thomas	85.0	−15.03	11.15	1.62
William	112.0	11.97	−7.06	−0.53
Mean	100.03	0	0	0
SS		9335.74	4211.25	
PRE			0.55	

On the whole, the errors are smaller so the sum of squared errors decreases by making predictions conditional on age. In Figure 6.8, the sum of squares for $Wt.0$ (i.e., the SSE for a simple model making unconditional predictions) is 9335.74, but decreases to 4211.25 for $Wt.0,Age$ (i.e., the SSE for a model making predictions conditional on age); the proportional reduction in error is 0.55. In other words, PRE = .55 for this model comparison:

$$\text{MODEL A: } Wt_i = \beta_0 + \beta_1 Age + \varepsilon_i$$

$$\text{MODEL C: } Wt_i = \beta_0 + \varepsilon_i$$

Verifying this value for PRE using a simple regression program is a useful exercise for the reader.

A Model of *Wt*.0,*Age* using *Ht*.0,*Age*

Why is, for example, Philip heavier than expected and Joyce lighter than expected for their age? Perhaps Philip is unusually tall and Joyce is unusually short for their age. We might be tempted simply to add Height to the model. However, our intuitions and visual inspection of the height by age scatterplot suggests that height and age are partially redundant. That is, height and age share some of the same information that might be used to predict weight. The only useful part of the height information is the part that is not redundant with age. We can easily find that part by using age to predict height in this model:

$$Ht = \beta_0 + \beta_1 Age + \varepsilon_i$$

and computing the residuals:

$$e_i = Ht.0, Age_i = Ht_i - \hat{Ht}$$

A simple regression reveals that:

$$\hat{Ht}_i = 25.22 + 2.79 Age_i$$

In other words, we predict that these students grow about 2.79 inches each year. In particular, the model predicts Philip's height as 25.22 + 2.79(16) = 69.9 inches, but he is actually 72 inches tall. Hence, he is indeed unusually tall for his age by $Ht.0, Age$ = 72 – 69.9 = 2.1 inches. Because he is taller for his age, we might want to consider a model in which we adjust his predicted weight upward. Joyce, based on her age of 12, is expected to be 25.22 + 2.79(11) = 55.9 inches tall. However, she is only 51.3 inches tall, or $Ht.0, Age$ = 51.3 – 55.9 = –4.6 inches shorter than we expected for her age. And so perhaps her weight prediction should be adjusted downward. In general, we want to ask whether students who are taller for their age (or when "adjusting for" or "controlling for" age) tend to be heavier than expected for their age. In other words, we are asking whether there is a relationship between $Ht.0, Age$ (the part of height that is unrelated to age) and $Wt.0, Age$ (the part of weight that is unrelated to age). The values for $Ht.0, Age$ for all 19 students are listed in Figure 6.8. We want to consider the model:

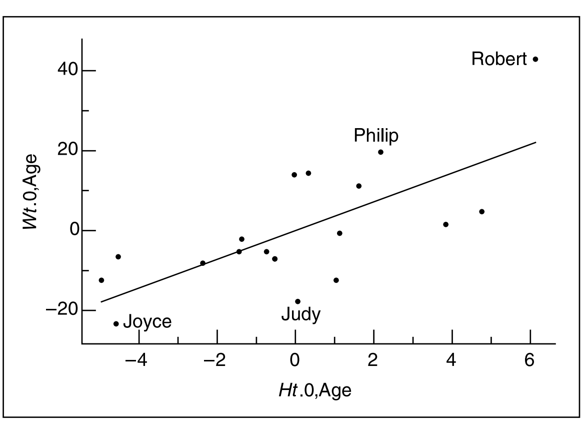
$$Wt.0, Age_i = \beta_1 Ht.0, Age_i + \varepsilon_i$$

Figure 6.9 shows the data to answer this question in a scatterplot. Indeed, those students who are taller than expected for their age ($Ht.0, Age$ is high) tend to be heavier than expected for their age ($Wt.0, Age$ is also high). The estimated model is:

$$\hat{Wt}.0, Age_i = 3.6 Ht.0, Age_i$$

As with any slope, this means that as $Ht.0, Age$ increases by one unit, the prediction of $Wt.0, Age$ increases by 3.6 pounds. But let us be more precise, even though it is a bit wordy: For each inch students are taller than expected for their age, we expect them to weigh an additional 3.6 pounds more than expected for their age. Another way to say this is that after adjusting height and weight for age, each additional inch predicts an increase of 3.6 pounds. Note that for those students for whom their heights were exactly

FIGURE 6.9 Relationship between *Ht*.0,*Age* and *Wt*.0,*Age*



what we expected based on their age (i.e., $Ht.0,Age = 0$), we would expect them to weigh exactly what we expected based on their age (i.e., $\hat{Wt}.0,Age = 0$). Philip is 2.1 inches taller than expected for his age so we expect him to weigh 3.6(2.1) = 7.6 pounds more than expected for his age, but he actually weighs 19.6 pounds more than expected for his age. Or, Philip weighs $Wt.0,Age,Ht = 19.6 - 7.6 = 12$ pounds more than we would expect based on a model using both age and height. Similarly, Joyce, 4.6 inches shorter than expected for her age, weighs $Wt.0,Age,Ht = -23 - 3.6(-4.6) = -23 - (-16.6)$ $= -6.4$ pounds less than we would expect based on her age and height. When we just considered weight, Philip and Joyce were by far the heaviest and lightest students, respectively; naively we may have even thought that they were unhealthily overweight and underweight because they were so far from the average weight of these students. However, after accounting for (or adjusting for) their age and height, both Philip and particularly Joyce are close to the average weight expected. On the other hand, Robert and Judy, both marked in Figure 6.9, neither of whom appeared extreme when we only considered average weight, are the furthest from the weights we would expect given their heights and ages. Note that Robert is unusually tall for his age of 12, but he is still far heavier (by almost 21 pounds) than we would expect because of his extra height. Judy is almost exactly the height we would expect for her age of 14, but she is about 16 pounds lighter than we would expect for her combination of age and height. In other words, it is only after adjusting our weight predictions for age and height that we are able to see who is truly overweight (far above the regression line in Figure 6.9) or underweight (far below the regression line).

For each of the 19 students, Figure 6.10 lists their weights and their residuals from the various models we have considered so far. As a touchstone, let us consider the row

FIGURE 6.10 Students' weights and residuals for each of three models

Name	*Wt*	*Wt.0*	*Wt.0,Age*	*Wt.0,Age,Ht*
Alfred	112.5	12.47	4.74	−12.38
Antonia	84.0	−16.03	−12.46	5.38
Barbara	98.0	−2.03	1.54	−12.27
Camella	102.5	2.47	−5.26	−0.08
Henry	102.5	2.47	−5.26	−2.60
Jamal	83.0	−17.03	−2.15	2.78
Jane	84.5	−15.53	−0.65	−4.71
Janet	112.5	12.47	−6.56	9.73
Jayden	84.0	−16.03	−12.46	−16.20
John	99.5	−0.53	14.35	13.16
Joyce	50.5	−49.53	−23.35	−6.88
Judy	90.0	−10.03	−17.76	−17.98
Louise	77.0	−23.03	−8.15	0.37
Mary	112.0	11.97	−7.06	−5.15
Philip	150.0	49.97	19.63	11.79
Robert	128.0	27.97	42.85	20.80
Sequan	133.0	32.97	13.94	14.05
Thomas	85.0	−15.03	11.15	5.32
William	112.0	11.97	−7.06	−5.15
Mean	100.03	0	0	0
SS		9335.74	4211.25	2121.04
PRE			0.55	0.50

of values for Philip. His weight of 150 pounds is about 50 (*Wt*.0) pounds heavier than average, about 19.6 (*Wt*.0,*Age*) pounds heavier than average for students of his age, and only about 12 (*Wt*.0,*Age*,*Ht*) pounds heavier than we would expect for someone of his age and height. Thus, the deviation or error between Philip's actual weight (the data) and our model of weight steadily decreases from 50 to 20 to 12. On the whole, although there are exceptions (such as Judy), the errors decrease as more predictors are included in the model (i.e., *Wt*.0,*Age* < *Wt*.0,*Age*,*Ht*). Most importantly, the sum of squared errors steadily decreases. As we saw before, 55% of the squared error for the simple model is reduced when we add age. Now, 50% of the squared error for the model with age is reduced when we add the part of height not redundant with age. Overall, relative to the simple model of weight, using both age and height reduces error by 77%:

$$\text{PRE} = \frac{9335.74 - 2121.04}{9335.74} = .77$$

of the original squared error. We will return to these values of PRE for different model comparisons when we consider statistical inference for multiple regression models later in this chapter.

Conclusion: Interpretation of Partial Regression Coefficients

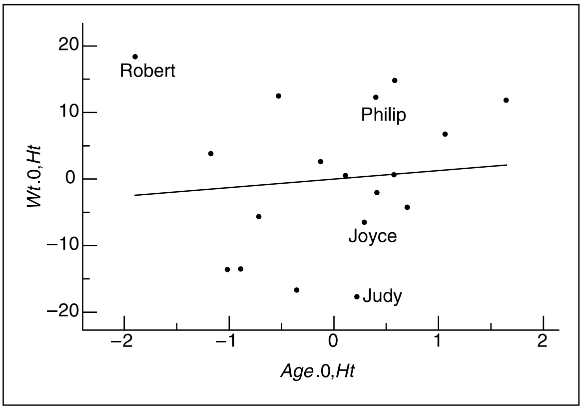
The goal of this series of simple regressions using residuals is to better understand regression coefficients. Now that we have a good understanding of the data from this detailed examination, it is time to return to the original goal. Earlier we noted that the estimate of the model from multiple regression is:

$$\hat{Wt}_i = -141.22 + 1.28Age_i + 3.60Ht_i$$

The coefficient of 3.6 for *Ht* is *exactly* the same as the slope in Figure 6.9 where it is the coefficient predicting *Wt*.0,*Age* from *Ht*.0,*Age*. Therefore, the meaning of the 3.6 in the multiple regression equation is exactly the same as above: for each inch taller students are for their age, we expect them to be 3.6 pounds heavier for their age. Or we might say, after adjusting both height and weight for age, for each inch taller, a student is expected to weigh 3.6 pounds more. In other words, the coefficient of 3.6 in the multiple regression model describes the relationship between that part of height that is not redundant or shared with age and that part of weight that is not redundant or shared with age. It is common in reports of multiple regression to write something simple like: "When controlling for age, weight increases 3.6 pounds for each additional inch of height." However, it is important to remember the more precise meaning of the regression coefficient that we developed in the previous section and to remember that it is the slope in the scatterplot of two residuals in Figure 6.9. Statisticians and statistics programs often refer to plots like Figure 6.9 as *partial regression plots* because of their basis for defining the partial regression slope or coefficient. The wise data analyst will routinely examine partial regression plots, whose many benefits include (a) a visual representation of the regression coefficient and (b) visual identification of any unusual patterns or observations (such as Robert and Judy).

To interpret the coefficient of 1.28 for age in the multiple regression equation, we do not need to do the complete series of simple regressions parallel to those that we did

FIGURE 6.11 Partial regression plot between *Age*.0,*Ht* and *Wt*.0,*Ht*



above. Instead, we can reverse the role of height and age in the above statements interpreting the coefficient for height and we can examine the computer-generated partial regression plot in Figure 6.11 relating *Age*.0,*Ht* to *Wt*.0,*Ht*. Thus, for each year older students are than we would guess from their heights, we expect them to be 1.28 pounds heavier for their height. Or we might say, after adjusting both age and weight for height, for each year older, a student is expected to weigh 1.28 pounds more. Still other language would be: After statistically equating two students on height or "holding height constant," we expect the student older by one year to weigh only 1.28 pounds more than the younger student. In that context, 1.28 pounds does not seem like much; indeed, the slope in Figure 6.11 appears as if it might not differ significantly from a slope of zero (we will check this later).

Note that, as must be the case, Robert and Judy have the largest residuals (i.e., greatest distance from the regression lines in both of the partial regression plots) because they are unexpectedly heavy and light, respectively, given their heights and ages. And again, Philip and Joyce are closer to their expected weights (i.e., closer to the regression lines) based on their heights and ages. In other words, *Wt*.0,*Age*,*Ht* = *Wt*.0,*Ht*,*Age*. Although overall predictions of weight have improved, considerable error remains. We might now ask whether there is a variable that could predict when students are unusually heavy (or light) for their heights and ages. Looking at the names above the regression line (i.e., unusually heavy for their height and ages) and those below (i.e., unusually light for their height and ages) suggests an obvious variable to consider.

STATISTICAL INFERENCE IN MULTIPLE REGRESSION

Now that we understand the meaning of the partial regression coefficients, the remaining issue is how to determine whether those coefficients are significantly different from zero and to ask other model comparison questions involving models with multiple predictors. The general strategy for doing statistical inference or asking questions about data is exactly the same as before. In particular, the calculations of PRE and F pose no new problems as they are defined exactly as before. The only change is that the extra parameters in the multiple regression equation give us lots of freedom in defining Model