24

MULTIPLE LINEAR REGRESSION

24.1 INTRODUCTION AND OBJECTIVES

In the previous chapter, simple linear regression was used when you have one independent variable and one dependent variable. This chapter presents multiple linear regression, which is used when you have two or more independent variables and one dependent variable. The research question for those using multiple regression concerns how the multiple independent variables, either by themselves or together, influence changes in the dependent variable. You use the same basic concepts as with simple linear regression, except that now you have multiple independent variables. The object of multiple linear regression is to develop a *prediction equation* that permits the estimation of the value of the dependent variable based on the knowledge of multiple independent variables.

The data requirements for multiple linear regression are the same as for simple linear regression. Sample size is always an issue with statistical methods, and the same is true for regression. The sample size in this example is small to facilitate the visualization and input of data. One should keep in mind, though, that larger samples are usually better when performing most statistical analysis. The multiple regression example used in this chapter is as basic as possible—small sample size and only two independent variables, the minimum number required when using multiple regression. The reader is asked to be cognizant of the fact that more independent variables (and a larger sample size) could be analyzed using exactly the same techniques described in this chapter.

OBJECTIVES

After completing this chapter, you will be able to

Describe the purpose of multiple linear regression

Input variable information and data for multiple linear regression

- Describe the data assumptions required for multiple linear regression
- Use SPSS to conduct multiple linear regression analysis
- Interpret scatterplots concerning the data assumptions for regression
- Interpret probability plots concerning the data assumptions for regression
- Describe and interpret the SPSS output from multiple linear regression
- Interpret ANOVA analysis as it relates to multiple linear regression
- Interpret the coefficients table for multiple linear regression
- Write the prediction equation and make predictions using SPSS and the calculator

24.2 RESEARCH SCENARIO AND TEST SELECTION

The researcher wants to understand how certain physical factors may influence an individual's weight. The research scenario centers on the belief that an individual's "height" and "age" (independent variables) are related to the individual's "weight" (dependent variable). Another way of stating the scenario is that age and height *influence* the weight of an individual. When attempting to select the analytic approach, an important consideration is the level of measurement. As with simple regression, the dependent variable must be measured at the scale level (interval or ratio). The independent variables are almost always continuous, although there are methods to accommodate discrete variables. In the example presented above, all data are measured at the scale level. What type of statistical analysis would you suggest to investigate the relationship of height and age to a person's weight?

Regression analysis comes to mind since we are attempting to estimate (predict) the value of one variable based on the knowledge of the others, which can be done with a prediction equation. Simple regression can be ruled out since we have two independent variables and one dependent variable. Let's consider multiple linear regression as a possible analytic approach.

We must first check to see if our variables are approximately normally distributed. Furthermore, it is required that the relationship between the variables be approximately linear. And we also look for *homoscedasticity*, which means that the variances in the dependent variable are the same for each level of the independent variables. Here's an example of homoscedasticity. A distribution of individuals who are 61 inches tall and aged 41 years would have the same variability in weight as those who are 72 inches tall and aged 31 years. In the sections that follow, some of these required data characteristics will be examined immediately, others when we get deeper into the analysis. Let's get started!

24.3 RESEARCH QUESTION AND NULL HYPOTHESIS

The null hypothesis (H_0) is that there is not a statistically significant relationship between an individual's weight and that person's age and height. The alternative hypothesis (H_A) states the opposite: There is a statistically significant relationship between an individual's weight and that person's age and height. Therefore, this research question involves two independent variables, "height" and "age," and one dependent variable, "weight." The investigator wishes to determine how height and age, taken together or individually, might explain the variation in an individual's weight. Such information could assist someone attempting to estimate an individual's weight based on the knowledge of his or her height and age. Another way of stating the research question uses the concept of prediction and error reduction. How successfully could we predict an individual's weight given that we know his or her age and height? How much error could be reduced in making the prediction when age and height are known? One final question: Is the relationship between weight and each of the two independent variables (age and height) statistically significant or is any observed relationship due to chance?

24.4 DATA INPUT

In this section, you enter hypothetical data for 12 randomly selected individuals measured on weight, height, and age. You then use SPSS to analyze these data using multiple linear regression. As in recent chapters, detailed instructions on entering the variable information and the data are not given. The Variable View screen, shown in Figure 24.1, serves as a guide for the entry of variable information. Figure 24.1 contains the material needed to successfully enter all the variable information. All data and information needed to conduct the analysis can be found on the companion website at *weight_expl_chap24.sav* but with restricted access.

Figure 24.2 contains the data for the three variables on the 12 individuals. The table in Figure 24.2 is a copy of the Data View screen and therefore shows exactly what your data entry should look like.

Follow the bullet points below and enter both the variable information and the data for the three variables; save the file as instructed.

- Start SPSS and click **Cancel** in the *SPSS Statistics* opening window.
- Click **File**, select **New**, and click **Data**.
- Click **Variable View** (enter all the variable information as presented in Figure 24.1).
- Click **Data View** (carefully enter all the data for *weight*, *height*, and *age* given in Figure 24.2).
- Click **File**, then click **Save As**; type *weight_expl_chap24.sav* in the *File Name* box, and then click **OK**.

FIGURE 24.1 ⬢ VARIABLE VIEW FOR MULTIPLE LINEAR REGRESSION FOR THREE VARIABLES

*weight.sav [DataSet3] - IBM SPSS S

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	weight	Numeric	8	0	Weight in pounds	None	None	8	Center	Scale	Input
2	height	Numeric	8	0	Height in inches	None	None	8	Center	Scale	Input
3	age	Numeric	8	0	Age in years	None	None	8	Center	Scale	Input

FIGURE 24.2 ⬢ DATA VIEW FOR MULTIPLE REGRESSION FOR THREE VARIABLES

	weight	height	age
1	115	62	41
2	140	62	21
3	125	62	31
4	125	64	21
5	145	64	31
6	135	64	41
7	165	72	41
8	190	72	31
9	175	72	21
10	150	66	31
11	155	66	31
12	140	64	21

You have now entered and saved the data for an individual's weight, height, and age. In the next section, we check the distributions for normality.

24.5 DATA ASSUMPTIONS (NORMALITY)

Watch the tutorial video at study.sagepub.com/aldrich3e

As was done in the previous chapter, we first check the data distributions for normality.

- Click **Analyze**, select **Nonparametric Tests**, and then click **One-Sample** (the *One-Sample Nonparametric Tests* window opens).
- Click the **Objective** tab, and then click **Customize analysis**.

- Click the **Fields** tab. (If your three variables are not in the *Test Fields* pane, then move them to it.)
- Click the **Settings** tab, click **Customize tests**, and then click **Kolmogorov-Smirnov test**.
- Click **Options**, make sure **Normal** is checked, then click **OK**.
- Click **Run** (the Output Viewer opens showing Figure 24.3).

The results of the Kolmogorov–Smirnov (K–S) test indicate that two of the three variables are indeed normally distributed (see Figure 24.3). "Weight" and "Age" pass the test even when subjected to the *Lilliefors* correction factor, which is automatically applied to the K–S test by the SPSS program. It was decided to take a closer look at the variable "Height," which did not pass the K–S test for normality. The first "look" was to do a P–P Plot, which looked fairly good. However, the appearance of the "Height" data in a histogram was not so good, but the mean (65.8) and standard deviation (3.95) were encouraging. It was next decided to conduct the K–S test without the Lilliefors correction. This procedure reduces the power of the K–S test to detect departures from normality, but the judgment call was to proceed. There are some who may criticize this approach (reducing the power of the K–S test); however, we did have confidence that the slight departure from normality for height would not negatively impact the prediction equation's performance. The bullet points needed to accomplish this procedure, which would improve the chances of the variable "Height" passing the K–S test, are given next.

FIGURE 24.3 ● THE KOLMOGOROV–SMIRNOV TEST FOR NORMALITY—THREE VARIABLES

Hypothesis Test Summary

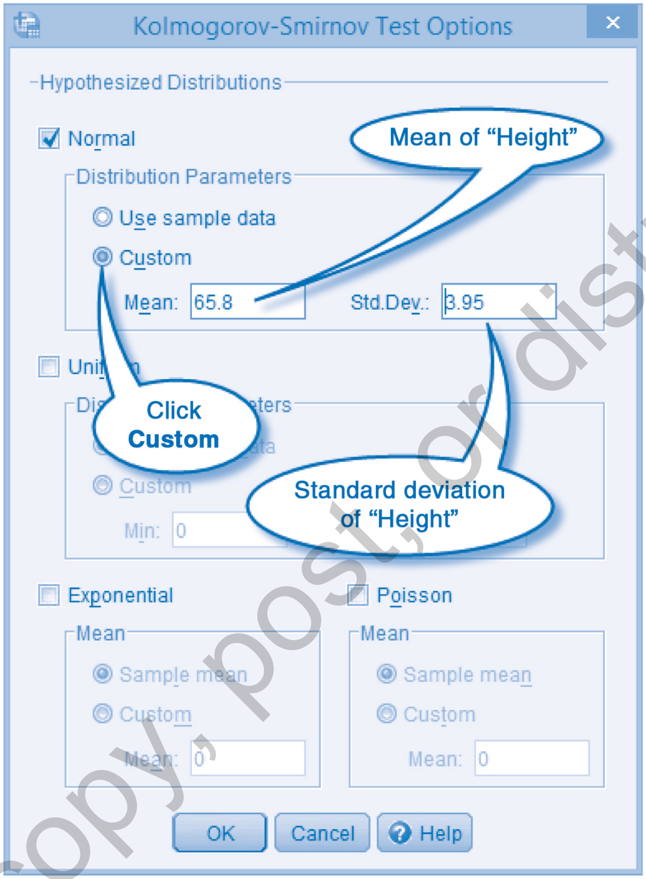
	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Weight in pounds is normal with mean 147 and standard deviation 21.881.	One-Sample Kolmogorov-Smirnov Test	.200[1,2]	Retain the null hypothesis.
2	The distribution of Height in inches is normal with mean 66 and standard deviation 3.950.	One-Sample Kolmogorov-Smirnov Test	.022[1]	Reject the null hypothesis.
3	The distribution of Age in years is normal with mean 30 and standard deviation 7.930.	One-Sample Kolmogorov-Smirnov Test	.153[1]	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

[1] Lilliefors Corrected

[2] This is a lower bound of the true significance.

FIGURE 24.4 ⬢ THE *KOLMOGOROV–SMIRNOV TEST OPTIONS* WINDOW



- Click **Analyze**, select **Nonparametric Tests**, and then click **One-Sample** (the *One-Sample Nonparametric Tests* window opens).
- Click the **Objective** tab, and then click **Customize analysis**.
- Click the **Fields** tab (make sure that *only* the variable "Height" is in the *Test Fields* pane).
- Click the **Settings** tab, click **Customize tests**, and then click **Kolmogorov-Smirnov Test**.
- Click the **Options** button just below **Kolmogorov-Smirnov Test**, and the *Kolmogorov-Smirnov Test Options* window opens (see Figure 24.4). Make sure **Normal** is clicked.
- In the *Distribution Parameters* pane, click **Custom**; in the *Mean* box, type **65.8**; and in the *Std. Dev.* box, type **3.95** (these values are easily obtained from the prior

FIGURE 24.5 ⬢ KOLMOGOROV–SMIRNOV TEST FOR "HEIGHT" WITHOUT LILLIEFORS CORRECTION

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Height in inches is normal with mean 65.800 and standard deviation 3.95.	One-Sample Kolmogorov-Smirnov Test	.396	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

K–S test shown in Figure 24.3) (the window should now look like Figure 24.4). Finally, click **OK**.

- Click **Run** (the Output Viewer opens, showing Figure 24.5).

Using the K–S test without the Lilliefors correction is successful, and our "height" variable passes the normality test. Based on this finding we make the decision to proceed with the regression analysis.

24.6 REGRESSION AND PREDICTION

As was done with the simple linear regression in Chapter 23, the data are next checked for linearity, equal variances, and normality of the error terms (residuals). If it's not already running, open *weight_expl_chap24.sav*, and follow the procedure presented next. Once you have completed all these analytic requests, you will see the output as presented in the following sections.

- Click **Analyze**, select **Regression**, and then click **Linear** (the *Linear Regression* window opens; see Figure 24.6 for its appearance after moving the variables).
- Click **Weight**, and then click the **arrow** next to the *Dependent:* box.
- Click **Height**, and then click the **arrow** next to the *Independent(s):* box.
- Click **Age**, and then click the **arrow** next to the *Independent(s):* box (at this point, your screen should look like Figure 24.6).
- Click **Statistics** (the *Linear Regression: Statistics* window opens; see Figure 24.7).
- Click **Estimates**, and then click **Model fit** (see Figure 24.7).
- Click **Continue** (returns to the *Linear Regression* window depicted in Figure 24.6).
- Click **Plots** (the *Linear Regression: Plots* window opens; see Figure 24.8; actually, this is the same analytic request you made when doing simple regression).

FIGURE 24.6 ● *LINEAR REGRESSION* WINDOW AFTER MOVING THE VARIABLES

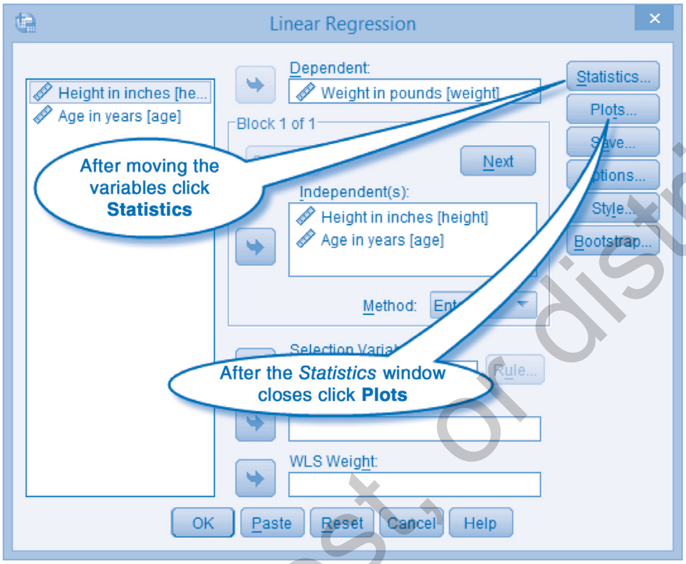
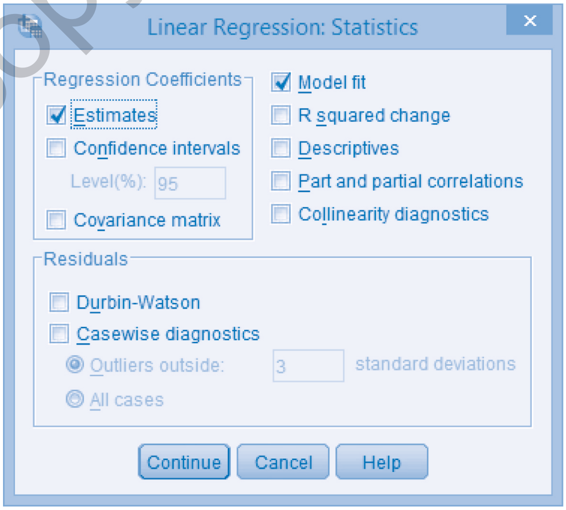


FIGURE 24.7 THE *LINEAR REGRESSION: STATISTICS* WINDOW



- Click ***ZPRED**, and then click the **arrow** beneath the *Y:* box.
- Click ***ZRESID**, and then click the **arrow** beneath the *X:* box.
- Click **Normal probability plot**.
- Click **Continue**, and then click **OK** (this final click produces all the output required to interpret our analysis).

You now have all the output required to finalize and interpret additional data assumptions and your multiple linear regression analysis.

24.7 INTERPRETATION OF OUTPUT (DATA ASSUMPTIONS)

Figure 24.8 shows that the *Normal probability plot* box was checked. The result is shown in Figure 24.9. The data markers are close to the diagonal line, which provides evidence that the *residuals* (error terms) are indeed normally distributed, which is a requirement of the linear regression procedure.

The final plot, Figure 24.10, results from the requests we made as shown in Figure 24.8. As with simple variable regression, the scatterplot combines the standardized predicted values (*ZPRED) with the values for the standardized residuals (*ZRESID). Since the data markers follow no pattern—they are randomly dispersed in the scatterplot—we assume equality

FIGURE 24.8 ⬢ THE *LINEAR REGRESSION: PLOTS* WINDOW: FURTHER DATA ASSUMPTION CHECKS

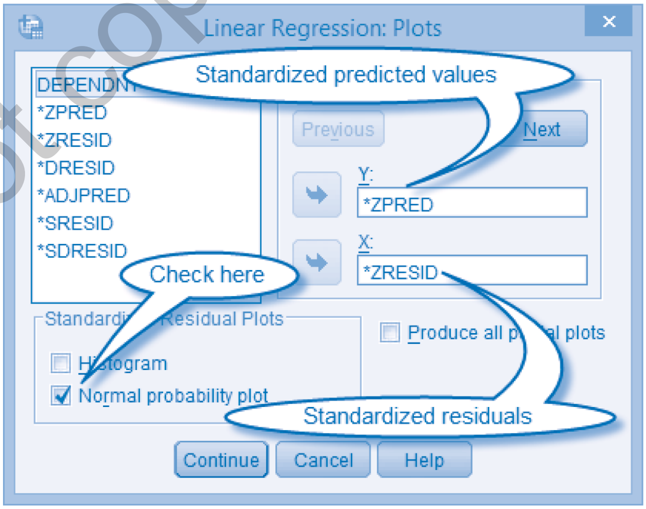


FIGURE 24.9 ⬢ *NORMAL P–P PLOT OF REGRESSION STANDARDIZED RESIDUAL* (ERROR TERMS FOR WEIGHT)

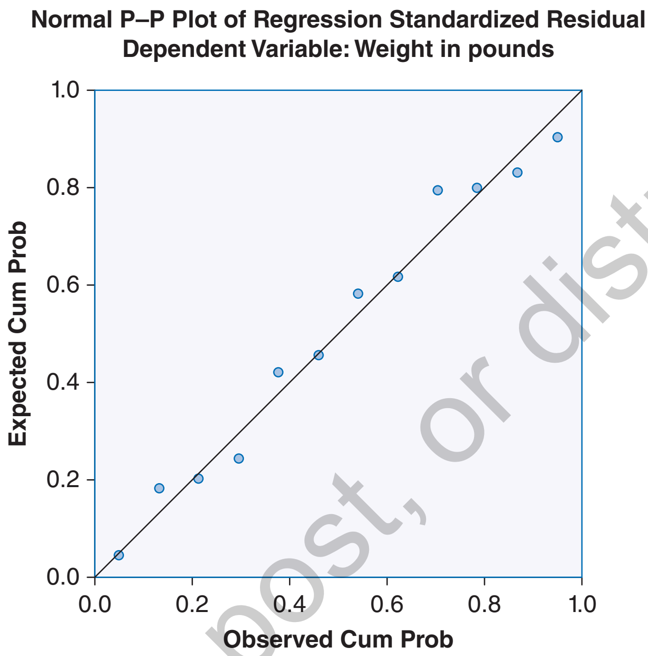


FIGURE 24.10 ⬢ SCATTERPLOT OF RESIDUALS FOR WEIGHT: LACK OF PATTERN INDICATES EQUAL VARIANCES

of variances. There are numerous appearances that Figure 24.10 may take on that would indicate unequal variances. One such appearance is referred to as the "bow tie" scatterplot. The bow tie scatterplot has the error terms bunched up along both verticals and tapering toward the middle, which is the zero point in Figure 24.10. There are many other shapes that indicate unequal variances.

24.8 INTERPRETATION OF OUTPUT (REGRESSION AND PREDICTION)

The *Model Summary* shown in Figure 24.11 resulted from you clicking **Model Fit**, as depicted in Figure 24.7. The information provided in the *Model Summary* gives us information regarding the strength of the relationship between our variables.

The value .919 shown in the *R* column of the table in Figure 24.11 shows a *strong* multiple correlation coefficient. It represents the correlation coefficient when both independent variables ("age" and "height") are taken together and compared with the dependent variable "weight." The *Model Summary* indicates that the amount of change in the dependent variable is determined by the two independent variables—not by one as in simple regression. From an "interpretation" standpoint, the value in the next column, *R Square*, is extremely important. The *R Square* of .845 indicates that 84.5% (.845 × 100) of the variance in an individual's "weight" (dependent variable) can be explained by both the independent variables, "height" and "age." It is safe to say that we have a "good" predictor of weight if an individual's height and age are known. We next examine the *ANOVA* table shown in Figure 24.12.

The *ANOVA* table indicates that the mathematical model (the regression equation) can accurately explain variation in the dependent variable. The value of .000 (which is less than .05) provides evidence that there is a low probability that the variation explained by the model is due to chance. We conclude that changes in the dependent variable result from changes in the independent variables. In this example, changes in height and age resulted in statistically significant changes in weight.

FIGURE 24.11 ⬢ *MODEL SUMMARY* FOR MULTIPLE LINEAR REGRESSION

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.919[a]	.845	.811	9.515

a. Predictors: (Constant), Age in years, Height in inches

b. Dependent Variable: Weight in pounds

FIGURE 24.12 ● *ANOVA* TABLE INDICATING A SIGNIFICANT RELATIONSHIP BETWEEN THE VARIABLES

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4451.886	2	2225.943	24.588	.000[b]
	Residual	814.780	9	90.531		
	Total	5266.667	11			

a. Dependent Variable: Weight in pounds

b. Predictors: (Constant), Age in years, Height in inches

24.8.1 Prediction

As in the prior chapter, we come to that interesting point in regression analysis where we seek to discover the unknown. We accomplish such a "discovery" by using a prediction equation to make estimates based on our 12 original observations. Let's see how the *Coefficients* table in Figure 24.13 can assist us in making such discovery possible.

As with simple linear regression, the *Coefficients* table shown in Figure 24.13 provides the essential values for the prediction equation. The prediction equation for this example takes the following form:

$$\hat{y} = a + b_1x_1 + b_2x_2,$$

where $\hat{y}$ is the predicted value, a the intercept, b_1 the slope for "height," x_1 the independent variable "height," b_2 the slope for "age," and x_2 the independent variable "age."

You may recall that in the previous chapter on simple regression, we defined the slope (b) and the intercept (a). Those definitions are the same for the multiple regression equation, except that we now have a slope for each of the independent variables.

The equation simply states that you multiply the individual slopes by the values of the independent variables and then add the products to the intercept—not too difficult. The slopes and intercepts can be found in the table shown in Figure 24.13. Look in the column labeled *B*. The intercept (the value for a in the above equation) is located in the (*Constant*) row and is –175.175. The value below this of 5.072 is the slope for "height," and below that is the value of –.399, the slope for "age." The values for x are found in SPSS's *Data View* for the dataset *weight_expl_chap24.sav*. Substituting the regression coefficients, the slope and the intercept, into the equation, we find the following:

$$\hat{y} = 175.175 + (5.072 * height) + (-.399 * age).$$

At this stage of the analysis you are ready to use the *Compute Variable* function of SPSS and make some predictions. The following bullet points assume that SPSS is running and the *weight* dataset is open.

- Click **Transform**, and then click **Compute Variable** (the *Compute Variable* window opens; see Figure 24.14, which shows the upper portion of the window, with the completed operations as described in the following bullet points).
- Click the **Target Variable** box, and then type *pred_weight*.
- Click the **Type & Label** box, and a window opens; then type *predicted weight* in the *Label* box.
- Click **Continue** (the *Compute Variable* window remains as shown in Figure 24.14).

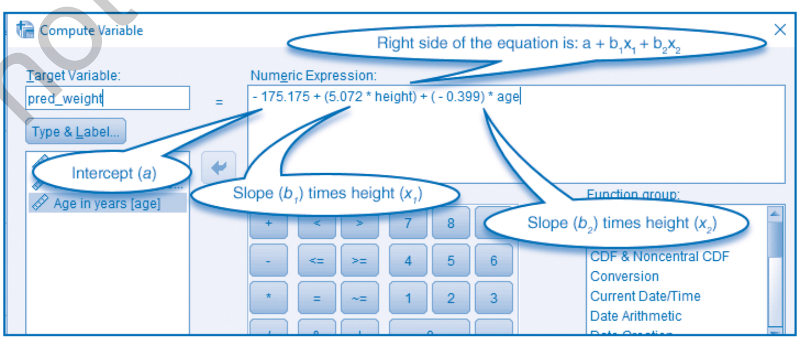
FIGURE 24.13 ● THE *COEFFICIENTS* TABLE SHOWING THE VALUES USED IN THE PREDICTION EQUATION

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-175.175	48.615		-3.603	.006
	Height in inches	5.072	.727	.916	6.974	.000
	Age in years	-.399	.362	-.145	-1.103	.299

a. Dependent Variable: Weight in pounds

FIGURE 24.14 ● COMPLETED *COMPUTE VARIABLE* WINDOW—PREDICTION EQUATION



- Click the **Numeric Expression** box, and type *−175.175 + (5.072 ∗ height) + (−.399 ∗ age)* (if preferred, you could use the keypad in the *Compute Variable* window and the variable list to write this expression—try it, you may find it easier).
- Click **OK** at the bottom of this window (the bottom portion of the window is not shown in Figure 24.14).

The new variable "pred_weight" is automatically inserted into the *weight* dataset once you click **OK** in the above series of bullet points. If the weight dataset (containing the new variable called pred_weight) does not appear, then minimize the Output Viewer screen. This will then show your new dataset containing the predicted values resulting from the application of the regression equation.

Figure 24.15 shows the new Data View screen with the just-created variable. Let's look at *Case 10* and interpret the newly created variable. *Case 10* shows that someone measured at 66 inches in height and aged 31 weighed 150 pounds. The prediction equation estimated that an individual possessing those two values (66 and 31) would weigh 147.21 pounds. If you read each of the cases for height and age, you can read the actual observed weight as well as the prediction.

As we did for simple regression, any values for x_1 and x_2 that might be of special interest could be plugged into the equation and then solved for the predicted y value. One note of caution is that the values chosen should be within the range of the original observations to maintain accuracy. You can accomplish such predictions by substituting values into the *Numeric expression* panel in the *Compute Variable* window and creating a new variable as described in the previous chapter. The other method is to use a handheld calculator, as was also done in the previous chapter.

24.9 RESEARCH QUESTION ANSWERED

At the beginning of this chapter, it was stated that the purpose of the research was to investigate whether a person's weight is influenced by his or her age and height. You might also recall that the null hypothesis was that age and height had no influence on a person's weight. How well did our multiple regression analysis answer these questions? And could we reject the null hypothesis and thereby provide evidence in support of our alternative hypothesis?

First, our questions concerning the required data assumptions for using multiple regression were answered in the affirmative. It was determined that multiple linear regression could be used with the randomly selected data provided for the analysis (see Figures 24.3, 24.5, 24.9, and 24.10). Next, the prediction equation, which was developed from previous observations, was found to reduce the error in predicting weight by 84.5% (see Figure 24.11). Additional statistical evidence supporting the value of our prediction equation was provided with the finding of a significant *F* test. The significance was less than .05, indicating a low probability that the explanation of the variation in weight, when using age and height, was the result of chance (see Figure 24.12). Additional empirical evidence in support of the prediction equation was observed when the *Compute Variable* function of SPSS was used to calculate predicted values. These predicted values then compared favorably with our observations (see Figure 24.15).

FIGURE 24.15 ● DATA VIEW SHOWING A NEW VARIABLE FROM USING THE PREDICTION EQUATION

	weight	height	age	pred_weight
1	115	62	41	122.93
2	140	62	21	130.91
3	125	62	31	126.92
4	125	64	21	141.05
5	[illegible]	[illegible]	31	137.06
[illegible]	[illegible]	[illegible]	41	133.07
7	[illegible]	[illegible]	41	173.65
8	190	72	31	177.64
9	175	72	21	181.63
10	150	66	31	147.21
11	155	66	31	147.21
12	140	64	21	141.05

New variable—results from the *Compute Variable* request

24.10 Summary

In this chapter, multiple linear regression was presented. With multiple linear regression, you have two or more independent variables and one dependent variable. The object was to write a prediction equation that would permit the estimation of the value of the dependent variable based on the knowledge of two or more independent variables. In the next chapter, you learn about a third type of regression known as logistic regression. With logistic regression, you attempt to predict a binary (categorical) dependent variable.

24.11 Review Exercises

Answers for Exercises 1, 2, and 3 can be found in Appendix C while answers for 4 and 5 can be found on the professor-facing portion of the companion website (study.sagepub.com/aldrich3e).

24.1 This exercise is an extension of the senior center manager's problem in the previous chapter (Review Exercise 23.1). You may recall that the manager developed a prediction equation that estimated the number of books checked out at the library using the "patrons' age" as the *single independent* variable. For the current exercise, used to illustrate multiple linear regression, we add a second independent variable—"total years of education." Using the single variable, the model developed was able to account for 86% of the variance in the number of books checked out. Although the senior center manager was happy with that result, she wishes to add total years of education in the hope of improving her model. The manager wants to use a new equation (using two independent variables) to make predictions and then compare those predictions with the observed data to see how well it works. She also wishes to predict the number of books checked out by someone aged 63 with 16 years of education, which was not directly observed in her data. Use multiple linear regression, write the null and alternative hypotheses, conduct the analysis, write the prediction equations, make the predictions, and interpret the results. Her data are presented in the SPSS *Data View* as given below or can be downloaded from the companion website at *prob_24.1_library_books.sav*.

	age	education	books
1	62	12	2
2	67	16	6
3	65	14	4
4	70	22	10
5	66	18	7
6	63	14	4
7	67	16	6
8	65	18	4
9	63	12	2
10	68	22	9
11	65	16	5
12	69	18	7
13	68	16	7
14	62	14	3
15	64	14	4
16	70	18	9
17	63	12	2
18	68	16	6

24.2 This problem is based on the simple regression you did in the previous chapter (Review Exercise 23.2). We just added another variable called "freedom index" to demonstrate an example of multiple regression. You now have two independent variables ("constitutional law score" and "freedom index") and one dependent variable that counts the "number of laws introduced" by the legislator.

The political consultant wants to determine if the scores on knowledge of constitutional law and score of the freedom index are related to the number of gun control laws introduced. He also wishes to extend any findings into the realm of prediction by using regression to estimate the number of laws introduced by a legislator rated average on both these independent variables. He also wishes to use the equation to predict values that can then be directly compared with the observed values. Use multiple linear regression, write the null and alternative hypotheses, conduct the analysis, write the prediction equations, make the predictions, and interpret the results. His data are presented in the SPSS *Data View* below or can be downloaded from the companion website at *prob_24.2_gun_law.sav*.

	const_score	gun_control	freeindex
1	98	1	28
2	86	2	23
3	74	3	26
4	63	4	24
5	51	6	14
6	97	1	25
7	85	2	21
8	77	5	23
9	65	6	18
10	53	8	12
11	94	2	21
12	83	2	26
13	74	4	15
14	69	4	19
15	55	6	10
16	97	2	21
17	84	4	21
18	71	4	18
19	64	5	17
20	57	8	18
21	99	1	30
22	82	3	23
23	75	4	14
24	63	4	28

24.3 As we have done in the previous two exercises, we bring forward from the previous chapter a simple linear regression problem and add an additional variable. In that exercise (Review Exercise 23.3), you had one independent variable, which was "the number of times an individual attended church during a month." For this current exercise, you will add another independent variable, which is "the number of times one prays in a day." The deacon of the church wants to see if the earlier prediction equation could be improved by adding this additional variable. As before, he wants to compare the performance of the new equation with the actual observed values. In addition, he wishes to predict the number of volunteer hours for those rated as average on the two independent variables. Use multiple linear regression, write the null and alternative hypotheses, do the analysis, and interpret the results. The deacon's new data are presented in the SPSS *Data View* below or may be downloaded from the companion website at *prob_24.3_church_attend.sav*.

	churchattend	pray	hrsvolunteer
1	10	6	16
2	6	5	9
3	2	1	4
4	3	2	6
5	5	5	10
6	9	6	11
7	10	7	16
8	2	2	2
9	7	3	5
10	8	6	10
11	3	2	7
12	6	4	10

24.4 In most of the world religions, there is the foundational belief that it is good to live by the Golden Rule (simply put, treat others as you wish to be treated). Living life according to the Golden Rule is the dependent variable for this work. Let's say it is measured by a valid and reliable scale quantifying the degree one chooses to live by the Golden Rule on a scale from 0 to 30. One independent variable is a measure of religious activity on a scale of 0 to 80. The other independent variable is volunteerism as measured on a scale of 0 to 120. A religious scholar believed there was a relationship between these variables. The scholar's hypothesis was that by knowing a person's score on the scales of religious activity and volunteerism, one could predict a person's level of living life by the Golden Rule. Assist the scholar in exploring regression techniques and the possibility of writing a prediction equation. Write the null and alternative hypotheses, select the statistical approach, do the necessary testing, draw conclusions, and report your findings. The data are presented in the SPSS *Data View* below or can be downloaded from the companion website at *prob_24.4_golden.sav*.

prob_24.4_golden.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities

	volunteer	religion	golden
1	108	25	28
2	96	29	23
3	84	35	21
4	61	39	19
5	107	42	9
6	107	48	19
7	95	51	16
8	87	59	18
9	75	61	13
10	63	64	7
11	93	69	16
12	93	72	21
13	84	67	10
14	79	66	14
15	65	61	10
16	107	58	21
17	94	55	2
18	81	48	5
19	74	42	12
20	67	39	13
21	109	31	25
22	92	29	18
23	85	24	9
24	73	21	18

24.5 For this exercise open SPSS sample file *behavior.sav* and select “Argue” and “Shout” as independent variables. Select “Sleep” as the dependent variable. These variables were the result of a 10-point scale ranging from 0 = extremely appropriate to 10 = extremely inappropriate. The selection of the variables for this exercise were mainly for a demonstration of the statistical procedure rather than a demonstration of social behavior. Your task is to investigate a possible relationship between the three scale variables of “Argue,” “Shout,” and “Sleep.” Write the null and alternative hypotheses for the relationship between variables and attempt to use the independent variables (“Shout” and “Argue”) to predict the “Sleep” variable.