Example 2 Shoe Print and Height: Finding a Coefficient of Determination

If we use the 40 paired shoe print lengths (cm) and heights (cm) from Data Set 2 in Appendix B, we find that the linear correlation coefficient is $r = 0.813$. Find the coefficient of determination. Also, find the percentage of the total variation in y (height) that can be explained by the linear correlation between shoe print length and height.

Solution

With $r = 0.813$, the coefficient of determination is $r^2 = 0.813^2 = 0.661$.

Interpretation

Because r^2 is the proportion of total variation that can be explained, we conclude that 66.1% of the total variation in height can be explained by shoe print length, and the other 33.9% cannot be explained by shoe print length. The other 33.9% might be explained by some other factors and/or random variation.

using TECHNOLOGY

STATDISK Enter the paired data in columns of the STATDISK Data Window, select **Analysis** from the main menu bar, then select **Correlation and Regression.** Enter a value for the significance level (such as 0.05), and select the two columns of data to be used. Click on **Evaluate.** The STATDISK display will include the linear correlation coefficient r, the coefficient of determination, the regression equation, the value of the standard error of estimate s_e, the total variation, the explained variation, and the unexplained variation.

MINITAB Minitab can be used to find the regression equation, the standard error of estimate s_e (labeled S), the value of the coefficient of determination (labeled R-sq), and the limits of a prediction interval. Enter the x data in column C1 and the y data in column C2, then select the options **Stat, Regression,** and **Regression.** Enter C2 in the box labeled "Response" and enter C1 in the box labeled "Predictors." If you want a prediction interval for some given value of x, click on the **Options** box and enter the desired value of x_0 in the box labeled "Prediction intervals for new observations."

EXCEL Excel can be used to find the regression equation, the standard error of estimate s_e, and the coefficient of determination (labeled R square). First enter the paired data in columns A and B and proceed to use either XLSTAT or Excel's Data Analysis add-n.

XLSTAT Use the same procedure given at the end of Section 10-3. See the results in the "Goodness of fit" section, where the value of s_e is identified as "RMSE" (for root mean square error). The value of r^2 is labeled as R^2.

Data Analysis add-in: If using Excel 2013, 2010, or 2007, click on **Data,** then click on **Data Analysis;** if using Excel 2003, click on **Tools,** then click on **Data Analysis.** Select **Regression,** and then click **OK.** Enter the range for the y values, such as B1:B6. Enter the range for the x values, such as A1:A6. Click **OK.**

TI-83/84 PLUS The TI-83/84 Plus calculator can be used to find the linear correlation coefficient r, the equation of the regression line, the standard error of estimate s_e, and the coefficient of determination (labeled as r^2). Enter the paired data in lists L1 and L2, then press STAT and select **TESTS,** and then choose the option **LinRegTTest.** For Xlist enter L1, for Ylist enter L2, use a Freq (frequency) value of 1, and select $\neq 0$. Scroll down to Calculate, then press ENTER.

STATCRUNCH Click on **Open StatCrunch.** Enter the columns of data or open a data set. Click on **Stat,** then select **Regression,** then select **Simple Linear.** Enter the columns to be used, then click on **Calculate.** Results include the regression equation, the standard error of estimate s_e (identified as "estimate of error standard deviation"), and the value of the coefficient of determination (labeled R-sq). A prediction interval can be obtained by clicking on **Next** and entering a value of x to be used for the prediction.

10-4 Basic Skills and Concepts

Statistical Literacy and Critical Thinking

1. s_e Notation Using Data Set 1 from Appendix B, if we let the predictor variable x represent heights of males and let the response variable y represent weights of males, the sample of