> **Note**
>
> See page 242 for a quick review of how to obtain the sample estimates for the linear regression model. The estimates can also be obtained using technology.

In order to perform inference on the linear model, some assumptions about the nature of the error terms are required.

Assumptions about the Error Term in the Linear Model

1. The ε_i are presumed to be normally distributed with a mean of 0 and a variance of σ_e^2.
2. The ε_i are presumed to be independent of each other.

PROPERTIES

> **Technology**
>
> The instructions for calculating the coefficients for the simple linear regression model using various technologies can be found on stat.hawkeslearning.com under **Discovering Statistics and Data, Fourth Edition > Technology Instructions > Regression > Simple Linear Regression**.

With the addition of the error term, the model's parameters are β_0, β_1, and σ_e^2. The estimation of these quantities was discussed in Chapter 5. The actual verification of these assumptions cannot be made prior to a regression analysis, but they can be validated by doing an analysis of the **residuals** (errors). A residual analysis is beyond the scope of this book, therefore, we will assume that the error terms satisfy the necessary assumptions in order to proceed with inference in the regression analysis.

In addition to the formal assumptions stated above, a linear model should be used to fit data that appears to be reasonably linearly related. Because of the wide availability of computer programs that calculate least squares estimates, you will not need to manually calculate estimates very often.

13.1 Exercises

Basic Concepts

1. Why is an error term incorporated in the simple linear model?
2. What does the error term represent?
3. What assumptions are made about the error term in the simple linear model?
4. What are the parameters of the simple linear regression model? Identify their estimates from the sample.

Exercises

5. Consider the following data and regression output relating a student's grade to the number of absences from class.

   a. Determine the independent and dependent variables and write the equation of the model we desire to estimate.

   b. Determine the estimated model (regression equation) for predicting a student's grade based on the number of absences from class.

Class Absences and Grade

Number of Absences	Grade
3	3.9
5	3.8
6	2.9
6	2.7
6	2.4
7	2.3
8	1.9

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.917589
R Square	0.84197
Adjusted R Square	0.810364
Standard Error	0.329598
Observations	7

ANOVA

	df	*SS*
Regression	1	2.89396978
Residual	5	0.543173077
Total	6	3.437142857

	Coefficients	*Standard Error*	*t Stat*
Intercept	5.427885	0.51610437	10.51703
Number of Absences	–0.44135	0.085509913	–5.16134

c. Determine the mean square error (MSE). (See Section 5.3.)

6. The following table of data gives the weeks of gestation and corresponding birth weights for a sample of ten babies.

Gestation Period and Birth Weight

Weeks of Gestation	Birth Weight
34	5.9
34	5.7
35	6.2
36	6.6
36	6.8
37	7.0
38	7.2
38	7.5
39	8.0
40	8.2

a. Determine the independent and dependent variables and write the equation of the model we desire to estimate.

b. Determine the estimated model (regression equation) for predicting a baby's birth weight based on the number of weeks of gestation.

c. Determine the mean square error (MSE). (See Section 5.3.)

d. Determine the coefficient of determination and explain its meaning in terms of the problem.

Data

The data set can be found on stat.hawkeslearning.com under **Discovering Statistics and Data, Fourth Edition > Data Sets > Global Statistics by Country**.

7. Use the Global Statistics by Country data set to answer the following questions.
   a. Determine the regression equation for predicting birth rate based on female literacy rate.
   b. Determine the coefficient of determination and explain its meaning in terms of the problem.
   c. Find the predicted birth rate for a female literacy rate of 70.
   d. Find the predicted birth rate for a female literacy rate of 90.
   e. As the female literacy rate rises what happens to the birth rate? Does this make sense? Explain your answer.

13.2 Inference Concerning β_1

Since β_1 specifies the rate of change between x and y, in most linear models the parameter of interest is β_1. Two inferential techniques are useful in evaluating the estimate of β_1. Confidence intervals, similar in structure to those used for means and proportions, will be developed. In addition, a hypothesis testing procedure will be presented to test whether β_1 is equal to some particular value.

The Confidence Interval for β_1

Developing a confidence interval for β_1 requires thinking about the estimate b_1 as a random variable. Each random sample from the population will produce different data and hence different least squares estimates of b_0 and b_1. The confidence interval will serve two purposes, to place bounds on the location of β_1 as well as to provide information about the quality of the point estimate b_1. The form of the confidence interval is familiar.

$$\begin{matrix}\text{Sample}\\ \text{estimate of}\\ \text{parameter}\end{matrix} \pm \left(\begin{matrix}\text{A certain number of standard}\\ \text{deviations units depending on}\\ \text{the desired confidence}\end{matrix}\right) \cdot \left(\begin{matrix}\text{The standard}\\ \text{deviation of the}\\ \text{sample estimate}\end{matrix}\right)$$

The **sample estimate** of β_1 is b_1. The variance of b_1 is given by

$$\sigma_{b_1}^2 = \frac{\sigma_e^2}{\sum(x_i - \bar{x})^2}$$

but like all population measurements, $\sigma_{b_1}^2$ usually has to be estimated from the data. The sample estimate of the variance of b_1 is given by

$$s_{b_1}^2 = \frac{s_e^2}{\sum(x_i - \bar{x})^2}.$$

The only difference in the computation of $\sigma_{b_1}^2$ and $s_{b_1}^2$ is the replacement of the population variance of the error terms, σ_e^2, with the corresponding sample statistic, s_e^2. The **standard deviation (standard error) of the sample estimate** b_1 is

$$s_{b_1} = \sqrt{\frac{s_e^2}{\sum(x_i - \bar{x})^2}}.$$