Introductory STATISTICS

9TH EDITION

Neil WEISS

Chapter 15

Inferential Methods in Regression and Correlation

Section 15.1

The Regression Model; Analysis of Residuals

Figure 15.1

Population regression line

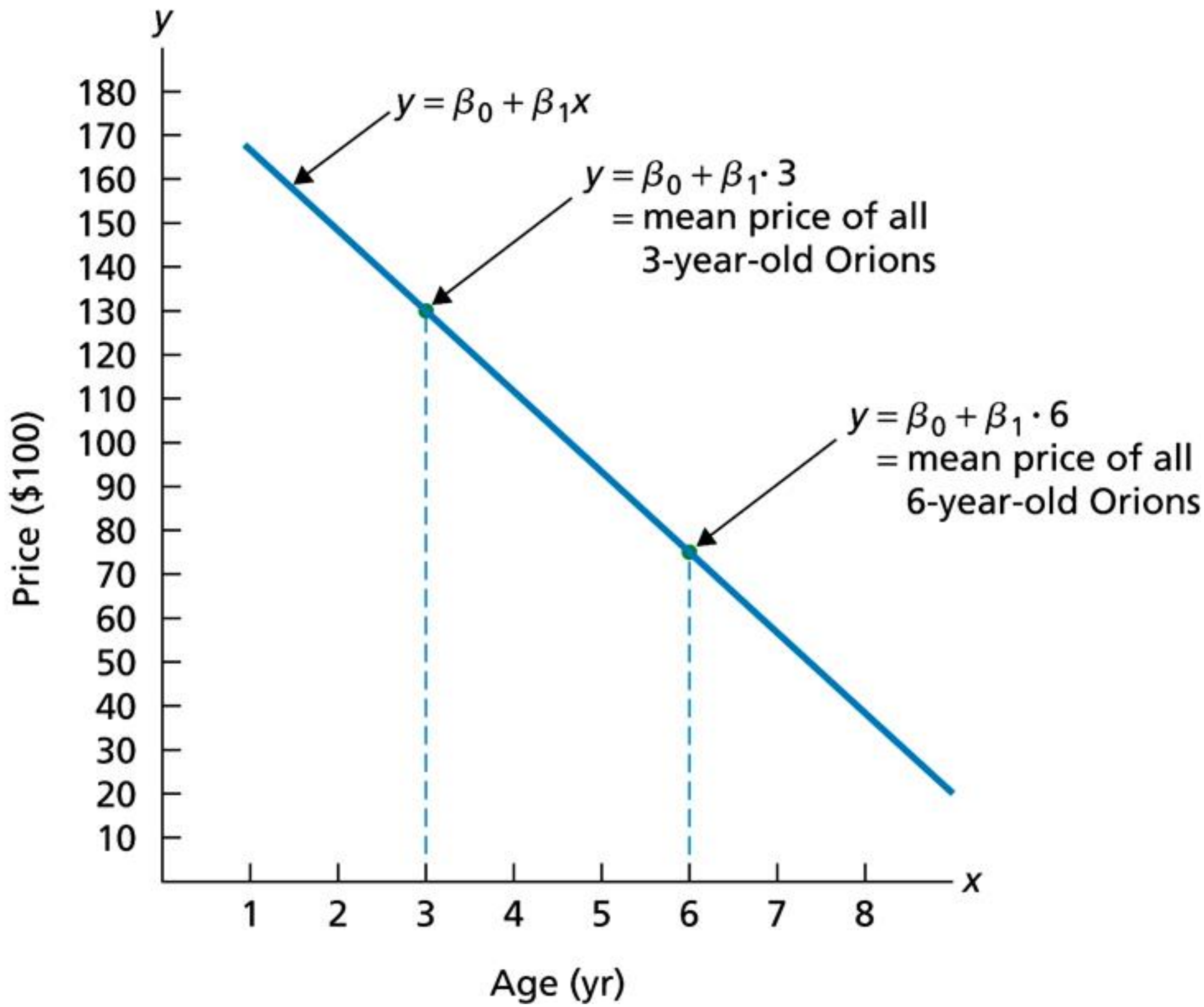


Figure 15.2

Price distributions for 2-, 5-, and 7-year-old Orions under Assumptions 2 and 3 (The means shown for the three normal distributions reflect Assumption 1)

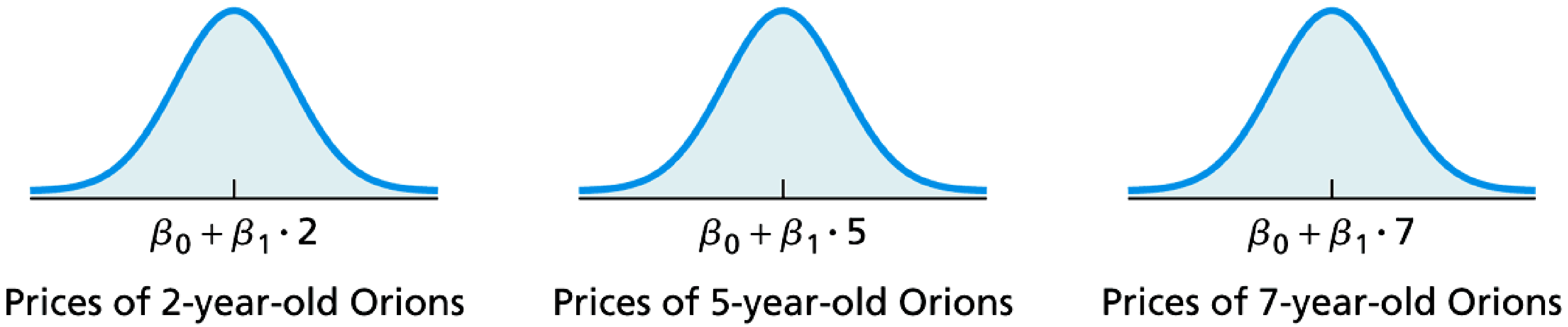


Figure 15.3

Graphical portrayal of Assumptions 1–3 for regression inferences pertaining to age and price of Orions

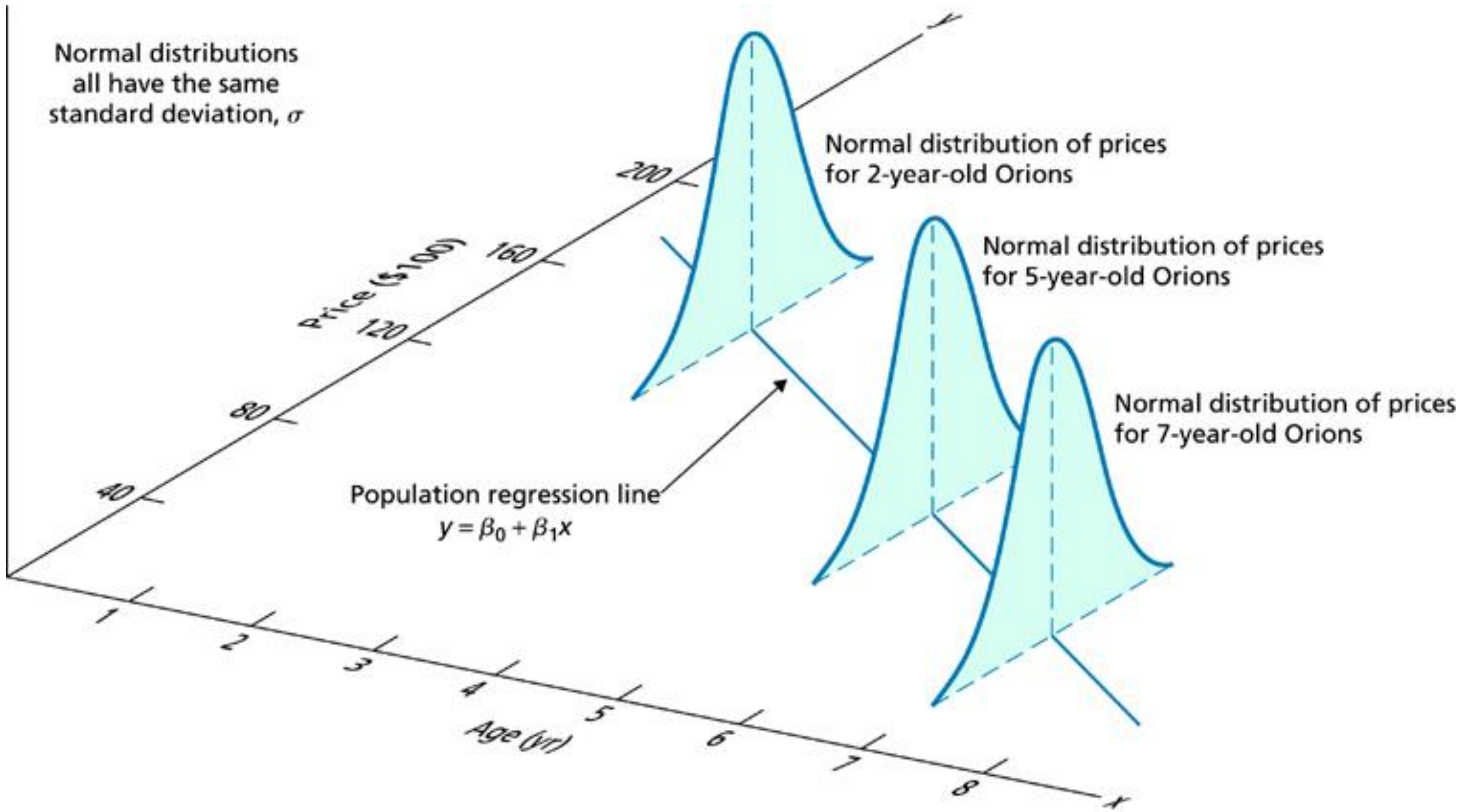
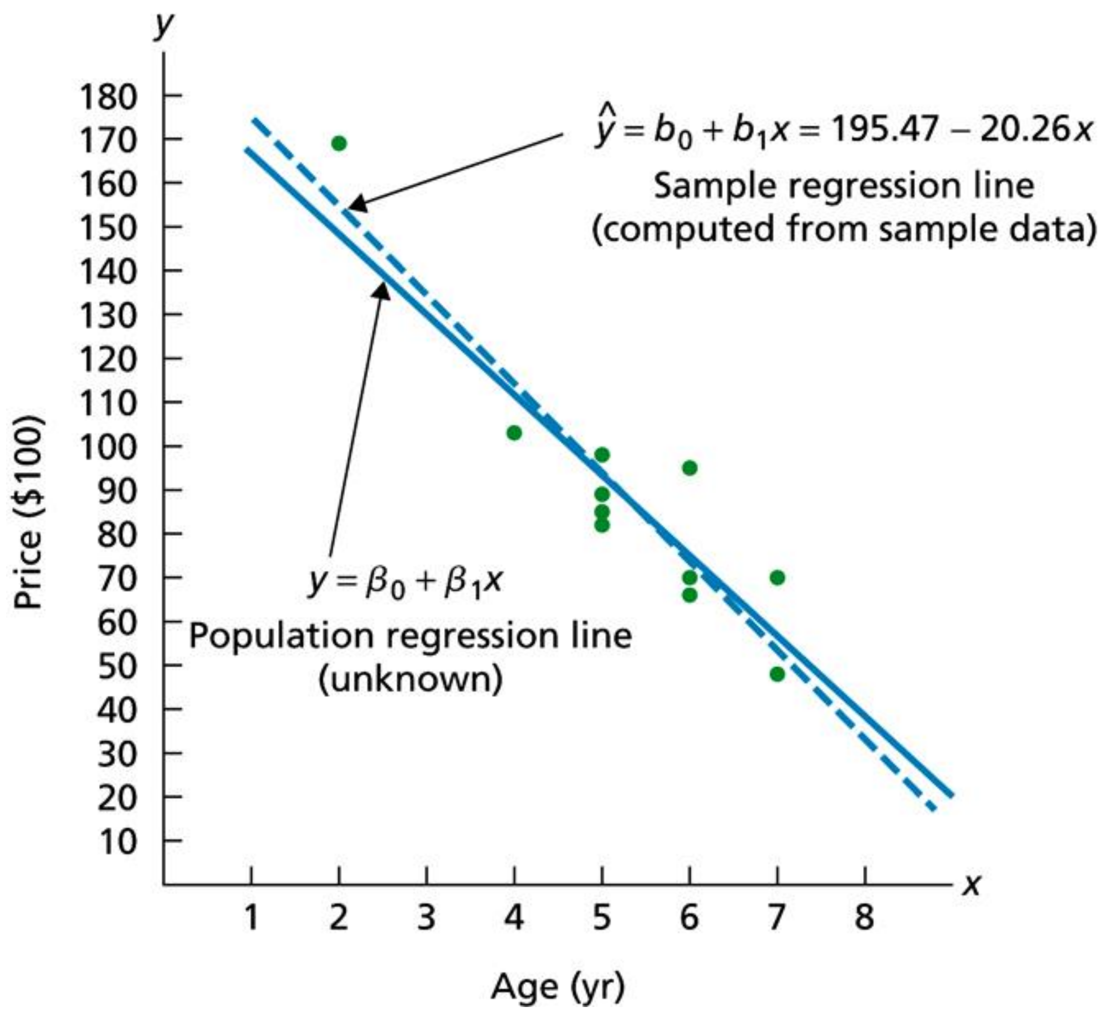


Figure 15.4

Population regression line and sample regression line for age and price of Orions



Definition 15.1

Standard Error of the Estimate

The **standard error of the estimate, s_e,** is defined by

$$s_e = \sqrt{\frac{SSE}{n-2}},$$

where *SSE* is the error sum of squares.

Figure 15.5

Residual of a data point

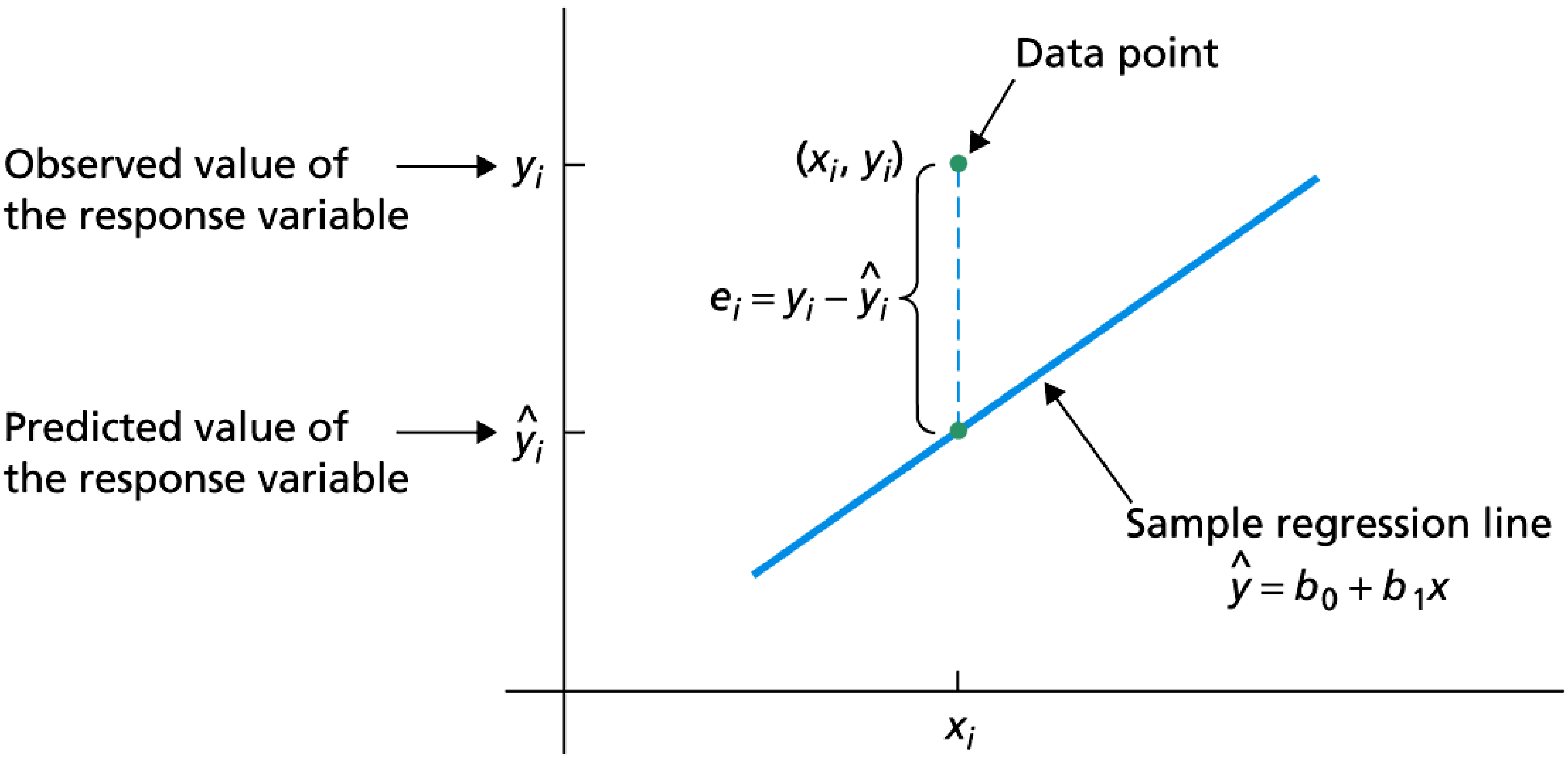


Figure 15.6

Residual plots suggesting (a) no violation of linearity or constant standard deviation, (b) violation of linearity, and (c) violation of constant standard deviation

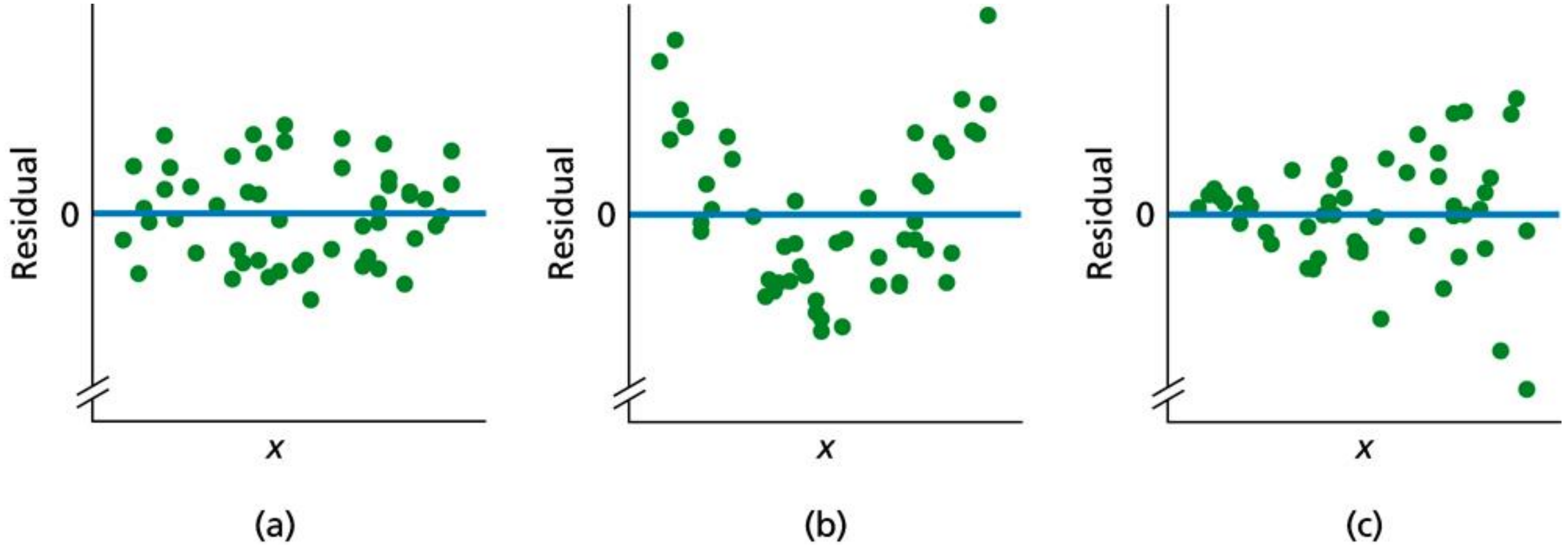
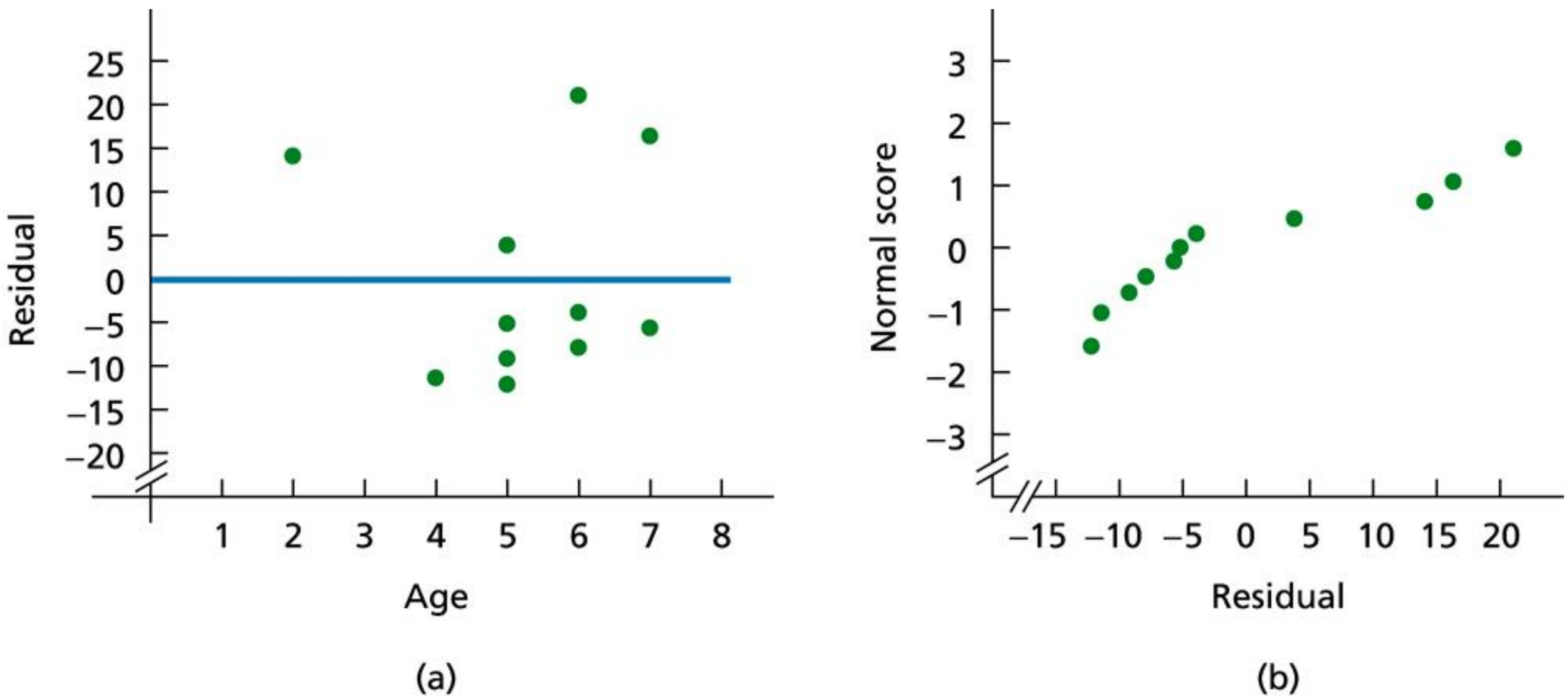


Figure 15.7

(a) Residual plot; (b) normal probability plot for residuals



Section 15.2

Inferences for the Slope of the Population Regression Line

Procedure 15.1

Regression t-Test

Purpose To perform a hypothesis test to decide whether a predictor variable is useful for making predictions

Assumptions The four assumptions for regression inferences

Step 1 The null and alternative hypotheses are, respectively,

H_0: $\beta_1 = 0$ (predictor variable is not useful for making predictions)

H_a: $\beta_1 \neq 0$ (predictor variable is useful for making predictions).

Step 2 Decide on the significance level, α.

Step 3 Compute the value of the test statistic

$$t = \frac{b_1}{s_e / \sqrt{S_{xx}}}$$

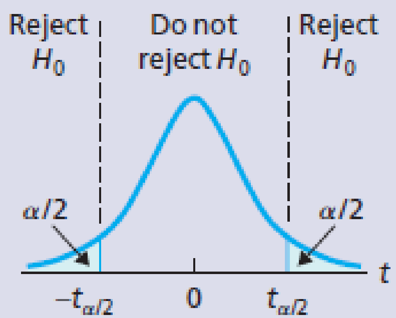
and denote that value t_0.

Procedure 15.1 (cont.)

CRITICAL-VALUE APPROACH	OR	P-VALUE APPROACH

Critical-value approach

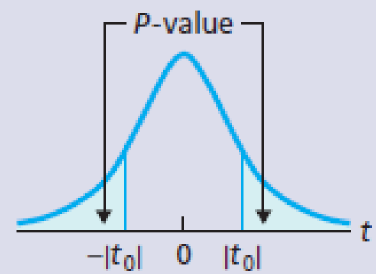
Step 4 The critical values are $\pm t_{\alpha/2}$ with df $= n - 2$. Use Table IV to find the critical values.



Step 5 If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.

P-value approach

Step 4 The t-statistic has df $= n - 2$. Use Table IV to estimate the P-value, or obtain it exactly by using technology.



Step 5 If $P \leq \alpha$, reject H_0; otherwise, do not reject H_0.

Step 6 Interpret the results of the hypothesis test.

Procedure 15.2

Regression *t*-Interval Procedure

Purpose To find a confidence interval for the slope, β_1, of the population regression line

Assumptions The four assumptions for regression inferences

Step 1 **For a confidence level of $1 - \alpha$, use Table IV to find $t_{\alpha/2}$ with $\text{df} = n - 2$.**

Step 2 **The endpoints of the confidence interval for β_1 are**

$$b_1 \pm t_{\alpha/2} \cdot \frac{s_e}{\sqrt{S_{xx}}}.$$

Step 3 **Interpret the confidence interval.**

Section 15.3

Estimation and Prediction

Procedure 15.3

Conditional Mean t-Interval Procedure

Purpose To find a confidence interval for the conditional mean of the response variable corresponding to a particular value of the predictor variable, x_p

Assumptions The four assumptions for regression inferences

Step 1 **For a confidence level of $1-\alpha$, use Table IV to find $t_{\alpha/2}$ with df $= n-2$.**

Step 2 **Compute the point estimate, $\hat{y}_p = b_0 + b_1 x_p$.**

Step 3 **The endpoints of the confidence interval for the conditional mean of the response variable are**

$$\hat{y}_p \pm t_{\alpha/2} \cdot s_e \sqrt{\frac{1}{n} + \frac{(x_p - \Sigma x_i / n)^2}{S_{xx}}}.$$

Step 4 **Interpret the confidence interval.**

Procedure 15.4

Predicted Value *t*-Interval Procedure

Purpose To find a prediction interval for the value of the response variable corresponding to a particular value of the predictor variable, x_p

Assumptions The four assumptions for regression inferences

Step 1 **For a prediction level of $1 - \alpha$, use Table IV to find $t_{\alpha/2}$ with $\text{df} = n - 2$.**

Step 2 **Compute the predicted value, $\hat{y}_p = b_0 + b_1 x_p$.**

Step 3 **The endpoints of the prediction interval for the value of the response variable are**

$$\hat{y}_p \pm t_{\alpha/2} \cdot s_e \sqrt{1 + \frac{1}{n} + \frac{(x_p - \Sigma x_i/n)^2}{S_{xx}}}.$$

Step 4 **Interpret the prediction interval.**

Section 15.4

Inferences in Correlation

Procedure 15.5

Correlation *t*-Test

Purpose To perform a hypothesis test for a population linear correlation coefficient, ρ

Assumptions The four assumptions for regression inferences

Step 1 **The null hypothesis is H_0: $\rho = 0$, and the alternative hypothesis is**

H_a: $\rho \neq 0$ **(Two tailed)** **or** H_a: $\rho < 0$ **(Left tailed)** **or** H_a: $\rho > 0$. **(Right tailed)**

Step 2 **Decide on the significance level, α.**

Step 3 **Compute the value of the test statistic**

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

and denote that value t_0.

Procedure 15.5 (cont.)

Correlation *t*-Test

Purpose To perform a hypothesis test for a population linear correlation coefficient, ρ

Assumptions The four assumptions for regression inferences

Step 1 **The null hypothesis is H_0: $\rho = 0$, and the alternative hypothesis is**

$$H_a: \rho \neq 0 \text{ (Two tailed)} \quad \text{or} \quad H_a: \rho < 0 \text{ (Left tailed)} \quad \text{or} \quad H_a: \rho > 0. \text{ (Right tailed)}$$

Step 2 **Decide on the significance level, α.**

Step 3 **Compute the value of the test statistic**

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

and denote that value t_0.

Section 15.5

Testing for Normality

Procedure 15.6

Correlation Test for Normality

Purpose To perform a hypothesis test to decide whether a variable is not normally distributed

Assumption Simple random sample

Step 1 **The null and alternative hypotheses are, respectively,**

$$H_0\text{: The variable is normally distributed}$$

$$H_a\text{: The variable is not normally distributed.}$$

Step 2 **Decide on the significance level, α.**

Step 3 **Compute the value of the test statistic**

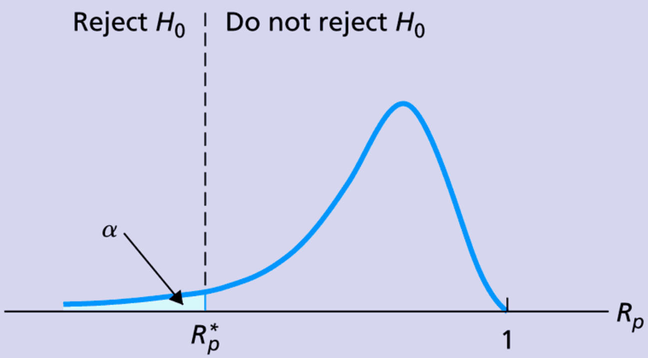
$$R_p = \frac{\Sigma x_i w_i}{\sqrt{S_{xx} \Sigma w_i^2}},$$

where x and w denote observations of the variable and the corresponding normal scores, respectively. Denote the value of the test statistic R_p^0.

Procedure 15.6 (cont.)

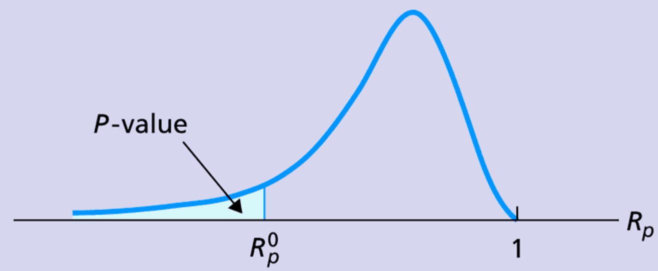
CRITICAL-VALUE APPROACH | OR | ***P*-VALUE APPROACH**

Step 4 The critical value is R_p^*. Use Table IX to find the critical value.



Step 5 If the value of the test statistic falls in the rejection region, reject H_0; otherwise, do not reject H_0.

Step 4 Use Table IX to estimate the *P*-value, or obtain it exactly by using technology.



Step 5 If $P \leq \alpha$, reject H_0; otherwise, do not reject H_0.

Step 6 Interpret the results of the hypothesis test.

Table 15.7

Table for computing R_p

Adjusted gross income x	Normal score w	xw	x^2	w^2
7.8	−1.64	−12.792	60.84	2.6896
9.7	−1.11	−10.767	94.09	1.2321
10.6	−0.79	−8.374	112.36	0.6241
12.7	−0.53	−6.731	161.29	0.2809
12.8	−0.31	−3.968	163.84	0.0961
18.1	−0.10	−1.810	327.61	0.0100
21.2	0.10	2.120	449.44	0.0100
33.0	0.31	10.230	1,089.00	0.0961
43.5	0.53	23.055	1,892.25	0.2809
51.1	0.79	40.369	2,611.21	0.6241
81.4	1.11	90.354	6,625.96	1.2321
93.1	1.64	152.684	8,667.61	2.6896
395.0	0.00	274.370	22,255.50	9.8656