CHAPTER 3

Exercise Solutions

EXERCISE 3.1

(a) The required interval estimator is $b_1 \pm t_c \text{se}(b_1)$. When $b_1 = 83.416$, $t_c = t_{(0.975,38)} = 2.024$ and $\text{se}(b_1) = 43.410$, we get the interval estimate:

$$83.416 \pm 2.024 \times 43.410 = (-4.46, 171.30)$$

We estimate that β_1 lies between –4.46 and 171.30. In repeated samples, 95% of similarly constructed intervals would contain the true β_1.

(b) To test $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$ we compute the *t*-value

$$t_1 = \frac{b_1 - \beta_1}{\text{se}(b_1)} = \frac{83.416 - 0}{43.410} = 1.92$$

Since the *t* = 1.92 value does not exceed the 5% critical value $t_c = t_{(0.975,38)} = 2.024$, we do not reject H_0. The data do not reject the zero-intercept hypothesis.

(c) The *p*-value 0.0622 represents the sum of the areas under the *t* distribution to the left of *t* = –1.92 and to the right of *t* = 1.92. Since the *t* distribution is symmetric, each of the tail areas that make up the *p*-value are $p/2 = 0.0622/2 = 0.0311$. The level of significance, α, is given by the sum of the areas under the PDF for $|t| > |t_c|$, so the area under the curve for $t > t_c$ is $\alpha/2 = .025$ and likewise for $t < -t_c$. Therefore not rejecting the null hypothesis because $\alpha/2 < p/2$, or $\alpha < p$, is the same as not rejecting the null hypothesis because $-t_c < t < t_c$. From Figure xr3.1(a) we can see that having a *p*-value > 0.05 is equivalent to having $-t_c < t < t_c$.

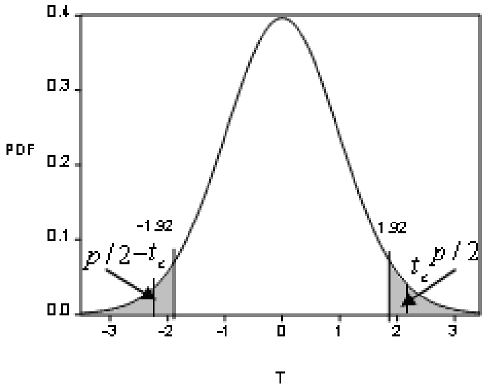

Figure xr3.1(a) Critical and observed *t* values for Exercise 3.1(c)

Exercise 3.1 (continued)

(d) Testing $H_0:\beta_1 = 0$ against $H_1:\beta_1 > 0$, uses the same *t*-value as in part (b), t = 1.92. Because it is a one-tailed test, the critical value is chosen such that there is a probability of 0.05 in the right tail. That is, $t_c = t_{(0.95,38)} = 1.686$. Since t = 1.92 > t_c = 1.69, H_0 is rejected, the alternative is accepted, and we conclude that the intercept is positive. In this case *p*-value = *P*(*t* > 1.92) = 0.0311. We see from Figure xr3.1(b) that having the *p*-value < 0.05 is equivalent to having *t* > 1.69.

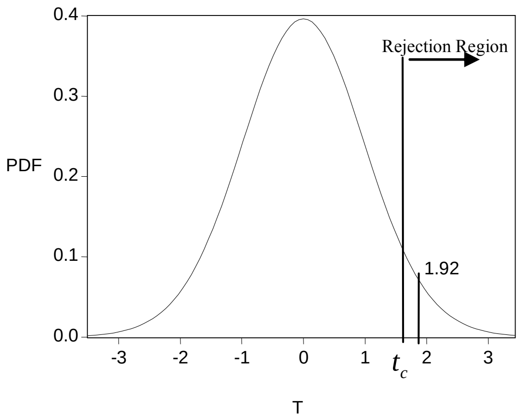

Figure xr3.1(b) Rejection region and observed *t* value for Exercise 3.1(d)

(e) The term "level of significance" is used to describe the probability of rejecting a true null hypothesis when carrying out a hypothesis test. The term "level of confidence" refers to the probability of an interval estimator yielding an interval that includes the true parameter. When carrying out a two-tailed test of the form $H_0:\beta_k = c$ versus $H_1:\beta_k \neq c$, non-rejection of H_0 implies c lies within the confidence interval, and vice versa, providing the level of significance is equal to one minus the level of confidence.

(f) False. The test in (d) uses the level of significance 5%, which is the probability of a Type I error. That is, in repeated samples we have a 5% chance of rejecting the null hypothesis when it is true. The 5% significance is a probability statement about a procedure not a probability statement about β_1. It is careless and dangerous to equate 5% level of significance with 95% confidence, which relates to interval estimation procedures, not hypothesis tests.

EXERCISE 3.2

(a) The coefficient of *EXPER* indicates that, on average, a technical artist's quality rating goes up by 0.076 for every additional year of experience.

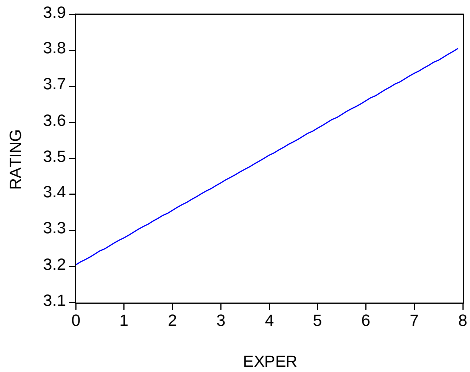

Figure xr3.2(a) Estimated regression function

(b) Using the value $t_c = t_{(0.975,22)} = 2.074$, the 95% confidence interval for β_2 is given by

$$b_2 \pm t_c \text{se}(b_2) = 0.076 \pm 2.074 \times 0.044 = (-0.015,\ 0.167)$$

We are 95% confident that the procedure we have used for constructing a confidence interval will yield an interval that includes the true parameter β_2.

(c) To test $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$, we use the test statistic $t = b_2/\text{se}(b_2) = 0.076/0.044 = 1.727$. The t critical value for a two tail test with $N - 2 = 22$ degrees of freedom is 2.074. Since $-2.074 < 1.727 < 2.074$ we fail to reject the null hypothesis.

(d) To test $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 > 0$, we use the t-value from part (c), namely $t = 1.727$, but the right-tail critical value $t_c = t_{(0.95,22)} = 1.717$. Since $1.727 > 1.717$, we reject H_0 and conclude that β_2 is positive. Experience has a positive effect on quality rating.

Exercise 3.2 (continued)

(e) The *p*-value of 0.0982 is given as the sum of the areas under the *t*-distribution to the left of −1.727 and to the right of 1.727. We do not reject H_0 because, for $\alpha = 0.05$, *p*-value > 0.05. We can reject, or fail to reject, the null hypothesis just based on an inspection of the *p*-value. Having the *p*-value > α is equivalent to having $|t| < t_c = 2.074$.

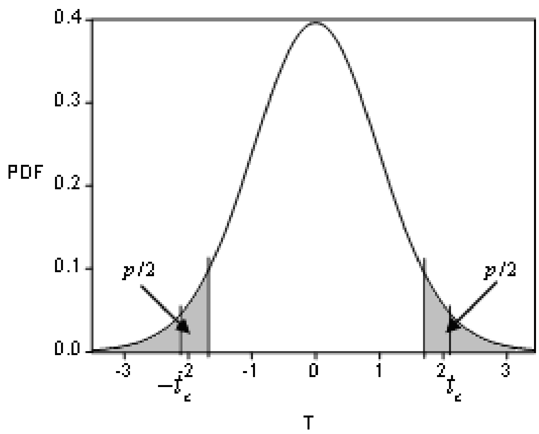

Figure xr3.2(b) *P*-value diagram

EXERCISE 3.3

(a) Hypotheses: $H_0: \beta_2 = 0$ against $H_1: \beta_2 \neq 0$

Calculated *t*-value: $t = 0.310/0.082 = 3.78$

Critical *t*-value: $\pm t_c = \pm t_{(0.995,22)} = \pm 2.819$

Decision: Reject H_0 because $t = 3.78 > t_c = 2.819$.

(b) Hypotheses: $H_0: \beta_2 = 0$ against $H_1: \beta_2 > 0$

Calculated *t*-value: $t = 0.310/0.082 = 3.78$

Critical *t*-value: $t_c = t_{(0.99,22)} = 2.508$

Decision: Reject H_0 because $t = 3.78 > t_c = 2.508$.

(c) Hypotheses: $H_0: \beta_2 = 0$ against $H_1: \beta_2 < 0$

Calculated *t*-value: $t = 0.310/0.082 = 3.78$

Critical *t*-value: $t_c = t_{(0.05,22)} = -1.717$

Decision: Do not reject H_0 because $t = 3.78 > t_c = -1.717$.

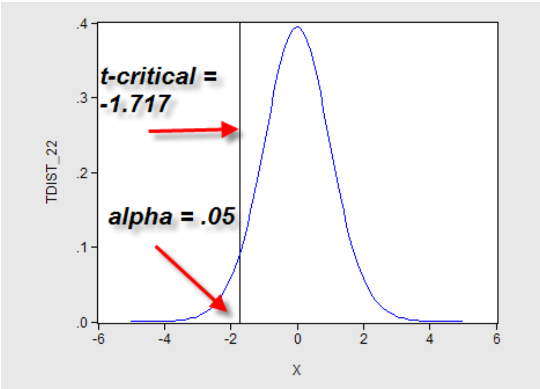

Figure xr3.3 One tail rejection region

(d) Hypotheses: $H_0: \beta_2 = 0.5$ against $H_1: \beta_2 \neq 0.5$

Calculated *t*-value: $t = (0.310 - 0.5)/0.082 = -2.32$

Critical *t*-value: $\pm t_c = \pm t_{(0.975,22)} = \pm 2.074$

Decision: Reject H_0 because $t = -2.32 < -t_c = -2.074$.

(e) A 99% interval estimate of the slope is given by

$$b_2 \pm t_c \text{se}(b_2) = 0.310 \pm 2.819 \times 0.082 = (0.079, 0.541)$$

We estimate β_2 to lie between 0.079 and 0.541 using a procedure that works 99% of the time in repeated samples.

EXERCISE 3.4

(a) $b_1 = t \times \text{se}(b_1)$ = 1.257 × 2.174 = 2.733

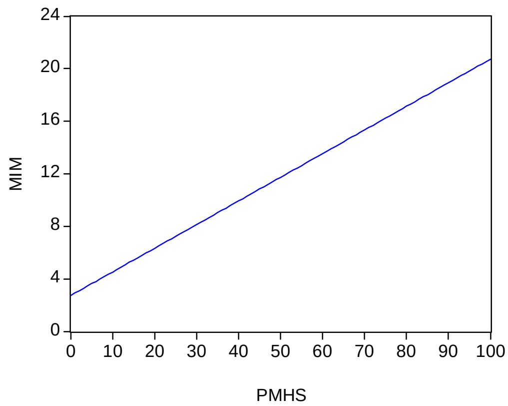

Figure xr3.4(a) Estimated regression function

(b) $\text{se}(b_2) = b_2/t = 0.180/5.754 = 0.0313$

(c) p-value = 2 × $(1 - P(t < 1.257))$ = 2 × (1 – 0.8926) = 0.2147

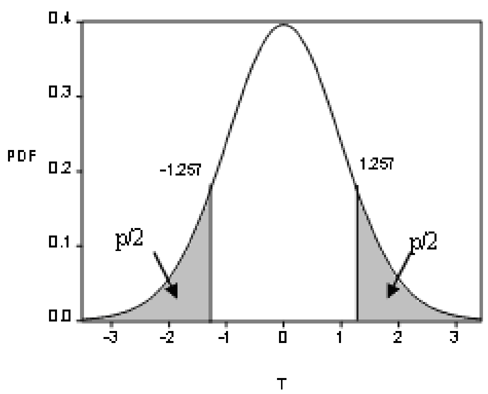

Figure xr3.4(b) *P*-value diagram

(d) The estimated slope $b_2 = 0.18$ indicates that a 1% increase in males 18 and older, who are high school graduates, increases average income of those males by $180. The positive sign is as expected; more education should lead to higher salaries.

(e) Using $t_c = t_{(0.995,49)} = 2.68$, a 99% confidence interval for the slope is given by

$$b_2 \pm t_c \text{se}(b_2) = 0.180 \pm 2.68 \times 0.0313 = (0.096, 0.264)$$

Exercise 3.4 (continued)

(f) For testing $H_0: \beta_2 = 0.2$ against $H_1: \beta_2 \neq 0.2,$ we calculate

$$t = \frac{0.180 - 0.2}{0.0313} = -0.639$$

The critical values for a two-tailed test with a 5% significance level and 49 degrees of freedom are $\pm t_c = \pm 2.01.$ Since t = –0.634 lies in the interval (–2.01, 2.01), we do not reject H_0. The null hypothesis suggests that a 1% increase in males 18 or older, who are high school graduates, leads to an increase in average income for those males of $200. Non-rejection of H_0 means that this claim is compatible with the sample of data.

EXERCISE 3.5

(a) The linear relationship between life insurance and income is estimated as

$$\widehat{INSURANCE} = 6.8550 + 3.8802 INCOME$$
$$\text{(se)} \qquad (7.3835)(0.1121)$$

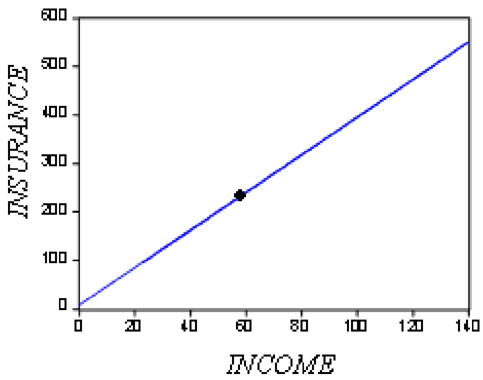

Figure xr3.5 Fitted regression line and mean

(b) The relationship in part (a) indicates that, as income increases, the amount of life insurance increases, as is expected. If taken literally, the value of b1 = 6.8550 implies that if a family has no income, then they would purchase \$6855 worth of insurance. However, given the lack of data in the region where $INCOME = 0$, this value is not reliable.

(i) If income increases by \$1000, then an estimate of the resulting change in the amount of life insurance is \$3880.20.

(ii) The standard error of b_2 is 0.1121. To test a hypothesis about β_2 the test statistic is

$$\frac{b_2 - \beta_2}{\text{se}(b_2)} \sim t_{(N-2)}$$

An interval estimator for β_2 is $\left[b_2 - t_c \text{se}(b_2), b_2 + t_c \text{se}(b_2)\right]$, where t_c is the critical value for t with $(N-2)$ degrees of freedom at the α level of significance.

(c) To test the claim, the relevant hypotheses are H_0: $\beta_2 = 5$ versus H_1: $\beta_2 \neq 5$. The alternative $\beta_2 \neq 5$ has been chosen because, before we sample, we have no reason to suspect $\beta_2 > 5$ or $\beta_2 < 5$. The test statistic is that given in part (b) (ii) with β_2 set equal to 5. The rejection region (18 degrees of freedom) is $|t| > 2.101$. The value of the test statistic is

$$t = \frac{b_2 - 5}{\text{se}(b_2)} = \frac{3.8802 - 5}{0.1121} = -9.99$$

As $t = -9.99 < -2.101$, we reject the null hypothesis and conclude that the estimated relationship does not support the claim.

Exercise 3.5 (continued)

(d) To test the hypothesis that the slope of the relationship is one, we proceed as we did in part (c), using 1 instead of 5. Thus, our hypotheses are H_0: $\beta_2 = 1$ versus H_1: $\beta_2 \neq 1$. The rejection region is $|t| > 2.101$. The value of the test statistic is

$$t = \frac{3.8802 - 1}{0.1121} = 25.7$$

Since $t = 25.7 > t_c = 2.101,$ we reject the null hypothesis. We conclude that the amount of life insurance does not increase at the same rate as income increases.

(e) Life insurance companies are interested in household characteristics that influence the amount of life insurance cover that is purchased by different households. One likely important determinant of life insurance cover is household income. To see if income is important, and to quantify its effect on insurance, we set up the model

$$INSURANCE_i = \beta_1 + \beta_2 INCOME_i + e_i$$

where $INSURANCE_i$ is life insurance cover by the *i*-th household, $INCOME_i$ is household income, β_1 and β_2 are unknown parameters that describe the relationship, and e_i is a random uncorrelated error that is assumed to have zero mean and constant variance σ^2.

To estimate our hypothesized relationship, we take a random sample of 20 households, collect observations on *INSURANCE* and *INCOME* and apply the least-squares estimation procedure. The estimated equation, with standard errors in parentheses, is

$$\widehat{INSURANCE} = 6.8550 + 3.8802 INCOME$$
$$(\text{se}) \qquad (7.3835)(0.1121)$$

The point estimate for the response of life-insurance coverage to an income increase of \$1000 (the slope) is \$3880 and a 95% interval estimate for this quantity is (\$3645, \$4116). This interval is a relatively narrow one, suggesting we have reliable information about the response. The intercept estimate is not significantly different from zero, but this fact by itself is not a matter for concern; as mentioned in part (b), we do not give this value a direct economic interpretation.

The estimated equation could be used to assess likely requests for life insurance and what changes may occur as a result of income changes.

EXERCISE 3.6

(a) A 95% interval estimator for β_2 is $b_2 \pm t_{(0.975,14)} \times \text{se}(b_2)$. Using our sample of data the corresponding interval estimate is

$$-0.3857 \pm 2.145 \times 0.03601 = (-0.4629, -0.3085)$$

If we used the interval estimator in repeated samples, then 95% of interval estimates like the above one would contain β_2. Thus, β_2 is likely to lie in the range given by the above interval.

(b) We set up the hypotheses H_0: $\beta_2 = 0$ versus H_1: $\beta_2 < 0$. The alternative $\beta_2 < 0$ is chosen because we would expect the unit costs of production to decline as cumulative production increases if there is learning. The test statistic, given H_0 is true, is

$$t = \frac{b_2}{\text{se}(b_2)} \sim t_{(14)}$$

The rejection region is $t < -1.761$. The value of the test statistic is

$$t = \frac{-0.3857}{0.03601} = -10.71$$

Since $t = -10.71 < -1.761$, we reject H_0 and conclude that learning does exist. We conclude in this way because -10.71 is an unlikely value to have come from the t distribution which is valid when there is no learning.

EXERCISE 3.7

(a) We set up the hypotheses $H_0: \beta_j = 1$ versus $H_1: \beta_j \neq 1$. The economic relevance of this test is to test whether the return on the firm's stock is risky relative to the market portfolio. Each beta measures the volatility of the stock relative to the market portfolio and volatility is often used to measure risk. A beta value of one indicates that the stock's volatility is the same as that of the market portfolio. The test statistic given H_0 is true, is

$$t = \frac{b_j - 1}{\text{se}(b_j)} \sim t_{(118)}$$

The rejection region is $t < -1.980$ and $t > 1.980$, where $t_{(0.975,118)} = 1.980$.

The results for each company are given in the following table:

Stock	*t*-value	Decision rule
Disney	$t = \frac{0.9593 - 1}{0.1420} = -0.287$	Since $-1.98 < t < 1.98$, fail to reject H_0
GE	$t = \frac{0.9830 - 1}{0.1047} = -0.162$	Since $-1.98 < t < 1.98$, fail to reject H_0
GM	$t = \frac{1.0744 - 1}{0.1558} = 0.478$	Since $-1.98 < t < 1.98$, fail to reject H_0
IBM	$t = \frac{1.2683 - 1}{0.1554} = 1.726$	Since $-1.98 < t < 1.98$, fail to reject H_0
Microsoft	$t = \frac{1.4299 - 1}{0.1882} = 2.284$	Since $t > 1.98$, reject H_0
Mobil-Exxon	$t = \frac{0.4030 - 1}{0.08228} = -7.256$	Since $t < -1.98$, reject H_0

For Disney, GE, GM and IBM, we failed to reject the null hypothesis, indicating that the sample data are consistent with the conjecture that the Disney, GE, GM and IBM stocks have the same volatility as the market portfolio. For Microsoft and Mobil-Exxon, we rejected the null hypothesis, and concluded that these two stocks do not have the same volatility as the market portfolio.

Exercise 3.7 (continued)

(b) We set up the hypotheses $H_0: \beta_j \geq 1$ versus $H_1: \beta_j < 1$. The relevant test statistic, given H_0 is true, is

$$t = \frac{b_j - 1}{\text{se}\left(b_j\right)} \sim t_{(118)}$$

The rejection region is $t < -1.658$ where $t_c = t_{(0.05,118)} = -1.658$. The value of the test statistic is

$$t = \frac{0.4030 - 1}{0.08228} = -7.256$$

Since $t = -7.256 < t_c = -1.658$, we reject H_0 and conclude that Mobil-Exxon's beta is less than 1. A beta equal to 1 suggests a stock's variation is the same as the market variation. A beta less than 1 implies the stock is less volatile than the market; it is a defensive stock.

(c) We set up the hypotheses $H_0: \beta_j \leq 1$ versus $H_1: \beta_j > 1$. The relevant test statistic, given H_0 is true, is

$$t = \frac{b_j - 1}{\text{se}\left(b_j\right)} \sim t_{(118)}$$

The rejection region is $t > 1.658$ where $t_c = t_{(0.95,118)} = 1.658$. The value of the test statistic is

$$t = \frac{1.4299 - 1}{0.1882} = 2.284$$

Since $t = 2.284 > t_c = 1.658$, we reject H_0 and conclude that Microsoft's beta is greater than 1. A beta equal to 1 suggests a stock's variation is the same as the market variation. A beta greater than 1 implies the stock is more volatile than the market; it is an aggressive stock.

(d) A 95% interval estimator for Microsoft's beta is $b_j \pm t_{(0.975,118)} \times \text{se}(b_j)$. Using our sample of data the corresponding interval estimate is

$$1.4299 \pm 1.980 \times 0.1882 = (1.057, 1.803)$$

Thus we estimate, with 95% confidence, that Microsoft's beta falls in the interval 1.057 to 1.803. It is possible that Microsoft's beta falls outside this interval, but we would be surprised if it did, because the procedure we used to create the interval works 95% of the time. The problem with the interval estimate is that it is wide. We feel sure that Microsoft is more volatile than the market, but how much more is not known precisely.

Exercise 3.7 (continued)

(e) The two hypotheses are H_0: $\alpha_j = 0$ versus H_1: $\alpha_j \neq 0$. The test statistic, given H_0 is true, is

$$t = \frac{a_j}{\text{se}(a_j)} \sim t_{(118)}$$

The rejection region is $t < -1.980$ and $t > 1.980$, where $t_{(0.975,118)} = 1.980$.
The results for each company are given in the following table:

Stock	*t*-value	Decision rule
Disney	$t = \frac{-0.0010}{0.0067} = -0.152$	Since $-1.98 < t < 1.98$, fail to reject H_0
GE	$t = \frac{0.0059}{0.0049} = 1.199$	Since $-1.98 < t < 1.98$, fail to reject H_0
GM	$t = \frac{-0.0023}{0.0073} = -0.317$	Since $-1.98 < t < 1.98$, fail to reject H_0
IBM	$t = \frac{0.0068}{0.0073} = 0.940$	Since $-1.98 < t < 1.98$, fail to reject H_0
Microsoft	$t = \frac{0.0102}{0.0088} = 1.156$	Since $-1.98 < t < 1.98$, fail to reject H_0
Mobil-Exxon	$t = \frac{0.0073}{0.0039} = 1.904$	Since $-1.98 < t < 1.98$, fail to reject H_0

We do not reject the null hypothesis for any of the stocks. This indicates that the sample data is consistent with the conjecture from economic theory that the intercept term equals 0.

EXERCISE 3.8

(a) We set up the hypotheses H_0: $\beta_2 = 0$ versus H_1: $\beta_2 < 0$. The alternative $\beta_2 < 0$ is chosen because an inverse relationship is one where the dependent variable increases as the independent variable decreases, and visa versa. Thus, a negative β_2 suggests an inverse relationship between variables. The test statistic, given H_0 is true, is

$$t = \frac{b_2}{\text{se}(b_2)} \sim t_{(182)}$$

The rejection region is $t < t_{(0.05,182)} = -1.653$. The value of the test statistic is

$$t = \frac{-194.233}{10.2061} = -19.031$$

Since $t = -19.03 < -1.653$, we reject the null hypothesis that $\beta_2 = 0$ and accept the alternative that $\beta_2 < 0$. We conclude that there is a statistically significant inverse relationship between the number of house starts and the 30-year fixed interest rate.

(b) We set up the hypotheses $H_0: \beta_2 = -150$ versus $H_1: \beta_2 \neq -150$. The test statistic, given H_0 is true, is

$$t = \frac{b_2 - \beta_2}{\text{se}(b_2)} \sim t_{(182)}$$

The rejection region is $t < -1.973$ and $t > 1.973$, with $t_{(0.975,182)} = 1.973$. The value of the test statistic is

$$t = \frac{-194.233 + 150}{10.2061} = -4.334$$

Since $t = -4.334 < -1.973$, we reject the null hypothesis $\beta_2 = -150$ and accept the alternative that $\beta_2 \neq -150$. The data indicate that, if the 30-year fixed interest rate increases by 1%, house starts will not fall by 150,000.

(c) A 95% interval estimate of the slope from the regression estimated in part (a) is:

$$-194.233 \pm 1.973 \times 10.2061 = (-214.4, -174.1)$$

This interval estimate suggests that, with 95% confidence, an increase in the 30-year fixed interest rate by 1% will result in a drop in house starts of between 174,100 to 214,400 houses. We would be surprised if the true value of β_2 did not lie in this interval.

In part (b) we tested, at a 5% level of significance, whether $\beta_2 = -150$, and we came to the conclusion that $\beta_2 \neq -150$. This conclusion is consistent with our interval estimate because at a 95% level of confidence, -150 lies outside the interval. Remember the relationship between confidence intervals and hypothesis testing: At a $(1-\alpha)$ level of confidence and an α level of significance, we will not reject a null hypothesis for a hypothesized value if it falls inside the confidence interval.

EXERCISE 3.9

(a) We set up the hypotheses H_0: $\beta_2 = 0$ versus H_1: $\beta_2 > 0$. The alternative $\beta_2 > 0$ is chosen because we assume that growth, if it does influence the vote, will do so in a positive way. The test statistic, given H_0 is true, is

$$t = \frac{b_2}{\text{se}(b_2)} \sim t_{(29)}$$

The rejection region is $t > 1.699 = t_{(0.95,29)}$. The value of the test statistic is

$$t = \frac{0.6599}{0.1631} = 4.0460$$

Since t = 4.0460 > 1.699, we reject the null hypothesis that β_2 = 0 and accept the alternative that $\beta_2 > 0$. We conclude that economic growth has a positive effect on the percentage vote.

(b) A 95% interval estimate for β_2 from the regression in part (a) is:

$$b_2 \pm t_{(0.975,29)}\text{se}(b_2) = 0.6599 \pm 2.045 \times 0.1631 = (0.3264, 0.9934)$$

This interval estimate suggests that, with 95% confidence, the true value of β_2 is between 0.3264 and 0.9934. Since β_2 represents the change in percentage vote due to economic growth, we expect that a 1% increase in the growth rate will increase the percentage vote by an amount between 0.3264 to 0.9934 percent.

(c) We set up the hypotheses H_0: $\beta_2 = 0$ versus H_1: $\beta_2 < 0$. The alternative $\beta_2 < 0$ is chosen because we assume that inflation, if it does influence the vote, will do so in a negative way. The test statistic, given H_0 is true, is

$$t = \frac{b_2}{\text{se}(b_2)} \sim t_{(29)}$$

The rejection region is $t < -1.699 = t_{(0.05,29)}$. The value of the test statistic is

$$t = \frac{-0.4450}{0.5197} = -0.856$$

Since $-0.856 > -2.045$, we do not reject the null hypothesis. There is not enough evidence to suggest inflation has a negative effect on the vote.

(d) A 95% interval estimate for β_2 from the regression in part (c) is:

$$b_2 \pm t_{(0.975,29)}\text{se}(b_2) = -0.4450 \pm 2.045 \times 0.5197 = (-1.508, 0.618)$$

This interval estimate suggests that, with 95% confidence, the true value of β_2 is between -1.508 and 0.618. It suggests that a 1% increase in the inflation rate could increase or decrease or have no effect on the percentage vote.

EXERCISE 3.10

(a) The coefficient β_2 represents the increase in price from an extra square foot of living area. We can refer to it as the marginal price per square foot.

(i) A 95% interval estimate of β_2 for all houses is:

$$b_2 \pm t_{(0.975,1078)} \times \text{se}(b_2) = 92.747 \pm 1.962 \times 2.4105 = (88.02, 97.48)$$

We estimate, with 95% confidence, that the marginal price per square foot for all houses lies between \$88.02 and \$97.48.

(ii) A 95% interval estimate of β_2 for town houses is:

$$b_2 \pm t_{(0.975,68)} \times \text{se}(b_2) = 55.585 \pm 1.995 \times 7.0999 = (41.42, 69.75)$$

We estimate, with 95% confidence, that the marginal price per square foot for town houses lies between \$41.42 and \$69.75.

(iii) A 95% interval estimate of β_2 for French style houses is:

$$b_2 \pm t_{(0.975,95)} \times \text{se}(b_2) = 184.167 \pm 1.985 \times 10.1626 = (163.99, 204.34)$$

We estimate, with 95% confidence, that the marginal price per square foot for French style houses lies between \$163.99 and \$204.34.

These confidence interval estimates tell us that town houses have a lower marginal price per square foot compared to the average, and also that French style houses have a much higher marginal price per square foot than all houses. Furthermore, we see that the narrowest confidence interval is that for all houses, reflecting the fact that the larger sample size provides more information, leading to a smaller standard error and more precise estimation.

(b) The results for testing the hypotheses H_0: $\beta_2 = 80$ versus H_1: $\beta_2 \neq 80$ are given in the following table. In each case the test statistic is $t = (b_2 - 80)/\text{se}(b_2)$ which has a $t_{(N-2)}$ distribution if H_0 is true. The rejection region is $t < -t_c$ and $t > t_c$ where $t_c = t_{(0.975,N-2)}$.

Sample	t-value	$N-2$	t_c	Decision rule
All houses	$t = \dfrac{92.7473 - 80}{2.4105} = 5.29$	1078	1.962	$t >$ 1.962, reject H_0
Town houses	$t = \dfrac{55.5853 - 80}{7.0999} = -3.44$	68	1.995	$t < -1.995$, reject H_0
French Style	$t = \dfrac{184.1667 - 80}{10.1626} = 10.25$	95	1.985	$t >$ 1.985, reject H_0

All cases lead to the rejection of the null hypothesis. We conclude that an additional square foot does not add \$80 to the average sale price of all houses, the sale price of town houses, nor the sale price of French style houses.

EXERCISE 3.11

(a) For all houses in sample:

Hypotheses: $H_0: \beta_2 = 80$ against $H_1: \beta_2 \neq 80$

Calculated *t*-value: $t = (81.3890 - 80)/1.9185 = 0.724$

Critical *t*-value: $\pm t_c = \pm t_{(0.975,878)} = \pm 1.963$

Decision: Do not reject H_0 because $-1.963 < 0.724 < 1.963$.

We conclude that the data is consistent with the conjecture that an additional square foot of living space is associated with an increase in the sale price of the house by \$80.

(b) For houses that are vacant at time of sale:

Hypotheses: $H_0: \beta_2 = 80$ against $H_1: \beta_2 \neq 80$

Calculated *t*-value: $t = (69.9080 - 80)/2.2675 = -4.45$

Critical *t*-value: $\pm t_c = \pm t_{(0.975,463)} = \pm 1.965$

Decision: Reject H_0 because $-4.45 < -1.965$

We conclude that, for houses that are vacant at time of sale, an additional square foot of living space is not associated with an increase in the sale price of the house by \$80.

(c) For houses that are occupied at time of sale:

Hypotheses: $H_0: \beta_2 = 80$ against $H_1: \beta_2 \neq 80$

Calculated *t*-value: $t = (89.2588 - 80)/3.0394 = 3.05$

Critical *t*-value: $\pm t_c = \pm t_{(0.975,413)} = \pm 1.966$

Decision: Reject H_0 because $3.05 > 1.966$.

We conclude that, for houses that are occupied at time of sale, an additional square foot of living space is not associated with an increase in the sale price of the house by \$80.

(d) For houses that are occupied at time of sale:

Hypotheses: $H_0: \beta_2 \leq 80$ against $H_1: \beta_2 > 80$

Calculated *t*-value: $t = (89.2588 - 80)/3.0394 = 3.05$

Critical *t*-value: $t_c = t_{(0.95,413)} = 1.649$

Decision: Reject H_0 because $3.05 > 1.649$

We conclude that, for houses that are occupied at time of sale, an additional square foot of living space increases the sale price of the house by more than \$80.

(e) For houses that are vacant at time of sale:

Hypotheses: $H_0: \beta_2 \geq 80$ against $H_1: \beta_2 < 80$

Calculated *t*-value: $t = (69.9080 - 80)/2.2675 = -4.45$

Critical *t*-value: $t_c = t_{(0.05,463)} = -1.648$

Decision: Reject H_0 because $-4.45 < -1.648$

We conclude that, for houses that are vacant at time of sale, an additional square foot of living space increases the sale price by less than \$80.

Exercise 3.11 (continued)

(f) (i) A 95% interval estimate for β_2 from the full sample is given by

$$b_2 \pm t_{(0.975,878)} \times \text{se}(b_2) = 81.389 \pm 1.963 \times 1.9185 = (77.62, 85.15)$$

(ii) A 95% interval estimate for β_2 for houses vacant at the time of sale is given by

$$b_2 \pm t_{(0.975,463)} \times \text{se}(b_2) = 69.908 \pm 1.965 \times 2.2675 = (65.45, 74.36)$$

(iii) A 95% interval estimate for β_2 for houses occupied at the time of sale is given by

$$b_2 \pm t_{(0.975,413)} \times \text{se}(b_2) = 89.259 \pm 1.966 \times 3.039 = (83.28, 95.23)$$

EXERCISE 3.12

(a) Estimated equation:

$$\widehat{WAGE} = 8.6658 + 0.0824 EXPER$$
$$\text{(se)} \quad (0.3787) \quad (0.0173)$$
$$(t) \quad (22.88) \quad (4.77)$$

The estimated equation tells us that with every year of experience the associated increase in hourly wage is \$0.0824. Furthermore, it tells us that the average wage for those without experience is \$8.6658. The relatively large t-values suggest that the least squares estimates are statistically significant at a 5% level of significance.

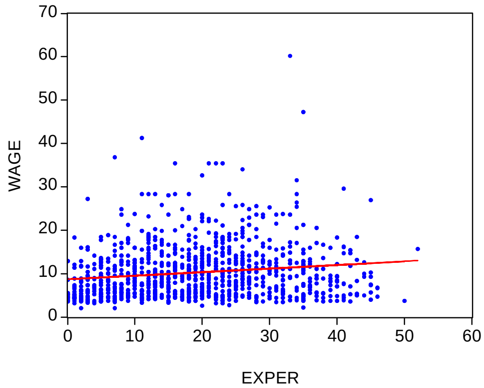

Figure xr3.12(a) Fitted regression line and observations

(b) Hypotheses: $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 > 0$
The test statistic, given H_0 is true, is

$$t = \frac{b_2}{\text{se}(b_2)} \sim t_{(998)}$$

Calculated t-value: $t = (0.0824)/0.0173 = 4.769$
Critical t-value: $t_c = t_{(0.95,998)} = 1.646$
Decision: Reject H_0 because $4.769 > 1.646$
We conclude that the slope of the relationship, β_2, is statistically significant. There is a positive relationship between the hourly wage and a worker's experience.

Exercise 3.12 (continued)

(c) (i) For females, the estimated equation is:

$$\widehat{WAGE} = 8.4747 + 0.0209 EXPER$$
$$(\text{se}) \quad (0.4797) \; (0.0218)$$
$$(t) \quad (17.67) \; (0.958)$$

With every extra year of experience the associated increase in average hourly wage for females is \$0.0209. This estimate is not significantly different from zero, however. The average wage for females without experience is \$8.4747.

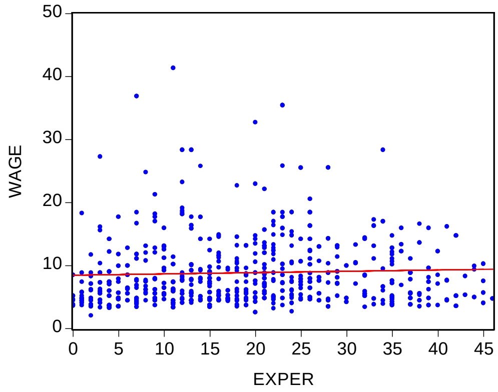

Figure xr3.12(b) Fitted regression line and observations for females

(c) (ii) For males, the estimated equation is:

$$\widehat{WAGE} = 8.8200 + 0.1448 EXPER$$
$$(\text{se}) \quad (0.5549) \; (0.0254)$$
$$(t) \quad (15.89) \; (5.698)$$

With every extra year of experience, the associated increase in average hourly wage for males is \$0.1448. The average wage for males without experience is \$8.8200.

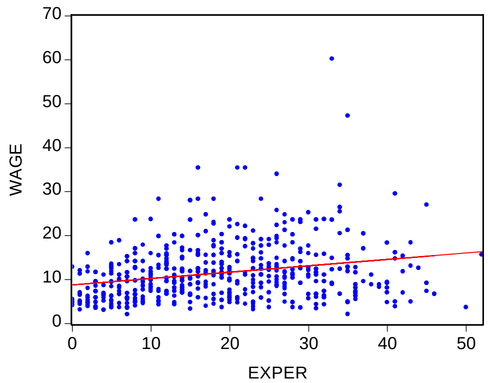

Figure xr3.12(c) Fitted regression line and observations for males

Exercise 3.12(c) (continued)

(c) (iii) For blacks, the estimated equation is:

$$\widehat{WAGE} = 6.0054 + 0.1197 EXPER$$
$$(\text{se}) \quad (0.9973) \quad (0.0461)$$
$$(t) \quad (6.022) \quad (2.594)$$

With every extra year of experience, the associated increase in average hourly wage for blacks is \$0.1197. The average wage for blacks without experience is \$6.0054.

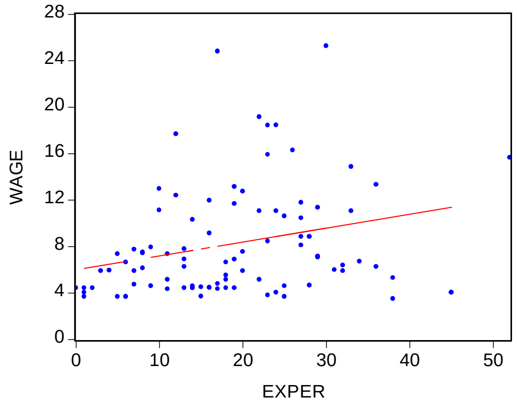

Figure xr3.12(d) Fitted regression line and observations for blacks

(c) (iv) For white males, the estimated equation is:

$$\widehat{WAGE} = 9.0315 + 0.1451 EXPER$$
$$(\text{se}) \quad (0.5808) \quad (0.0266)$$
$$(t) \quad (15.55) \quad (5.452)$$

With every extra year of experience the associated increase in average hourly wage for white males is \$0.1451. The average wage for white males without experience is \$9.0315.

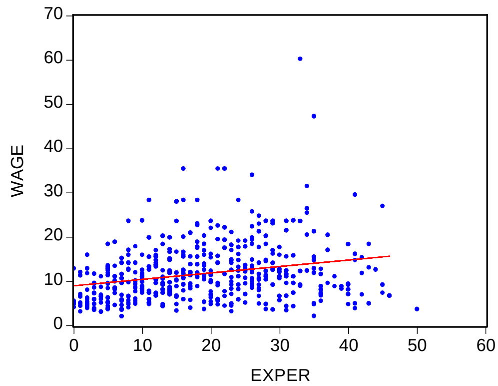

Figure xr3.12(e) Fitted regression line and observations for white males

Exercise 3.12(c) (continued)

(c) Comparing the estimated wage equations for the four categories, we find that experience counts the most, or leads to the largest increase in wages, for white males. The effect is only slightly less for males in general. It is less for blacks and very small for females. For those with no experience the wage ranking is white males, males, females, blacks.

(d) Residual plots

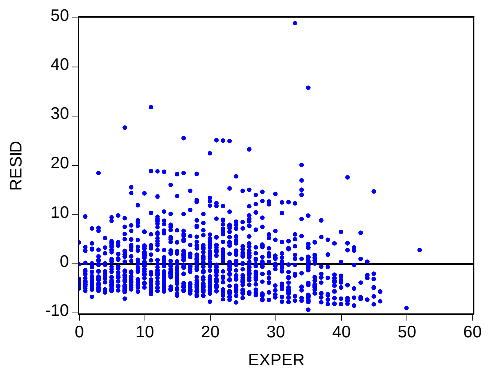

Figure xr3.12(f) Plotted residuals for full sample regression

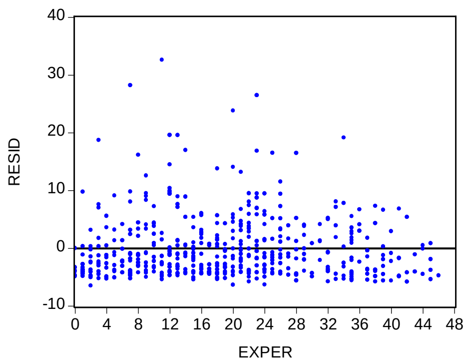

Figure xr3.12(g) Plotted residuals for female regression

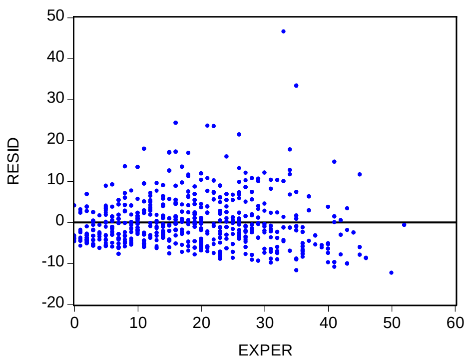

Figure xr3.12(h) Plotted residuals for male regression

Exercise 3.12(d) (continued)

(d)

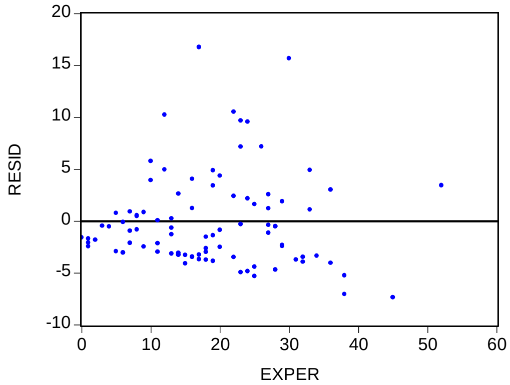

Figure xr3.12(i) Plotted residuals for black regression

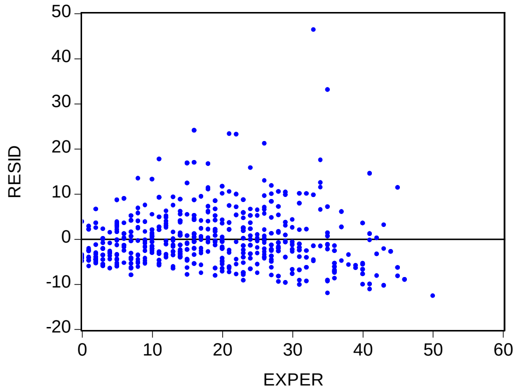

Figure 3.12(j) Plotted residuals for white male regression

The main observation that can be made from all the residual plots is that the pattern of positive residuals is quite different from the pattern of negative residuals. There are very few negative residuals with an absolute magnitude larger than 10, whereas the positive residuals are often larger than 10, with a few very large ones, and one over 40. These characteristics suggest a distribution of the errors that is not normally distributed, but skewed to the right.

EXERCISE 3.13

(a) Estimated equation:

$$\widehat{WAGE} = 8.5837 + 0.0842 EXPER$$
$$\text{(se)} \quad (0.1738) \; (0.0078)$$
$$(t) \quad (49.40) \; (10.76)$$

With every extra year of experience the associated increase in hourly wage is \$0.0842. The average wage for those without experience is \$8.5837. The relatively large *t*-values imply the least squares estimates are statistically significant at a 5% level of significance.

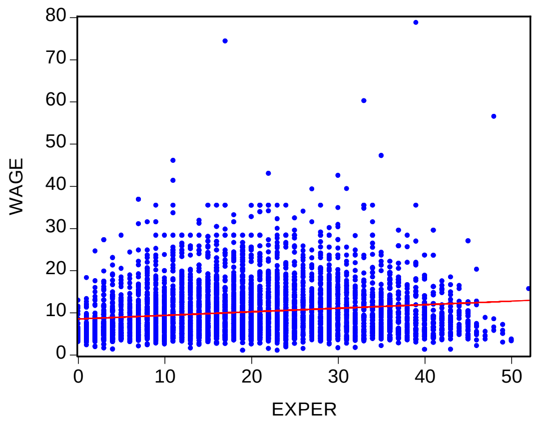

Figure xr3.13(a) Fitted regression line and observations using all data

(b) Hypotheses: $H_0: \beta_2 = 0$ against $H_1: \beta_2 > 0$. The test statistic, given H_0 is true, is

$$t = \frac{b_2}{\text{se}(b_2)} \sim t_{(4731)}$$

Calculated *t*-value: $t = (0.0842)/0.0078 = 10.76$

Critical *t*-value: $t_c = t_{(0.95,4731)} = 1.645$

Decision: Reject H_0 because $t = 10.76 > t_c = 1.645$

We conclude that the slope of the relationship, β_2, is statistically significant. There is a positive relationship between the hourly wage and a worker's experience.

Exercise 3.13 (continued)

(c) (i) For females, the estimated equation is:

$$\widehat{WAGE} = 8.0375 + 0.0501EXPER$$
$$\text{(se)} \quad (0.2285) \; (0.0103)$$
$$(t) \quad (35.18) \; (4.856)$$

With every extra year of experience the associated increase in average hourly wage for females is \$0.0501. The average wage for females without experience is \$8.0375.

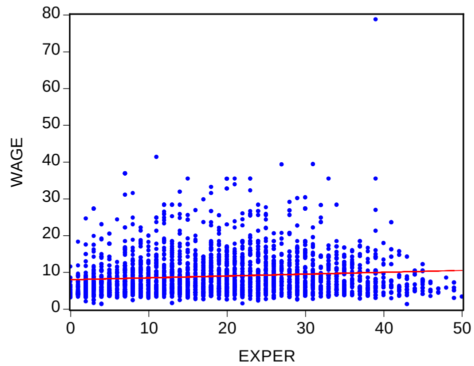

Figure 3.13(b) Fitted regression line and observations for females

(c) (ii) For males, the estimated equation is:

$$\widehat{WAGE} = 9.1170 + 0.1153EXPER$$
$$\text{(se)} \quad (0.2510) \; (0.0113)$$
$$(t) \quad (36.32) \; (10.216)$$

With every extra year of experience the associated increase in average hourly wage for males is \$0.1153. The average wage for males without experience is \$9.1170.

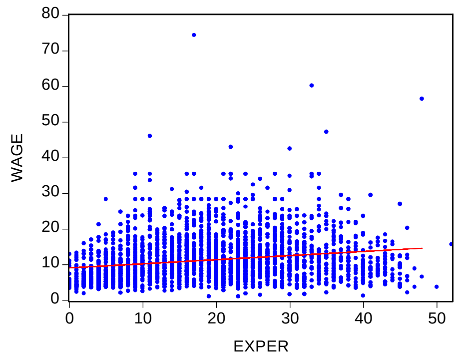

Figure xr3.13(c) Fitted regression line and observations for males

Exercise 3.13(c) (continued)

(c) (iii) For blacks, the estimated equation is:

$$\widehat{WAGE} = 7.3825 + 0.0667 EXPER$$
$$\text{(se)} \quad (0.5002) \; (0.0233)$$
$$(t) \quad (14.76) \; (2.860)$$

With every extra unit of experience the associated increase in average hourly wage for blacks is \$0.0667. The average wage for blacks without experience is \$7.3825.

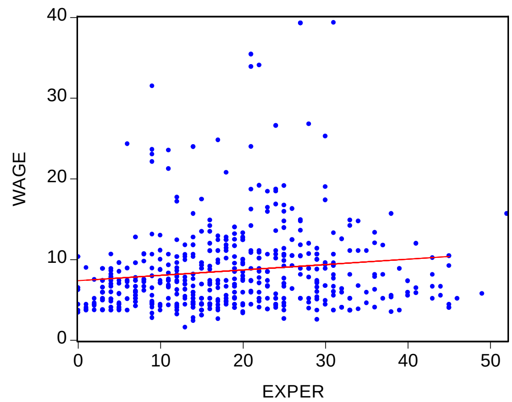

Figure xr3.13(d) Fitted regression line and observations for blacks

(c) (iv) For white males, the estimated equation is:

$$\widehat{WAGE} = 9.2606 + 0.1164 EXPER$$
$$\text{(se)} \quad (0.2644) \; (0.0118)$$
$$(t) \quad (35.02) \; (9.847)$$

With every extra year of experience the associated increase in average hourly wage for white males is \$0.1164. The average wage for white males without experience is \$9.2606.

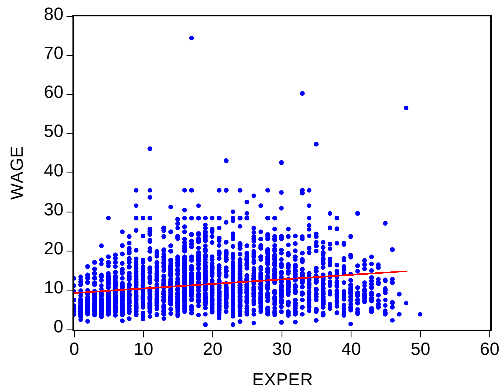

Figure xr3.13(e) Fitted regression line and observations for white males

Exercise 3.13(c) (continued)

(c) Comparing the estimated wage equations for the four categories, we find that experience counts the most, or leads to the largest increase in wages, for white males. The effect is only slightly less for males in general. For blacks experience is worth slightly more than half of what it is for white males. For females experience is worth slightly less than half of what it is for white males. For those with no experience the wage ranking is white males, males, females, blacks.

(d) Residual plots

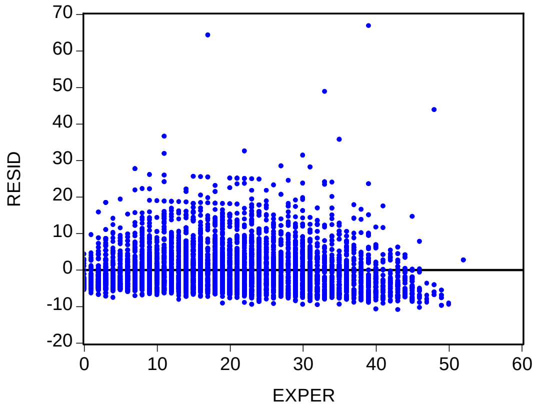

Figure xr3.13(f) Plotted residuals for full sample regression

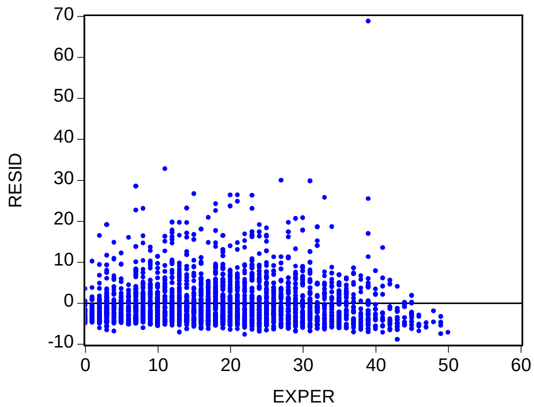

Figure xr3.13(g) Plotted residuals for female regression

Exercise 3.13(d) (continued)

(d)

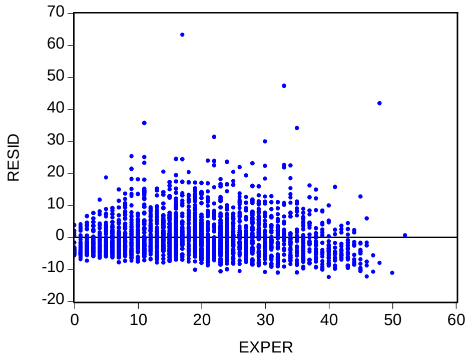

Figure xr3.13(h) Plotted residuals for male regression

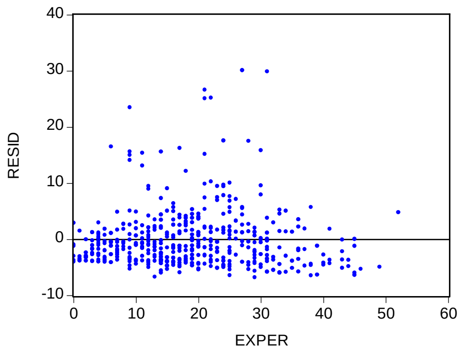

Figure xr3.13(i) Plotted residuals for black regression

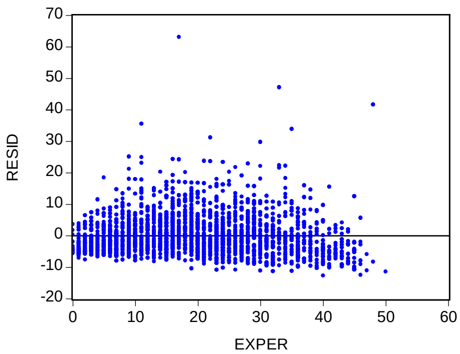

Figure xr3.13(j) Plotted residuals for white male regression

In all residual plots the pattern of positive residuals is quite different from the pattern of negative residuals. There are very few negative residuals with an absolute magnitude larger than 10, whereas the positive residuals are often larger than 10, with a few very large ones, and one over 40. These characteristics suggest a distribution of the errors that is not normally distributed, but skewed to the right.