Solutions to Exercises in Chapter 5

5.1 (a) The required interval is $b_1 \pm t_c \text{se}(b_1)$ where $b_1 = 40.768$, $t_c = 2.024$ and $\text{se}(b_1) = 22.139$. That is

$$40.768 \pm 2.024 \times 22.139 = (-4.04, 88.57)$$

We estimate that β_1 lies between –4.04 and 85.57. In repeated samples 95% of similarly constructed intervals would contain β_1.

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

$$t_1 = \frac{b_1 - \beta_1}{\text{se}(b_1)} = \frac{40.768 - 0}{22.139} = 1.84$$

Since the 5% critical value $t_c = 2.024$ exceeds 1.84, we do not reject H_0. The data do not reject the zero-intercept hypothesis.

(c) The *p*-value 0.0734 represents the sum of the areas under the *t* distribution to the left of –1.84 and to the right of 1.84. Since the *t* distribution is symmetric, each of the tail areas will be $0.0734/2 = 0.0367$. Each of the areas in the tails beyond the critical values $\pm t_c = \pm 2.02$ is 0.025. Since 0.025 < 0.0367, H_0 is not rejected. From Figure 5.1 we can see that having a *p*-value > 0.05 is equivalent to having $-t_c < t < t_c$.

(d) Testing $H_0: \beta_1 = 0$ against $H_1: \beta_1 > 0$, requires the same *t*-value as in part (b), *t* = 1.84. 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, $t_c = 1.686$. Since *t* = 1.84 > t_c = 1.69, H_0 is rejected and we conclude that the intercept is positive. In this case *p*-value = *P*(*t* > 1.84) = 0.0367. We see from Figure 5.2 that having the *p*-value < 0.05 is equivalent to having *t* > 1.69.

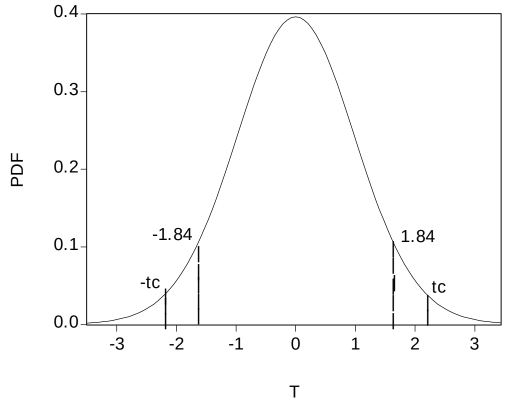

Figure 5.1 Critical and Observed *t* Values for Two-Tailed Test in Question 5.1(c)

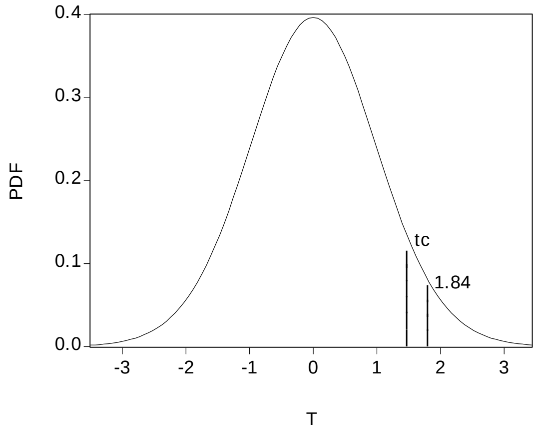

Figure 5.2 Observed and Critical *t* Vlaues for One-Tailed Test in Question 1(d)

(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 $H_1 : \beta_k \neq c$, nonrejection 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. Strictly speaking, we cannot make probability statements about constant unknown parameters like β_1. Thus, if 95% confident is regarded as synonymous with a 95% probability, the statement is false. However, if we treat the term "confident" more loosely, the statement could be regarded as true. The probability of accepting $H_1 : \beta_1 > 0$ when it is false is 0.05. Thus, after we have accepted H_1, in this sense we can say we are 95% confident that β_1 is positive.

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

(b) The 95% confidence interval for β_2 is given by

$$b_2 \pm t_c \text{se}(b_2) = 0.0761 \pm 2.074 \times 0.04449 = (-0.016, 0.168)$$

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

(c) For testing $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$, the p-value is 0.1012 It is given as the sum of the areas under the t-distribution to the left of -1.711 and to the right of 1.711. The area in each of these tails is $0.1012/2 = 0.0506$. We do not reject H_0 because, for $\alpha = 0.05$, p-value > 0.05.

(d) The predicted quality rating of a draftsman with 5 years experience is

$$\widehat{\text{rating}} = 3.2038 + 0.076118 \times 5 = 3.58$$

The steps required to compute a prediction interval will depend on the software you are using. Most software will give you a standard error of the forecast error $\text{se}(f)$, obtained as the square root of

$$\hat{\text{var}}(\hat{y}_0 - y_0) = \hat{\sigma}^2\left(1 + \frac{1}{T} + \frac{(5-\bar{x})^2}{\sum(x_t - \bar{x})^2}\right)$$

Then, a 95% prediction interval can be obtained from

$$\hat{y}_0 \pm t_c \text{se}(f) = 3.58 \pm 2.074\,\text{se}(f)$$

5.3 (a) The estimated slope coefficient indicates that, on average, a 1% increase in real total expenditure leads to a 0.322% increase in real food expenditure. It is the elasticity of food expenditure with respect to total expenditure.

(b) For testing $H_0 : \beta_2 = 0.25$ against the alternative $H_1 : \beta_2 \neq 0.25$, we compute the *t* value, assuming H_0 is true, as

$$t = \frac{b_2 - \beta_2}{\text{se}(b_2)} = \frac{0.3224 - 0.25}{0.01945} = 3.72$$

The critical value for a two-tailed test, a 0.01 significance level and 23 degrees of freedom is $t_c = 2.807$. Since $t = 3.72 > t_c = 2.807$, we reject H_0 and conclude the elasticity for food expenditure is not equal to 0.25.

(c) A 95% confidence interval for β_2 is given by

$$b_2 \pm t_c \text{se}(b_2) = 0.3224 \pm 2.0687 \times 0.019449 = (0.282,\ 0.363)$$

(d) The error terms must be normally and independently distributed with zero mean and constant variance. This assumption is necessary for the ratio $(b_2 - \beta_2)/\text{se}(b_2)$ to have a *t*-distribution. If the sample size was 100 we could dispense with the assumption of a normally distributed error and rely on a central limit theorem to show that $(b_2 - \beta_2)/\text{se}(b_2)$ has an approximate *t* or normal distribution.

(d) Omitting an important variable will bias the estimate of β_2 and make the formulas for computing the test statistic and confidence interval incorrect.

5.4 Since the reported *t*-statistic is given by $t = b/\text{se}(b_2)$ and the estimated variance is $\hat{\text{var}}(b) = [\text{se}(b)]^2$, in this case we have

$$\hat{\text{var}}(b) = (b/t)^2 = (-3782.196/-6.607)^2 = 32{,}7702$$

5.5 (a) For *p* = 0.005, the null hypothesis would be rejected at both the 5% and 1% levels of significance.

(b) For *p* = 0.0108, the null hypothesis would be rejected at the 5% level of significance, but not at the 1% level of significance.

5.6 (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: $\pm t_c = \pm 2.819$
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: $t_c = 2.508$
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: $t_c = -1.717$
Decision: Do not reject H_0 because $t = 3.78 > t_c = -1.717$.

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

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

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

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

5.7 (a) When estimating $E(y_0)$, we are estimating the average value of *y* for all observational units with an *x*-value of x_0. When predicting y_0, we are predicting the value of *y* for one observational unit with an *x*-value of x_0. The first exercise does not involve the random error e_0; the second does.

(b)
$$E(b_1 + b_2 x_0) = E(b_1) + E(b_2)x_0 = \beta_1 + \beta_2 x_0$$

$$\mathrm{var}(b_1 + b_2 x_0) = \mathrm{var}(b_1) + x_0^2\,\mathrm{var}(b_2) + 2x_0\,\mathrm{cov}(b_1, b_2)$$

$$= \frac{\sigma^2 \sum x_t^2}{T\sum(x_t - \bar{x})^2} + \frac{\sigma^2 x_0^2}{\sum(x_t - \bar{x})^2} - \frac{2\sigma^2 x_0 \bar{x}}{\sum(x_t - \bar{x})^2}$$

$$= \frac{\sigma^2\left(\sum(x_t - \bar{x})^2 + T\bar{x}^2\right)}{T\sum(x_t - \bar{x})^2} + \frac{\sigma^2(x_0^2 - 2x_0\bar{x})}{\sum(x_t - \bar{x})^2}$$

$$= \sigma^2\left(\frac{1}{T} + \frac{x_0^2 - 2x_0\bar{x} + \bar{x}^2}{\sum(x_t - \bar{x})^2}\right) = \sigma^2\left(\frac{1}{T} + \frac{(x_0 - \bar{x})^2}{\sum(x_t - \bar{x})^2}\right)$$

5.8 It is not appropriate to say that $E(\hat{y}_0) = y_0$ because y_0 is a random variable.

$$E(\hat{y}_0) = \beta_1 + \beta_2 x_0 \neq \beta_1 + \beta_2 x_0 + e_0 = y_0$$

We need to include y_0 in the expectation so that

$$E(\hat{y}_0 - y_0) = E(\hat{y}_0) - E(y_0) = \beta_1 + \beta_2 x_0 - \left(\beta_1 + \beta_2 x_0 + E(e_0)\right) = 0.$$

5.9 The estimated equation is

$$\hat{price}_t = -426.7 + 46.005\, sqft_t$$
$$(5061.2) \quad (2.803) \quad \text{(se)}$$

(a) A 95% confidence interval for β_2 is

$$b_2 \pm t_c \text{se}(b_2) = 46.005 \pm 1.97 \times 2.803 = (40.48, 51.53)$$

(b) To test $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 > 0,$ we compute the *t*-value $t = 46.01/2.803 = 16.41$. At a 5% significance level the critical value for a one-tailed test and 211 degrees of freedom is $t_c = 1.652$. Since t = 16.41 > t_c = 1.65, H_0 is rejected. We conclude there is a positive relationship between house size and price.

(c) To test $H_0 : \beta_2 = 50$ against $H_1 : \beta_2 \neq 50,$ we compute the *t*-value

$$t = (46.005 - 50)/2.803 = -1.43.$$

At a 5% significance level the critical values for a two-tailed test and 211 degrees of freedom are $\pm t_c = \pm 1.97$. Since t = –1.43 lies between –1.97 and 1.97, we do not reject H_0. The data are not in conflict with the hypothesis that says the value of a square foot of housing space is $50.

(d) The point prediction for house price for a house with 2000 square feet is

$$\hat{price}_0 = -426.7 + 46.005 \times 2000 = 91{,}583$$

A 95% interval prediction for house price for a house with 2000 square feet is

$$\hat{price}_0 \pm t_c \text{se}(f) = 91583 \pm 1.97 \times 8202.6 = (75424, 107742)$$

5.10 $\hat{y}_0 = b_1 + b_2 x_0 = 1 + 1 \times 5 = 6$

$$\hat{\text{var}}(f) = \hat{\sigma}^2 \left(1 + \frac{1}{T} + \frac{(x_0 - \bar{x})^2}{\sum (x_t - \bar{x})^2} \right) = 5.3333 \left(1 + \frac{1}{5} + \frac{(5-1)^2}{10} \right) = 14.9332$$

$$\text{se}(f) = \sqrt{14.9332} = 3.864$$

5.11 Using appropriate computer software we find that

$b_1 = 0.46562$ $\widehat{\text{var}}(b_1) = 0.0138097$ $\text{se}(b_1) = 0.1175$

$b_2 = 0.29246$ $\widehat{\text{var}}(b_2) = 0.00016705$ $\text{se}(b_2) = 0.01292$

(a) The interval estimators for β_1 and β_2 are given by $b_1 \pm t_c\text{se}(b_1)$ and $b_2 \pm t_c\text{se}(b_2)$ where $t_c = 2.16$ is the 5% critical value with 13 degrees of freedom. Therefore, the interval estimate for β_1 is

$$0.46562 \pm 2.16(0.1175) = (0.2118, 0.7195)$$

The interval estimate for β_2 is

$$0.29246 \pm 2.16(0.01292) = (0.2645, 0.3204)$$

If we use the interval estimators to compute a large number of interval estimates like these, in repeated samples, 95% of these intervals will contain β_1 and β_2.

(b) To test the hypothesis that $\beta_1 = 0$ against the alternative it is positive, we set up the hypotheses H_0: $\beta_1 = 0$ vs H_1: $\beta_1 > 0$. The test statistic is $t = b_1 / \text{se}(b_1)$. Since the test is a one-tailed test, at a 5% significance level the rejection region is $t > 1.771$. The value of the test statistic is $t = 0.46562/0.1175 = 3.962$. Since $t = 3.962 > t_c = 1.771$, we reject the null hypothesis indicating that the data are not compatible with $\beta_1 = 0$; they support the hypothesis $\beta_1 > 0$.

(c) The hypotheses are H_0: $\beta_2 = 0$ vs H_1: $\beta_2 > 0$. The test statistic is $t = b_2 / \text{se}(b_2)$. For a 5% significance level and a one-tailed test, the rejection region is $t_c > 1.771$. The value of the test statistic is $t = 0.29246/0.01292 = 22.628$. Since $t = 22.628 > t_c = 1.771$, we reject the null hypothesis and conclude that the data are not compatible with $\beta_2 = 0$; they support the alternative hypothesis that β_2 is positive.

(d) The marginal product of the input is dy/dx which is equal to β_2. Thus, the hypotheses are H_0: $\beta_2 = 0.35$ vs H_1: $\beta_2 \neq 0.35$. The test statistic is $t = (b_2 - 0.35)/\text{se}(b_2)$. At a 5% significance level, the rejection region is $|\,t\,| > 2.160$. The value of the test statistic is $t = (0.29246 - 0.35)/0.01292 = -4.452$. Since $t = -4.452 < -t_c = -2.160$, we reject the null hypothesis and conclude that the data are not compatible with $\beta_2 = 0.35$. The data do not support the hypothesis that the marginal product of the input is 0.35.

(e) The sampling variability for the input level 8 is

$$\widehat{\text{var}}(\hat{y}_0 - y_0) = \hat{\sigma}^2\left[1 + \frac{1}{15} + \frac{(8-\bar{x})^2}{\sum(x_t - \bar{x})^2}\right] = 0.04677\left(1 + \frac{1}{15} + \frac{(8-8)^2}{280}\right) = 0.04989$$

The sampling variability for the input level 16 is

$$\widehat{\text{var}}(\hat{y}_0 - y_0) = \hat{\sigma}^2\left[1 + \frac{1}{15} + \frac{(16-\bar{x})^2}{\sum(x_t - \bar{x})^2}\right] = 0.04677\left(1 + \frac{1}{15} + \frac{(16-8)^2}{280}\right) = 0.06058$$

The prediction error variance is smallest at the sample mean $\bar{x} = 8$ and becomes larger the further x_0 is from $\bar{x}$. Since $x_0 = 16$ is outside the sample range, the prediction error variance in this case is greater than the squares of all the standard errors in the table in part (b). The variance of the prediction error refers to the variance of $(\hat{y}_0 - y_0)$ in repeated samples, where, for each sample, we have different least squares estimates b_1 and b_2, and hence a different predictor $\hat{y}_0$, as well as a different realized future value y_0.

5.12 The least squares estimated demand equation is

$$\widehat{\ln q_t} = 7.1528 - 1.9273 \ln p_t$$
$$(0.0442) \quad (0.2241)$$

The figures in parentheses are standard errors.

(a) To test the hypothesis that the elasticity of demand is equal to −1, we set up the hypotheses H_0: $\beta_2 = -1$ versus H_1: $\beta_2 \neq -1$. The test statistic is $t = \left[b_2 - (-1)\right]/\text{se}(b_2)$. With 10 degrees of freedom and a 5% significance level the rejection region is $|t| > 2.228$. The value of the test statistic is

$$t = \frac{-1.9273 + 1}{0.2241} = -4.138.$$

Since $t = -4.138 < -2.228$, we reject the null hypothesis and conclude that the elasticity of demand for hamburgers is not equal to −1.

(b) The predicted logarithm of the number of hamburgers sold when price is \$2 is

$$\ln(\hat{q}_0) = 7.1527 - 1.9269\ \ln(2) = 5.8168$$

and so a point prediction for the number of hamburgers is

$$\hat{q}_0 = \exp(5.8168) = 335.9$$

Thus, if the price is \$2, it is predicted that 336 hamburgers will be sold.

To find an interval prediction for the number of hamburgers, we first find an interval prediction for the logarithm of the number of hamburgers. A 95% interval predictor for the logarithm is

$$\ln(\hat{q}_0) \pm 2.228\ \text{se}(f)$$

Now, $\text{se}(f) = 0.135783$, and so a 95% interval prediction for $\ln(q_0)$ when $\ln(p_0) = \ln(2) = 0.693147$ is

$$5.8168 \pm 2.228(0.13578) = (5.5143, 6.1194)$$

Given $\exp(5.5143) = 248$ and $\exp(6.1194) = 455$, a 95% interval prediction for the number of hamburgers sold is (248, 455).

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

$$\hat{y}_t = \underset{(7.3835)}{6.8550} + \underset{(0.1121)}{3.8802}\, x_t$$

where the numbers in parentheses are corresponding standard errors.

(b) The relationship in part (a) indicates that, as income increases, the amount of life insurance increases, as is expected. The value of $b_1 = 6.8550$ implies that if a family has no income, then they would purchase \$6855 worth of insurance. It is necessary to be careful of this interpretation because there is no data for families with an income close to zero. Parts (i), (ii) and (iii) discuss the slope coefficient.

(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

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

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 $(T–2)$ degrees of freedom at the α level of significance.

(iii) 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 (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.

(iv) 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 $y_t = \beta_1 + \beta_2 x_t + e_t$ where y_t is life insurance cover by the t-th household, x_t is household income, β_1 and β_2 are unknown parameters that describe the relationship, and e_t 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 y and x, and apply the least-squares estimation procedure. The estimated equation, with standard errors in parentheses, is given in part (a). The point estimate for the response of life-insurance cover to an income increase of \$1000 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 where changes may occur as a result of income changes.

(c) To test the hypothesis that the slope of the relationship is one, we proceed as we did in part (b)(iii), 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 hypothesis that the amount of life insurance increases at the same rate as income increases.

(d) If income = \$100,000, then the predicted amount of life insurance is

$$\hat{y}_0 = 6.8550 + 3.8802(100) = 394.875.$$

That is, the predicted life insurance is \$394,875 for an income of \$100,000.

5.14 (a) A 95% interval estimator for β_2 is $b_2 \pm 2.145\ \text{se}(b_2)$. 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, if there is learning, that unit costs of production would decline as cumulative production increased. 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.

(c) The prediction of the log of unit cost when $q_0 = 2000$ is

$$\ln(\hat{u}_0) = 6.0191 - 0.3857 \ln(2000) = 3.0875$$

The 95% prediction interval for the unit cost of production is

$$\exp\left(\ln(\hat{u}_0) \pm t_c \text{se}(f)\right) = \exp(3.0875 \pm 2.1448 \times 0.051474) = (19.63, 24.48)$$

(d) How quickly workers learn to perform their tasks, and hence the speed with which unit costs of production fall as production proceeds, are important pieces of information to managers of production plants. To investigate this relationship for the production of titanium dioxide by the DuPont Corporation, we set up the economic model $u = u_1 q^a$ where u is the unit cost of production after producing q units, u_1 is the unit cost of production for the first unit and a is the elasticity of unit costs with respect to cumulative production. A corresponding statistical model is

$$\ln(u_t) = \beta_1 + \beta_2 \ln(q_t) + e_t$$

where the subscript t denotes the year for which observations u_t and q_t were recorded, $\beta_1 = \ln(u_1)$, $\beta_2 = a$ and e_t is assumed to be an uncorrelated random error with zero mean and constant variance.

Using 16 observations from 1955 to 1970, the estimated relationship is

$$\ln(\hat{u}_t) = 6.019 - 0.3859 \ln(q_t)$$
$$(0.275) \quad (0.0360)$$

Both coefficients have the expected signs and are significantly different from zero at a 0.01 level of significance. The estimated cost of the first unit produced is $\hat{u}_1 = \exp(\hat{\beta}_1) = \exp(6.019) = 411.2$. A 1% increase in production decreases unit costs by 0.386%. Using a 95% interval estimate to assess the reliability of this point estimate, we estimate that the percentage decline in unit costs lies between 0.463 and 0.308. The DuPont management can use this information to predict future unit costs. For example, after producing 2000 units, the unit cost of production is predicted to fall to a value within the 95% interval (19.63, 24.48).

5.15 (a) We set up the hypotheses H_0: $\beta_2 = 1$ versus H_1: $\beta_2 < 1$. The relevant test statistic, given H_0 is true, is

$$t = \frac{b_2 - 1}{\text{se}(b_2)} \sim t_{(118)}$$

The rejection region is $t < -1.658$. The value of the test statistic is

$$t = \frac{0.7147 - 1}{0.08562} = -3.332$$

Since $t = -3.332 < t_c = -1.658$, we reject H_0 and conclude that Mobil Oil'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.

(b) The estimated model is given by $\hat{y}_t = 0.004241 + 0.7147\, x$ where x is the risk premium of the market portfolio and y is Mobil's risk premium. Predicting Mobil's premium when $x = 0.01$, we have

$$\hat{y}_0 = 0.004241 + 0.7147 \times 0.01 = 0.01139$$

When $x = 0.1$, the prediction is

$\hat{y}_0$ = 0.004241 + 0.7147 × 0.1 = 0.07571

Interval estimates for each value of x are given by $\hat{y}_0 \pm t_c \text{se}(f)$ where, for a 95% interval (and 118 degrees of freedom), t_c = 1.98. Also, for x = 0.01, $\text{se}(f)$ = 0.06434 and for x = 0.1, $\text{se}(f)$ = 0.06483. The two 95% interval estimates are:

for x = 0.01: 0.01139 ± 1.98 × 0.06434 = (–0.1160, 0.1388)
for x = 0.1: 0.07571 ± 1.98 × 0.06483 = (–0.0527, 0.2041)

In the context of the problem (predicting Mobil's risk premium), these intervals are very wide and not very informative.

(c) The two hypotheses are H_0: $\beta_1 = 0$ versus H_1: $\beta_1 \neq 0$. The test statistic, given H_0 is true, is

$$t = \frac{b_1}{\text{se}(b_1)} \sim t_{(118)}$$

The rejection region is $|t| > 1.98$. The value of the test statistic is

$$t = \frac{0.0042408}{0.005881} = 0.7211$$

Since $t = 0.7211 < t_c = 1.98$, we do not reject H_0. The data are compatible with a zero intercept.

(d) Without an intercept the estimated model is

$$\underset{}{\hat{y}_t} = \underset{(0.0850)}{0.7211}\, x_t$$

with the number in parentheses being the standard error. Testing H_0: $\beta_2 = 1$ against H_1: $\beta_2 < 1$, the test statistic, given H_0 is true, is

$$t = \frac{b_2 - 1}{\text{se}(b_2)} \sim t_{(119)}$$

The rejection region is $t < -1.658$. The value of the test statistic is

$$t = \frac{0.7211 - 1}{0.08498} = -3.282$$

Since $t = -3.282 < -1.658$, we reject H_0 and conclude that Mobil Oil's beta is less than 1.

Predicting Mobil's risk premium for x = 0.01 and x = 0.10, we have

for x = 0.01: $\hat{y}_0$ = 0.7211 × 0.01 = 0.007211
for x = 0.1: $\hat{y}_0$ = 0.7211 × 0.1 = 0.072112

Before turning to interval predictions for these two values of x, note that the formula we have been using for the variance of the prediction error is only valid when the model has an intercept. Your computer software will recognize the change and give the right answer. However, it is instructive to derive the correct expression for models without an intercept. The prediction error is given by

$$f = \hat{y}_0 - y_0 = b_2 x_0 - \beta_2 x_0 - e_0 = (b_2 - \beta_2)x_0 - e_0$$

$$\text{var}(f) = x_0^2 \text{ var}(b_2 - \beta_2) + \text{var}(e_0) = \frac{x_0^2 \sigma^2}{\sum x_t^2} + \sigma^2$$

(The covariance between (b_2 − β_2) and e_0 is zero.) To show that $\text{var}(b_2) = \sigma^2 / \sum x_t^2$, note that, from Exercise 3.7,

$$b_2 = \frac{\sum x_t y_t}{\sum x_t^2}$$

and $\text{var}(b_2) = \left(\frac{1}{\sum x_t^2}\right)^2 \sum x_t^2 \text{ var}(y_t) = \left(\frac{1}{\sum x_t^2}\right)^2 \sigma^2 \sum x_t^2 = \frac{\sigma^2}{\sum x_t^2}$.

Returning to the standard error of the prediction error, we have

$$\text{se}(f) = \hat{\sigma}\left(\frac{x_0^2}{\sum x_t^2} + 1\right)^{1/2} = 0.063945\left(\frac{x_0^2}{0.56624} + 1\right)^{1/2}$$

When x = 0.01, $\text{se}(f)$ = 0.06395 and the 95% prediction interval is

$$0.00721 \pm 1.98 \times 0.06395 = (-0.1194,\ 0.1338)$$

When x = 0.1, $\text{se}(f)$ = 0.06451 and the 95% prediction interval is

$$0.07211 \pm 1.98 \times 0.06451 = (-0.05561,\ 0.1998).$$

(e) Before investing on the stock market, investors appreciate an indication of the riskiness of alternative stocks. Some investors may be prepared to buy a stock with a low expected return providing its variance is also low. Others may go for risky stocks in the hope of a big gain. And, some might develop a portfolio of stocks that have a variety of risks. Whatever the situation, it is important to be able to assess the riskiness of different stocks. This riskiness can be examined by looking at the magnitude of β_j in the model

$$(r_j - r_f) = \alpha_j + \beta_j (r_m - r_f) + e_j$$

where r_j, r_f and r_m are the return on security j, the risk free rate, and the market rate, respectively. Values of β_j less than 1 suggest stock j is less volatile than the market and not a risky stock. Values of β_j greater than 1 are an indication that stock j is risky; its variation is very sensitive to variation in the market.

To assess the characteristics of Mobil Oil's stock 120 monthly observations on r_j, r_f and r_m, for the period 1978 to 1987, are collected. The least-squares estimated equation is

$$\underset{}{(\hat{r}_j - r_f)} = \underset{(0.00588)}{0.00424} + \underset{(0.086)}{0.715}\,(r_m - r_f)$$

A 95% interval estimate for Mobil's β_j is (0.545, 0.884). Thus, we can conclude that Mobil's stock is less volatile than the overall market. It is a good choice for a risk averse investor.

However, reduced volatility can bring with it the cost of a reduced rate of return. As we discovered in part (b), when the market risk premium is 10%, the predicted risk premium for Mobil is only 7.57%. With a low market risk premium, such as 1%, the prediction for Mobil is comparatively higher (1.14%). This higher value is a consequence of the positive intercept estimate. In both cases, it must be recognized that our model is not a good one for predicting Mobil's risk premium. The wide prediction intervals mean that there is a great deal of uncertainty associated with the realized value of the risk premium. When the market risk premium is 10%, we predict that Mobil's risk premium will lie between –5.27% and 20.41%; for a market risk premium of 1%, the corresponding prediction is between –16.6% and 13.88%.

Thus, while we have been able to confidently conclude that Mobil's stock is less volatile than the market, we have not been able to give a reliable prediction of Mobil's risk premium or rate of return.

5.16 (a) (a) $b_1 = t \times \text{se}(b_1)$ = 1.257 × 2.1738 = 2.732

(b) p-value = 2 × $\left(1 - P(t < 1.257)\right)$ = 2 × (1 – 0.8926) = 0.2148

(c) $\text{se}(b_2) = b_2/t = 0.18014/5.754 = 0.0313$

(d) $\widehat{\text{var}}(b_1) = \left[\text{se}(b_1)\right]^2 = 2.1738^2 = 4.725$

(b) 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.

(c) A 99% confidence interval for the slope is given by

$$b_2 \pm t_c \text{se}(b_2) = 0.1801 \pm 2.68 \times 0.0313 = (0.096, 0.264)$$

(d) For testing $H_0 : \beta_2 = 0.2$ against $H_1 : \beta_2 \neq 0.2,$ we calculate $t = (0.1801 - 0.2)/0.0313 =$ –0.634. 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. Nonrejection of H_0 means that this claim is compatible with the sample of data.

(e) The Louisiana residual is

$$\hat{e}_0 = 15.365 - 2.732 - 0.18014 \times 61.3 = 1.59.$$

(f) The prediction is

$$M\hat{I}M_0 = 2.732 + 0.18014 \times 75 = 16.24$$

5.17 (a) Let y_t be the quantity of soda consumed and x_t be the maximum temperature. The linear relationship between y_t and x_t is $y_t = \beta_1 + \beta_2 x_t + e_t$. Using the data given, the least squares estimates of the equation are given by

$$\hat{y}_t = -771.26 + 25.761\, x_t \qquad R^2 = 0.9338$$
$$(127.13) \quad (1.714)$$

where standard errors are in parenthesis.

(b) To test whether increases in temperature increase the quantity consumed, we test the hypothesis that $H_0: \beta_2 = 0$ against $H_1: \beta_2 > 0$. Given H_0 is true, the test statistic is $t = b_2/\text{se}(b_2)$. Using a 5% significance level, and noting we have 16 degrees of freedom, the rejection region is $t > 1.746$. The value of the test statistic is

$$t = \frac{25.761}{1.7141} = 15.029$$

Since 15.029 > 1.746, we reject H_0 and conclude that there is enough data evidence to suggest that higher temperatures do increase the quantity consumed.

(c) At $x_0 = 70$, the point prediction for the amount of soda sold is

$$\hat{y}_0 = -771.26 + 25.761(70) = 1032.0$$

To compute a prediction interval we need the standard error of the prediction error. Using computer software, it is found to be $\text{se}(f) = 60.974$. A 95% prediction interval is given by

$$\hat{y}_0 \pm t_c \text{se}(f) = 1032 \pm 2.21 \times 60.974 = (902.7, 1161.3)$$

(d) The temperature for which we predict zero sodas to be sold is that value of x_0 which satisfies the equation

$$0 = -771.26 + 25.761\, x_0$$

or, $$x_0 = 771.26/25.761 = 29.9$$

5.18 (a) The relationship shows how an increase (or decrease) in apprehensions of people entering the U.S. illegally depends on an increase (or decrease) in time spent policing the borders. The slope coefficient gives the elasticity of (A_t/A_{t-1}) with respect to changes in (E_t/E_{t-1}). Since the variables are measured in terms of the logs of ratios or "log differences", they represent relative changes rather than original magnitudes.

To test the significance of the estimated slope, we test $H_0: \beta_2 = 0$ against $H_1: \beta_2 \neq 0$. The calculated t-value is $t = 0.510/0.126 = 4.05$. With such a large sample size we can take $t_c = 1.96$ as the 5% critical value. Since 4.05 > 1.96 we reject H_0 and conclude that the estimated slope is significant.

(b) This relationship describes how the change in apprehensions of illegal entrants depends on changes in the Mexican wage rate in its manufacturing sector. The slope coefficient gives the elasticity of (A_t/A_{t-1}) with respect to changes in (MW_t/MW_{t-1}). Again, note that the variables represent changes in the logs.

To test the significance of the estimated slope, we test $H_0:\beta_2=0$ against $H_1:\beta_2\neq 0$. The calculated t-value is $t=-0.550/0.169=-3.25$. Since $-3.25<-1.96$ we reject H_0 at the 5% level of significance. The estimated slope is significant.

5.19 (a) The estimated slopes in the table show how the growth rate of a country is expected to change when there is a change in life expectancy. According to the theory of Swanson and Kopecky, the signs should be positive. Thus, the sign for the OECD countries is not what you would expect.

(b) The test results for $H_0:\beta_2=0$ against $H_1:\beta_2>0$ appear in the following table.

Group	*d* of *f*	*t*-value	t_c	Decision
Africa	36	3.36	1.69	Reject H_0
OECD	21	–0.70	1.72	Do not reject H_0
Latin America	20	0.34	1.72	Do not reject H_0
Asia	15	2.51	1.75	Reject H_0
All Countries	102	5.77	1.66	Reject H_0

(c) A 95% interval estimate of the slope for Latin America is given by

$$b_2 \pm t_c \text{se}(b_2) = 0.012 \pm 2.086 \times 0.03529 = (-0.062,\ 0.086)$$

A 95% interval estimate of the slope for Asia is

$$b_2 \pm t_c \text{se}(b_2) = 0.113 \pm 2.131 \times 0.04502 = (0.017,\ 0.209)$$