PRINCIPLES OF ECONOMETRICS

5TH EDITION

ANSWERS TO ODD-NUMBERED EXERCISES IN CHAPTER 5

EXERCISE 5.1

(a)

x_{i2}^*	x_{i3}^*	y_i^*
0	1	0
1	− 2	0
2	2	3
− 2	1	− 1
1	− 2	0
− 2	− 1	− 3
0	1	1

(b) $\sum y_i^* x_{i2}^* = 14, \quad \sum x_{i2}^{*2} = 14, \quad \sum y_i^* x_{i3}^* = 9, \quad \sum x_{i3}^{*2} = 16, \quad \sum x_{i2}^* x_{i3}^* = 0$

(c) $b_2 = 1 \quad b_3 = 0.5625 \quad b_1 = 1$

(d) $\hat{e} = (-0.5625,\ 0.125,\ -0.125,\ 0.4375, 0.125,\ -0.4375,\ 0.4375)$

(e) $\hat{\sigma}^2 = 0.234375$

(f) $r_{23} = 0$

(g) $\text{se}(b_2) = 0.129387$

(h) $SSE = 0.9375 \quad SST = 20 \quad SSR = 19.0625 \quad R^2 = 0.9531$

EXERCISE 5.3

(a) (i) $t = 0.659$.

(ii) $\text{se}(b_2) = 0.4842$.

(iii) $b_3 = -1.45494$.

(iv) $R^2 = 0.0575$

(v) $\hat{\sigma} = 6.217$

(b) Assuming the strict exogeneity condition holds so that we can interpret the coefficients as causal, we have the following interpretations:

The value $b_2 = 2.765$ suggests that a 1% increase in total expenditure will increase the share of expenditure going to alcohol by approximately 0.02765 percentage points.

The value $b_3 = -1.45494$ suggests that, if the household has one more child, the share of alcohol expenditure of that household decreases by 1.45494 percentage points.

The value $b_4 = -0.1503$ suggests that, if the age of the household head increases by 1 year, the share of alcohol expenditure decreases by 0.1503 percentage points.

(c) A 95% confidence interval for β_4 is $(-0.1964, -0.1042)$. This interval tells us that, if the age of the household head increases by 1 year, the share of the alcohol expenditure is estimated to decrease by an amount between 0.1042 and 0.1964 percentage points.

(d) With the exception of the intercept, all coefficient estimates are significantly different from zero at a 5% level because their *p*-values are all less than 0.05.

(e) The null and alternative hypotheses are $H_0 : \beta_3 = -2,\ H_1 : \beta_3 \neq -2$. The calculated *t*-value is $t = 1.475$. We fail to reject H_0 because $|1.475| < 1.96 = t_{(0.975,\,1196)}$. There is no evidence to suggest that having an extra child leads to a decline in the alcohol budget share that is different from two percentage points.

EXERCISE 5.5

(a) $$\operatorname{var}(b_2 \mid \mathbf{x}) = \frac{12\sigma^2}{T(T+1)(T-1)}$$

As $T \to \infty$, $\operatorname{var}(b_2 \mid \mathbf{x}) \to 0$, and thus the least squares estimator is consistent.

(b) $$\operatorname{var}(b_2 \mid \mathbf{x}) = \frac{\sigma^2}{\frac{1}{3}\left(1-0.5^{2T}\right) - \frac{\left(1-0.5^T\right)^2}{T}}$$

As $T \to \infty$, $\operatorname{var}(b_2 \mid \mathbf{x}) \to 3\sigma^2$. Since $\lim_{T\to\infty} \operatorname{var}(b_2 \mid \mathbf{x}) \neq 0$**,** the least squares estimator b_2 is not consistent.

(c) In part (a), increasing *T* provides more and more information about β_2 and the relationship between *y* and *x*. However, in part (b), $x_t = 0.5^t$ becomes zero for moderately sized *T*, and increasing *T* further does not provide any more information about β_2.

EXERCISE 5.7

(a) $\operatorname{se}(b_2) = 0.1959$ $\quad\operatorname{se}(b_3) = 0.1406$

(b) The *t*-value for testing $H_0 : \beta_2 = 0$ against the alternative $H_1 : \beta_2 \neq 0$ is $t = 3.66$. The 5% critical values are $\pm t_{(0.975,\,200)} = \pm 1.972$. Since $3.66 > 1.972$, we reject $H_0 : \beta_2 = 0$ and conclude $\beta_2 \neq 0$.

(c) The *t*-value for testing $H_0 : \beta_3 \leq 0.9$ against the alternative $H_1 : \beta_2 > 0.9$ is $t = 1.08$. The 10% critical value is $t_{(0.9,\,200)} = 1.286$. Since $1.08 < 1.286$, we fail to reject $H_0 : \beta_3 \leq 0.9$; there is insufficient evidence to conclude that $\beta_2 > 0.9$.

(d) The standard error for $b_2 - b_3$ is $\operatorname{se}(b_2 - b_3) = 0.3117$. The *t*-value for testing $H_0 : \beta_2 = \beta_3$ against the alternative $H_1 : \beta_2 \neq \beta_3$ is $t = -1.07$. The 1% critical values are $\pm t_{(0.995,\,200)} = \pm 2.601$. Since $-2.601 < -1.07 < 2.601$, we fail to reject $H_0 : \beta_2 = \beta_3$; there is insufficient evidence to conclude that $\beta_2 \neq \beta_3$.

EXERCISE 5.9

(a) We estimate that a 1% increase in population is associated with a 0.02674 increase in the expected number of medals won, holding all else fixed. We estimate that a 1% increase in GDP is associated with a 0.0427 increase in the expected number of medals won, holding all else fixed

(b) 27.5% of the variation in the number of medals won, about its sample mean, is explained by the variation in $\ln(POPM)$ and $\ln(GDPB)$.

(c) $H_0:\beta_3=0$ and $H_1:\beta_3>0$. The test statistic is $b_3/\text{se}(b_3)\sim t_{(N-K=61)}$ if the null hypothesis is true. The $\alpha=0.10$ *t*-critical value is 1.29558. The calculated value of the test statistic is $t=2.485$. Since this is greater than the critical value we conclude that there is a positive relationship between the expected number of medals won and $\ln(GDPB)$, at the 10% significance level. If we change the significance level to 5% the same conclusion holds since the *t*-critical value is 1.6702.

(d) $H_0:\beta_2=0$ and $H_1:\beta_2>0$. The test statistic is $b_2/\text{se}(b_2)\sim t_{(N-K=61)}$ if the null hypothesis is true. The $\alpha=0.10$ *t*-critical value is 1.29558. The calculated value of the test statistic is $t=1.335$. Since this is greater than the critical value we conclude that there is a positive relationship between the expected number of medals won and $\ln(POPM)$, at the 10% significance level. If we change the significance level to 5% the critical value is 1.6702, so that at a 5% significance level we cannot conclude that there is a positive relationship between the expected number of medals won and population.

(e) $E(MEDALS\,|\,POPM=58, GDPB=1010)=27.60867$. A 95% interval estimate is $(19.165, 36.053)$.

(f) 20 medals is within the 95% interval estimate, but the interval is rather wide. The required *p*-value is 0.0765.

(g)
$$\text{se}\left(b_1+b_2\ln(58)+b_3\ln(1010)\right)=\sqrt{\begin{array}{l}\widehat{\text{var}}(b_1)+[\ln(58)]^2\,\widehat{\text{var}}(b_2)+[\ln(1010)]^2\,\widehat{\text{var}}(b_3)\\ +2\ln(58)\widehat{\text{cov}}(b_1,b_2)+2\ln(1010)\widehat{\text{cov}}(b_1,b_3)\\ +2\ln(58)\ln(1010)\widehat{\text{cov}}(b_2,b_3)\end{array}}$$

EXERCISE 5.11

(a) $\dfrac{\partial WAGE}{\partial EXPER}=\beta_3+2\beta_4 EXPER$

(b) We expect β_2 to be positive as workers with a higher level of education should receive higher wages. Also, we expect β_3 and β_4 to be positive and negative, respectively. When workers are relatively inexperienced, additional experience leads to a larger increase in their wages than it does after they become relatively more experienced. Also, eventually we expect wages to decline with experience as a worker gets older and their productivity declines. A positive β_3 and a negative β_4 gives a quadratic function with these properties.

(c) The number of years of experience at which wages start to decline is $EXPER^* = -\beta_3/2\beta_4$.

(d) (i) (2.009, 2.649)

(ii) (0.253, 0.673)

(iii) (0.070, 0.220)

(iv) (26.0, 43.2)

EXERCISE 5.13

(a) $DU_t = -\gamma G_t + \gamma G_N = \gamma G_N - \gamma G_t = \beta_1 + \beta_2 G_t$ where $\beta_1 = \gamma G_N$, and $\beta_2 = -\gamma$.

(b) $\hat{\gamma} = 0.2713$ and $\hat{G}_N = 0.7331$.

(c) The required standard errors and *t*-values for zero null hypotheses are

$$\begin{array}{ll} \text{se}(b_1) = 0.02686 & t = 7.405 \\ \text{se}(b_2) = 0.02472 & t = -10.975 \\ \text{se}(\hat{\gamma}) = 0.02472 & t = 10.975 \\ \text{se}\left(\hat{G}_N\right) = 0.07578 & t = 9.674 \end{array}$$

The *t*-values are all large – much greater than $t_{(0.975,178)} = 1.973$ – and so we conclude that the estimates are all significantly different from zero at a 5% level.

(d) For testing $H_0: G_N = 0.8$ against the alternative $H_1: G_N \neq 0.8$, the critical values are $\pm t_{(0.975,178)} = \pm 1.973$. The test statistic value is $t = -0.883$. Since $-1.973 < -0.883 < 1.973$, we fail to reject H_0. There is insufficient evidence to conclude that the quarterly growth rate is different from 0.8%.

(e) A 95% interval estimate for γ is $[0.223, 0.320]$.

(f) A 95% interval estimate for EU_f is $[5.930, 6.032]$.

(g) The 95% prediction interval is $[5.436, 6.526]$.

The interval in (f) relates to an estimate of the mean or average level of unemployment given U_{2014Q4} and G_{2015Q1}; its standard error does not include uncertainty about the error term. The interval in (g) is for a predicted future value of unemployment. Its realized value will include the error term; the standard error reflects the extra uncertainty from not knowing the error term and leads to a wider interval.

EXERCISE 5.15

(a) $E(e_i) = E_x\left[E(e_i \mid x_i)\right] = E_x\left[c\left(x_i^2 - 1\right)\right] = c\left[E_x\left(x_i^2\right) - 1\right] = c(1-1) = 0$

(b) $\text{cov}(e_i, x_i) = E_x\left[x_i E(e_i \mid x_i)\right] = E_x\left[x_i c\left(x_i^2 - 1\right)\right] = c\left[E_x\left(x_i^3\right) - E_x\left(x_i\right)\right] = c(0-0) = 0$

(c) Because $E(e_i \mid x_i) \neq 0$, the least squares estimator will not be unbiased. However, it will be consistent because $\text{cov}(e_i, x_i) = 0$.

EXERCISE 5.17

The answers confirmed by results from your software are

$$b_1 = 1 \qquad b_2 = 1 \qquad b_3 = 0.5625 \qquad \hat{\sigma}^2 = 0.234375 \qquad r_{23} = 0 \qquad \text{se}(b_2) = 0.129387$$

$$SSE = 0.9375 \qquad SST = 20 \qquad SSR = 19.0625 \qquad R^2 = 0.9531$$

EXERCISE 5.19

(a) First, assume $\varepsilon < 1$, then $1 + \dfrac{\beta_2}{\beta_1 + \beta_2 \ln(TOTEXP)} < 1$ and $\dfrac{\beta_2}{\beta_1 + \beta_2 \ln(TOTEXP)} < 0$

Now, $\beta_1 + \beta_2 \ln(TOTEXP)$ is the expected budget share for food which must be positive. Hence, $\beta_2 < 0$.

Now, assume $\beta_2 < 0$, then $(\beta_1 + \beta_2 \ln(TOTEXP))(\varepsilon - 1) < 0$ and $(\varepsilon - 1) < 0$

It follows that $\varepsilon < 1$.

(b) The estimated equation is

$$\begin{aligned} \widehat{WFOOD} &= 0.9938 - 0.1405 \ln(TOTEXP) \\ \text{(se)} \quad & \quad (0.0371) \quad (0.0083) \\ (p\text{-value}) & \quad (0.0000) \quad (0.0000) \end{aligned}$$

We estimate that a 1% change in total expenditure will decrease the expected budget share for food by 0.0014. This estimate is relatively precise; a 95% interval estimate for β_2 is $[-0.1567, -0.1243]$.

(c) Mean budget share estimates

	Point Estimate	Standard Error	Interval Estimate
$TOTEXP = 50$	0.44416	0.00558	(0.4332, 0.4551)
$TOTEXP = 90$	0.36158	0.00306	(0.3556, 0.3676)
$TOTEXP = 170$	0.27221	0.00624	(0.2600, 0.2845)

(d) Elasticity estimates

	Point Estimate	Standard Error	Interval Estimate
$TOTEXP = 50$	0.68366	0.01543	(0.6534, 0.7139)
$TOTEXP = 90$	0.61140	0.02328	(0.5657, 0.6571)
$TOTEXP = 170$	0.48383	0.04107	(0.4032, 0.5644)

(e) As total expenditure increases, the mean budget share for food and the expenditure elasticity for food decline, in line with expectations for a commodity which is a necessity. For the mean budget share, estimation is more precise at median total expenditure and less precise at the outer percentiles. For the elasticity, estimation becomes less reliable as total expenditure increases.

EXERCISE 5.21

(a)

Variable	Coefficient	Std. Error	*t*-value	*p*-value
C	-2.674652	0.996569	-2.683862	0.0103
TREND	0.678101	0.191928	3.533095	0.0010
RAIN	1.508177	0.430729	3.501450	0.0011
RAIN^2	-0.157079	0.046622	-3.369233	0.0016
*RAIN*TREND*	-0.092838	0.046761	-1.985389	0.0535

(b) All coefficient estimates, with the exception of that for $RAIN \times TREND$, are significantly different from zero at both the 5% and 10% levels of significance. The estimated coefficient for $RAIN \times TREND$ is significantly different from zero at a 10% level, but not a 5% level.

(c) The positive sign on *RAIN* and the negative sign on $RAIN^2$ imply a positive response of yield to rainfall, but with diminishing returns to more rainfall. This is the expected relationship. Also, the negative sign on $RAIN \times TREND$ implies the extra yield from an increase in rainfall is less in the later years, a consequence of the development of drought resistant varieties. The positive sign on *TREND* indicates that technological improvements have increased yield, as expected; the negative sign on $RAIN \times TREND$ shows there is a lesser impact from technological improvements in years when rainfall is higher. All the signs are as expected.

(d) (i) In 1959, $RAIN = 2.98$ and $TREND = 0.9$, and $\widehat{ME}_{1959} = 0.4884$.

A 95% interval estimate is $[0.1948, 0.7820]$

(ii) In 1995, $RAIN = 4.797$ and $TREND = 4.5$, and $\widehat{ME}_{1995} = -0.4166$.

A 95% interval estimate is $[-0.7284, -0.1048]$

There are two main observations. First, the interval estimates are relatively wide when viewed relative to the magnitude of yield which, in the sample, varies from 0.363 to 2.967. This large width is partially attributable to the large units of measurement for rainfall – one decimeter (4 inches). It would be more realistic to examine the marginal effect of an extra centimeter (0.4 inches) of rainfall in which case we divide the marginal effect estimates by 10. The second main observation is the negative interval when rainfall is high, and we are towards the end of the sample period. Rain is valuable if there has been little rain and drought resistant varieties have not been developed. It has a negative effect if there has been much rain and we have drought resistant varieties.

(e) (i) When $TREND = 0.9$, $\widehat{RAIN}_{max} = 4.5347$

A 95% interval estimate is given by $[4.059, 5.010]$

(ii) When $TREND = 4.5$, $\widehat{RAIN}_{\max} = 3.4709$

A 95% interval estimate is given by $[2.679, 4.263]$

The point estimate for the yield-maximizing rainfall in 1959 is over 1dm larger than that in 1995, reflecting the absence of drought resistant varieties in the earlier year. However, the interval estimates do overlap slightly, by approximately 0.2dm. The actual rainfall in 1995 of 4.797dm was well above the yield-maximizing estimate of 3.471dm, and also outside the interval estimate (2.679, 4.263). This outcome explains the estimated negative marginal effect in part (d)(ii). In contrast, in 1959 the observed rainfall of 0.9dm was well below the yield-maximizing estimate of 4.53dm, and outside the interval estimate (4.059, 5.010), leading to a definite positive marginal effect for rainfall.

EXERCISE 5.23

(a) The expected sign for β_2 is negative because, as the number of grams in a given sale increases, the price per gram should decrease, implying a discount for larger sales. We expect β_3 to be positive; the purer the cocaine, the higher the price. The sign for β_4 will depend on how demand and supply are changing over time. For example, a fixed demand and an increasing supply will lead to a fall in price. A fixed supply and increased demand would lead to a rise in price.

(b)

$$\begin{array}{lllll} \widehat{PRICE} = 90.8467 & -\ 0.0600 QUANT & +\ 0.1162 QUAL & -\ 2.3546 TREND & R^2 = 0.5097 \\ \text{(se)} \quad (8.5803) & (0.0102) & (0.2033) & (1.3861) & \\ (t) \quad (10.588) & (-5.892) & (0.5717) & (-1.6987) & \end{array}$$

The estimated values for β_2, β_3 and β_4 are -0.0600, 0.1162 and -2.3546, respectively. They imply that as quantity (number of grams in one sale) increases by 1 unit, the mean price will go down by 0.0600. Also, as the quality increases by 1 unit the mean price goes up by 0.1162. As time increases by 1 year, the mean price decreases by 2.3546. All the signs turn out according to our expectations, with β_4 implying supply has been increasing faster than demand.

(c) 0.5097.

(d) $H_0 : \beta_2 \geq 0$ against $H_1 : \beta_2 < 0$. The calculated *t*-value of -5.892 is less than the critical *t* value, $t_{(0.95,52)} = -1.675$. We reject H_0 and conclude that sellers are willing to accept a lower price if they can make sales in larger quantities.

(e) $H_0 : \beta_3 \leq 0$ against $H_1 : \beta_3 > 0$. The calculated *t*-value of 0.5717 is not greater than the critical *t* = 1.675 We do not reject H_0. We cannot conclude that a premium is paid for better quality cocaine.

(f) The average annual change in the cocaine price is given by the value $b_4 = -2.3546$. It has a negative sign suggesting that the price decreases over time. A possible reason for a decreasing price is the development of improved technology for producing cocaine, such that suppliers can produce more at the same cost.

EXERCISE 5.25

(a)

$$\widehat{\ln(PRICE)} = 5.04536 + 0.056639SQFT - 0.235847SQFT^{1/2}$$
$$\begin{array}{cccc} \text{(se)} & (0.33222) & (0.010517) & (0.119247) \\ (p\text{-value}) & (0.0000) & (0.0000) & (0.0485) \end{array}$$

All coefficients are significantly different from zero at a 5% level of significance. The intercept and the coefficient of *SQFT* are also significant at a 1% level but the coefficient of $SQFT^{1/2}$ is not.

(b)

$$ME_{SQFT} = \beta_2 + \frac{\beta_3}{2SQFT^{1/2}}$$

After multiplying by 100, this marginal effect gives the percentage change in price from an extra 100 square feet of floor space. As the size of the house (*SQFT*) increases, this percentage change approaches the constant $100\beta_2$. Thus, we expect $\beta_2 > 0$. The sign of the coefficient β_3 determines whether the percentage change in price is increasing or decreasing as it approaches $100\beta_2$. For $\beta_3 > 0$, the percentage change in price is decreasing as it approaches $100\beta_2$; for $\beta_3 < 0$, it is increasing. Either sign is possible for β_3. However, we do expect the magnitudes of β_2 and β_3 to be such that $ME_{SQFT} > 0$, for all values of *SQFT*.

(c) (i) For $SQFT = 15$, $\widehat{ME}_{SQFT=15} = 0.026191$. A 95% interval estimate is $[0.01598, 0.03640]$
With 95% confidence, we estimate, that, for houses of 1500 square feet, an extra 100 square feet will increase price by a percentage amount between 1.598% and 3.640%

(ii) For $SQFT = 30$, $\widehat{ME}_{SQFT=30} = 0.035109$. A 95% interval estimate is $[0.03205, 0.03816]$
With 95% confidence, we estimate, that, for houses of 3000 square feet, an extra 100 square feet will increase price by a percentage amount between 3.205% and 3.816%.

(iii) For $SQFT = 45$, $\widehat{ME}_{SQFT=45} = 0.03906$. A 95% interval estimate is $[0.03488, 0.04324]$
With 95% confidence, we estimate, that, for houses of 4500 square feet, an extra 100 square feet will increase price by a percentage amount between 3.488% and 4.324%.

As *SQFT* increases, the point estimates for the marginal effect increase at a decreasing rate. The precision of these estimates is greatest, and the interval estimate narrowest, when $SQFT = 30$.

(d) From $\ln(PRICE) = \beta_1 + \beta_2 SQFT + \beta_3 SQFT^{1/2} + e$, we have

$$PRICE = \exp\left\{\beta_1 + \beta_2 SQFT + \beta_3 SQFT^{1/2} + e\right\}$$

Then, recognizing that when $\left(e \mid SQFT\right) \sim N\left(0, \sigma^2\right)$, $E\left(\exp(e) \mid SQFT\right) = \sigma^2/2$, we have

$$E\left(PRICE \mid SQFT\right) = \exp\left\{\beta_1 + \beta_2 SQFT + \beta_3 SQFT^{1/2} + \sigma^2/2\right\}$$

Differentiating this expression with respect to *SQFT*, yields

$$\frac{\partial E\left(PRICE \mid SQFT\right)}{\partial SQFT} = \left(\beta_2 + \frac{\beta_3}{2SQFT^{1/2}}\right) \times \exp\left\{\beta_1 + \beta_2 SQFT + \beta_3 SQFT^{1/2} + \sigma^2/2\right\}$$

(e) $\hat{C} = 1.05915$.

$$\hat{S}_{SQFT=15} = 3.815848 \qquad \hat{S}_{SQFT=30} = 8.194208 \qquad \hat{S}_{SQFT=45} = 15.947805$$

$$\hat{S}_{SQFT=15}\hat{C} = 4.04155 \qquad \hat{S}_{SQFT=30}\hat{C} = 8.67889 \qquad \hat{S}_{SQFT=45}\hat{C} = 16.89111$$

(f) $\text{se}\left(\hat{S}_{SQFT=15}\right) = 0.654444 \qquad \text{se}\left(\hat{S}_{SQFT=30}\right) = 0.447515 \qquad \text{se}\left(\hat{S}_{SQFT=45}\right) = 1.206289$

(g) The test statistic is

$$t = \frac{\hat{S}\hat{C} - 9}{\text{se}\left(\hat{S}\right)}$$

We reject H_0 if the calculated value t_{calc} is such that $t_{calc} \geq 1.965$ or $t_{calc} \leq -1.965$. The calculated values for each of the three cases are:

(i) $t_{calc(15)} = -7.577$ (ii) $t_{calc(30)} = -0.718$ (iii) $t_{calc(45)} = 6.542$

Thus, for houses of sizes 1500 and 4500 square feet, we reject the null hypothesis that an extra 100 square feet increases the expected price by an extra $9,000. However, for houses of size 3000, this hypothesis cannot be rejected.

(h) $\text{se}\left(\hat{S}_{SQFT=15} \times \hat{C}\right) = 0.6933 \qquad \text{se}\left(\hat{S}_{SQFT=30} \times \hat{C}\right) = 0.4750 \qquad \text{se}\left(\hat{S}_{SQFT=45} \times \hat{C}\right) = 1.2791$

These standard errors are slightly larger than those in part (f) which ignored the uncertainty from estimating σ^2 , but the difference is not sufficiently great to change the outcome of the hypothesis tests in part (g).

EXERCISE 5.27

(a)

$$\widehat{RHAMMER} = -217.85 + 2.7560 YEARS\_OLD + 1.9941 INCHSQ10$$
$$(\text{se}) \qquad (82.37) \quad (0.9493) \qquad\qquad (0.4835)$$

(b) We estimate that an addition year of age increases expected price by $2,756, holding other things constant. The 95% interval estimate is [0.8899, 4.6220].

(c) We estimate that an addition 10 square inches increases expected price by $1,994, holding other things constant. The 95% interval estimate is [1.0436, 2.9446].

(d)

$$\widehat{RHAMMER} = -273.85 + 2.8621 YEARS\_OLD + 4.3120 INCHSQ10 - 0.01460 INCHSQ10^2$$
$$(\text{se}) \qquad (84.14) \quad (0.9425) \qquad\qquad (0.9591) \qquad\qquad (0.00523)$$

The model in part (a) is unrealistic. It suggests that increasing the painting size will continue to increase the expected price by the same amount. Artists would create infinitely large paintings under this scenario. By adding the quadratic term we allow for diminishing returns to painting size, which is realized through the negative coefficient on the squared term.

(e) The interval estimate is [1.0095, 4.7147]. It is only slightly different due to a slight change in the point estimate and a slightly smaller standard error.

(f) The point and interval estimates for each of the painting sizes are

*INCHSQ*10	Estimate	Std. Err	Lower bound	Upper bound
5	4.1660	0.9142	2.3691	5.9630
25	3.5822	0.7442	2.1193	5.0451
90	1.6849	0.4923	0.7172	2.6526

The marginal value of additional size declines as the painting size gets larger.

(g) The 95% interval estimate is [89.77, 205.67].

(h) The 95% interval estimate is [−61.98, 26.91].

(i) $YEARS\_OLD = 81.13$

EXERCISE 5.29

$$\widehat{LCRMRTE} = -3.364 - 2.044PRBARR - 0.9932PRBCONV \qquad R^2 = 0.737$$
$$\text{(se)} \quad (0.317) \quad (0.265) \qquad (0.0908)$$
$$+0.3004PRBPRIS + 150.2POLPC + 0.002114WCON$$
$$(0.4027) \qquad (13.5) \qquad (0.000761)$$

All five variables are expected to have negative effects on the crime rate. We expect each of them to act as a deterrent to crime. In the estimated equation the probability of an arrest and the probability of conviction have negative signs as expected, and both coefficients are significantly less than zero with *p*-values of 0.0000. On the other hand, the coefficients of the other three variables, the probability of a prison sentence, the number of police and the weekly wage in construction have positive signs, which is contrary to our expectations. Of these three variables, the coefficient of *PRBPRIS* is not significantly different from zero, but the other two, *POLPC* and *WCON*, are significantly different from zero, and have unexpected positive signs. Thus, it appears that the variables, *PRBARR* and *PRBCONV* are the most important for crime deterrence. The positive sign for the coefficient of *POLPC* may have been caused by endogeneity, a concept considered in Chapter 10. In the context of this example, high crime rates may be more likely to exist in counties with greater numbers of police because more police are employed to counter high crime rates. It is less clear why *WCON* should have a positive sign. Perhaps construction companies have to pay higher wages to attract workers to counties with higher crime rates.

EXERCISE 5.31

(a)

$$\widehat{TIME} = 20.8701 + 0.36813DEPART + 1.5219REDS + 3.0237TRAINS$$
$$\text{(se)} \quad (1.6758) \ (0.03510) \qquad (0.1850) \qquad (0.6340)$$

β_1: Bill's expected commute time when he leaves Carnegie at 6:30AM and encounters no red lights and no trains is estimated to be 20.87 minutes.

β_2: If Bill leaves later than 6:30AM, the increase in his expected traveling time is estimated to be 3.7 minutes for every 10 minutes that his departure time is later than 6:30AM (assuming the number of red lights and trains are constant).

β_3: The expected increase in traveling time from each red light, with departure time and number of trains held constant, is estimated to be 1.52 minutes.

β_4: The expected increase in traveling time from each train, with departure time and number of red lights held constant, is estimated to be 3.02 minutes.

(b) The 95% confidence intervals for the coefficients are:

β_1: $(17.57, 24.17)$

β_2: $(0.299, 0.437)$

β_3: $(1.16, 1.89)$

β_4: $(1.77, 4.27)$

In the context of driving time, these intervals are relatively narrow ones. We have obtained precise estimates of each of the coefficients.

(c) The hypotheses are $H_0: \beta_3 \geq 2$ and $H_1: \beta_3 < 2$. The 5% critical value is $t_{(0.05,245)} = -1.651$. The calculated *t*-value is -2.584. Since $-2.584 < -1.651$, we reject H_0. We conclude that the expected delay from each red light is less than 2 minutes.

(d) The hypotheses are $H_0: \beta_4 = 3$ and $H_1: \beta_4 \neq 3$. The 10% critical values are $t_{(0.05,245)} = -1.651$ and $t_{(0.95,245)} = 1.651$. The calculated *t*-value is 0.037. Since $-1.651 < 0.037 < 1.651$, we fail to reject H_0. The data are consistent with the hypothesis that the expected delay from each train is 3 minutes.

(e) The hypotheses are $H_0: \beta_2 \geq 1/3$ and $H_1: \beta_2 < 1/3$. We reject H_0 if $t \leq t_{(0.05,245)} = -1.651$, where the calculated *t*-value is 0.991. Since $0.991 > -1.651$, we fail to reject H_0. The data are consistent with the hypothesis that delaying departure time by 30 minutes increases expected travel time by at least 10 minutes.

(f) The hypotheses are $H_0: \beta_4 \geq 3\beta_3$ and $H_1: \beta_4 < 3\beta_3$. The rejection region for the *t*-test is $t \leq t_{(0.05,245)} = -1.651$, and the calculated *t*-value is -1.825. Since $-1.825 < -1.651$, we reject H_0. The expected delay from a train is less than three times the delay from a red light.

(g) The hypotheses are $H_0: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 \leq 45$ and $H_1: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 > 45$. The rejection region for the *t*-test is $t \geq t_{(0.95,245)} = 1.651$, where the calculated *t*-value is -1.726. Since $-1.726 < 1.651$, we fail to reject H_0 at a 5% significance level. Alternatively, we fail to reject H_0 because the *p*-value = $\Pr\left(t_{(245)} > -1.726\right) =$ 0.9572, which is greater than 0.05. There is insufficient evidence to conclude that Bill will get to the University after 7:45AM.

(h) If it is imperative that Bill is not late for his meeting, he will wish to establish, with a high degree of probability, that his commute time will be less than 45 minutes. To do so, having a commute time as less than 45 minutes should be the alternative hypothesis. Having it as the null hypothesis, and failing to reject the null hypothesis, does not imply the commute time will necessarily be less than 45 minutes. If the hypotheses are reversed as

$$H_0: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 \geq 45 \qquad H_1: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 < 45$$

then the 5% rejection region is $t \leq t_{(0.05,245)} = -1.651$. In this case we reject H_0 because $-1.726 < -1.651$. Bill's expected commute time is such that he can expect to be on time for the meeting.

EXERCISE 5.33

Variable	Coefficient	Std. Error	*t*-value	*p*-value
C	1.037933	0.275741	3.764159	0.0002
EDUC	0.089539	0.031082	2.880742	0.0040
EDUC^2	0.001458	0.000924	1.577892	0.1149
EXPER	0.044879	0.007297	6.149930	0.0000
EXPER^2	-0.000468	7.60E-05	-6.157303	0.0000
EDUC*EXPER	-0.001010	0.000379	-2.664994	0.0078

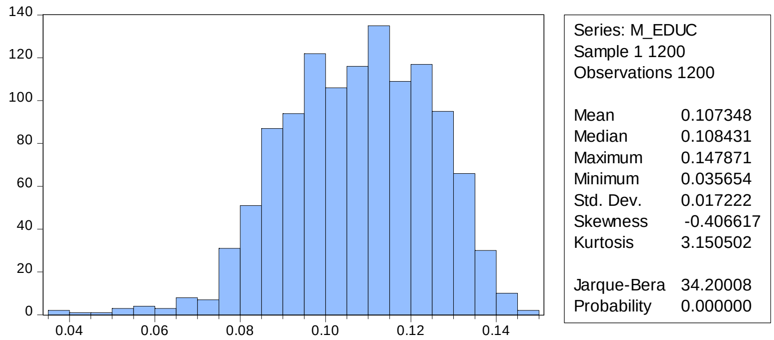
(a) All coefficient estimates are significantly different from zero at a 1% level of significance with the exception of that for $EDUC^2$ which is significant at a 12% significance level.

(b) $ME_{EDUC} = \beta_2 + 2\beta_3 EDUC + \beta_6 EXPER$

Its estimate is $\widehat{ME}_{EDUC} = 0.089539 + 0.002916\, EDUC - 0.001010\, EXPER$

The marginal effect of education increases as the level of education increases, but decreases with the level of experience.

(c) A histogram of the marginal effects and their summary statistics are



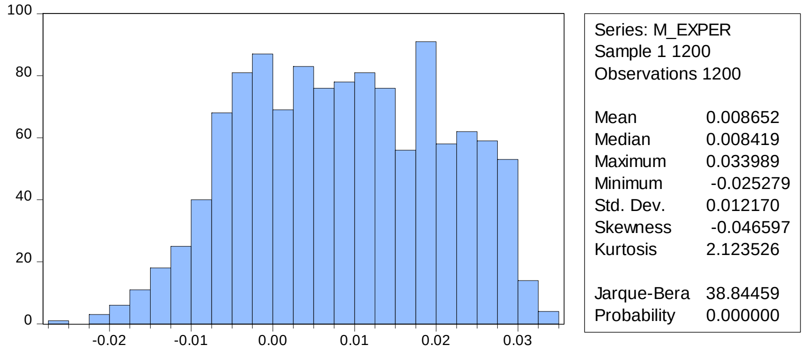
We observe that the marginal effects range from 0.036 to 0.148 with most of them concentrated between 0.085 and 0.13. The 5th, 50th (median) and 95th percentiles are, respectively, $\widehat{ME}_{(EDUC,0.05)} = 0.080$, $\widehat{ME}_{(EDUC,0.50)} = 0.108$, and $\widehat{ME}_{(EDUC,0.95)} = 0.134$.

(d) $ME_{EXPER} = \beta_4 + 2\beta_5 EXPER + \beta_6 EDUC$.

Its estimate is $\widehat{ME}_{EXPER} = 0.044879 - 0.000936\, EXPER - 0.001010\, EDUC$

The marginal effect of experience decreases as the level of education increases, and as the years of experience increase.

(e) A histogram of the marginal effects and their summary statistics are



Although most of the marginal effects of experience are positive, there is a large proportion (28.3%) that are negative. Overall, the values range from -0.025 to 0.034. The 5th, 50th (median) and 95th percentiles are, respectively, $\widehat{ME}_{(EXPER,\,0.05)} = -0.010$, $\widehat{ME}_{(EXPER,\,0.50)} = 0.008$, and $\widehat{ME}_{(EXPER,\,0.95)} = 0.028$.

(f) The null and alternative hypotheses are

$$H_0: -\beta_2 - 33\beta_3 + 10\beta_4 + 260\beta_5 + 152\beta_6 \geq 0$$

$$H_1: -\beta_2 - 33\beta_3 + 10\beta_4 + 260\beta_5 + 152\beta_6 < 0$$

We reject H_0 and conclude that David's expected log-wage is greater if

$$t = \frac{-b_2 - 33b_3 + 10b_4 + 260b_5 + 152b_6}{\text{se}\left(-b_2 - 33b_3 + 10b_4 + 260b_5 + 152b_6\right)} \leq t_{(0.05,\,1194)} = -1.646$$

The calculated *t*-value is $t = 1.670$. Since $1.670 > -1.646$, we fail to reject H_0. There is insufficient evidence to conclude that David's log-wage is greater.

(g) The null and alternative hypotheses are

$$H_0: -\beta_2 - 33\beta_3 + 10\beta_4 + 420\beta_5 + 144\beta_6 \geq 0$$

$$H_1: -\beta_2 - 33\beta_3 + 10\beta_4 + 420\beta_5 + 144\beta_6 < 0$$

We reject H_0 and conclude that David's expected log-wage is greater if

$$t = \frac{-b_2 - 33b_3 + 10b_4 + 420b_5 + 144b_6}{\text{se}\left(-b_2 - 33b_3 + 10b_4 + 420b_5 + 144b_6\right)} \leq t_{(0.05,\,1194)} = -1.646$$

The calculated *t*-value is $t = -2.062$. Since $-2.062 < -1.646$, we reject H_0, and conclude that David's log-wage is greater. This test result is not the same as in part (f). The difference in outcomes is attributable to diminishing returns to experience. Because Svetlana initially had 18 years of experience, her extra years of experience had a relatively small impact on her log-wage. Because David had only eight years of experience in the first instance, the extra eight years had a relatively large impact on his log-wage.

(h) The null and alternative hypotheses are

$$H_0: 12\beta_5 - 4\beta_6 = 0 \qquad H_1: 12\beta_5 - 4\beta_6 \neq 0$$

We reject H_0 if

$$t = \frac{12b_5 - 4b_6}{\text{se}(12b_5 - 4b_6)} \geq t_{(0.975,1194)} = 1.962 \quad \text{or} \quad t = \frac{12b_5 - 4b_6}{\text{se}(12b_5 - 4b_6)} \leq t_{(0.025,1194)} = -1.962$$

The calculated *t*-value is $t = -1.027$. Since $-1.962 < -1.027 < 1.962$, we fail to reject H_0. There is no evidence to suggest the marginal effects from extra experience are different for Jill and Wendy.

(i) We assume that, as time goes on, Jill gains more experience, but no more education. We estimate it will be 19.667 more years before her marginal effect becomes negative. A 95% interval estimate for the number of years before her marginal effect become negative is $[15.96, 23.40]$.