

# CHAPTER 7

## Exercise Answers

### EXERCISE 7.2

- (a) *Intercept*: At the beginning of the time period over which observations were taken, on a day which is not Friday, Saturday or a holiday, and a day which has neither a full moon nor a half moon, the estimated average number of emergency room cases was 93.69.

*T*: We estimate that the average number of emergency room cases has been increasing by 0.0338 per day, other factors held constant. The  $t$ -value is 3.06 and  $p$ -value = 0.003 < 0.01.

*HOLIDAY*: The average number of emergency room cases is estimated to go up by 13.86 on holidays, holding all else constant. The “holiday effect” is significant at the 0.05 level.

*FRI* and *SAT*: The average number of emergency room cases is estimated to go up by 6.9 and 10.6 on Fridays and Saturdays, respectively, holding all else constant. These estimated coefficients are both significant at the 0.01 level.

*FULLMOON*: The average number of emergency room cases is estimated to go up by 2.45 on days when there is a full moon, all else constant. However, a null hypothesis stating that a full moon has no influence on the number of emergency room cases would not be rejected at any reasonable level of significance.

*NEWMOON*: The average number of emergency room cases is estimated to go up by 6.4 on days when there is a new moon, all else held constant. However, a null hypothesis stating that a new moon has no influence on the number of emergency room cases would not be rejected at the usual 10% level, or smaller.

- (b) There are very small changes in the remaining coefficients, and their standard errors, when *FULLMOON* and *NEWMOON* are omitted.
- (c) Testing  $H_0 : \beta_6 = \beta_7 = 0$  against  $H_1 : \beta_6$  or  $\beta_7$  is nonzero, we find  $F = 1.29$ . The 0.05 critical value is  $F_{(0.95, 2, 222)} = 3.307$ , and corresponding  $p$ -value is 0.277. Thus, we do not reject the null hypothesis that new and full moons have no impact on the number of emergency room cases.

**EXERCISE 7.5**

- (a) The estimated equation, with standard errors in parentheses, is

$$\begin{aligned} \widehat{\ln(\text{PRICE})} = & 4.4638 + 0.3334\text{UTOWN} + 0.03596\text{SQFT} - 0.003428(\text{SQFT} \times \text{UTOWN}) \\ & \text{(se)} \quad (0.0264)(0.0359) \quad (0.00104) \quad (0.001414) \\ & -0.000904\text{AGE} + 0.01899\text{POOL} + 0.006556\text{FPLACE} \quad R^2 = 0.8619 \\ & (0.000218) \quad (0.00510) \quad (0.004140) \end{aligned}$$

- (b) Using this result for the coefficients of *SQFT* and *AGE*, we estimate that an additional 100 square feet of floor space is estimated to increase price by 3.6% for a house not in University town and 3.25% for a house in University town, holding all else fixed. A house which is a year older is estimated to sell for 0.0904% less, holding all else constant. The estimated coefficients of *UTOWN*, *AGE*, and the slope-indicator variable *SQFT\_UTOWN* are significantly different from zero at the 5% level of significance.
- (c) An approximation of the percentage change in price due to the presence of a pool is 1.90%. The exact percentage change in price due to the presence of a pool is estimated to be 1.92%.
- (d) An approximation of the percentage change in price due to the presence of a fireplace is 0.66%. The exact percentage change in price due to the presence of a fireplace is also 0.66%.
- (e) The percentage change in price attributable to being near the university, for a 2500 square-foot home, is 28.11%.

**EXERCISE 7.9**

- (a) The estimated average test scores are:

regular sized class with no aide = 918.0429  
 regular sized class with aide = 918.3568  
 small class = 931.9419

From the above figures, the average scores are higher with the small class than the regular class. The effect of having a teacher aide is negligible.

The results of the estimated models for parts (b)-(g) are summarized in the table on page 38.

- (b) The coefficient of *SMALL* is the difference between the average of the scores in the regular sized classes (918.36) and the average of the scores in small classes (931.94). That is  $b_2 = 931.9419 - 918.0429 = 13.899$ . Similarly the coefficient of *AIDE* is the difference between the average score in classes with an aide and regular classes. The *t*-value for the significance of  $\beta_3$  is  $t = 0.136$ . The critical value at the 5% significance level is 1.96. We cannot conclude that there is a significant difference between test scores in a regular class and a class with an aide.

**Exercise 7.9 (continued)**

Exercise 7-9

	(1)	(2)	(3)	(4)	(5)
	(b)	(c)	(d)	(e)	(g)
<i>C</i>	918.043*** (1.641)	904.721*** (2.228)	923.250*** (3.121)	931.755*** (3.940)	918.272*** (4.357)
<i>SMALL</i>	13.899*** (2.409)	14.006*** (2.395)	13.896*** (2.294)	13.980*** (2.302)	15.746*** (2.096)
<i>AIDE</i>	0.314 (2.310)	-0.601 (2.306)	0.698 (2.209)	1.002 (2.217)	1.782 (2.025)
<i>TCHEXPER</i>		1.469*** (0.167)	1.114*** (0.161)	1.156*** (0.166)	0.720*** (0.167)
<i>BOY</i>			-14.045*** (1.846)	-14.008*** (1.843)	-12.121*** (1.662)
<i>FREELUNCH</i>			-34.117*** (2.064)	-32.532*** (2.126)	-34.481*** (2.011)
<i>WHITE_ASIAN</i>			11.837*** (2.211)	16.233*** (2.780)	25.315*** (3.510)
<i>TCHWHITE</i>				-7.668*** (2.842)	-1.538 (3.284)
<i>TCHMASTERS</i>				-3.560* (2.019)	-2.621 (2.184)
<i>SCHURBAN</i>				-5.750** (2.858)	. .
<i>SCHRURAL</i>				-7.006*** (2.559)	. .
<i>N</i>	5786	5766	5766	5766	5766
adj. R-sq	0.007	0.020	0.101	0.104	0.280
BIC	66169.500	65884.807	65407.272	65418.626	64062.970
SSE	31232400.314	30777099.287	28203498.965	28089837.947	22271314.955

Standard errors in parentheses

\* p&lt;0.10, \*\* p&lt;0.05, \*\*\* p&lt;0.01

- (c) The *t*-statistic for the significance of the coefficient of *TCHEXPER* is 8.78 and we reject the null hypothesis that a teacher's experience has no effect on total test scores. The inclusion of this variable has a small impact on the coefficient of *SMALL*, and the coefficient of *AIDE* has gone from positive to negative. However *AIDE*'s coefficient is not significantly different from zero and this change is of negligible magnitude, so the sign change is not important.
- (d) The inclusion of *BOY*, *FREELUNCH* and *WHITE\_ASIAN* has little impact on the coefficients of *SMALL* and *AIDE*. The variables themselves are statistically significant at the  $\alpha = 0.01$  level of significance.

**Exercise 7.9 (continued)**

- (e) The regression result suggests that *TCHWHITE*, *SCHRURAL* and *SCHURBAN* are significant at the 5% level and *TCHMASTERS* is significant at the 10% level. The inclusion of these variables has only a very small and negligible effect on the estimated coefficients of *AIDE* and *SMALL*.
- (f) The results found in parts (c), (d) and (e) suggest that while some additional variables were found to have a significant impact on total scores, the estimated advantage of being in small classes, and the insignificance of the presence of a teacher aide, is unaffected. The fact that the estimates of the key coefficients did not change is support for the randomization of student assignments to the different class sizes. The addition or deletion of uncorrelated factors does not affect the estimated effect of the key variables.
- (g) We find that inclusion of the school effects increases the estimates of the benefits of small classes and the presence of a teacher aide, although the latter effect is still insignificant statistically. The *F*-test of the joint significance of the school indicators is 19.15. The 5% *F*-critical value for 78 numerator and 5679 denominator degrees of freedom is 1.28, thus we reject the null hypothesis that all the school effects are zero, and conclude that at least some are not zero.

The variables *SCHURBAN* and *SCHRURAL* drop out of this model because they are exactly collinear with the included 78 indicator variables.

**EXERCISE 7.14**

- (a) We expect the parameter estimate for the dummy variable *PERSON* to be positive because of reputation and knowledge of the incumbent. However, it could be negative if the incumbent was, on average, unpopular and/or ineffective. We expect the parameter estimate for *WAR* to be positive reflecting national feeling during and immediately after first and second world wars.
- (b) The regression functions for each value of *PARTY* are:

$$E(\text{VOTE} \mid \text{PARTY} = 1) = (\beta_1 + \beta_7) + \beta_2 \text{GROWTH} + \beta_3 \text{INFLATION} + \beta_4 \text{GOODNEWS} \\ + \beta_5 \text{PERSON} + \beta_6 \text{DURATION} + \beta_8 \text{WAR}$$

$$E(\text{VOTE} \mid \text{PARTY} = -1) = (\beta_1 - \beta_7) + \beta_2 \text{GROWTH} + \beta_3 \text{INFLATION} + \beta_4 \text{GOODNEWS} \\ + \beta_5 \text{PERSON} + \beta_6 \text{DURATION} + \beta_8 \text{WAR}$$

The intercept when there is a Democrat incumbent is  $\beta_1 + \beta_7$ . When there is a Republican incumbent it is  $\beta_1 - \beta_7$ . Thus, the effect of *PARTY* on the vote is  $2\beta_7$  with the sign of  $\beta_7$  indicating whether incumbency favors Democrats ( $\beta_7 > 0$ ) or Republicans ( $\beta_7 < 0$ ).

**Exercise 7.14 (continued)**

- (c) The estimated regression using observations for 1916-2004 is

$$\widehat{VOTE} = 47.2628 + 0.6797GROWTH - 0.6572INFLATION + 1.0749GOODNEWS$$

$$\begin{array}{cccc} \text{(se)} & (2.5384) & (0.1107) & (0.2914) & (0.2493) \\ & & & & \\ & + 3.2983PERSON & - 3.3300DURATION & - 2.6763PARTY & + 5.6149WAR \\ & (1.4081) & (1.2124) & (0.6264) & (2.6879) \end{array}$$

The signs are as expected. Can you explain why? All the estimates are statistically significant at a 1% level of significance except for *INFLATION*, *PERSON*, *DURATION* and *WAR*. The coefficients of *INFLATION*, *DURATION* and *PERSON* are statistically significant at a 5% level, however. The coefficient of *WAR* is statistically insignificant at a level of 5%. Lastly, an  $R^2$  of 0.9052 suggests that the model fits the data very well.

- (d) Using the data for 2008, and based on the estimates from part (c), we summarize the actual and predicted vote as follows, along with a listing of the values of the explanatory variables.

vote	growth	inflation	goodnews	person	duration	party	war	votehat
46.6	.22	2.88	3	0	1	-1	0	48.09079

Thus, we predict that the Republicans, as the incumbent party, will lose the 2008 election with 48.091% of the vote. This prediction was correct, with Democrat Barack Obama defeating Republican John McCain with 52.9% of the popular vote to 45.7%.

- (e) A 95% confidence interval for the vote in the 2008 election is

$$\widehat{VOTE}_{2012} \pm t_{(0.975,15)} \times \text{se}(f) = (42.09, 54.09)$$

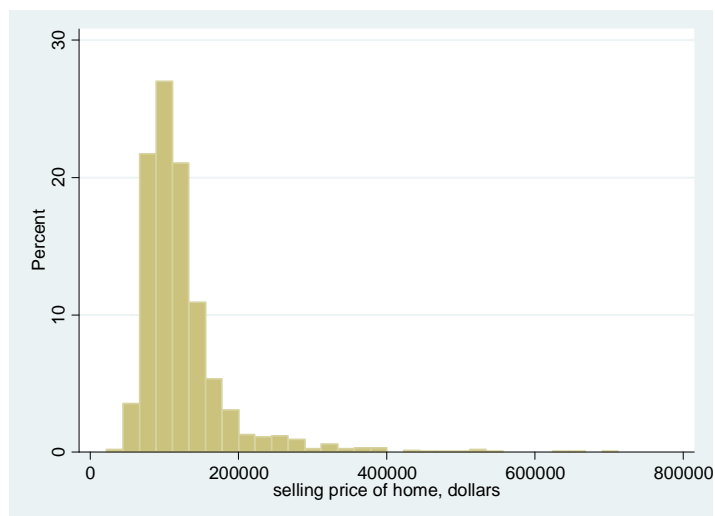
- (f) For the 2012 election the Democratic party will have been in power for one term and so we set
- DURATION*
- = 1 and
- PARTY*
- = 1. Also, the incumbent, Barack Obama, is running for election and so we set
- PERSON*
- = 1.
- WAR*
- = 0. We use the value of inflation 3.0% anticipating higher rates of inflation after the policy stimulus. We consider 3 scenarios for
- GROWTH*
- and
- GOODNEWS*
- representing good economic outcomes, moderate and poor, if there is a “double-dip” recession. The values and the prediction intervals based on regression estimates with data from 1916-2008, are

<i>GROWTH</i>	<i>INFLATION</i>	<i>GOODNEWS</i>	lb	vote	ub
3.5	3	6	45.6	51.5	57.3
1	3	3	40.4	46.5	52.5
-3	3	1	35.0	41.5	48.0

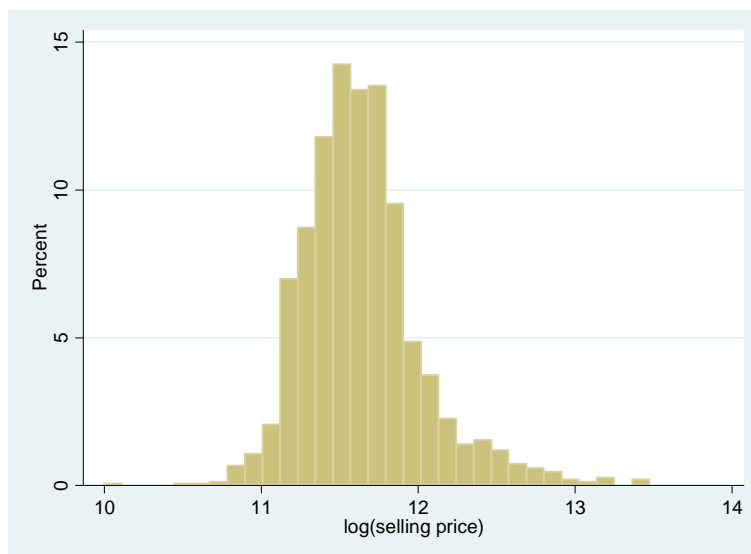
We see that if there is good economic performance, then President Obama can expect to be re-elected. If there is poor economic performance, then we predict he will lose the election with the upper bound of the 95% prediction interval for a vote in his favor being only 48%. In the intermediate case, with only modest growth and less good news, then we predict he will lose the election, though the interval estimate upper bound is greater than 50%, meaning that anything could happen.

**EXERCISE 7.16**

- (a) The histogram for *PRICE* is positively skewed. On the other hand, the logarithm of *PRICE* is much less skewed and is more symmetrical. Thus, the histogram of the logarithm of *PRICE* is closer in shape to a normal distribution than the histogram of *PRICE*.



**Figure xr7.16(a) Histogram of *PRICE***



**Figure xr7.16(b) Histogram of  $\ln(\textit{PRICE})$**

**Exercise 7.16 (continued)**

- (b) The estimated equation is

$$\begin{aligned} \widehat{\ln(\text{PRICE}/1000)} &= 3.9860 + 0.0539\text{LIVAREA} - 0.0382\text{BEDS} - 0.0103\text{BATHS} \\ &\quad \text{(se)} \quad (0.0373) \quad (0.0017) \quad (0.0114) \quad (0.0165) \\ &\quad + 0.2531\text{LGELOT} - 0.0013\text{AGE} + 0.0787\text{POOL} \\ &\quad (0.0255) \quad (0.0005) \quad (0.0231) \end{aligned}$$

All coefficients are significant with the exception of that for *BATHS*. All signs are reasonable: increases in living area, larger lot sizes and the presence of a pool are associated with higher selling prices. Older homes depreciate and have lower prices. Increases in the number of bedrooms, holding all else fixed, implies smaller bedrooms which are less valued by the market. The number of baths is statistically insignificant, so its negative sign cannot be reliably interpreted.

- (c) The price of houses on lot sizes greater than 0.5 acres is approximately  $100(\exp(-0.2531) - 1) = 28.8\%$  larger than the price of houses on lot sizes less than 0.5 acres.
- (d) The estimated regression after including the interaction term is:

$$\begin{aligned} \widehat{\ln(\text{PRICE}/1000)} &= 3.9649 + 0.0589\text{LIVAREA} - 0.0480\text{BEDS} - 0.0201\text{BATHS} \\ &\quad \text{(se)} \quad (0.0370) \quad (0.0019) \quad (0.0113) \quad (0.0164) \\ &\quad + 0.6134\text{LGELOT} - 0.0016\text{AGE} + 0.0853\text{POOL} \\ &\quad (0.0632) \quad (0.0005) \quad (0.0228) \\ &\quad - 0.0161\text{LGELOT} \times \text{LIVAREA} \\ &\quad (0.0026) \end{aligned}$$

Interpretation of the coefficient of  $\text{LGELOT} \times \text{LIVAREA}$ :

The estimated marginal effect of an increase in living area of 100 square feet in a house on a lot of less than 0.5 acres is 5.89%, all else constant. The same increase for a house on a large lot is estimated to increase the house selling price by 1.61% less, or 4.27%. However, note that by adding this interaction variable into the model, the coefficient of *LGELOT* increases dramatically. The inclusion of the interaction variable separates the effect of the larger lot from the fact that larger lots usually contain larger homes.

- (e) The value of the
- F*
- statistic is

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N - K)} = \frac{(72.0633 - 65.4712)/6}{65.4712/(1488)} = 24.97$$

The 5% critical *F* value is  $F_{(0.95, 6, 1488)} = 2.10$ . Thus, we conclude that the pricing structure for houses on large lots is not the same as that on smaller lots.

**Exercise 7.16 (continued)**

A summary of the alternative model estimations follows.

Exercise 7-16

	(1) <i>LGELOT=1</i>	(2) <i>LGELOT=0</i>	(3) Rest	(4) Unrest
<i>C</i>	4.4121*** (0.183)	3.9828*** (0.037)	3.9794*** (0.039)	3.9828*** (0.038)
<i>LIVAREA</i>	0.0337*** (0.005)	0.0604*** (0.002)	0.0607*** (0.002)	0.0604*** (0.002)
<i>BEDS</i>	-0.0088 (0.048)	-0.0522*** (0.012)	-0.0594*** (0.012)	-0.0522*** (0.012)
<i>BATHS</i>	0.0827 (0.066)	-0.0334** (0.017)	-0.0262 (0.017)	-0.0334* (0.017)
<i>AGE</i>	-0.0018 (0.002)	-0.0016*** (0.000)	-0.0008* (0.000)	-0.0016*** (0.000)
<i>POOL</i>	0.1259* (0.074)	0.0697*** (0.024)	0.0989*** (0.024)	0.0697*** (0.025)
<i>LGELOT</i>				0.4293*** (0.141)
<i>LOT_AREA</i>				-0.0266*** (0.004)
<i>LOT_BEDS</i>				0.0434 (0.037)
<i>LOT_BATHS</i>				0.1161** (0.052)
<i>LOT_AGE</i>				-0.0002 (0.001)
<i>LOT_POOL</i>				0.0562 (0.060)
<i>N</i>	95	1405	1500	1500
adj. R-sq	0.676	0.608	0.667	0.696
BIC	50.8699	-439.2028	-252.8181	-352.8402
SSE	7.1268	58.3445	72.0633	65.4712

Standard errors in parentheses

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

\*\* *LOT\_X* indicates interaction between *LGELOT* and *X*