Exercises 5.24   p.243

5.24 The file **collegetown** contains data on 500 single-family houses sold in Baton Rouge, Louisiana during 2009–2013. We will be concerned with the selling price in thousands of dollars (PRICE), the size of the house in hundreds of square feet (SQFT), and the age of the house measured as a categorical variable (AGE), with 1 representing the newest and 11 the oldest. Let X denote all observations on SQFT and AGE. Use all observations to estimate the following regression model:

$$PRICE_i = \beta_1 + \beta_2 SQFT_i + \beta_3 (SQFT_i \times AGE_i) + e$$

Question: a.-e.

a. Report the results. Are the estimated coefficients significantly different from zero?

b. Write down expressions for the marginal effects ∂E(PRICE|X)/∂SQFT and ∂E(PRICE|X)/∂AGE. Interpret each of the coefficients. Given the categorical nature of the variable AGE, what assumptions are being made?

c. Find point and 95% interval estimates for the marginal effect ∂E(PRICE|X)/∂SQFT for houses that are (i) 5 years old, (ii) 20 years old, and (iii) 40 years old. How do these estimates change as AGE increases? (Refer to the file collegetown.def for the definition of AGE.)

d. As a house gets older and moves from one age category to the next, the expected price declines by $6000. Set up this statement as a null hypothesis for houses with (i) 1500 square feet, (ii) 3000 square feet, and (iii) 4500 square feet. Using a 5% significance level, test each of the null hypotheses against an alternative that the price decline is not $6000. Discuss the outcomes.

e. Find a 95% prediction interval for the price of a 60-year old house with 2500 square feet. In the data set there are four 60-year old houses with floor space between 2450 and 2550 square feet. What prices did they sell for? How many of these prices fall within your prediction interval? Is the model a good one for forecasting price? 5.25 The file collegetown contains data on 500

dataset: **collegetown (**500 houses sold)

Variable

dependent variable Y: PRICE = 1,000 US$

independent variable X: SQFT=100 ft^2 AGE= 1~11 newest to oldest

The regression model specified in Exercise 5.24 is:

equation eq02.ls PRICE = C(1) + C(2)*SQFT + C(3)*AGE*SQFT

PRICE = -105.261 + 14.916*SQFT - 0.234*AGE*SQFT

Dependent Variable: PRICE
Method: Least Squares
Date: 12/14/24 Time: 20:59
Sample: 1 500
Included observations: 500

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-105.2613	13.35759	-7.880261	0.0000
SQFT	14.91562	0.651346	22.89969	0.0000
AGE*SQFT	-0.233562	0.073423	-3.181034	0.0016

R-squared	0.648474	Mean dependent var	250.2369
Adjusted R-squared	0.647059	S.D. dependent var	171.4765
S.E. of regression	101.8721	Akaike info criterion	12.09130
Sum squared resid	5157833.	Schwarz criterion	12.11658
Log likelihood	-3019.824	Hannan-Quinn criter.	12.10122
F-statistic	458.4175	Durbin-Watson stat	0.942106
Prob(F-statistic)	0.000000		

Question a: Estimation of Coefficients

To estimate the coefficients using EViews 8.1: The regression equation as

PRICE=β1+β2SQFT+β3(SQFT×AGE)

β_1=-105.261 β_2= 14.916 β_3= - 0.234

Question b: Interpretation of Coefficients

- $$\frac{\partial E(PRICE|X)}{\partial SQFT} = \beta_2 + \beta_3 AGE_i$$
- β 2: This coefficient represents the expected price change for each unit increase in SQFT when AGE is hold.
- β 3: This coefficient shows how the impact of an additional hundred square feet on price changes as the age of the unit increases. As the age of a house increases, it will have a negative impact on the price.

Question c: Marginal Effect Calculation

1.From the regression model:

$$PRICE_i = \beta_1 + \beta_2 SQFT_i + \beta_3 (SQFT_i \times AGE_i) + e$$

The marginal effect of SQFT on PRICE is given by:

- $$\frac{\partial E(PRICE|X)}{\partial SQFT} = \beta_2 + \beta_3 AGE_i$$

2. Point Estimates

To find the point estimates for the marginal effect at different ages:

1)Substitute the values of AGE into the expression for the marginal effect.

2)Use the estimated coefficients β2 and β3 obtained from the above regression analysis.

-For Different Ages:

At 5 years old: $ME_5 = \beta_2 + \beta_3 * 5 = 14.916 + (-0.234)*5 = 13.746$

At 20 years old: $ME_{20} = \beta_2 + \beta_3 * 20 = 14.916 + (-0.234)*20 = 10.236$

At 40 years old: $ME_{40} = \beta_2 + \beta_3 * 40 = 14.916 + (-0.234)*40 = 5.556$

3.Interval Estimates

To calculate the 95% confidence interval for the marginal effects:

1).Calculate the standard error of the marginal effect using the delta method or obtain it directly from EViews.

2).The confidence interval can be calculated as:

$$CI = ME_{age} \pm Z_{\alpha/2} \times SE(ME_{age})$$

Where $Z_{\alpha/2}$ is the critical value from the standard normal distribution (approximately 1.96 for a 95% confidence level).

ME_{age} CI = 13.746 +- 1.96 * $SE(ME_{age})$

ME_{age} CI = 13.746 +- {1.96 * 0.07342} = [13.6020968, 13.8899032]

Question d:

Hypothesis Setup

Null and Alternative Hypotheses

For each house size, we will set up the null hypothesis (H_0) and the alternative hypothesis (H_a) as follows:

Null Hypothesis (H_0) : The expected price decline as a house gets older is $6000.

Alternative Hypothesis (Ha): The expected price decline is not $6000.

so,

For 1500 square feet: $H_0: \beta_3 = 6000$ $H_a: \beta_3 \neq 6000$	For 3000 square feet: $H_0: \beta_3 = 6000$ $H_a: \beta_3 \neq 6000$	For 4500 square feet: $H_0: \beta_3 = 6000$ $H_a: \beta_3 \neq 6000$

To test these hypotheses at a 5% significance level, should perform the steps:

1.Using EViews, run the regression model that includes the interaction term between SQFT and AGE:

$$PRICE_i = \beta_1 + \beta_2 SQFT_i + \beta_3(SQFT_i \times AGE_i) + e$$

2: obtain Coefficient Estimates

Δprice = β_3 * sqft = -0.233562 * 1500 = -350.343
Δprice = β_3 * sqft = -0.233562 * 3000 = -700.686
Δprice = β_3 * sqft = -0.233562 * 4500 = -1051.029

SE(β_3) = 0.073423

3. Then, for each house size calculate the t-statistic by:

$$\frac{\Delta\, price - (-6000)}{se(\Delta\, price)}$$

t = (-350.343+6000) / 0.073423 * 1500 = 51.298
t = (-700.686 +6000) / 0.073423 * 3000 = 24.058
t = (-1051.029+6000) / 0.073423 *4500 = 14.979

4. Decision Rule: If the absolute value of calculated t-statistic exceeds the critical t-value, reject. If not, fail to reject.

Answer: All Three t-statistic are much greater than 1.96. Reject the null hypothesis. The expected price decline is significantly different from $6000 for a house with 1500 or 3000 or 4500 square feet.

Question e:

Predicted PRICE= −105.2613 + 14.91562 * 2500 + (−0.233562)*(2500×60) = **$2149.4887**

Standard Error of Prediction:

SE(prediction) = $\sqrt{se^2 + var(predicted\ price)}$

As S.E. of prediction is the same as S.E. of regression, so SE(prediction)=101.8721

Prediction Interval:

PI = Predicted PRICE ± $t_{\alpha/2}$, n−2×SE(prediction)

- tα/2,n−2 is the critical value from the t-distribution for a 95% confidence level.

the Interval:
SE(prediction) is approximately the standard error of the regression for simplicity. -
PI=2149.4887 ± 1.96 × 101.8721
PI=2149.4887 ± 199.6733

- Lower Bound: 2149.4887−199.6733=1949.8154
- Upper Bound: 2149.4887+199.6733=2349.162

Answer: Prediction Interval: [1949.8154, 2349.162]