3.23 The file *collegetown.dat* contains data on 500 single-family houses sold in Baton Rouge, Louisiana, during 2009-2013. The data include sale price in $1000 units, *PRICE*, and total interior area in hundreds of square feet, *SQFT*.

(a) Using the quadratic regression model, $PRICE = \alpha_1 + \alpha_2 SQFT^2 + e$, test the hypothesis that the marginal effect on expected house price of increasing the size of a 2000 square foot house by 100 square feet is less than or equal to $13000 against the alternative that that the marginal effect will be greater than $13000. Use the 5% level of significance. Clearly state the test statistic used, the rejection region, and the test *p*-value. What do you conclude?

The estimated model is

$$\widehat{PRICE} = 93.5659 \; + 0.1845 SQFT^2$$
$$\text{(se)} \quad (6.0722) \quad (0.005256)$$

The marginal effect is $dE[PRICE]/dSQFT = 2\alpha_2 SQFT$. The estimate of this marginal effect is for a 2000 square foot house is $2\hat{\alpha}_2(20) = 7.3808$ with standard error $\sqrt{\widehat{\text{var}}(40\hat{\alpha}_2)} = \sqrt{40^2\,\widehat{\text{var}}(\hat{\alpha}_2)} = \sqrt{40^2(0.00002762)} = 0.2102$. The hypothesis is $H_0: 2\alpha_2(20) \le 13$ and the alternative is $H_1: 2\alpha_2(20) > 13$. This is a right-tail test. The test statistic is $t = (40\hat{\alpha}_2 - 13)/\sqrt{40\hat{\alpha}_2} \sim t_{(498)}$ if the null hypothesis is true. The right-tail critical value for the 5% level of significance is $t_{(0.95, N-2=498)} = 1.6479$. The calculated *t*-value is −26.7288. This falls in the non-rejection region. The *p*-value is $P[t_{(498)} \ge -26.7288] = 1.0$. We cannot conclude that the marginal effect is greater than $13,000.

(b) Using the quadratic regression model, $PRICE = \alpha_1 + \alpha_2 SQFT^2 + e$, test the hypothesis that the marginal effect on expected house price of increasing the size of a 4000 square foot house by 100 square feet is less than or equal to $13000 against the alternative that that the marginal effect will be greater than $13000. Use the 5% level of significance. Clearly state the test statistic used, the rejection region, and the test *p*-value. What do you conclude?

The marginal effect is $dE[PRICE]/dSQFT = 2\alpha_2 SQFT$. The estimate of this marginal effect is for a 4000 square foot house is $2\hat{\alpha}_2(40) = 14.7615$ with standard error $\sqrt{\widehat{\text{var}}(80\hat{\alpha}_2)} = \sqrt{80^2\,\widehat{\text{var}}(\hat{\alpha}_2)} = \sqrt{80^2(0.00002762)} = 0.4205$. The hypothesis is

$H_0: 2\alpha_2(40) \le 13$ and the alternative is $H_1: 2\alpha_2(40) > 13$. This is a right-tail test. The test statistic is $t = (80\hat{\alpha}_2 - 13)\big/\sqrt{80\hat{\alpha}_2} \sim t_{(498)}$ if the null hypothesis is true. The right-tail critical value for the 5% level of significance is $t_{(0.95, N-2=498)} = 1.6479$. The calculated *t*-value is 4.1895. This falls in the rejection region. The *p*-value is $P[t_{(498)} \ge 4.1895] = 0.000$. We conclude, at the 5% level, that the marginal effect is greater than \$13,000.

(c) Using the quadratic regression model, $PRICE = \alpha_1 + \alpha_2 SQFT^2 + e$, estimate the expected price $E(PRICE \mid SQFT) = \alpha_1 + \alpha_2 SQFT^2$ for a house of 2000 square feet. Construct a 95% interval estimate of the expected price. Describe your interval estimate to a general audience.

The estimated expected price is $\hat{E}(PRICE \mid SQFT = 20) = \hat{\alpha}_1 + \hat{\alpha}_2 (20)^2 = 167.3735$. To construct a 95% interval estimate we need the standard error of this quantity, which is

$$\begin{aligned}\sqrt{\operatorname{var}(\hat{\alpha}_1 + 400\hat{\alpha}_2)} &= \sqrt{\operatorname{var}(\hat{\alpha}_1) + 400^2 \operatorname{var}(\hat{\alpha}_2) + 2(400)\operatorname{cov}(\hat{\alpha}_1, \hat{\alpha}_2)} \\ &= \sqrt{36.871924 + 400^2 (0.0000276) + 2(400)(-0.02345446)} \\ &= 4.746378\end{aligned}$$

The required critical value for a 95% interval estimate is $t_{(0.975, N-2=498)} = 1.964739$. The resulting interval estimate is $167.3735 \pm (1.964739)4.746378$ or [158.0481, 176.6988]. Using this quadratic model we estimate with 95% confidence that the average price of a 2000 square foot house is between \$158,048.10 and \$176,698.80.

(d) Locate houses in the sample with 2000 square feet of living area. Calculate the sample mean (average) of their selling prices. Is the sample average of the selling price for houses with *SQFT* = 20 compatible with the result in part (c)? Explain.

In the sample there are 3 houses with 2000 square feet. They sold for \$138,000, \$169,000 and \$183,000. The average is 163.3333 which is inside the interval estimate. The interval estimate in (c) was for the expected, or population average, price for homes with 2000 square feet of living area. The average of the 3 observed house prices is not a population average, but we should not be surprised by the outcome.