3.22 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 linear regression $PRICE = \beta_1 + \beta_2 SQFT + e$ estimate the elasticity of expected house *PRICE* with respect to *SQFT*, evaluated at the sample means. Construct a 95% interval estimate for the elasticity, treating the sample means as if they are given (not random) numbers. What is the interpretation of the interval?

The estimated equation is

$$\widehat{PRICE} = -115.4236 + 13.40294 SQFT$$
$$\text{(se)} \qquad (13.0882) \quad (0.4492)$$

For the linear regression model the elasticity is $\varepsilon = \beta_2 x/y$. The estimated elasticity is evaluated at a point on the fitted line, here the sample means of *PRICE* (250.2369) and *SQFT* (27.28212). The estimated elasticity is $\hat{\varepsilon} = 13.4029(27.28212)/250.2369 = 1.4613$. A 95% interval requires the t critical value $t_{(0.975, N-2=498)} = 1.96$ and the standard error of the elasticity which is $\text{se}(\hat{\varepsilon}) = \text{se}(b_2)\bar{x}/\bar{y} = 0.4491(27.2821)/250.2369 = 0.04897$. The resulting interval estimate is [1.365, 1.557]. We estimate that a 1% increase in the total interior living space of a house will increase expected price by 1.37% to 1.56%.

(b) Test the null hypothesis that the elasticity, calculated in part (a), is one against the alternative that the elasticity is not one. Use the 1% level of significance. Clearly state the test statistic used, the rejection region, and the test *p*-value. What do you conclude?

$H_0: \varepsilon = \beta_2 \bar{x}/\bar{y} = 1$ versus $H_1: \varepsilon = \beta_2 \bar{x}/\bar{y} \neq 1$. The test statistic is $t = (\hat{\varepsilon} - 1)/\text{se}(\hat{\varepsilon}) \sim t_{(498)}$ if the null hypothesis is true. For this two tail test at the 1% level the test critical values are $t_{(0.995, N-2=498)} = 2.5857$ and $t_{(0.005, N-2=498)} = -2.5857$. The calculated value of t is 9.419 which is greater than the upper critical value 2.5857 so that we reject the null hypothesis that the elasticity is 1 and accept the alternative that it is not. The *p*-value is the probability that $P[t_{(498)} \geq 9.419] = 0.000$ plus $P[t_{(498)} \leq -9.419] = 0.000$, which is 0.000.

(c) Using the linear regression model $PRICE = \beta_1 + \beta_2 SQFT + e$ test the hypothesis that the marginal effect on expected house price of increasing house size 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 null hypothesis is $H_0 : \beta_2 \le 13$ versus $H_1 : \beta_2 > 13$. This is a one-tail, right-tail test. For the 5% level of significance the critical value is $t_{(0.95, N-2=498)} = 1.6479$, thus we reject the null hypothesis if the calculated $t = (b_2 - 13)/\text{se}(b_2)$ is greater than this value. The calculated *t* = 0.8970, which is less than the critical value, so that we fail to reject the null hypothesis at the 5% level. The *p*-value is $P[t_{(498)} \ge .8971] = 0.1851$. We are unable to conclude that a 100 square foot increase in living area will increase price by more than \$13,000.

(d) Using the linear regression $PRICE = \beta_1 + \beta_2 SQFT + e$, estimate the expected price, $E(PRICE \mid SQFT) = \beta_1 + \beta_2 SQFT$, 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 $\hat{E}(PRICE \mid SQFT) = b_1 + b_2 SQFT$. For a 2000 square foot house this is $\hat{E}(PRICE \mid SQFT = 20) = b_1 + b_2(20) = 152.635$. The required standard error is the square root of the estimated variance, which is

$$\begin{aligned}\sqrt{\widehat{\text{var}}\left[b_1 + b_2(20)\right]} &= \sqrt{\widehat{\text{var}}(b_1) + 20^2\,\widehat{\text{var}}(b_2) + 2(20)\widehat{\text{cov}}(b_1, b_2)} \\ &= \sqrt{171.29978 + 20^2(0.20174797) + 2(20)(-5.5041123)} \\ &= 5.642205\end{aligned}$$

The 95% interval estimate is

$$152.6351 \pm t_{(0.975,498)}(5.642205) = [141.5496,\ 163.7206]$$

We estimate, with 95% confidence, that the average price of a home of 2000 square feet of interior living space is between \$141,549.60 and \$163.720.60.

(e) 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 (d)? 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 (d) 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.