PRINCIPLES OF ECONOMETRICS

5TH EDITION

ANSWERS TO ODD-NUMBERED EXERCISES IN CHAPTER 2

EXERCISE 2.1

(a)

$\sum x_i = 5$	$\sum y_i = 10$	$\sum (x_i - \bar{x}) = 0$	$\sum (x_i - \bar{x})^2 = 10$	$\sum (y - \bar{y}) = 0$	$\sum (x - \bar{x})(y - \bar{y}) = 8$

$\bar{x} = 1, \bar{y} = 2$

(b)
$$b_2 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} = \frac{8}{10} = 0.8$$

$$b_1 = \bar{y} - b_2 \bar{x} = 2 - 0.8 \times 1 = 1.2$$

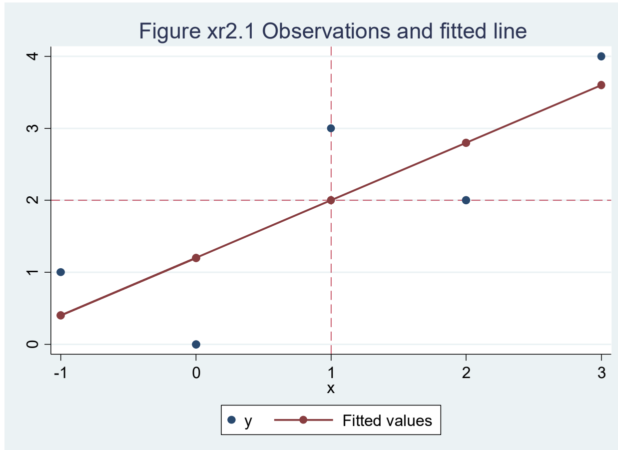
(c)
$$\sum_{i=1}^{5} x_i^2 = 15$$
$$\sum_{i=1}^{5} x_i y_i = 18$$
$$\sum_{i=1}^{5} x_i^2 - N\bar{x}^2 = 10$$
$$\sum_{i=1}^{5} x_i y_i - N\bar{x}\bar{y} = 8$$

(d)

$\sum x_i = 5$	$\sum y_i = 10$	$\sum \hat{y}_i = 10$	$\sum \hat{e}_i = 0$	$\sum \hat{e}_i^2 = 3.6$	$\sum x_i \hat{e}_i = 0$

$$s_y^2 = 2.5$$
$$s_x^2 = 2.5$$
$$s_{xy} = 2$$
$$r_{xy} = 0.8$$
$$CV_x = 158.11388$$
$$\text{median}(x) = 1$$

(e)



(f) See figure above. The fitted line passes through the point of the means, $\bar{x} = 1, \bar{y} = 2$.

(g) $\bar{y} = 2$, $b_1 + b_2\bar{x} = 2$

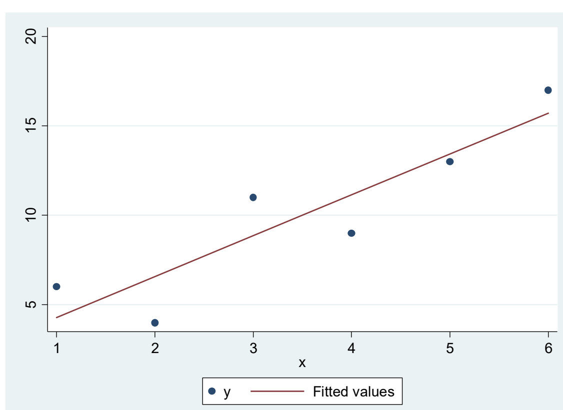
(h) $\bar{\hat{y}} = 2$

(i) $\hat{\sigma}^2 = 1.2$

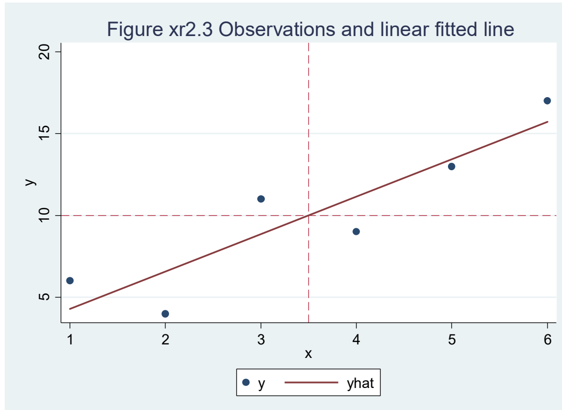
(j) $\widehat{var}(b_2|\mathbf{x}) = 0.12$ and $se(b_2) = 0.34641$

EXERCISE 2.3

(a) We show the least squares fitted line.



(b) $b_2 = 2.285714$, $b_1 = 2$



(c) $\bar{y} = 10$ and $\bar{x} = 3.5$ and $\hat{y} = b_1 + b_2\bar{x} = 10$

(d)

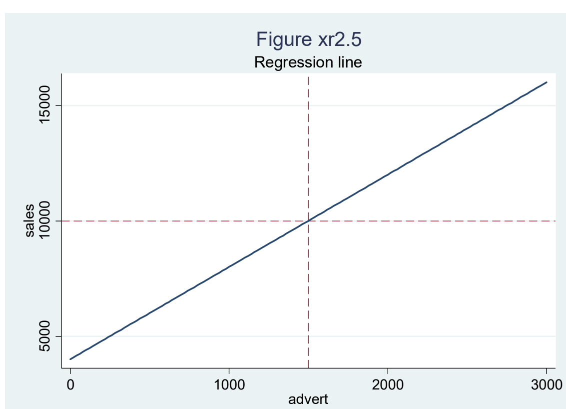
$\hat{e}_i$
1.71429
−2.57143
2.14286
−2.14286
−0.42857
1.28571

(e) $\sum \hat{e}_i = 0$, and $\sum \hat{e}_i^2 = 20.57143$

(f) $\sum x_i \hat{e}_i = 0$

EXERCISE 2.5

(a) $SALES = 4000 + 4 \times ADVERT$



EXERCISE 2.7

(a) $\sum \hat{e}_i^2 = 697.82566$

(b) $\sum(x_i - \bar{x})^2 = 1553.8833$

(c) $b_2 = 1.02896$ suggests that a 1% increase in the percentage of the population with a bachelor's degree or more will lead to an increase of \$1028.96 in the mean income per capita.

(d) $b_1 = 11.519745$

(e) $\sum x_i^2 = 39722.17$

(f) $\hat{e}_i = -4.9231453$

EXERCISE 2.9

(a) $E\left(\hat{\beta}_{2,mean} \mid \mathbf{x}\right) = E\left[\left(\bar{y}_2 - \bar{y}_1\right)/\left(\bar{x}_2 - \bar{x}_1\right) \mid \mathbf{x}\right] = \left[1/\left(\bar{x}_2 - \bar{x}_1\right)\right] E\left[\left(\bar{y}_2 - \bar{y}_1\right) \mid \mathbf{x}\right]$

$$E\left[\left(\bar{y}_2 - \bar{y}_1\right) \mid \mathbf{x}\right] = E\left[\bar{y}_2 \mid \mathbf{x}\right] - E\left[\bar{y}_1 \mid \mathbf{x}\right]$$

$$E\left[\bar{y}_2 \mid \mathbf{x}\right] = E\left[\frac{1}{3}\sum_{i=4}^{6} y_i \mid \mathbf{x}\right] = \frac{1}{3}\sum_{i=4}^{6} E\left(y_i \mid \mathbf{x}\right) = \frac{1}{3}\sum_{i=4}^{6}\left(\beta_1 + \beta_2 x_i\right)$$
$$= \frac{1}{3}\left[3\beta_1 + \beta_2 \sum_{i=4}^{6} x_i\right] = \beta_1 + \beta_2 \frac{1}{3}\left(\sum_{i=4}^{6} x_i\right) = \beta_1 + \beta_2 \bar{x}_2$$

Similarly $E\left[\bar{y}_1 \mid \mathbf{x}\right] = \beta_1 + \beta_2 \bar{x}_1$. Then

$$E\left[\left(\bar{y}_2 - \bar{y}_1\right) \mid \mathbf{x}\right] = E\left[\bar{y}_2 \mid \mathbf{x}\right] - E\left[\bar{y}_1 \mid \mathbf{x}\right] = \left(\beta_1 + \beta_2 \bar{x}_2\right) - \left(\beta_1 + \beta_2 \bar{x}_1\right) = \beta_2\left(\bar{x}_2 - \bar{x}_1\right)$$

Finally,

$$E\left(\hat{\beta}_{2,mean} \mid \mathbf{x}\right) = E\left[\left(\bar{y}_2 - \bar{y}_1\right)/\left(\bar{x}_2 - \bar{x}_1\right) \mid \mathbf{x}\right] = \left[1/\left(\bar{x}_2 - \bar{x}_1\right)\right] E\left[\left(\bar{y}_2 - \bar{y}_1\right) \mid \mathbf{x}\right]$$
$$= \left[1/\left(\bar{x}_2 - \bar{x}_1\right)\right]\beta_2\left(\bar{x}_2 - \bar{x}_1\right) = \beta_2$$

(b) $E\left(\hat{\beta}_{2,mean}\right) = E_{\mathbf{x}}\left[E\left(\hat{\beta}_{2,mean} \mid \mathbf{x}\right)\right] = E_{\mathbf{x}}\left(\beta_2\right) = \beta_2$

(c) $\operatorname{var}\left(\hat{\beta}_{2,mean} \mid \mathbf{x}\right) = \left[1/(\bar{x}_2 - \bar{x}_1)\right]^2 \operatorname{var}\left[(\bar{y}_2 - \bar{y}_1) \mid \mathbf{x}\right] = \left[1/(\bar{x}_2 - \bar{x}_1)\right]^2 \left\{\operatorname{var}[\bar{y}_2 \mid \mathbf{x}] + \operatorname{var}[\bar{y}_1 \mid \mathbf{x}]\right\}$

$$\operatorname{var}[\bar{y}_2 \mid \mathbf{x}] = \operatorname{var}\left[\frac{1}{3}\sum_{i=4}^{6} y_i \mid \mathbf{x}\right] = \frac{1}{9}\left[\sum_{i=4}^{6} \operatorname{var}(y_i \mid \mathbf{x})\right] = \frac{1}{9}\left(3\sigma^2\right) = \sigma^2/3$$

Similarly $\operatorname{var}[\bar{y}_1 \mid \mathbf{x}] = \sigma^2/3$. So that

$$\operatorname{var}\left(\hat{\beta}_{2,mean} \mid \mathbf{x}\right) = \left[1/(\bar{x}_2 - \bar{x}_1)\right]^2 \left\{\operatorname{var}[\bar{y}_2 \mid \mathbf{x}] + \operatorname{var}[\bar{y}_1 \mid \mathbf{x}]\right\} = \left[1/(\bar{x}_2 - \bar{x}_1)\right]^2 \left[\frac{\sigma^2}{3} + \frac{\sigma^2}{3}\right] = \frac{2\sigma^2}{3(\bar{x}_2 - \bar{x}_1)^2}$$

We know that $\operatorname{var}\left(\hat{\beta}_{2,mean} \mid \mathbf{x}\right)$ is larger than the variance of the least squares estimator because $\hat{\beta}_{2,mean}$ is a linear estimator. To show this note that

$$\hat{\beta}_{2,mean} = (\bar{y}_2 - \bar{y}_1)/(\bar{x}_2 - \bar{x}_1) = \frac{1}{(\bar{x}_2 - \bar{x}_1)}\left[\left(\frac{\sum_{i=4}^{6} y_i}{3}\right) - \left(\frac{\sum_{i=1}^{3} y_i}{3}\right)\right] = \left[\frac{\sum_{i=4}^{6} y_i}{3(\bar{x}_2 - \bar{x}_1)} - \frac{\sum_{i=1}^{3} y_i}{3(\bar{x}_2 - \bar{x}_1)}\right]$$
$$= \sum_{i=1}^{6} a_i y_i$$

Where $a_1 = a_2 = a_3 = \dfrac{-1}{3(\bar{x}_2 - \bar{x}_1)}$ and $a_4 = a_5 = a_6 = \dfrac{1}{3(\bar{x}_2 - \bar{x}_1)}$

Furthermore $\hat{\beta}_{2,mean}$ is an unbiased estimator. From the Gauss-Markov theorem we know that the least squares estimator is the "best" linear unbiased estimator, the one with the smallest variance. Therefore, we know that $\operatorname{var}\left(\hat{\beta}_{2,mean} \mid \mathbf{x}\right)$ is larger than the variance of the least squares estimator.

EXERCISE 2.11

(a) We estimate that each additional \$100 per month income is associated with an additional 52 cents per person expenditure, on average, on food away from home. If monthly income is zero, we estimate that household will spend an average of \$13.77 per person on food away from home.

(b) $\hat{y} = 24.17$.

(c) $\hat{\varepsilon} = 0.43$.

(d) In this log-linear relationship, the elasticity is $\hat{\varepsilon} = 0.007(20) = 0.14$.

(e) For x = 20, $d\hat{y}/dx = 0.1860$. For x = 30, $d\hat{y}/dx = 0.1995$. It is increasing at an increasing rate. Also, the second derivative, the rate of change of the first derivative is $d^2\hat{y}/dx^2 = \exp(3.14 + 0.007x)(0.007)^2 > 0$. A positive second derivative means that the function is increasing at an increasing rate for all values of x.

(f) The number of zeros is 2334 – 2005 = 329. The reason for the reduction in the number of observations is that the logarithm of zero is undefined and creates a missing data value. The software throws out the row of data when it encounters a missing value when doing its calculations.

EXERCISE 2.13

(a) We estimate that each additional 1000 *FTE* students increase real total academic cost per student by \$266, holding all else constant. The intercept suggests if there were no students the real total academic cost per student would be \$14,656.

(b) $\widehat{ACA\_LSU} = 22.0907$.

(c) $\hat{e} = -0.6877$.

(d) $\overline{ACA} = 20.732975$.

EXERCISE 2.15

(a) $$b_{EZ} = \frac{y_2 - y_1}{x_2 - x_1} = \left(\frac{1}{x_2 - x_1}\right) y_2 - \left(\frac{1}{x_2 - x_1}\right) y_1 = \sum k_i y_i$$

where $k_1 = \dfrac{-1}{x_2 - x_1}$, $k_2 = \dfrac{1}{x_2 - x_1}$, and $k_3 = k_4 = ... = k_N = 0$

Thus, b_{EZ} is a linear estimator.

(b) $$E\left(b_{EZ} \mid \mathbf{x}\right) = E\left[\frac{y_2 - y_1}{x_2 - x_1} \mid \mathbf{x}\right] = \frac{1}{x_2 - x_1}\left(\beta_1 + \beta_2 x_2\right) - \frac{1}{x_2 - x_1}\left(\beta_1 + \beta_2 x_1\right) = \beta_2 = E\left(b_{EZ}\right)$$

(c) $$\text{var}\left(b_{EZ} \mid \mathbf{x}\right) = \text{var}(\sum k_i y_i \mid \mathbf{x}) = \sum k_i^2 \,\text{var}\left(e_i \mid \mathbf{x}\right) = \sigma^2 \sum k_i^2 = \frac{2\sigma^2}{\left(x_2 - x_1\right)^2}$$

(d) If $e_i \sim N\left(0, \sigma^2\right)$, then $b_{EZ} \mid \mathbf{x} \sim N\left[\beta_2, \dfrac{2\sigma^2}{\left(x_2 - x_1\right)^2}\right]$

(e) To convince E.Z. Stuff that var(b_2|**x**) < var(b_{EZ}|**x**), we need to show that

$$\frac{2\sigma^2}{\left(x_2 - x_1\right)^2} > \frac{\sigma^2}{\sum\left(x_i - \bar{x}\right)^2} \quad \text{or that} \quad \sum\left(x_i - \bar{x}\right)^2 > \frac{\left(x_2 - x_1\right)^2}{2}$$

Consider

$$\frac{\left(x_2 - x_1\right)^2}{2} = \frac{\left[\left(x_2 - \bar{x}\right) - \left(x_1 - \bar{x}\right)\right]^2}{2} = \frac{\left(x_2 - \bar{x}\right)^2 + \left(x_1 - \bar{x}\right)^2 - 2\left(x_2 - \bar{x}\right)\left(x_1 - \bar{x}\right)}{2}$$

Thus, we need to show that

$$2\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2 > \left(x_2 - \bar{x}\right)^2 + \left(x_1 - \bar{x}\right)^2 - 2\left(x_2 - \bar{x}\right)\left(x_1 - \bar{x}\right)$$

or that

$$(x_1-\bar{x})^2+(x_2-\bar{x})^2+2(x_2-\bar{x})(x_1-\bar{x})+2\sum_{i=3}^{N}(x_i-\bar{x})^2>0$$

or that

$$\left[(x_1-\bar{x})+(x_2-\bar{x})\right]^2+2\sum_{i=3}^{N}(x_i-\bar{x})^2>0.$$

This last inequality clearly holds. Thus, b_{EZ} is not as good as the least squares estimator. Rather than prove the result directly, as we have done above, we could also refer Professor E.Z. Stuff to the Gauss Markov theorem.

EXERCISE 2.17

(a)

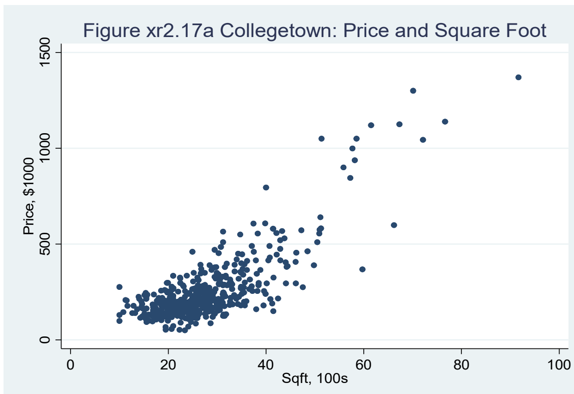


Figure xr2.11(a) Price (in $1,000s) against square feet for houses (in 100s)

(b) The fitted linear relationship is

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

We estimate that an additional 100 square feet of living area will increase the expected home price by $13,402.94 holding all else constant. The estimated intercept −115.4236 would imply that a house with zero square feet has an expected price of $−115,423.60.

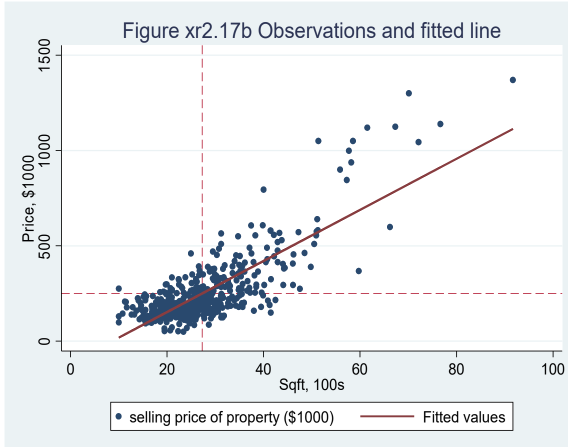


Figure xr2.17(b) Observations and fitted line

(c) The fitted quadratic model is

$$\widehat{PRICE} = 93.5659 + 0.1845 SQFT^2$$
$$(se) \qquad (6.0722) \quad (0.00525)$$

We estimate that an additional 100 square feet of living area for a 2000 square foot home will increase the expected home price by $7,380.80 holding all else constant.

(d)

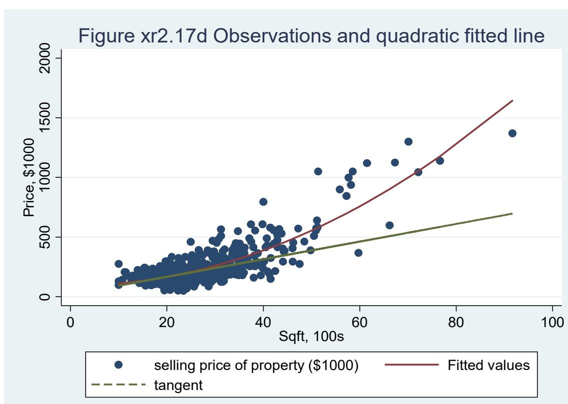
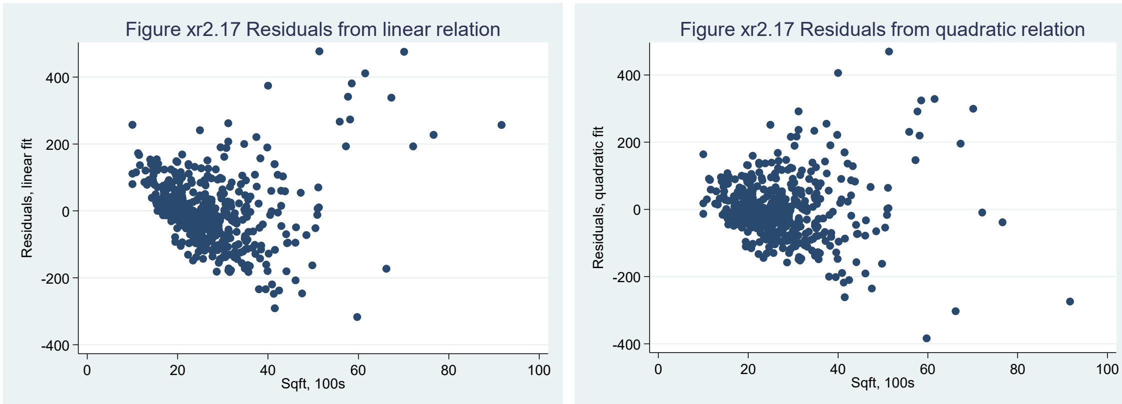


Figure xr2.17(d) Observations and quadratic fitted line

(e) $\hat{\varepsilon} = 0.882$

(f) The residual plots are



Figures xr2.17(f) Residuals from linear and quadratic relations

In both models, the residual patterns do not appear random. The variation in the residuals increases as *SQFT* increases, suggesting that the homoskedasticity assumption may be violated.

(g) The sum of square residuals linear relationship is 5,262,846.9. The sum of square residuals for the quadratic relationship is 4,222,356.3. In this case the quadratic model has the lower *SSE*. The lower *SSE* means that the data values are closer to the fitted line for the quadratic model than for the linear model.

EXERCISE 2.19

(a)

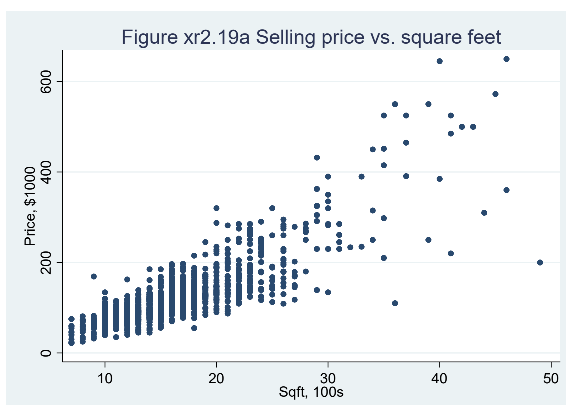


Figure xr2.19(a) Scatter plot of selling price and living area

(b) The estimated linear relationship is

$$\widehat{SPRICE} = -35.9664 + 9.8934 LIVAREA$$
$$(se) \qquad (3.3085) \quad (0.1912)$$

We estimate that an additional 100 square feet of living area will increase the expected home price by \$9,893.40 holding all else constant. The estimated intercept −35.9664 would imply that a house with zero square feet has an expected price of \$−35,966.40. This estimate is not meaningful in this example. The reason is that there are no data values with a house size near zero.

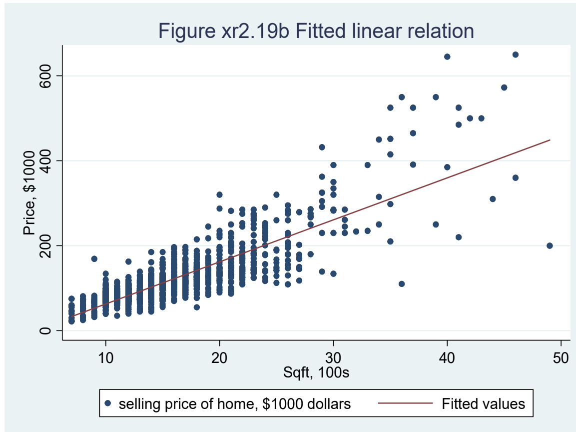


Figure xr2.19(b) Fitted linear relation

(c) The estimated quadratic equation is

$$\widehat{SPRICE} = 56.4572 + 0.2278 LIVAREA^2$$
$$(se) \qquad (1.6955) \quad (0.0043)$$

We estimate that an additional 100 square feet of living area for a 1500 square foot home will increase the expected home price by $6,834 holding all else constant.

(d)

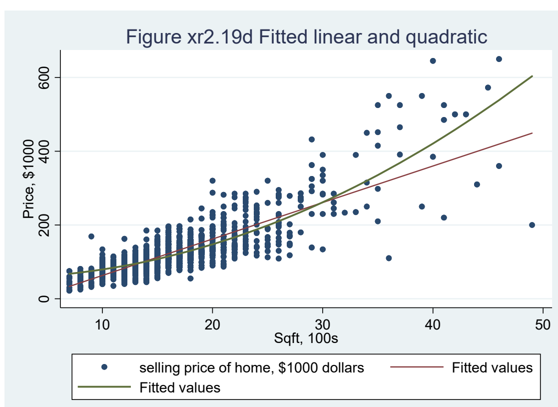


Figure xr2.19(d) Fitted linear and quadratic relations

The sum of squared residuals for the linear relation is *SSE* = 1,879,826.9948. For the quadratic model the sum of squared residuals is *SSE* = 1,795,092.2112. In this instance, the sum of squared residuals is smaller for the quadratic model, one indicator of a better fit.

(e) If the quadratic model is in fact "true," then the results and interpretations we obtain for the linear relationship are incorrect, and may be misleading.

EXERCISE 2.21

(a)

$$\widehat{SPRICE} = 152.6144 - 0.9812AGE$$
$$(se) \qquad (3.3473) \quad (0.0949)$$

We estimate that a house that is new, $AGE = 0$, will have expected price \$152,614.40. We estimate that each additional year of age will reduce expected price by \$981.20, other things held constant. The expected selling price for a 30-year-old house is $\widehat{SPRICE} = \$123{,}177.70$.

(b)

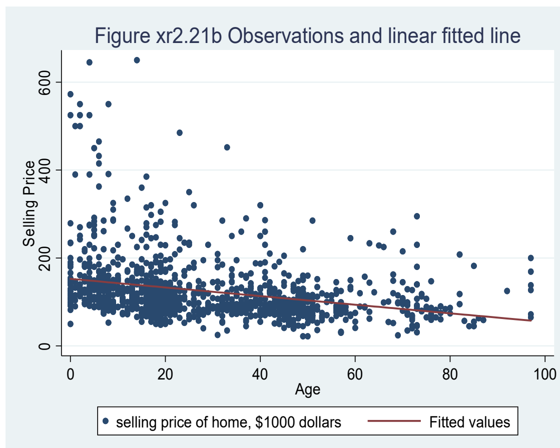


Figure xr2.21(b) Observations and linear fitted line

The data show an inverse relationship between house prices and age. The data on newer houses is not as close to the fitted regression line as the data for older homes.

(c)

$$\ln(\widehat{PRICE}) = 4.9283 - 0.0075AGE$$
$$(se) \qquad (0.0205) \; (0.0006)$$

We estimate that each additional year of age reduces expected price by about 0.75%, holding all else constant.

(d)

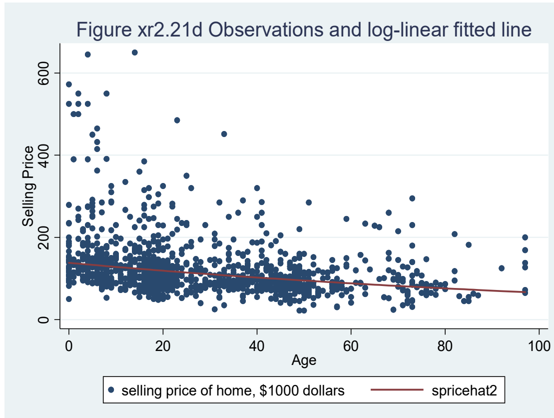


Figure xr2.21(c) Observations and log-linear fitted line

The fitted log-linear model is not too much different than the fitted linear relationship.

(e) The expected selling price of a house that is 30 years old is $\widehat{SPRICE} = \$110{,}370.32$.

(f) For the estimated linear relationship $\sum_{i=1}^{1200}\left(SPRICE - \widehat{SPRICE}\right)^2 = 5{,}580{,}871$. For the log-linear model $\sum_{i=1}^{1200}\left(SPRICE - \widehat{SPRICE}\right)^2 = 5{,}727.332$. The sum of squared differences between the data and fitted values is smaller for the estimated linear relationship, by a small margin. In this case, based on fit alone, we might choose the linear relationship rather than the log-linear relationship.

EXERCISE 2.23

(a)

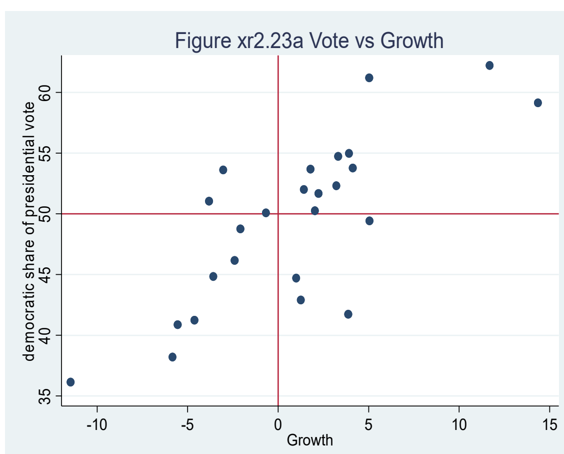


Figure xr2.23(a) Vote against Growth

There appears to be a positive association between *VOTE* and *GROWTH*.

(b) The estimated equation for 1916 to 2012 is

$$\widehat{VOTE} = 48.6160 + 0.9639GROWTH$$
$$(se) \quad (0.9043) \quad (0.1658)$$

The coefficient 0.9639 suggests that for a 1 percentage point increase in a favorable growth rate of *GDP* in the 3 quarters before the election there is an estimated increase in the share of votes of the democratic party of 0.9639 percentage points.

We estimate, based on the fitted regression intercept, that that the Democratic party's expected vote is 48.62% when the growth rate in *GDP* is zero. This suggests that when there is no real *GDP* growth, the Democratic party is expected to lose the popular vote. A graph of the fitted line and data is shown in the following figure.

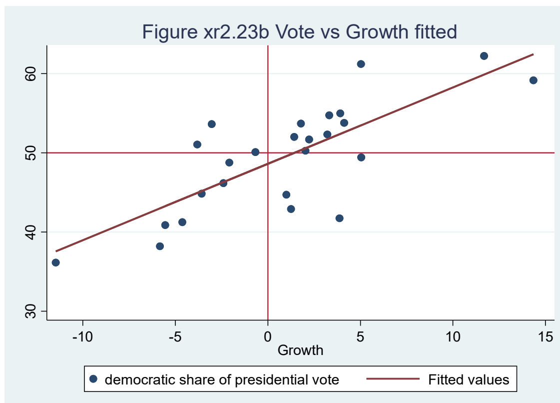


Figure xr2.23(a) Vote vs Growth fitted

(c) In 2016 the actual growth rate in GDP was 0.97% and the predicted expected vote in favor of the Democratic party was $\widehat{VOTE} = 49.55$, or 49.55%. The actual popular vote in favor of the Democratic party was 50.82%.

(d) The figure below shows a plot of *VOTE* against *INFLATION*. It is difficult to see if there is positive or inverse relationship.

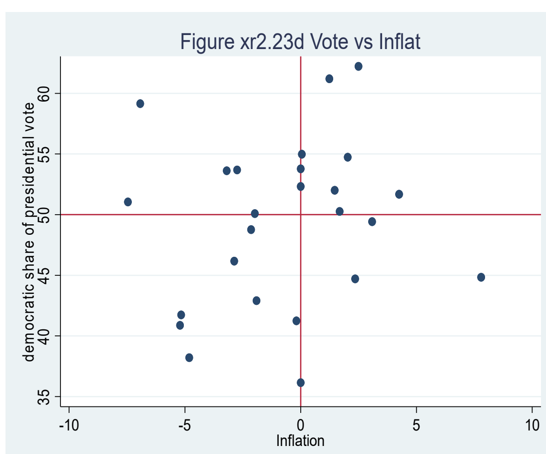


Figure xr2.23(d) Vote against Inflat

(e) The estimated equation (plotted in the figure below) is

$$\widehat{VOTE} = 49.6229 + 0.2616 INFLATION$$
$$(se) \quad (1.4188) \quad (0.3907)$$

We estimate that a 1 percentage point increase in inflation during the party's first 15 quarters increases the share of Democratic party's vote by 0.2616 percentage points. The estimated intercept suggests that when inflation is at 0% for that party's first 15 quarters, the expected share of votes won by the Democratic party is 49.6%.

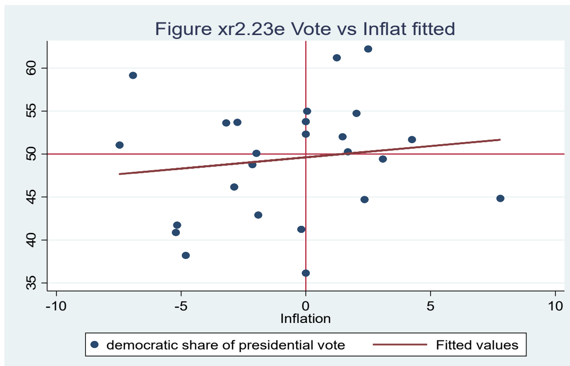


Figure xr2.23(e) Vote vs Inflat fitted

(f) The actual inflation value in the 2016 election was 1.42%. The predicted vote in favor of the Democratic candidate (Clinton) was $\widehat{VOTE} = 49.99$, or 49.99%.

EXERCISE 2.25

(a)

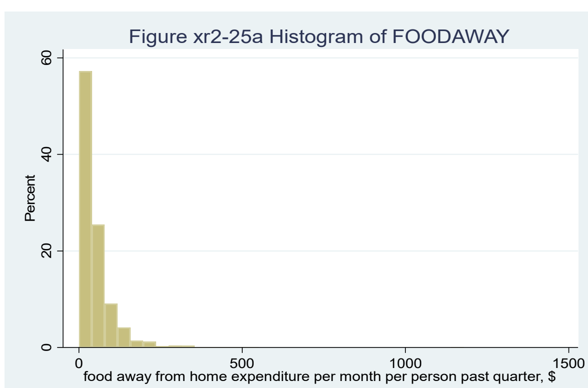


Figure xr2.25(a) Histogram of foodaway

The mean of the 1200 observations is 49.27, the 25th, 50th and 75th percentiles are 12.04, 32.56 and 67.60.

(b)

	N	Mean	Median
ADVANCED = 1	257	73.15	48.15
COLLEGE = 1	369	48.60	36.11
NONE	574	39.01	26.02

(c)

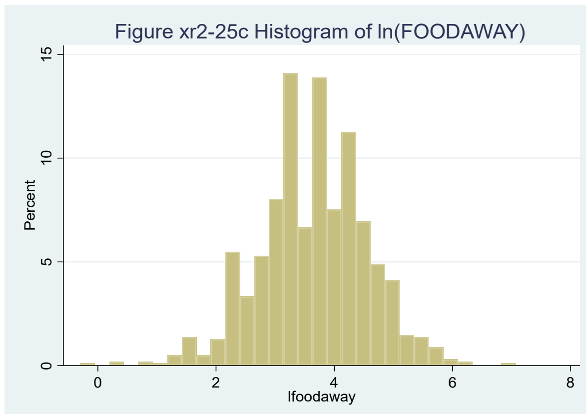


Figure xr2.25(c) Histogram of ln(foodaway)

There are 178 fewer values of ln(*FOODAWAY*) because 178 households reported spending \$0 on food away from home per person, and ln(0) is undefined. It creates a "missing value" which software cannot use in the regression.

(d) The estimated model is

$$\ln(\widehat{FOODAWAY}) = 3.1293 + 0.0069INCOME$$
$$(se) \qquad (0.0566) \quad (0.0007)$$

We estimate that each additional \$100 household income increases food away expenditures per person of about 0.69%, other factors held constant.

(e)

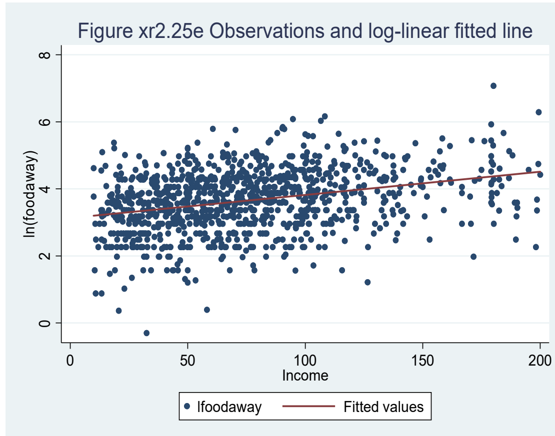


Figure xr2.25(e) Observations and log-linear fitted line

The plot shows a positive association between ln(*FOODAWAY*) and *INCOME*s.

(f)

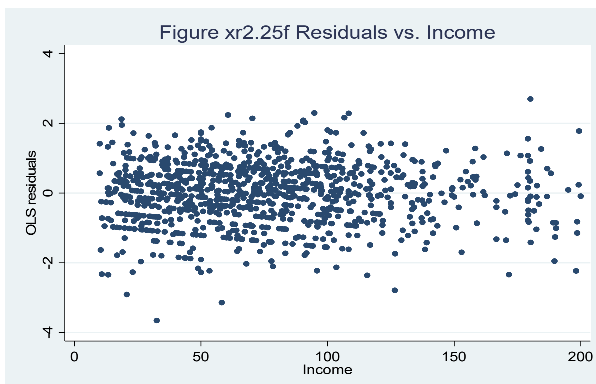


Figure xr2.25(f) Residuals vs. income

The OLS residuals do appear randomly distributed with no obvious patterns. There are fewer observations at higher incomes, so there is more “white space.”

EXERCISE 2.27

(a)

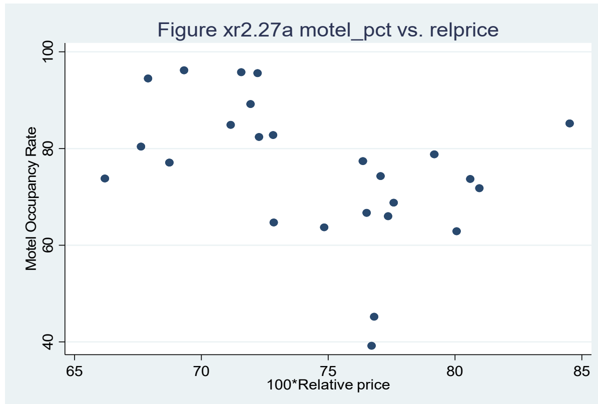


Figure xr2.27(a) Motel_pct vs. 100relprice

There seems to be an inverse association between relative price and occupancy rate.

(b)

$$\widehat{MOTEL\_PCT}_t = 166.6560 - 1.2212 RELPRICE_t$$
$$(se) \qquad (43.5709) \quad (0.5835)$$

Based economic reasoning we anticipate a negative coefficient for RELPRICE. The slope estimate is interpreted as saying, the expected model occupancy rate falls by 1.22% given a 1% increase in relative price, other factors held constant.

(c)

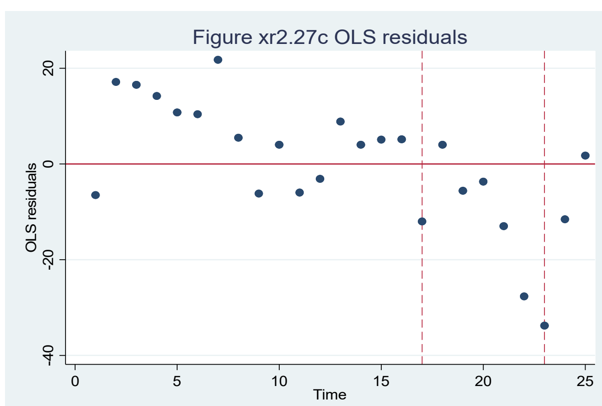


Figure xr2.27(c) OLS residuals

The residuals are scattered about zero for the first 16 observations but for observations 17-23 all but one of the residuals is negative. This suggests that the occupancy rate was lower than predicted by the regression model for these dates.

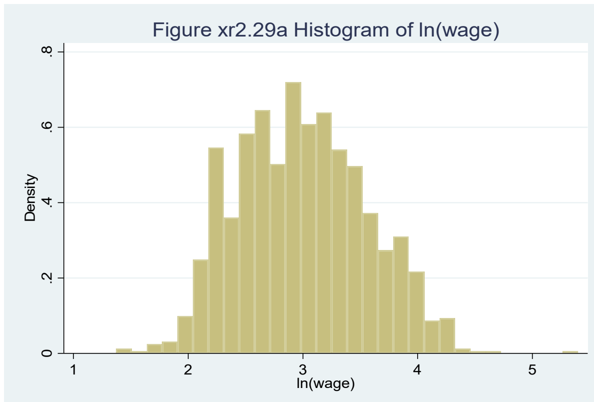
(d)

$$\widehat{MOTEL\_PCT}_t = 79.3500 - 13.2357 REPAIR_t$$
$$(se) \qquad (3.1541) \quad (5.9606)$$

We estimate that during the non-repair period the expected occupancy rate is 79.35%. During the repair period, the expected occupancy rate is estimated to fall by 13.24%, other things held constant, to 66.11%.

EXERCISE 2.29

(a)



variable	N	mean	median	min	max	skewness	kurtosis
ln(*WAGE*)	1200	2.9994	2.9601	1.3712	5.3986	0.2306	2.6846

Figure xr2.29(a) Histogram and statistics for ln(*WAGE*)

The histogram shows the distribution of ln(*WAGE*) to be almost symmetrical. Note that the mean and median are similar, which is not the case for skewed distributions. The skewness coefficient is not quite zero. Similarly, the kurtosis is not quite three, as it should be for a normal distribution.

(b) The OLS estimates are

$$\ln(\widehat{WAGE}) = 1.5968 + 0.0987EDUC$$
$$(se) \qquad (0.0702) \quad (0.0048)$$

We estimate that each additional year of education predicts a 9.87% higher wage, all else held constant.

(c) For someone with 12 years of education the predicted value is $\widehat{WAGE} = 16.1493$ and for someone with 16 years of education it is $\widehat{WAGE} = 23.9721$.

(d) For individuals with 12 and 16 years of education, respectively, these values are \$1.1850 and \$1.5801.

(e)

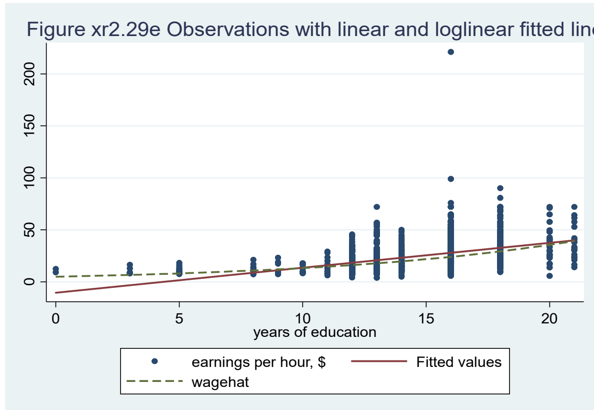


Figure xr2.29(e) Observations with linear and loglinear fitted lines

The log-linear model fits the data better at low levels of education.

(f) For the log-linear model this value is 228,573.5 and for the linear model 220,062.3. Based on this measure the linear model fits the data better than the linear model.