EXERCISE 2.1

(a)

x	y	$x-\overline{x}$	$(x-\overline{x})^2$	$y-\overline{y}$	$(x-\overline{x})(y-\overline{y})$
3	4	2	4	2	4
2	2	1	1	0	0
1	3	0	0	1	0
-1	1	-2	4	-1	2
0	0	-1	1	-2	2
$\sum x_i =$ 5	$\sum y_i =$ 10	$\sum(x_i-\overline{x}) =$ 0	$\sum(x_i-\overline{x})^2 =$ 10	$\sum(y-\overline{y}) =$ 0	$\sum(x-\overline{x})(y-\overline{y}) =$ 8

$\overline{x}=1, \quad \overline{y}=2$

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

b_2 is the estimated slope of the fitted line.

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

b_1 is the estimated value of $E(y)$ when $x=0$; it is the intercept of the fitted line.

(c)
$$\sum_{i=1}^{5} x_i^2 = 3^2 + 2^2 + 1^2 + (-1)^2 + 0^2 = 15$$
$$\sum_{i=1}^{5} x_i y_i = 3\times 4 + 2\times 2 + 1\times 3 + (-1)\times 1 + 0\times 0 = 18$$
$$\sum_{i=1}^{5} x_i^2 - N\overline{x}^2 = 15 - 5\times 1^2 = 10 = \sum_{i=1}^{5}(x_i-\overline{x})^2$$
$$\sum_{i=1}^{5} x_i y_i - N\overline{x}\overline{y} = 18 - 5\times 1\times 2 = 8 = \sum_{i=1}^{5}(x_i-\overline{x})(y_i-\overline{y})$$

(d)

x_i	y_i	$\hat{y}_i$	$\hat{e}_i$	$\hat{e}_i^2$	$x_i\hat{e}_i$
3	4	3.6	0.4	0.16	1.2
2	2	2.8	-0.8	0.64	-1.6
1	3	2	1	1	1
-1	1	0.4	0.6	0.36	-0.6
0	0	1.2	-1.2	1.44	0
$\sum x_i =$ 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

$$s_y^2 = \sum_{i=1}^{N} (y_i - \bar{y})^2 / (N-1) = 10/4 = 2.5$$
$$s_x^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2 / (N-1) = 10/4 = 2.5$$
$$s_{xy} = \sum_{i=1}^{N} (y_i - \bar{y})(x_i - \bar{x}) / (N-1) = 8/4 = 2$$
$$r_{xy} = s_{xy} / (s_x s_y) = 2 / \left(\sqrt{2.5}\sqrt{2.5}\right) = 0.8$$
$$CV_x = 100(s_x/\bar{x}) = 100\sqrt{2.5}/1 = 158.11388$$
$$\text{median}(x) = 1$$

(e)

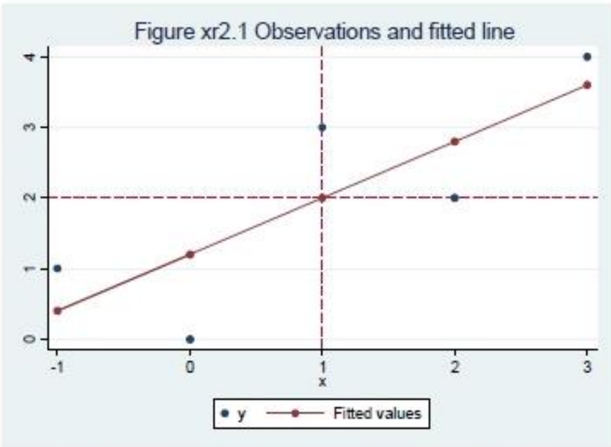

Figure xr2.1 Observations and fitted line

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

(g) Given $b_1 = 1.2$, $b_2 = 0.8$, $\bar{x} = 1$, $\bar{y} = 2$ and $\bar{y} = b_1 + b_2\bar{x}$, we have
$\bar{y} = 2 = b_1 + b_2\bar{x} = 1.2 + 0.8(1) = 2$

(h) $\bar{\hat{y}} = \sum \hat{y}_i / N = (3.6 + 2.8 + 2 + 0.4 + 1.2)/5 = 2 = \bar{y}$

(i) $$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{N-2} = \frac{3.6}{3} = 1.2$$

(j) $$\widehat{\text{var}}(b_2) = \frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2} = \frac{1.2}{10} = 0.12 \text{ and } se(b_2) = \sqrt{\widehat{\text{var}}(b_2)} = \sqrt{0.12} = 0.34641$$

EXERCISE 2.22

(a) The regression model is $TOTALSCORE = \beta_1 + \beta_2 SMALL + e$. Under the model assumptions

$$E(TOTALSCORE \mid SMALL) = \beta_1 + \beta_2 SMALL = \begin{cases} \beta_1 + \beta_2 & \text{if} \quad SMALL = 1 \\ \beta_1 & \text{if} \quad SMALL = 0 \end{cases}$$

Thus β_1 is the expected total score in regular sized classes, and $\beta_1 + \beta_2$ is the expected total score in small classes. The difference β_2 is an estimate of the difference in performance in small and regular sized classes. The model estimates are given in Table xr2-22a, Model (1).

Table xr2-22a

		C	$SMALL$	N	SSE
(1) *TOTALSCORE*	Coeff	916.4417	12.1753	775	4300389
	Std. err.	(3.6746)	(5.3692)		
(2) *READSCORE*	Coeff	432.6650	6.9245	775	705200
	Std. err.	(1.4881)	(2.1743)		
(3) *MATHSCORE*	Coeff	483.7767	5.2508	775	1910009
	Std. err.	(2.4489)	(3.5783)		

The estimated equation using a sample of small and regular classes (where $AIDE = 0$) is

$$\widehat{TOTALSCORE} = 916.442 + 12.175 SMALL$$

Comparing a sample of small and regular classes, we find students in regular classes achieve an average total score of 916.442 while students in small classes achieve an average of $916.442 + 12.175 = 928.617$. This is a 1.33% increase. This result suggests that small classes have a positive impact on learning, as measured by higher totals of all achievement test scores.

(b) The estimated equations using a sample of small and regular classes are given in Table xr2-22a as Models (2) and (3)

$$\widehat{READSCORE} = 432.665 + 6.925 SMALL$$

$$\widehat{MATHSCORE} = 483.77 + 5.251 SMALL$$

Students in regular classes achieve an average reading score of 432.7 while students in small classes achieve an average of 439.6. This is a 1.60% increase. In math students in regular classes achieve an average score of 483.77 while students in small classes achieve an average of 489.0. This is a 1.08% increase. These results suggests that small class sizes also have a positive impact on learning math and reading.

Exercise 2.22 (continued)

(c) The estimated equations using a sample of regular classes and regular classes with a full-time teacher aide (when $SMALL = 0$) are given in Table xr2-22b

Table xr2-22b

		C	$AIDE$	N	SSE
(4) $TOTALSCORE$	Coeff Std. err.	916.4417 (3.5586)	4.3065 (4.9940)	837	4356550
(5) $READSCORE$	Coeff Std. err.	432.6650 (1.4600)	2.8714 (2.0489)	837	733335
(6) $MATHSCORE$	Coeff Std. err.	483.7767 (2.3546)	1.4351 (3.3043)	837	1907234

$$\widehat{TOTALSCORE} = 916.442 + 4.31 AIDE$$

Students in regular classes without a teacher aide achieve an average total score of 916.4 while students in regular classes with a teacher aide achieve an average total score of 920.7. This is an increase of 0.47%. These results suggest that having a full-time teacher aide has a small impact on learning outcomes as measured by totals of all achievement test scores.

(d) The estimated equations using a sample of regular classes and regular classes with a full-time teacher aide are

$$\widehat{READSCORE} = 432.67 + 2.87 AIDE$$

$$\widehat{MATHSCORE} = 483.78 + 1.44 AIDE$$

The effect of having a teacher aide on learning is 0.66% for reading and 0.30% for math. These increases are smaller than the increases provided by smaller classes.

```
#R code

#a #b
star5_small = read.csv("C:/Users/Hsien/Downloads/star5_small.csv")
head(star5_small)
lm(totalscore ~ 1 + small, subset(star5_small, aide==0))
lm(readscore ~ 1 + small, subset(star5_small, aide==0))
lm(mathscore ~ 1 + small, subset(star5_small, aide==0))
#c
lm(totalscore ~ 1 + aide, subset(star5_small, small==0))
lm(readscore ~ 1 + aide, subset(star5_small, small==0))
lm(mathscore ~ 1 + aide, subset(star5_small, small==0))
```

EXERCISE 2.25

(a)

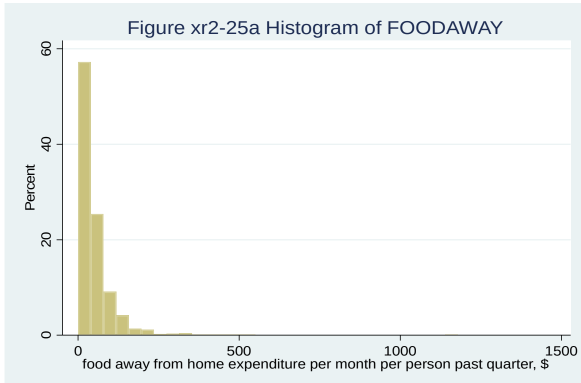

Figure xr2.25(a) Histogram of foodaway

The mean of the 1200 observations is 49.27, the 25th, 50th and 75th percentiles are 12.04, 32.56 and 67.60. The histogram figure shows a very skewed distribution, with a mean that is larger than the median. 50% of households spend $32.56 per person or less during a quarter.

(b) Households with a member with an advanced degree spend an average of about $25 more per person than households with a member with a college degree, but not advanced degree. Households with a member with a college degree, but not advanced degree, spend an average of about $9 more per person than households with no members with a college or advanced degree.

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

Exercise 2.25 (continued)

(c)

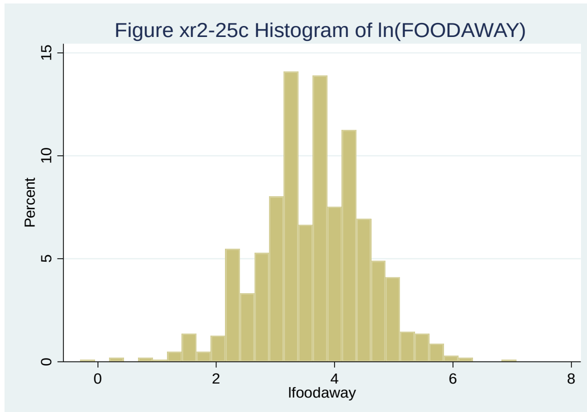

Figure xr2.25(c) Histogram of ln(foodaway)

The histogram of ln(*FOODAWAY*) is much less skewed. 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. If any variable has a missing value in either y_i or x_i the entire observation is deleted from regression calculations.

(d) The estimated model is

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

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

(e)

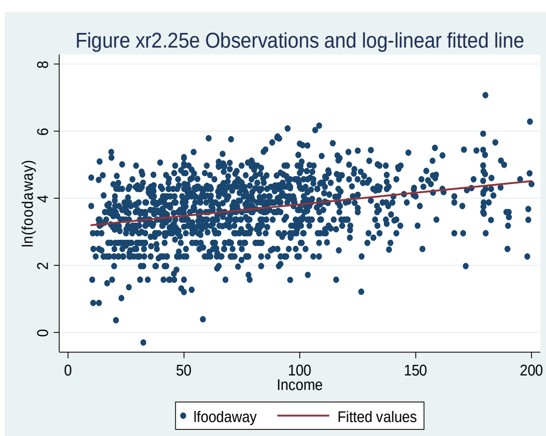

Figure xr2.25(e) Observations and log-linear fitted line

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

Exercise 2.25 (continued)

(f)

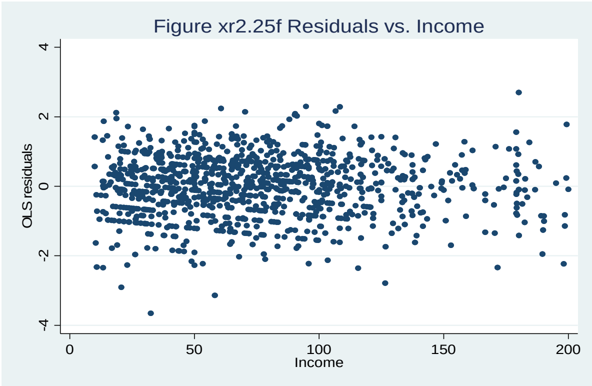

Figure xr2.25(f) Residuals vs. income

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

Prove

a. $\min_{\beta} \sum_{i=1}^{n} (y_i - \beta x_i)^2 = S(\beta)$

F.O.C. $\frac{\partial S(\beta)}{\partial \beta} = 2\sum_{i=1}^{n} x_i (y_i - \beta x_i) = 0$

$$\sum_{i=1}^{n} x_i y_i - \beta \sum_{i=1}^{n} x_i^2 = 0$$

$$\hat{\beta} = \sum_{i=1}^{n} x_i y_i \Big/ \sum_{i=1}^{n} x_i^2$$

b. $\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$, assume $w_i = \frac{x_i}{\sum x_i^2}$, then $\hat{\beta} = \sum w_i y_i$

$$\Rightarrow \hat{\beta} = \sum w_i (\beta x_i + e_i) = \beta \sum w_i x_i + \sum w_i e_i$$

$$= \beta + \sum w_i e_i$$

$$E(\hat{\beta}) = E(\beta + \sum w_i e_i) = \beta + \sum E(w_i e_i) = \beta \ \#$$

$$Var(\hat{\beta}) = Var(\beta + \sum w_i e_i) = \left(\frac{\sum x_i}{\sum x_i^2}\right)^2 Var(e_i) = \frac{\sigma^2}{\sum x_i^2} \ \#$$

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum x_i^2}\right) \ \#$$