2.11.1 Problems

2.1 Consider the following five observations. You are to do all the parts of this exercise using only a calculator.

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

x^2	x	y	$x-\bar{x}$	$(x-\bar{x})^2$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	xy
9	3	4	2	4	2	4	12
4	2	2	1	1	0	0	4
1	1	3	0	0	1	0	3
1	–1	1	-2	4	-1	2	-1
0	0	0	-1	1	-2	2	0
$\Sigma x^2 = 15$	$\sum x_i =$	$\sum y_i =$	$\sum(x_i-\bar{x}) =$ 0	$\sum(x_i-\bar{x})^2 =$ 10	$\sum(y_i-\bar{y}) =$ 0	$\sum(x_i-\bar{x})(y_i-\bar{y}) =$ 8	$\Sigma xy = 18$

a. Complete the entries in the table. Put the sums in the last row. What are the sample means $\bar{x}$ and $\bar{y}$?

b. Calculate b_1 and b_2 using (2.7) and (2.8) and state their interpretation.

c. Compute $\sum_{i=1}^{5} x_i^2$, $\sum_{i=1}^{5} x_i y_i$. Using these numerical values, show that $\sum(x_i-\bar{x})^2 = \sum x_i^2 - N\bar{x}^2$ and $\sum(x_i-\bar{x})(y_i-\bar{y}) = \sum x_i y_i - N\bar{x}\bar{y}$.

d. Use the least squares estimates from part (b) to compute the fitted values of y, and complete the remainder of the table below. Put the sums in the last row.

Calculate the sample variance of y, $s_y^2 = \sum_{i=1}^{N}(y_i-\bar{y})^2/(N-1)$, the sample variance of x, $s_x^2 = \sum_{i=1}^{N}(x_i-\bar{x})^2/(N-1)$, the sample covariance between x and y, $s_{xy} = \sum_{i=1}^{N}(y_i-\bar{y})(x_i-\bar{x})/(N-1)$, the sample correlation between x and y, $r_{xy} = s_{xy}/(s_x s_y)$ and the coefficient of variation of x, $CV_x = 100(s_x/\bar{x})$. What is the median, 50th percentile, of x?

x_i	y_i	$\hat{y}_i$	$\hat{e}_i$	$\hat{e}_i^2$	$x_i\hat{e}_i$
3	4				
2	2				
1	3				
–1	1				

d. Use the least squares estimates from part (b) to compute the fitted values of y, and complete the remainder of the table below. Put the sums in the last row.
Calculate the sample variance of y, $s_y^2 = \sum_{i=1}^N (y_i - \bar{y})^2/(N-1)$, the sample variance of x, $s_x^2 = \sum_{i=1}^N (x_i - \bar{x})^2/(N-1)$, the sample covariance between x and y, $s_{xy} = \sum_{i=1}^N (y_i - \bar{y})(x_i - \bar{x})/(N-1)$, the sample correlation between x and y, $r_{xy} = s_{xy}/(s_x s_y)$ and the coefficient of variation of x, $CV_x = 100(s_x/\bar{x})$. What is the median, 50th percentile, of x?

$\hat{y}_i = b_1 + b_2 x_i$ $\hat{e}_i = y_i - \hat{y}_i$
$= 1.2 + 0.8 x_i$

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	$\sum x_i\hat{e}_i =$ 0.8

e. On graph paper, plot the data points and sketch the fitted regression line $\hat{y}_i = b_1 + b_2 x_i$.
f. On the sketch in part (e), locate the point of the means $(\bar{x}, \bar{y})$. Does your fitted line pass through that point? If not, go back to the drawing board, literally.
g. Show that for these numerical values $\bar{y} = b_1 + b_2\bar{x}$.
h. Show that for these numerical values $\bar{\hat{y}} = \bar{y}$, where $\bar{\hat{y}} = \sum \hat{y}_i/N$.
i. Compute $\hat{\sigma}^2$.
j. Compute $\widehat{\text{var}}(b_2|\mathbf{x})$ and $\text{se}(b_2)$.

2.2 A household has weekly income of \$2000. The mean weekly expenditure for households with this income is $E(y|x = \$2000) = \mu_{y|x=\$2000} = \$220$, and expenditures exhibit variance $\text{var}(y|x = \$2,000) = \sigma^2_{y|x=\$2,000} = \$121$.

a. Assuming that weekly food expenditures are normally distributed, find the probability that a household with this income spends between \$200 and \$215 on food in a week. Include a sketch with your solution.

fitted related in part (a), compute the predicted value of *ACA*.

c. The actual value of *ACA* for LSU that year was 21.403. Calculate the least squares residual for LSU? Does the model overpredict or underpredict *ACA* for LSU?

d. The sample mean (average) full-time enrollment in U.S. public research universities in 2011 was 22,845.77. What was the sample mean of academic cost per student?

2.14 Consider the regression model $WAGE = \beta_1 + \beta_2 EDUC + e$, where *WAGE* is hourly wage rate in U.S. 2013 dollars and *EDUC* is years of education, or schooling. The regression model is estimated twice using the least squares estimator, once using individuals from an urban area, and again for individuals in a rural area.

$$\begin{array}{llll} \text{Urban} & \widehat{WAGE} = -10.76 + 2.46\, EDUC, & N = 986 \\ & \text{(se)} \qquad\quad (2.27) \quad (0.16) & \end{array}$$

$$\begin{array}{llll} \text{Rural} & \widehat{WAGE} = -4.88 + 1.80\, EDUC, & N = 214 \\ & \text{(se)} \qquad\quad (3.29) \quad (0.24) & \end{array}$$

a. Using the estimated rural regression, compute the elasticity of wages with respect to education at the "point of the means." The sample mean of *WAGE* is \$19.74.

b. The sample mean of *EDUC* in the urban area is 13.68 years. Using the estimated urban regression, compute the standard error of the elasticity of wages with respect to education at the "point of the means." Assume that the mean values are "givens" and not random.

c. What is the predicted wage for an individual with 12 years of education in each area? With 16 years of education?

2.15 Professor E.Z. Stuff has decided that the least squares estimator is too much trouble. Noting that two points determine a line, Dr. Stuff chooses two points from a sample of size *N* and draws a line between them, calling the slope of this line the EZ estimator of β_2 in the simple regression model. Algebraically,

2.11.2 Computer Exercises

2.16 The capital asset pricing model (CAPM) is an important model in the field of finance. It explains variations in the rate of return on a security as a function of the rate of return on a portfolio consisting of all publicly traded stocks, which is called the *market* portfolio. Generally, the rate of return on any investment is measured relative to its opportunity cost, which is the return on a risk-free asset. The resulting difference is called the *risk premium*, since it is the reward or punishment for making a risky investment. The CAPM says that the risk premium on security j is *proportional* to the risk premium on the market portfolio. That is,

$$r_j - r_f = \beta_j(r_m - r_f)$$

where r_j and r_f are the returns to security j and the risk-free rate, respectively, r_m is the return on the market portfolio, and β_j is the jth security's "*beta*" value. A stock's *beta* is important to investors since it reveals the stock's volatility. It measures the sensitivity of security j's return to variation in the whole stock market. As such, values of *beta* less than one indicate that the stock is "defensive" since its variation is less than the market's. A *beta* greater than one indicates an "aggressive stock." Investors usually want an estimate of a stock's *beta* before purchasing it. The CAPM model shown above is the "economic model" in this case. The "econometric model" is obtained by including an intercept in the model (even though theory says it should be zero) and an error term

$$r_j - r_f = \alpha_j + \beta_j(r_m - r_f) + e_j$$

a. Explain why the econometric model above is a simple regression model like those discussed in this chapter.

b. In the data file *capm5* are data on the monthly returns of six firms (GE, IBM, Ford, Microsoft, Disney, and Exxon-Mobil), the rate of return on the market portfolio (*MKT*), and the rate of return on the risk-free asset (*RISKFREE*). The 180 observations cover January 1998 to December 2012. Estimate the CAPM model for each firm, and comment on their estimated *beta* values. Which firm appears most aggressive? Which firm appears most defensive?

c. Finance theory says that the intercept parameter α_j should be zero. Does this seem correct given your estimates? For the Microsoft stock, plot the fitted regression line along with the data scatter.

d. Estimate the model for each firm under the assumption that $\alpha_j = 0$. Do the estimates of the *beta* values change much?

2.17 The data file *collegetown* contains observations on 500 single-family houses sold in Baton Rouge, Louisiana, during 2009–2013. The data include sale price (in thousands of dollars), *PRICE*, and total interior area of the house in hundreds of square feet, *SQFT*.

2.1

a. $\bar{x} = \dfrac{3+2+1-1+0}{5} = 1$

$\bar{y} = \dfrac{4+2+3+1+0}{5} = 2$

b. $b_2 = \dfrac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2} = \dfrac{8}{10} = 0.8$

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

b_1 為斜率，代表当 x 增加 1 个單位時，y 会增加 0.8 个單位

b_2 為截距項，代表当 $x=0$ 時，y 会 $= 1.2$

c. $\sum_{i=1}^{5} x_i^2 = 9+4+1+1 = 15$

$\sum_{i=1}^{5} x_i y_i = 12+4+3-1 = 18$

$\Sigma(x_i-\bar{x})^2 = 10$

$\Sigma x_i^2 - 5\bar{x}^2 = 15 - 5(1) = 10$

$\Sigma(x_i-\bar{x})(y_i-\bar{y}) = 8$

$\Sigma x_i y_i - 5(\bar{x}\bar{y}) = 18 - 5(2) = 8$

d. $s_y^2 = \dfrac{\Sigma(y_i-\bar{y})^2}{N-1} = \dfrac{2^2+0+1^2+(-1)^2+(-2)^2}{4} = 2.5$

$s_x^2 = \dfrac{\Sigma(x_i-\bar{x})^2}{N-1} = \dfrac{2^2+1^2+0+(-2)^2+(-1)^2}{4} = 2.5$

$s_{xy} = \dfrac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{N-1} = \dfrac{8}{4} = 2$

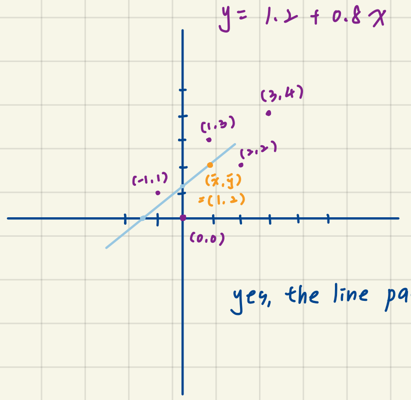
$r_{xy} = \dfrac{s_{xy}}{s_x s_y} = 0.8$

$CV_x = 100\left(\dfrac{s_x}{\bar{x}}\right) = 100\left(\dfrac{\sqrt{2.5}}{1}\right) = 158.1139$

Median of $x = 1$

e.

f.



yes, the line pass the point $(\bar{x},\bar{y}) = (1,2)$

g. $\bar{y} = 2 \quad \bar{x} = 1$

$\bar{y} = b_1 + b_2\bar{x} = 1.2 + 0.8(1) = 2$

h. $\bar{\hat{y}} = \dfrac{\Sigma \hat{y}_i}{N} = 2 = \bar{y}$

i. $\hat{\sigma}^2 = \dfrac{\Sigma \hat{e}_i^2}{N} = \dfrac{3.6}{5} = 0.72$

j. $\widehat{var}(b_2|x) = \dfrac{\hat{\sigma}^2}{\Sigma(x_i-\bar{x})^2} = \dfrac{0.72}{10} = 0.072$

$se(b_2) = \sqrt{\widehat{var}(b_2|x)} = \sqrt{0.072} = 0.2683$

2.16

a. the econometric model is:

$(r_j - r_f) = \alpha + \beta(r_m - r_f) + e_j$

and this is also a simple regression model where dependent var. (y) is $(r_j - r_f)$ and independent var. (x) is $(r_m - r_f)$

$\beta_1 = \alpha, \ \beta_2 = \beta_j$

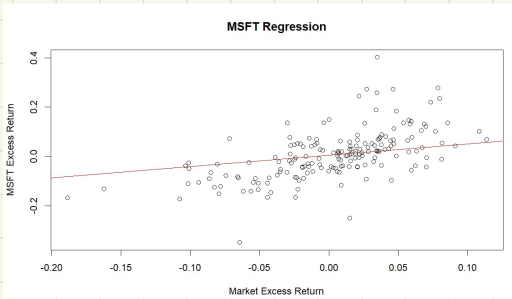
b.

公司	α(截距項)	β(市場風險)	t 值(β)	P 值(β)	R^2
GE	-0.0009586	1.1479521	12.821	<2e-16	0.4801
IBM	0.0060525	0.9768898	9.985	<2e-16	0.359
FORD	0.0037789	1.662031	8.032	1.27e-13	0.266
MSFT	0.0032496	1.20184	9.839	<2e-16	0.3523
DIS	0.0010469	1.011521	10.688	<2e-16	0.3909
XOM	0.0052835	0.4565208	6.38	1.48e-09	0.1861

Ford is the most aggressive because it has the highest $\beta = 1.662$. It means when market move 1 unit, Ford would move 1.66 units

ExxonMobil is the most defensive because it has the lowest $\beta = 0.456$. It means when market move 1 unit, Exxonmbil would move 0.456 units

c.



在現實中大部分公司的 $\alpha \neq 0$，也代表 ① CAPM 可能不完美，市場 risk premium 不能完美解釋个股報酬 ② 有其他影响的因子

d. Let $\alpha = 0 \Rightarrow (r_j - r_f) = \beta(r_m - r_f) + e_j$

```
Company: ge
Slope: 1.146763
------------------------
Company: ibm
Slope: 0.9843954
------------------------
Company: ford
Slope: 1.666717
------------------------
Company: msft
Slope: 1.205869
------------------------
Company: dis
Slope: 1.012819
------------------------
Company: xom
Slope: 0.4630727
------------------------
```

β 变动幅度不大，表示市場风險溢酬 $(r_m - r_f)$ 仍是最重要的解釋变数

2.14

a. $\bar{w}_R = 19.74$

$19.74 = -4.88 + 1.8\,\bar{E}_R$

$\bar{E}_R = 13.6778$

$\hat{\varepsilon} = b_2 \dfrac{\bar{E}_R}{\bar{w}_R} = 1.8 \dfrac{13.6778}{19.74} = 1.2472$

b. $\bar{E}_U = 13.68$

$\bar{w}_U = -10.76 + 2.46(13.68) = 22.89$

$se(\hat{\varepsilon}) = \sqrt{var(\hat{\varepsilon})}$

$= \sqrt{var(b_2 \times \frac{\bar{x}}{\bar{y}})}$

$= \sqrt{\left(\frac{\bar{x}}{\bar{y}}\right)^2} \times \sqrt{var(b_2)}$

$= \dfrac{\bar{x}}{\bar{y}} \times se(b_2) = \dfrac{13.68}{22.89}(0.16) = 0.0956$

c.

Urban: $-10.76 + 2.46(12) = 18.76$

$-10.76 + 2.46(16) = 28.6$

Rural: $-4.88 + 1.8(12) = 16.72$

$-4.88 + 1.8(16) = 23.92$