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.

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

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

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	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

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 $\text{se}(b_2)$.

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

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

⇒ 當x增加1單位，估計的y增加0.8單位

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

⇒ 當$x=0$時，估計的y為1.2

c. $\sum_{i=1}^{5} x_i^2 = 3^2+2^2+1^2+(-1)^2 = 15$

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

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

$\Sigma(x_i-\bar{x})(y_i-\bar{y}) = 8 = \Sigma x_i y_i - N\bar{x}\bar{y} = 18 - 5\times1\times2 = 8$

d.

$\hat{y}_i = 1.2 + 0.8x_i$	$\hat{e}_i = y_i - \hat{y}_i$
$\hat{y}_1 = 1.2 + 0.8\times3 = 3.6$	$\hat{e}_1 = 4 - 3.6 = 0.4$
$\hat{y}_2 = 1.2 + 0.8\times2 = 2.8$	$\hat{e}_2 = 2 - 2.8 = -0.8$
$\hat{y}_3 = 1.2 + 0.8\times1 = 2$	$\hat{e}_3 = 3 - 2 = 1$
$\hat{y}_4 = 1.2 \times 0.8 \times -1 = 0.4$	$\hat{e}_4 = 1 - 0.4 = 0.6$
$\hat{y}_5 = 1.2 + 0.8\times0 = 1.2$	$\hat{e}_5 = 0 - 1.2 = -1.2$

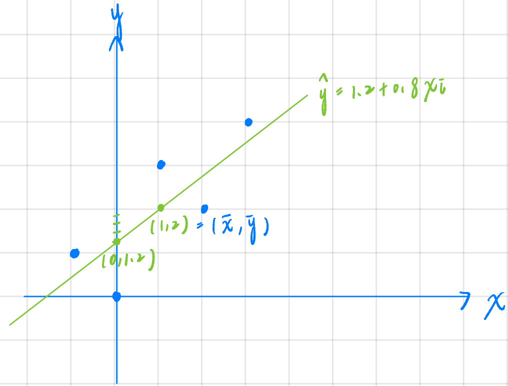
$s_y^2 = \dfrac{\Sigma(y_i-\bar{y})^2}{N-1} = \dfrac{10}{4} = 2.5 \qquad s_x^2 = \dfrac{\Sigma(x_i-\bar{x})^2}{N-1} = \dfrac{10}{4} = 2.5$

$s_{xy} = \dfrac{\Sigma(y_i-\bar{y})(x_i-\bar{x})}{N-1} = \dfrac{8}{4} = 2 \qquad r_{xy} = \dfrac{s_{xy}}{s_x s_y} = \dfrac{2}{\sqrt{2.5}\times\sqrt{2.5}} = 0.8$

$CV_x = 100\left(\dfrac{s_x}{\bar{x}}\right) = 100\times\dfrac{\sqrt{2.5}}{1} = 50\sqrt{10}$

$x_i = \{0,-1,1,2,3\}$ ⇒ median of x is 1.

e.



f. Yes, the fitted line pass through $(\bar{x},\bar{y}) = (1,2)$

g. $\bar{y} = 2 = b_1 + b_2\bar{x} = 1.2 + 0.8\times1 = 2$

h. $\bar{\hat{y}} = 10/5 = 2 = \bar{y} = 2$

i. $\hat{\sigma}^2 = \dfrac{\Sigma\hat{e}_i^2}{n-2} = \dfrac{3.6}{5-2} = 1.2$

j. $Var(b_2|x) = \dfrac{\hat{\sigma}^2}{SST_x} = \dfrac{1.2}{10} = 0.12$

$se(b_2) = \sqrt{Var(b_2|x)} = \sqrt{0.12} = 0.3464$

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.

$$\text{Urban} \quad \widehat{WAGE} = -10.76 + 2.46\, EDUC, \quad N = 986$$
$$\text{(se)} \quad (2.27) \quad (0.16)$$

$$\text{Rural} \quad \widehat{WAGE} = -4.88 + 1.80\, EDUC, \quad N = 214$$
$$\text{(se)} \quad (3.29) \quad (0.24)$$

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?

a. $\varepsilon = b_2 \dfrac{\bar{x}}{\bar{y}}$

$\bar{y} = 19.74 \Rightarrow 19.74 = -4.88 + 1.8 \times \overline{EDUC} \Rightarrow \overline{EDUC} = 13.6778$

$\varepsilon = 1.8 \times \dfrac{13.6778}{19.74} = 1.2472$

b. $SE(\varepsilon) = SE(b_2)\dfrac{\bar{x}}{\bar{y}}$

$\bar{y} = -10.76 + 2.46 \times 13.68 = 22.8928$

$SE(\varepsilon) = 0.16 \times \dfrac{13.68}{22.8928} = 0.0956$

c.

1. EDUC = 12

① Urban Wage $= -10.76 + 2.46 \times 12 = 18.76$

② Rural Wage $= -4.88 + 1.8 \times 12 = 16.72$

2. EDUC = 16

① Urban Wage $= -10.76 + 2.46 \times 16 = 28.6$

② Rural Wage $= -4.88 + 1.8 \times 16 = 23.92$

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 *j*th 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?

a ∵ $r_j - r_f$ ⇒ dependent variable

$r_m - r_f$ ⇒ independent variable

β_j ⇒ Regression coefficient

α_j ⇒ Intercept

e_j ⇒ error

∴ $r_j - r_f = \alpha_j + \beta_j(r_m - r_f) + e_j$ can be considered as linear regression model

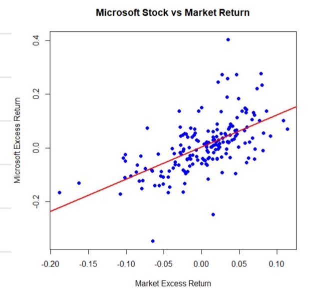
b. 詳細迴歸結果 ⇒ 下一頁

```
print(beta_values)
ge_execss  ibm_execss ford_execss msft_execss  dis_execss  xom_execss
1.1479521   0.9768898   1.6620307   1.2018398   1.0115207   0.4565208
```

Aggressive stock ⇒ larger β ⇒ ford

Defensive stock ⇒ smaller β ⇒ Exxon-Mobile

c.



	公司	intercept_values	intercept_T	intercept_P	beta_values
ge_execss	GE	-0.0009586682	-0.2166783	0.8287072	1.1479521
ibm_execss	IBM	0.0060525497	1.2519510	0.2122303	0.9768898
ford_execss	Ford	0.0037789112	0.3695640	0.7121467	1.6620307
msft_execss	MSFT	0.0032496009	0.5383843	0.5909844	1.2018398
dis_execss	DIS	0.0010469237	0.2238779	0.8231091	1.0115207
xom_execss	XOM	0.0052835329	1.4944284	0.1368343	0.4565208

⇒ T value < 2 & p value > 0.05

⇒ Don't reject $H_0: \alpha_j = 0$

d.

	公司	ori_Beta	alpha_0_Beta
ge_execss	GE	1.1479521	1.1467633
ibm_execss	IBM	0.9768898	0.9843954
ford_execss	Ford	1.6620307	1.6667168
msft_execss	MSFT	1.2018398	1.2058695
dis_execss	DIS	1.0115207	1.0128190
xom_execss	XOM	0.4565208	0.4630727

⇒ The results of assumption that $\alpha_j = 0$ is similar to the results of original α_j.

回歸結果: ge_execss

```
Call:
lm(formula = as.formula(paste(company, "~ mkt_execss")), data = capm5)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.173801 -0.033907 -0.003789  0.038858  0.202748 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.0009587  0.0044244  -0.217    0.829    
mkt_execss   1.1479521  0.0895394  12.821   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05925 on 178 degrees of freedom
Multiple R-squared:  0.4801,	Adjusted R-squared:  0.4772 
F-statistic: 164.4 on 1 and 178 DF,  p-value: < 2.2e-16
```

回歸結果: ibm_execss

```
Call:
lm(formula = as.formula(paste(company, "~ mkt_execss")), data = capm5)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.254880 -0.035266 -0.005322  0.031490  0.274520 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.006053   0.004834   1.252    0.212    
mkt_execss  0.976890   0.097839   9.985   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06474 on 178 degrees of freedom
Multiple R-squared:  0.359,	Adjusted R-squared:  0.3554 
F-statistic: 99.69 on 1 and 178 DF,  p-value: < 2.2e-16
```

回歸結果: ford_execss

```
Call:
lm(formula = as.formula(paste(company, "~ mkt_execss")), data = capm5)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.27315 -0.07875 -0.01198  0.04720  1.08874 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.003779   0.010225   0.370    0.712    
mkt_execss  1.662031   0.206937   8.032 1.27e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1369 on 178 degrees of freedom
Multiple R-squared:  0.266,	Adjusted R-squared:  0.2619 
F-statistic: 64.51 on 1 and 178 DF,  p-value: 1.271e-13
```

回歸結果: msft_execss

```
Call:
lm(formula = as.formula(paste(company, "~ mkt_execss")), data = capm5)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.27424 -0.04744 -0.00820  0.03869  0.35801 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.003250   0.006036   0.538    0.591    
mkt_execss  1.201840   0.122152   9.839   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08083 on 178 degrees of freedom
Multiple R-squared:  0.3523,	Adjusted R-squared:  0.3486 
F-statistic:  96.8 on 1 and 178 DF,  p-value: < 2.2e-16
```

回歸結果: dis_execss

```
Call:
lm(formula = as.formula(paste(company, "~ mkt_execss")), data = capm5)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.176233 -0.030021 -0.004232  0.029179  0.280528 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.001047   0.004676   0.224    0.823    
mkt_execss  1.011521   0.094638  10.688   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06262 on 178 degrees of freedom
Multiple R-squared:  0.3909,	Adjusted R-squared:  0.3875 
F-statistic: 114.2 on 1 and 178 DF,  p-value: < 2.2e-16
```

回歸結果: xom_execss

```
Call:
lm(formula = as.formula(paste(company, "~ mkt_execss")), data = capm5)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.114474 -0.030596 -0.001884  0.026849  0.215396 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.005284   0.003535   1.494    0.137    
mkt_execss  0.456521   0.071550   6.380 1.48e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04734 on 178 degrees of freedom
Multiple R-squared:  0.1861,	Adjusted R-squared:  0.1816 
F-statistic: 40.71 on 1 and 178 DF,  p-value: 1.48e-09
```