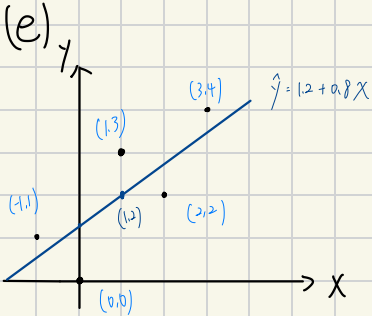**2.1** Consider the following five observations. You are to do all the parts of this exercise using only a calculator.

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

$\bar{X} = 1 \quad \bar{Y} = 2$

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 $\hat{\bar{y}} = \bar{y}$, where $\hat{\bar{y}} = \sum \hat{y}_i/N$.
i. Compute $\hat{\sigma}^2$.
j. Compute $\widehat{\text{var}}(b_2|\mathbf{x})$ and $se(b_2)$.

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

$Y = b_1 + b_2 \bar{X}$

$2 = b_1 + 0.8 \times 1$

$b_1 = 1.2$

$b_2 = 0.8$, 表示 X 每增加 1 單位, 預測的 Y 增加 0.8

$b_1 = 1.2$, 表示當 $X = 0$, 預測的 Y 為 1.2

(c) $\sum_{i=1}^{5} X_i^2 = 15 \qquad \sum_{i=1}^{5} X_i Y_i = 18$

$\sum(X_i - \bar{X})^2 = \sum X_i^2 - N\bar{X}^2$
$$= 15 - 5 \times 1^2 = 10$$

$\sum(X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - N\bar{X}\bar{Y}$
$$= 18 - 5 \times 1 \times 2$$
$$= 8$$

(d)
$$\hat{Y}_i = 1.2 + 0.8 X_i$$

(e)

(f)
$(\bar{X}, \bar{Y}) = (1, 2)$

迴歸線通過 $(\bar{X}, \bar{Y}) = (1, 2)$

(g)
$\hat{Y} = 1.2 + 0.8 \times 1 = 2$

(h)
$\sum \hat{Y}_i = 10 \qquad \bar{\hat{Y}}_i = \frac{10}{5} = 2 = \bar{Y}$

(i)
$$\hat{\sigma}^2 = \frac{\sum e_i^2}{N-2} = \frac{3.6}{5-2} = 1.2$$

(j)
$$\widehat{Var}(b_2|X) = \frac{\hat{\sigma}^2}{\sum(X_i - \bar{X})^2} = \frac{1.2}{10} = 0.12$$

$$SE(b_2) = \sqrt{Var(b_2|X)} = \sqrt{0.12} = \frac{\sqrt{3}}{5} = 0.3464$$

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.

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

Rural $\quad \widehat{WAGE} = -4.88 + 1.80\, EDUC, \quad N = 214$
$\quad\quad\quad$ (se) $\quad\quad$ (3.29) (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) $\varepsilon = \dfrac{dy/y}{dx/x} = 1.8\,\dfrac{EDUC}{WAGE} = 1.8\,\dfrac{(19.74+4.88)/1.8}{19.74}$

$\quad = \dfrac{\cancel{1.8}(19.74+4.88)}{\cancel{1.8}\times 19.74} = 1.2472$

(b) $\varepsilon = \dfrac{dy/y}{dx/x} = \beta_2\,\dfrac{EDUC}{WAGE}$

$\quad \hat{\varepsilon} = b_2\,\dfrac{EDUC}{WAGE}$

$\quad SE(\hat{\varepsilon}) = \sqrt{Var(\hat{\varepsilon})} = \dfrac{EDUC}{WAGE}\times 0.16 = \dfrac{13.68}{(-10.76+2.46\times13.68)}\times 0.16 = 0.0956$

(C) $EDUC = 12$

Urban $\quad \widehat{Wage} = -10.76 + 2.46 \times 12 = 18.76$

Rural $\quad \widehat{Wage} = -4.88 + 1.8 \times 12 = 16.72$

$EDOC = 16$

Urban $\quad \widehat{Wage} = -10.76 + 2.46 \times 16 = 28.6$

$\quad\quad\quad \widehat{Wage} = -4.88 + 1.8 \times 16 = 23.92$

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?

(c.)

```
> summary(modmsft)

Call:
lm(formula = capm5$msft - capm5$mkt ~ capm5$msft, data = capm5)

Residuals:
     Min        1Q    Median        3Q       Max
-0.090849 -0.021605 -0.006033  0.022099  0.139477

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.002678   0.002975   -0.90    0.369
capm5$msft   0.707370   0.029646   23.86   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03975 on 178 degrees of freedom
Multiple R-squared:  0.7618,    Adjusted R-squared:  0.7605
F-statistic: 569.3 on 1 and 178 DF,  p-value: < 2.2e-16
```
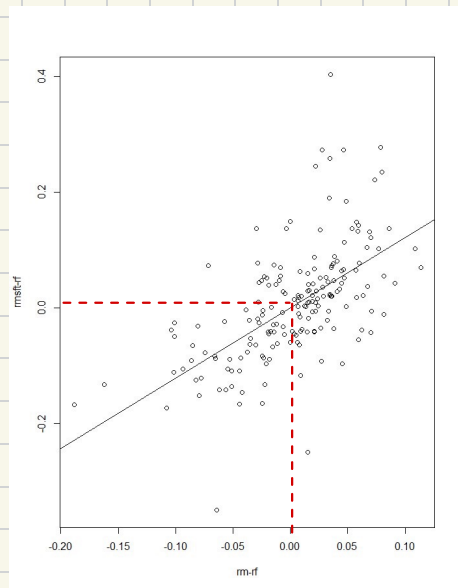
截距項 p-value 為 0.951 不顯著
∴不拒絕 H₀：$\alpha_{microsoft} = 0$
無充分証據說明 $\alpha \neq 0$，因此此模型估計
符合財務理論假設

a.

簡迴歸模型：$Y = \beta_0 + \beta_1 X + e$

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

可將 $r_j - r_f$ 視為應變數 Y　$(r_m - r_f)$ 視為自變數 X
β_j 代表資產相對於市場的 beta 值，即該資產對市場
報酬變動的敏感度
截距項 α_j 代表市場報酬無法解釋的額外報酬
誤差項 e_j 表影響個股報酬但未被市場報酬
解釋的隨機因素

(b.)
$\widehat{\beta}_{GE} = 1.1536$　　$\widehat{\beta}_{Ford} = 1.6610$　　$\widehat{\beta}_{Disney} = 1.0110$

$\widehat{\beta}_{IBM} = 0.9820$　　$\widehat{\beta}_{Microsoft} = 1.2113$　　$\widehat{\beta}_{xom} = 0.4598$

Ford 的 $\widehat{\beta}$ 值最大 ∴ most aggresive
Exxon-Mobil 的 $\widehat{\beta}$ 值最小 ∴ most defensive

(d)
$\widehat{\beta}^{*}_{GE} = 1.1454$　　$\widehat{\beta}^{*}_{Ford} = 1.6610$　　$\widehat{\beta}^{*}_{Disney} = 1.0080$

$\widehat{\beta}^{*}_{IBM} = 0.9903$　　$\widehat{\beta}^{*}_{Microsoft} = 1.2121$　　$\widehat{\beta}^{*}_{xom} = 0.4689$

在假定 $a_j = 0$ 情況下，與原先計算 $\widehat{\beta}_j$ 差異
不大