2.1

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	2				
1	3				
–1	1				
0	0				
$\sum x_i =$	$\sum y_i =$	$\sum(x_i-\bar{x}) =$	$\sum(x_i-\bar{x})^2 =$	$\sum(y_i-\bar{y}) =$	$\sum(x_i-\bar{x})(y_i-\bar{y}) =$

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

a.

	x	y	x-xbar	(x-xbar)^2	y-ybar	(x-xbar)(y-ybar)
	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	5	10	0	10	0	8

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

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 $\sum(x_i-\bar{x})(y_i-\bar{y}) = \sum x_i y_i - N\bar{x}\bar{y}$.

b.

$$b_2 = \frac{8}{10} = 0.8, b_1 = 2 - 0.8 * 1 = 1.2$$

The Ordinary Least Squares (OLS) Estimators

$$b_2 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} \tag{2.7}$$

$$b_1 = \bar{y} - b_2\bar{x} \tag{2.8}$$

where $\bar{y} = \sum y_i/N$ and $\bar{x} = \sum x_i/N$ are the sample means of the observations on y and x.

c.

	x^2	y^2		(x-xbar)^2	sumx^2-Nxbar		sum(x-xbar)(y-ybar)	xy	N*xbar*ybar
	9	16		4				12	
	4	4		1				4	
	1	9		0				3	
	1	1		4				-1	
	0	0		1				0	
sum	15	30		10	10		8	18	10

As u can see, $\sum x_i^2 - N\bar{x}^2 = 15 - 5 * 1 = 10 = \sum(x_i - \bar{x})^2$

$\sum x_i y_i - N\bar{x}\bar{y} = 18 - 5 * 1 * 2 = 8 = \sum(x_i - \bar{x})\,(y_i - \bar{y})$

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

d.

	x	y	yhat	ehat	ehat^2	x*ehat
	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	5	10	10	0	3.6	0

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?

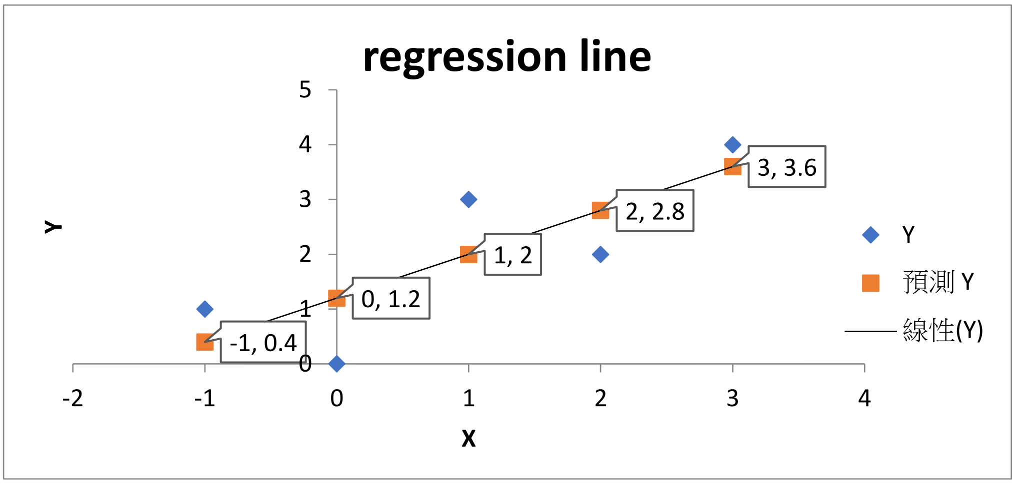
Variance of y= $\frac{10}{4} = 2.5$, Variance of x= $\frac{10}{4} = 2.5$

covariance= $\frac{8}{4} = 2$ 、correlation= $\frac{2}{2.5} = 0.8$

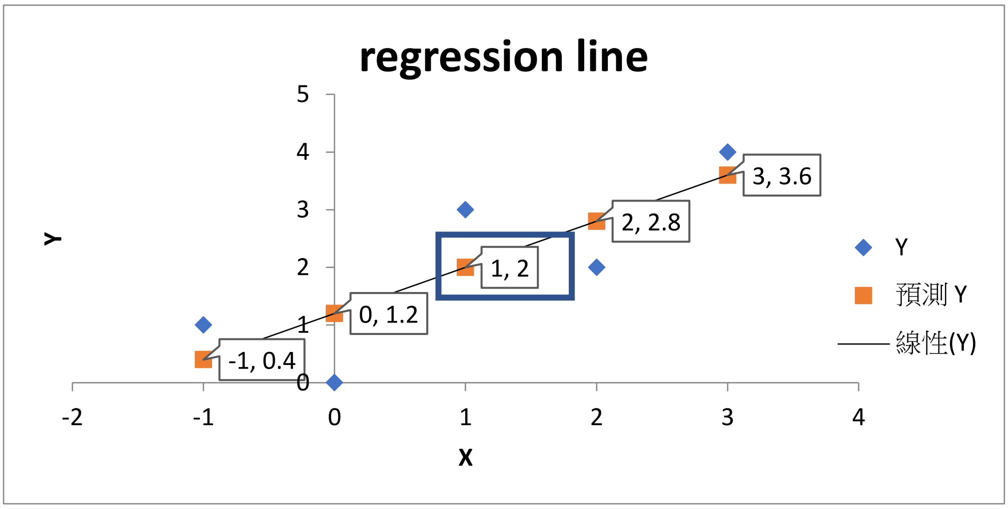
coefficient= $100 * \frac{\sqrt{2.5}}{1} = 158.11$ 、mean= 1

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 $\bar{y} = b_1 + b_2 \bar{x}$.

e.



f.



g.

$2 = 1.2 + 0.8 * 1$

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 $\text{se}(b_2)$.

h.

$(3.6 + 2.8 + 2 + 0.4 + 1.2)/5 = 2$

i.

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{N} \qquad = \frac{3.6}{5} = 0.72$$

j.

$$\widehat{\operatorname{var}}(b_2|\mathbf{x}) = \frac{\hat{\sigma}^2}{\sum(x_i - \bar{x})^2} = 0.72/10 = 0.072$$

$$\operatorname{se}(b_2) = \sqrt{\widehat{\operatorname{var}}(b_2|\mathbf{x})} = \sqrt{0.072} = 0.268328$$

2.14

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.

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

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

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.

a.

elasticity$= 1.8 * \frac{13.67778}{19.74} = 1.247214$

a.	1.8	wage	19.74	Educ	13.67778	elasticity	1.247214
b.	2.46	wage	22.8928	Educ	13.68	elasticity	1.470017

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?

b.

elasticity$= 2.46 * \frac{13.68}{22.8928} = 1.470017$

standard error of elasticity = var(elasticity)=var(beta_2)*13.68/22.8928

$$= 0.16 * 0.16 * \frac{13.68}{22.8928} = 0.015298$$

c.

	12	16
Urban	18.76	28.6
Rural	16.72	23.92

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

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

將 12、16 分別帶入 EDUC 即可求

2.16

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?

a.

簡單迴歸模型為$y = alpha + beta * x + error$, 其中 CAPM 中的$r_j - r_f$ 就是 y；$r_m - r_f$就是 x，因此 CAPM 為簡單迴歸

b.

```
Most Aggressive Firm (highest beta): Ford MKT_excess with beta = 1.662031
> cat("Most Defensive Firm (lowest beta):", most_defensive,
+     "with beta =", betas[most_defensive], "\n")
Most Defensive Firm (lowest beta): ExxonMobil MKT_excess with beta = 0.4565208
```

Code:

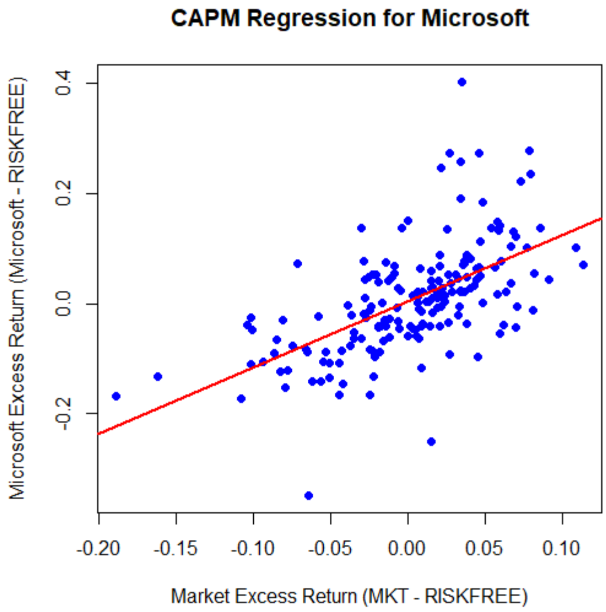
```
# b小題
# 計算超額回報
capm5$GE_excess <- capm5$ge - capm5$riskfree
capm5$IBM_excess <- capm5$ibm - capm5$riskfree
capm5$Ford_excess <- capm5$ford - capm5$riskfree
capm5$Microsoft_excess <- capm5$msft - capm5$riskfree
capm5$Disney_excess <- capm5$dis - capm5$riskfree
capm5$ExxonMobil_excess <- capm5$xom - capm5$riskfree
capm5$MKT_excess <- capm5$mkt - capm5$riskfree

# 為每個公司運行 CAPM 回歸
model_GE <- lm(GE_excess ~ MKT_excess, data = capm5)
beta_GE <- coef(model_GE)[2]

model_IBM <- lm(IBM_excess ~ MKT_excess, data = capm5)
beta_IBM <- coef(model_IBM)[2]

model_Ford <- lm(Ford_excess ~ MKT_excess, data = capm5)
beta_Ford <- coef(model_Ford)[2]

model_Microsoft <- lm(Microsoft_excess ~ MKT_excess, data = capm5)
beta_Microsoft <- coef(model_Microsoft)[2]

model_Disney <- lm(Disney_excess ~ MKT_excess, data = capm5)
beta_Disney <- coef(model_Disney)[2]

model_ExxonMobil <- lm(ExxonMobil_excess ~ MKT_excess, data = capm5)
beta_ExxonMobil <- coef(model_ExxonMobil)[2]

# 收集 beta 值
betas <- c(GE = beta_GE, IBM = beta_IBM, Ford = beta_Ford,
           Microsoft = beta_Microsoft, Disney = beta_Disney,
           ExxonMobil = beta_ExxonMobil)

# 打印 beta 值
cat("Estimated Beta Values:\n")
print(betas)

# 找出最激進和最保守的公司
max_beta_idx <- which.max(betas)
min_beta_idx <- which.min(betas)

most_aggressive <- names(betas)[which.max(betas)]
most_defensive <- names(betas)[which.min(betas)]

cat("\nMost Aggressive Firm (highest beta):", most_aggressive,
    "with beta =", betas[most_aggressive], "\n")
cat("Most Defensive Firm (lowest beta):", most_defensive,
    "with beta =", betas[most_defensive], "\n")
```

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.

The intercept (alpha) for Microsoft is 0.0032 with a p-value of 0.591 , indicating it is not statistically different from zero at the 5% significance level.



Code:

```r
# c小題
# Print the summary to check the intercept (alpha)
summary(model_Microsoft)

# Extract the intercept (alpha) and its p-value
alpha_Microsoft <- coef(model_Microsoft)[1]
p_value_alpha <- summary(model_Microsoft)$coefficients[1, 4]

# Comment on whether alpha is statistically zero
if (p_value_alpha < 0.05) {
  cat("The intercept (alpha) for Microsoft is", round(alpha_Microsoft, 4),
      "with a p-value of", round(p_value_alpha, 4),
      ", indicating it is statistically different from zero at the 5% significance level.\n")
} else {
  cat("The intercept (alpha) for Microsoft is", round(alpha_Microsoft, 4),
      "with a p-value of", round(p_value_alpha, 4),
      ", indicating it is not statistically different from zero at the 5% significance level.\n")
}

# Plot the scatter plot with the fitted regression line
plot(capm5$MKT_excess, capm5$Microsoft_excess,
     xlab = "Market Excess Return (MKT - RISKFREE)",
     ylab = "Microsoft Excess Return (Microsoft - RISKFREE)",
     main = "CAPM Regression for Microsoft",
     pch = 16, col = "blue")
abline(model_Microsoft, col = "red", lwd = 2)  # Add regression line
```

d.

```
GE.MKT_excess: The beta changes by -0.1%, which is a small change (less than 5%).
IBM.MKT_excess: The beta changes by 0.77%, which is a small change (less than 5%).
Ford.MKT_excess: The beta changes by 0.28%, which is a small change (less than 5%).
Microsoft.MKT_excess: The beta changes by 0.34%, which is a small change (less than 5%).
Disney.MKT_excess: The beta changes by 0.13%, which is a small change (less than 5%).
ExxonMobil.MKT_excess: The beta changes by 1.44%, which is a small change (less than 5%).
```

All of the beta values change small

```r
# d小題
# Use -1 in the formula to remove the intercept
model_GE_no_intercept <- lm(GE_excess ~ MKT_excess - 1, data = capm5)
beta_GE_no_intercept <- coef(model_GE_no_intercept)[1]

model_IBM_no_intercept <- lm(IBM_excess ~ MKT_excess - 1, data = capm5)
beta_IBM_no_intercept <- coef(model_IBM_no_intercept)[1]

model_Ford_no_intercept <- lm(Ford_excess ~ MKT_excess - 1, data = capm5)
beta_Ford_no_intercept <- coef(model_Ford_no_intercept)[1]

model_Microsoft_no_intercept <- lm(Microsoft_excess ~ MKT_excess - 1, data = capm5)
beta_Microsoft_no_intercept <- coef(model_Microsoft_no_intercept)[1]

model_Disney_no_intercept <- lm(Disney_excess ~ MKT_excess - 1, data = capm5)
beta_Disney_no_intercept <- coef(model_Disney_no_intercept)[1]

model_ExxonMobil_no_intercept <- lm(ExxonMobil_excess ~ MKT_excess - 1, data = capm5)
beta_ExxonMobil_no_intercept <- coef(model_ExxonMobil_no_intercept)[1]

# Collect betas without intercept
betas_no_intercept <- c(GE = beta_GE_no_intercept, IBM = beta_IBM_no_intercept,
                        Ford = beta_Ford_no_intercept, Microsoft = beta_Microsoft_no_intercept,
                        Disney = beta_Disney_no_intercept, ExxonMobil = beta_ExxonMobil_no_intercept)

# Compare the betas
cat("Betas with Intercept:\n")
print(betas)
cat("\nBetas without Intercept (alpha = 0):\n")
print(betas_no_intercept)

# Calculate the percentage change in betas
percent_change <- ((betas_no_intercept - betas) / betas) * 100
cat("\nPercentage Change in Betas (without intercept vs with intercept):\n")
print(percent_change)

# Comment on whether the betas change much
cat("\nComment on Beta Changes:\n")
for (firm in names(percent_change)) {
  change <- percent_change[firm]
  if (abs(change) < 5) {
    cat(firm, ": The beta changes by ", round(change, 2),
        "%, which is a small change (less than 5%).\n", sep = "")
  } else {
    cat(firm, ": The beta changes by ", round(change, 2),
        "%, which is a noticeable change (more than 5%).\n", sep = "")
  }
}
```