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	y	$x - \bar{x}$	$(x - \bar{x})^2$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$	$(y-\bar{y})^2$
3	4	2	4	2	4	4
2	2	1	1	0	0	0
1	3	0	0	1	0	1
–1	1	-2	4	-1	2	1
0	0	-1	1	-2	2	4
$\sum x_i =$ 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	$\sum(y-\bar{y})^2 =$ 10

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?

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

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

a. sample mean $\bar{x} = \frac{1}{5}(3+2+1-1+0) = \frac{1}{5} \times 5 = 1$

sample mean $\bar{y} = \frac{1}{5}(4+2+3+1+0) = \frac{1}{5} \times 10 = 2$

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

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

estimated regression line: $\hat{y}_i = 1.2 + 0.8x_i$

b_2 (slope) $= 0.8$, it means that for every 1-unit increase in x, the predicted value of y increase by 0.8 unit.

b_1 (intercept) $= 1.2$, it means that when $x = 0$, the predicted value of y is 1.2.

c. $\sum_{i=1}^{5} x_i^2 = 3^2 + 2^2 + 1^2 + (-1)^2 + 0^2 = 15$

$\sum_{i=1}^{5} x_i y_i = 3 \times 4 + 2 \times 2 + 1 \times 3 + (-1) \times 1 + 0 \times 0 = 18$

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

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

d. sample variance

$S_y^2 = \sum_{i=1}^{N}(y_i - \bar{y})^2/N-1 = 10/4 = 2.5$

$S_x^2 = \sum_{i=1}^{N}(x_i - \bar{x})^2/N-1 = 10/4 = 2.5$

sample covariance

$S_{xy} = \sum_{i=1}^{5}(x_i - \bar{x})(y_i - \bar{y})/N-1 = 8/4 = 2$

sample correlation

$r_{xy} = S_{xy}/S_x S_y = 2/(\sqrt{2.5} \times \sqrt{2.5}) = 0.8$

coefficient of variation

$CV_x = 100(S_x/\bar{x}) = 100 \times \sqrt{2.5}/1 = 50\sqrt{10} \approx 158.11$

$x: \{-1, 0, 1, 2, 3\}$, median of $x = 1$

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.

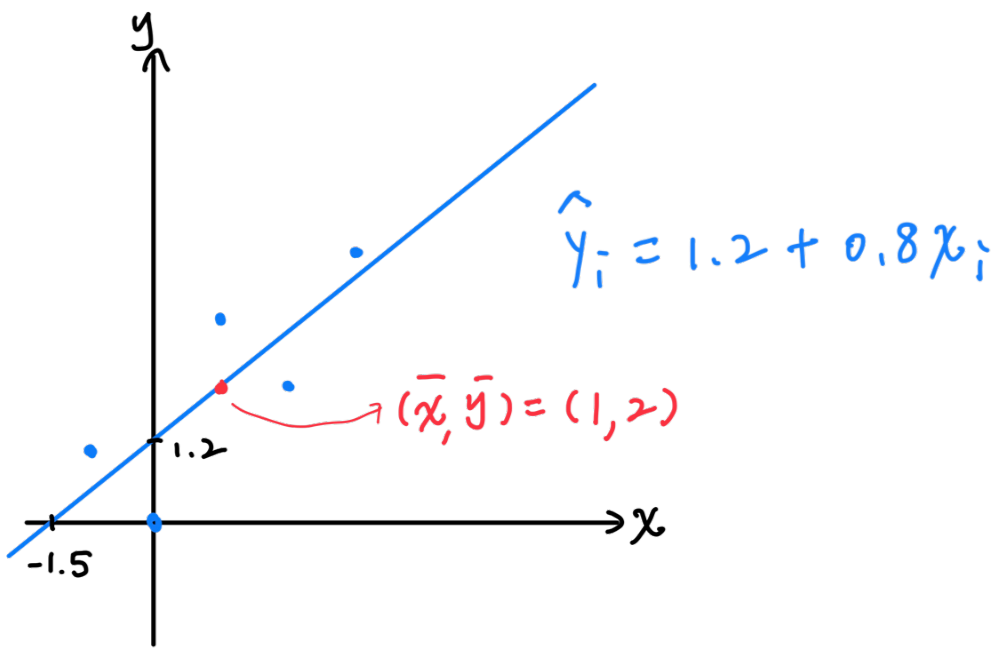
g. Show that for these numerical values $\bar{y} = b_1 + b_2 \bar{x}$.

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 $\widehat{\operatorname{var}}(b_2|\mathbf{x})$ and $\operatorname{se}(b_2)$.

e.



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

g. $\bar{y} = 2 = 1.2 + 0.8 \times 1 = 1.2 \times 0.8\,\bar{x}$

h. $\bar{\hat{y}} = \frac{1}{5}(3.6 + 2.8 + 2 + 0.4 + 1.2) = \frac{1}{5} \times 10 = 2 = \bar{y}$

i. $\hat{\sigma}^2 = \dfrac{\sum \hat{e}_i^2}{N-2} = 3.6 \div (5-2) = 1.2$

j. $\widehat{\operatorname{var}}(b_2 | x) = \dfrac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2} = 1.2/10 = 0.12$

$\operatorname{se}(b_2) = \sqrt{\widehat{\operatorname{var}}(b_2|x)} = \sqrt{0.12} = \dfrac{\sqrt{3}}{5} \approx 0.3464$

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.

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

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

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?

a. elasticity $(\varepsilon) = b_2 \times \dfrac{\bar{x}}{\bar{y}}$

$b_2 = 1.80, \quad \bar{y} = 19.74$

$19.74 = -4.88 + 1.80 \times \bar{x}, \quad \bar{x} \approx 13.6778$

$\varepsilon = 1.80 \times \dfrac{13.6778}{19.74} \approx 1.2472$

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

$SE(b_2) = 0.16, \quad \bar{x} = 13.68$

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

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

c. ① EDUC = 12

Urban wage = $-10.76 + 2.46 \times 12 = 18.76$

Rural wage = $-4.88 + 1.80 \times 12 = 16.72$

② EDUC = 16

Urban wage = $-10.76 + 2.46 \times 16 = 28.6$

Rural wage = $-4.88 + 1.80 \times 16 = 23.92$

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.

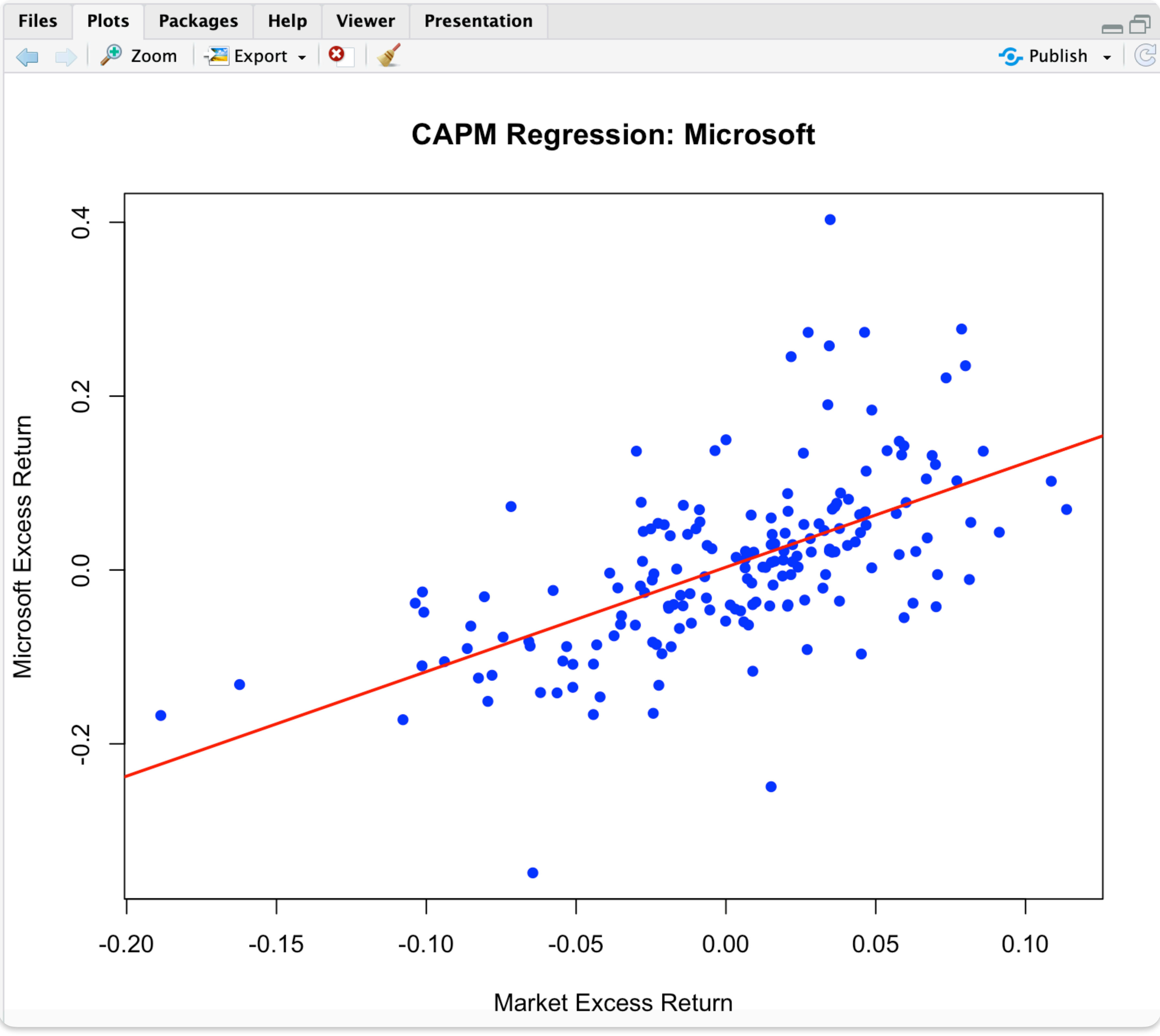
d. Estimate the model for each firm under the assumption that $\alpha_j = 0$. Do the estimates of the *beta* values change much?

a. Dependent variable: $r_j - r_f$, Independent variable: $r_m - r_f$
Regression coefficient: β_j , Intercept: α_j , Error: e_j
∴ "$r_j - r_f = \alpha_j + \beta_j(r_m - r_f) + e_j$" can be considered as a simple regression model ($y = b_1 + b_2 x + e$)

b. 看beta大小，most aggressive: Ford , most defensive: ExxonMobil

	Firm	Alpha	Beta	t(Alpha)	t(Beta)	p(Alpha)	p(Beta)	R-Squared
1	GE	-0.0009586682	1.1479521	-0.2166783	12.820635	0.8287072	4.469869e-27	0.4800926
2	IBM	0.0060525497	0.9768898	1.2519510	9.984657	0.2122303	6.373991e-19	0.3590052
3	Ford	0.0037789112	1.6620307	0.3695640	8.031573	0.7121467	1.271483e-13	0.2659980
4	Microsoft	0.0032496009	1.2018398	0.5383843	9.838921	0.5909844	1.631663e-18	0.3522665
5	Disney	0.0010469237	1.0115207	0.2238779	10.688323	0.8231091	6.500019e-21	0.3909121
6	ExxonMobil	0.0052835329	0.4565208	1.4944284	6.380428	0.1368343	1.480192e-09	0.1861364

c. 見上圖，t-value < 2 & p-value > 0.05，皆不顯著
Do not reject H_0 ($\alpha_j = 0$)



d. 比較原CAPM與新CAPM ($\alpha_j = 0$)，兩者beta相似

```
> # 比較新的 β 值與原本的 β 值
> comparison <- merge(results, results_no_intercept, by = "Firm")
> print(comparison[, c("Firm", "Beta", "Beta_no_Alpha")])
        Firm      Beta Beta_no_Alpha
1     Disney 1.0115207     1.0128190
2 ExxonMobil 0.4565208     0.4630727
3       Ford 1.6620307     1.6667168
4         GE 1.1479521     1.1467633
5        IBM 0.9768898     0.9843954
6  Microsoft 1.2018398     1.2058695
```