Chapter 2

1.

(a)

x	y	$x - \bar{x}$ 1	$(x-\bar{x})^2$	$y - \bar{y}$ 2	$(x-\bar{x})(y-\bar{y})$
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
$\sum x_i = 5$	$\sum y_i = 10$	$\sum(x_i - \bar{x}) = 0$	$\sum(x_i-\bar{x})^2 = 10$	$\sum(y_i - \bar{y}) = 0$	$\sum(x_i-\bar{x})(y_i-\bar{y}) = 8$

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

(b)

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

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

$\rightarrow$ The fitted regression line for these five points is $\hat{y}_i = 1.2 + 0.8x_i$.

Therefore, b_2 represents the slope of this line, which is 0.8 and b_1 represents the interpret of this line, which is 1.2.

(c)

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

$$\sum_{i=1}^{5} x_i y_i = 3\times4 + 2\times2 + 1\times3 + (-1)\times1 + 0\times0 = 18$$

$$\sum_{i=1}^{5} (x_i - \bar{x})^2 = 10 \text{ (from the above table)}$$

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

$$\rightarrow \sum_{i=1}^{5}(x_i-\bar{x})^2 = \sum_{i=1}^{5} x_i^2 - N\bar{x}^2 \ *$$

$$\sum_{i=1}^{5}(x_i-\bar{x})(y_i-\bar{y}) = 8 \text{ (from the above table)}$$

$$\sum_{i=1}^{5} x_i y_i - N\bar{x}\bar{y} = 18 - 5\times1\times2 = 8$$

$$\rightarrow \sum_{i=1}^{5}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{5} x_i y_i - N\bar{x}\bar{y}$$

(d)

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
$\sum x_i = 5$	$\sum y_i = 10$	$\sum \hat{y}_i = 10$	$\sum \hat{e}_i = 0$	$\sum \hat{e}_i^2 = 3.6$	$\sum x_i\hat{e}_i = 0$

- $\hat{y}_i = 1.2 + 0.8X_i$

$\rightarrow S_y^2 = \dfrac{\sum_{i=1}^{N}(y_i-\bar{y})^2}{N-1} = \dfrac{\sum_{i=1}^{5}(y_i-2)^2}{5-1} = \dfrac{2^2+0^2+1^2+(-1)^2+(-2)^2}{4} = 2.5$

$\rightarrow S_x^2 = \dfrac{\sum_{i=1}^{N}(X_i-\bar{X})^2}{N-1} = \dfrac{\sum_{i=1}^{5}(X_i-1)^2}{5-1} = \dfrac{2^2+1^2+0^2+(-2)^2+(-1)^2}{4} = 2.5$

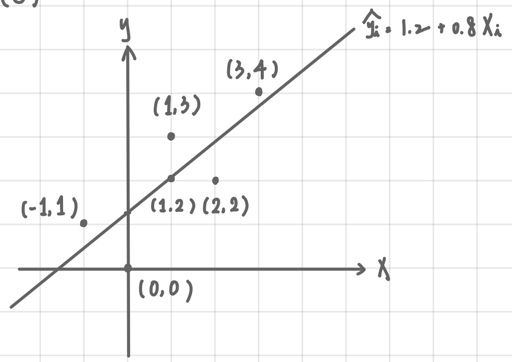
$\rightarrow S_{xy} = \dfrac{\sum_{i=1}^{N}(y_i-\bar{y})(X_i-\bar{X})}{N-1} = \dfrac{\sum_{i=1}^{5}(y_i-2)(X_i-1)}{5-1} = \dfrac{2\times2+0\times1+1\times0+(-1)\times(-2)+(-2)\times(-1)}{4} = 2$

$\rightarrow r_{xy} = \dfrac{S_{xy}}{S_x S_y} = \dfrac{2}{\sqrt{2.5\times2.5}} = 0.8$

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

$\rightarrow$ Median of $X = 1$, 50^{th} percentile of $X = 1$

(e)



(f)

It does pass the point $(\bar{X}, \bar{y}) = (1, 2)$.

(g)

From simple linear regression, the estimated regression equation is : $\hat{y}_i = b_1 + b_2 X_i$.

Taking the expectation on both sides : $E[\hat{y}_i] = E[b_1 + b_2 X_i] = b_1 + b_2 E[X_i]$

Since $E[\hat{y}_i] = \bar{y}$ and $E[X_i] = \bar{X}$, we get $\bar{y} = b_1 + b_2 \bar{X}$.

(h)

By the def., $\hat{y}$ is the mean of predicted values : $\bar{\hat{y}} = \frac{1}{N} \Sigma \hat{y}_i = \frac{1}{N} \Sigma (b_1 + b_2 X_i)$.

Since summation distributes over addition : $\bar{\hat{y}} = b_1 + b_2 \frac{1}{N} \Sigma X_i$.

By the def., $\frac{1}{N} \Sigma X_i = \bar{X}$, so $\bar{\hat{y}} = b_1 + b_2 \bar{X}$.

From part (g), we know $\bar{y} = b_1 + b_2 \bar{X}$.

Hence, $\bar{\hat{y}} = \bar{y}$.

(i)

By the def., $\hat{\sigma}^2 = \frac{1}{N-2} \Sigma e_i^2$, where $e_i = y_i - \hat{y}_i$.

Therefore, by part (d), $\hat{\sigma}^2 = \frac{1}{5-2} \cdot 3.6 = 1.2$.

(j)

By the def., $var(\hat{b}_2 | X) = \frac{\hat{\sigma}^2}{\Sigma (X_i - \bar{X})^2}$.

Thus, by part (a) and (i), $var(\hat{b}_2 | X) = \frac{1.2}{10} = 0.12$

By the def., $se(\hat{b}_2) = (var(\hat{b}_2 | X))^{\frac{1}{2}} = \sqrt{0.12} = 0.3464$

14.

(a)

Given that $\widehat{WAGE} = -4.88 + 1.8 EDUC$ and $\overline{WAGE} = 19.74$,

$$\varepsilon = \frac{d(WAGE)}{d(EDUC)} \times \frac{\overline{EDUC}}{\overline{WAGE}} = 1.8 \times \frac{(19.74 + 4.88) \div 1.8}{19.74} \approx 1.25$$

(b)

For the urban area,

$$\begin{cases} b_2 = 2.46 \\ se(b_2) = 0.16 \\ \overline{EDUC} = 13.68 \\ \overline{WAGE} = 19.74 \end{cases} \longrightarrow se(\varepsilon) = se(b_2) \times \frac{\overline{EDUC}}{\overline{WAGE}} = 0.16 \times \frac{13.68}{19.74} = 0.11$$

(c)

For $EDUC = 12$: $\begin{cases} \text{Urban} = -10.76 + 2.46 \times 12 = 18.76 \\ \text{Rural} = -4.88 + 1.8 \times 12 = 16.72 \end{cases}$

For $EDUC = 16$: $\begin{cases} \text{Urban} = -10.76 + 2.46 \times 16 = 28.6 \\ \text{Rural} = -4.88 + 1.8 \times 16 = 23.92 \end{cases}$

16.

(a)

The econometric model given is : $r_j - r_f = \alpha_j + \beta_j(r_m - r_f) + e_j$.

This is a simple linear regression equation of the form : $Y = \beta_0 + \beta_1 X + \varepsilon$

where $Y = r_j - r_f$, $X = r_m - r_f$, $\beta_1 = \beta_j$, $\beta_0 = \alpha_j$, $\varepsilon = e_j$.

This follows the basic structure of a simple linear regression where one independent variable (X) explains the dependent variable (Y) with some error. Since CAPM suggests that stock returns are linearly related to market returns, it naturally fits the regression framework.

(b)

The CAPM model is $r_j - r_f = \alpha_j + \beta_j(r_m - r_f) + e_j$.

	Firm	Alpha	Beta	R_Squared	P_Value	T_Stat_Beta
1	ge	-0.0009586682	1.1479521	0.4800926	4.469869e-27	12.820635
2	ibm	0.0060525497	0.9768898	0.3590052	6.373991e-19	9.984657
3	ford	0.0037789112	1.6620307	0.2659980	1.271483e-13	8.031573
4	msft	0.0032496009	1.2018398	0.3522665	1.631663e-18	9.838921
5	dis	0.0010469237	1.0115207	0.3909121	6.500019e-21	10.688323
6	xom	0.0052835329	0.4565208	0.1861364	1.480192e-09	6.380428

→ Most aggressive : Ford (Its beta is the largest.)

→ Most defensive : Exxon — Mobil (Its beta is the smallest.)

(c)

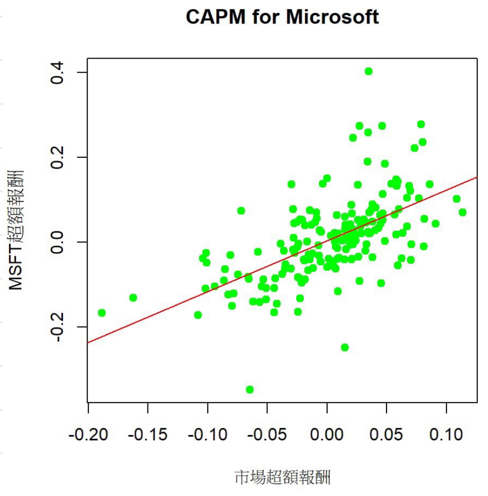
Finance theory predicts $\alpha_j = 0$,

meaning no systematic excess return beyond the CAPM prediction.

However, in our computation,

$\alpha_j \neq 0$ due to market inefficiencies in the real world or

that there are other variables primarily explain the model.



(d)

The result with $d_j = 0$:

	Firm	Beta_No_Intercept	R_Squared	P_Value_Beta
1	ge	1.1467633	0.4804485	3.021948e-27
2	ibm	0.9843954	0.3613794	3.639354e-19
3	ford	1.6667168	0.2676855	8.814439e-14
4	msft	1.2058695	0.3542837	9.878625e-19
5	dis	1.0128190	0.3923525	4.085622e-21
6	xom	0.4630727	0.1891534	9.479131e-10

The difference of beta is not much, showing that market excess return is the primary explanatory variable.

	Firm	Beta_With_Alpha	Beta_No_Intercept	Difference
1	ge	1.1479521	1.1467633	0.001188808
2	ibm	0.9768898	0.9843954	-0.007505539
3	ford	1.6620307	1.6667168	-0.004686085
4	msft	1.2018398	1.2058695	-0.004029708
5	dis	1.0115207	1.0128190	-0.001298251
6	xom	0.4565208	0.4630727	-0.006551910