CO2 Q01

(a)

| | $x$ | $y$ | $x-\bar{x}$ | $(x-\bar{x})^2$ | $y-\bar{y}$ | $(x-\bar{x})(y-\bar{y})$ |
|---|---|---|---|---|---|---|
| | 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 |
| $\Sigma$ | 5 | 10 | 0 | 10 | 0 | 8 |

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

**b.** Calculate $b_1$ and $b_2$ using (2.7) and (2.8) and state their interpretation.

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

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

$b_2 = 0.8$，當 $x$ 增加 1 unit，$y$ 增加 0.8 unit

$b_1 = 1.2$，當 $x=0$ 時，預測 $y$ 值為 1.2

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

$$\sum_{i=1}^{5} x_i^2 = 9+4+1+1+0 = 15$$

$$\sum_{i=1}^{5} x_i y_i = 12+4+3-1+0 = 18$$

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

$$\Sigma(x_i-\bar{x})(y_i-\bar{y}) = 8 = 18 - 5\cdot 1.2 = \Sigma x_i y_i - N\bar{x}\bar{y}$$

**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 | $\sum \hat{y}_i =$ | $\sum \hat{e}_i =$ | $\sum \hat{e}_i^2 =$ | $\sum x_i\hat{e}_i =$ |

$\hat{y}_i = b_1 + b_2 X = 1.2 + 0.8X$

$\Sigma \hat{y}_i = 3.6 + 2.8 + 2 + 0.4 + 1.2 = 10$

$\Sigma \hat{e}_i = 0.4 - 0.8 + 1 + 0.6 - 1.2 = 0$

$\Sigma \hat{e}_i^2 = 0.16 + 0.64 + 1 + 0.36 + 1.44 = 3.6$

$\Sigma x_i \hat{e}_i = 1.2 - 1.6 + 1 - 0.6 + 0 = 0$

$s_x^2 = 10/4 = 2.5 \quad s_{xy} = 8/4 = 2$

$s_y^2 = 10/4 = 2.5$

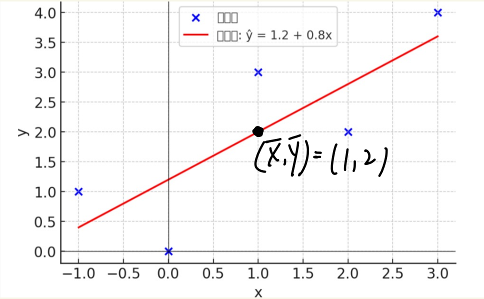
$r_{xy} = \sqrt{\frac{4}{(2.5)^2}} = 0.8$

$CV_x = 100(\sqrt{2.5}) \approx 100 \times 1.58 = 158\%$

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 $\bar{\hat{y}} = \sum \hat{y}_i / N$.

$\hat{y} = 1.2 + 0.8X$

$2 = 1.2 + 0.8 \cdot 1 \Rightarrow \bar{Y} = b_1 + b_2 \bar{X}$

$\bar{\hat{y}} = \sum \hat{y}_i / N = 10/5 = 2 = \bar{y}$

**i.** Compute $\hat{\sigma}^2$.

$\hat{\sigma}^2 = \frac{SSE}{n-2} = \frac{\sum \hat{e}_i^2}{n-2} = \frac{3.6}{3} = 1.2$

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

$$\widehat{var}(b_2|X) = \frac{\hat{\sigma}^2}{\Sigma(x_i-\bar{X})^2} = \frac{1.2}{10} = 0.12$$

$$se(b_2) = \sqrt{\widehat{var}(b_2|X)} = \sqrt{0.12} = 0.346$$

**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.

$\widehat{WAGE} = -4.88 + 1.8\, EDUC \qquad \overline{WAGE} = 19.74$

$\because (\overline{EDUC}, \overline{WAGE})$ 必在迴歸線上

$\therefore 19.74 = -4.88 + 1.8\,\overline{EDUC}, \quad \overline{EDUC} = 13.68$

$$elasticity = \beta_2 \frac{\overline{EDUC}}{\overline{WAGE}} = 1.8 \cdot \frac{13.68}{19.74} \approx 1.248$$

**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.

$$E = \beta_2 \cdot \frac{\overline{EDUC}}{\overline{WAGE}}$$

$$SE[E] = SE[\beta_2] \frac{\overline{EDUC}}{\overline{WAGE}}$$

$$\widehat{WAGE} = -10.76 + 2.46 \cdot (13.68) = 22.89$$

$$se(E) = 0.16 \cdot \frac{13.68}{22.89} \approx 0.0957$$

**c.** What is the predicted wage for an individual with 12 years of education in each area? With 16 years of education?

$E = 12$

Urban: $\widehat{WAGE} = -10.76 + 2.46 \cdot 12 = 18.76$

Rural: $\widehat{WAGE} = -4.88 + 1.8 \cdot 12 = 16.72$

$E = 16$

Urban: $\widehat{WAGE} = -10.76 + 2.46 \cdot 16 = 28.6$

Rural: $\widehat{WAGE} = -4.88 + 1.8 \cdot 16 = 23.92$

**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,

$$r_j - r_f = \beta_j \left(r_m - r_f\right)$$

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 $\beta_j$ 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

$$r_j - r_f = \alpha_j + \beta_j \left(r_m - r_f\right) + e_j$$

**a.** Explain why the econometric model above is a simple regression model like those discussed in this chapter.

應變數(Y)：$r_j - r_f$

自變數(X)：$r_m - r_f$

intercept($\beta_1$)：$\alpha_j$

迴歸係數($\beta_2$)：$\beta_j$

誤差項(e)：$e_j$

∵只有一個自變數($r_m - r_f$)，且$\beta_j$是一個線性係數 ∴這是一個SLR

**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?

計算每支股票的超額報酬：$r_j - r_f$

市場的超額報酬：$r_m - r_f$

之後根據最小平方法來估計模型

使用Python得到$\beta_j$的估計值

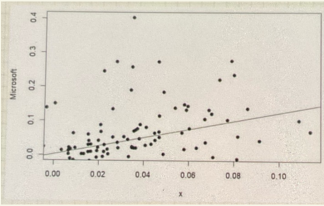
| 公司 | Beta 值 |
|---|---|
| GE | 1.148 |
| IBM | 0.977 |
| Ford | 1.662 |
| Microsoft | 1.202 |
| Disney | 1.012 |
| ExxonMobil | 0.457 |

由β值判斷 Ford最aggressive
ExxonMobil最defensive

**c.** Finance theory says that the intercept parameter $\alpha_j$ should be zero. Does this seem correct given your estimates? For the Microsoft stock, plot the fitted regression line along with the data scatter.

| 公司 | $\alpha_j$ (截距) |
|---|---|
| GE | -0.00096 |
| IBM | 0.00605 |
| Ford | 0.00378 |
| MSFT | 0.00325 |
| DIS | 0.00105 |
| XOM | 0.00528 |

根據資料,α值都非常接近0,所以CAPM理論正確




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

進行無截距迴歸模型分析,
得到結果與(b)小題幾乎一致

```r
# 計算個股與市場的超額報酬
df$ge_excess <- df$ge - df$riskfree
df$ibm_excess <- df$ibm - df$riskfree
df$ford_excess <- df$ford - df$riskfree
df$msft_excess <- df$msft - df$riskfree
df$dis_excess <- df$dis - df$riskfree
df$xom_excess <- df$xom - df$riskfree
df$mkt_excess <- df$mkt - df$riskfree
```

```r
# 執行 OLS 回歸：msft_excess ~ mkt_excess
msft_model <- lm(msft_excess ~ mkt_excess, data = df)

# 顯示回歸結果
summary(msft_model)
```

```r
# 繪製散點圖
plot(df$mkt_excess, df$msft_excess,
     main="CAPM Regression for Microsoft",
     xlab="Market Excess Return (r_m - r_f)",
     ylab="Microsoft Excess Return (r_msft - r_f)",
     col="blue", pch=16)

# 繪製回歸線
abline(msft_model, col="red", lwd=2)
```

```r
# 建立一個函數來計算 Beta 值
capm_regression <- function(stock) {
  model <- lm(df[[paste0(stock, "_excess")]] ~ df$mkt_excess)
  return(summary(model)$coefficients[2, 1])  # 提取 Beta 估計值
}

# 計算所有公司的 Beta 值
beta_values <- sapply(c("ge", "ibm", "ford", "msft", "dis", "xom"), capm_regression)

# 顯示 Beta 值
beta_values
```