HW1

1.

a.

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

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

b.

since $b_2=\dfrac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2}$ and $b_1=\bar{y}-b_2\bar{x}$

$\Rightarrow b_2=\frac{8}{10}=\frac{4}{5}$ and $b_1=2-\frac{4}{5}\times1=\frac{6}{5}$

c.

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

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

prove $\Sigma(x_i-\bar{x})^2=\Sigma x_i^2-N\bar{x}^2$:

$\Sigma x_i^2-N\bar{x}^2=15-5\times1^2$

$=10=\Sigma(x_i-\bar{x})^2$

prove $\Sigma(x_i-\bar{x})(y_i-\bar{y})=\Sigma x_iy_i-N\bar{x}\bar{y}$:

$\Sigma x_iy_i-N\bar{x}\bar{y}=18-5\times1\times2$

$=8=\Sigma(x_i-\bar{x})(y_i-\bar{y})$

d.

since $\hat{y}_i=b_1+b_2x_i$ and $\hat{e}_i=y_i-\hat{y}_i$

x_i	y_i	$\hat{y}_i$	$\hat{e}_i$	$\hat{e}_i^2$	$x_i\hat{e}_i$
3	4	$\frac{18}{5}$	$\frac{2}{5}$	$\frac{24}{25}$	$\frac{6}{5}$
2	2	$\frac{14}{5}$	$\frac{-4}{5}$	$\frac{16}{25}$	$\frac{-8}{5}$
1	3	2	1	1	1
–1	1	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{9}{25}$	$\frac{-3}{5}$
0	0	$\frac{6}{5}$	$\frac{-6}{5}$	$\frac{36}{25}$	0
$\sum x_i=$ 5	$\sum y_i=$ 10	$\sum\hat{y}_i=$ 10	$\sum\hat{e}_i=$ 0	$\sum\hat{e}_i^2=$ $\frac{18}{5}$	$\sum x_i\hat{e}_i=$ 0

$s_y^2=\sum_{i=1}^{N}\frac{(y_i-\bar{y})^2}{N-1}=\frac{10}{5-1}=\frac{5}{2}$

median of $x=1$

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

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

$r_{xy}=\frac{s_{xy}}{s_xs_y}=\frac{2}{\sqrt{\frac{5}{2}\times\frac{5}{2}}}=\frac{4}{5}$

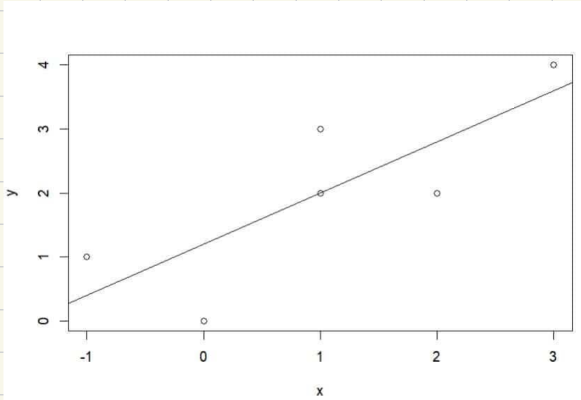
$CV_x=100\left(\frac{s_x}{\bar{x}}\right)=\frac{100\times\sqrt{\frac{5}{2}}}{1}=\frac{100\sqrt{5}}{\sqrt{2}}=158.113$

e.

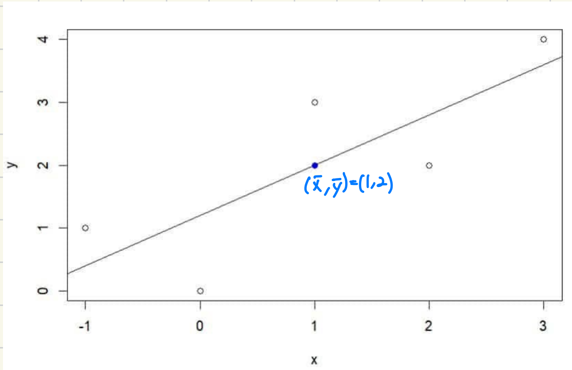
```
x<-c(3, 2, 1, -1, 0)
y<-c(4, 2, 3, 1, 0)
xp<-sum(x)/5
yp<-sum(y)/5
x<-c(x, xp)
y<-c(y, yp)

plot(x, y, ylim=c(min(y), max(y)), xlim=c(min(x), max(x)),
     xlab="x", ylab="y", type="p")
points(xp, yp, col="blue", pch=16)

xy=data.frame(x, y)
mod1<-lm(y~x, data=xy)
b1<-coef(mod1)[[1]]
b2<-coef(mod1)[[2]]
abline(b1, b2)
```



f.



g. $b_1+b_2\bar{x}=\frac{6}{5}+\frac{4}{5}\times1=2=\bar{y}$

h. $\hat{\bar{y}}=\Sigma\frac{\hat{y}_i}{N}=\frac{10}{5}=2=\bar{y}$

i. since $\hat{\sigma}^2=\frac{\Sigma e_i^2}{N-2}$

$\Rightarrow\hat{\sigma}^2=\frac{18/5}{3}=1.2$

j. since $\widehat{var}(b_2|x)=\frac{\hat{\sigma}^2}{\Sigma(x_i-\bar{x})^2}=\frac{1.2}{10}=\frac{3}{25}=0.12$

$se(b_2)=\sqrt{\widehat{var}(b_2|x)}=\frac{\sqrt{3}}{5}=0.34641$

2.

a.

since $\overline{WAGE}=19.74$

$\Rightarrow\overline{EDUC}=\frac{19.74+4.88}{1.8}=13.67$

Thus, the point of the means: (13.67, 19.74)

$\Rightarrow\varepsilon=1.8\times\frac{13.67}{19.74}=1.2465$

b.

$se(\varepsilon)=se(\beta_2\cdot\frac{\bar{x}}{\bar{y}})=se(\beta_2)\cdot\frac{\bar{x}}{\bar{y}}$

since $\overline{EDUC}=13.68$

$\Rightarrow\overline{WAGE}=13.68\times2.46-10.76$

$=22.8928$

Thus, $se(\varepsilon)=0.16\times\frac{13.68}{22.8928}=0.09561$

c.

$\boxed{\text{12 years}}$ In Urban: $12\times2.46-10.76=18.76$

In Rural: $12\times1.8-4.88=16.72$

$\boxed{\text{16 years}}$ In Urban: $16\times2.46-10.76=28.6$

In Rural: $16\times1.8-4.88=23.92$

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 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?

a. $r_j = r_f + \alpha_j + e_j + \beta_j(r_m - r_f)$, the variable $r_j - r_f$ only depends on one variable $r_m - r_f$, so the model is a simple regression model.

b. calculate $r_m - r_f$ and each of $r_j - r_f$, then by the simple regression model, find their parameters β_1, which is α_j in the formula, and β_2, which is β_j in the formula.

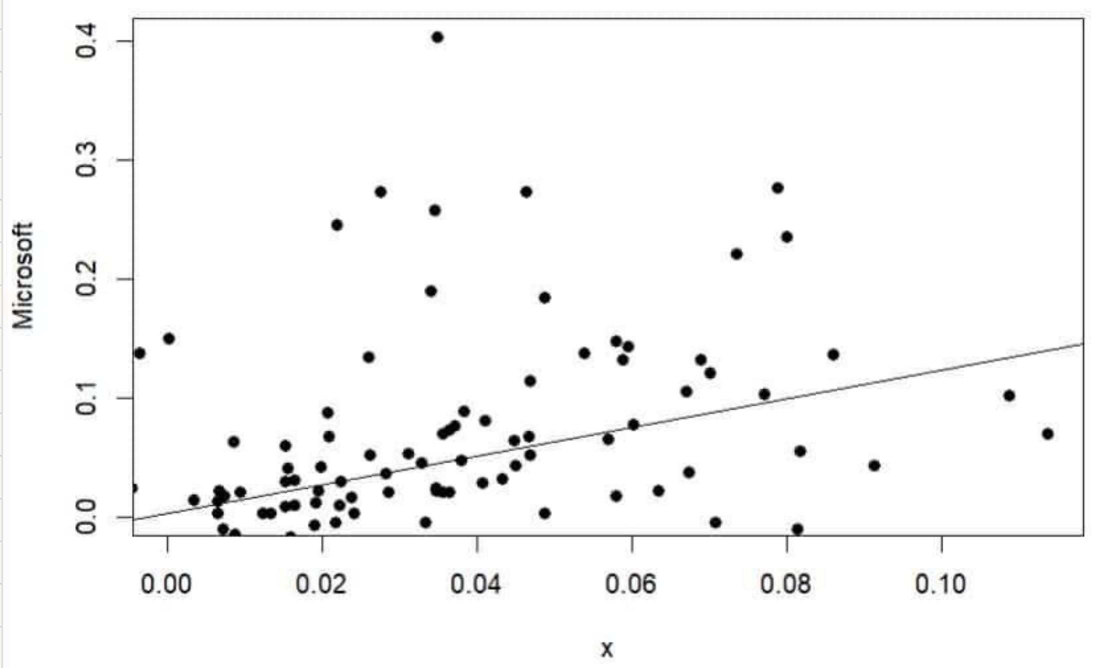
```
> theB2
[1] 1.1479521 0.9768898 1.6620307 1.2018398 1.0115207 0.4565208
```

according to the picture, ford appears most aggresive and Exxon appears most defensive.

c.

```
> theB1
[1] -0.0009586682  0.0060525497  0.0037789112  0.0032496009  0.0010469237  0.0052835329
```

According to the picture, each of their α is very close to zero, the Finance theory seems correct.



d. Assume each of $\alpha_j = 0$, and $\beta_j = \dfrac{cov(r_j - r_f, r_m - r_f)}{var(r_m - r_f)}$

```
> theSecondB
[1] 1.1479521 0.9768898 1.6620307 1.2018398 1.0115207 0.4565208
```

Compared with the β_j in (a) problem, the beta value does not change much.

rfD is the riskfree

```
ge<-c(capm5$ge)
ibm<-c(capm5$ibm)
ford<-c(capm5$ford)
msft<-c(capm5$msft)
dis<-c(capm5$dis)
xom<-c(capm5$xom)


x<-mkt-rfD
y1<-ge-rfD
y2<-ibm-rfD
y3<-ford-rfD
y4<-msft-rfD
y5<-dis-rfD
y6<-xom-rfD

v<-var(x)

xy1<-data.frame(x, y1)
mod1<-lm(y1~x, data=xy1)
b11<-coef(mod1)[[1]]
b12<-coef(mod1)[[2]]
plot(x, y1, ylim=c(min(y1), max(y1)), xlim=c(min(x), max(x)), xlab="x", ylab="ge", type="p")
cov1=cov(x, y1)
theSecondB1<-(cov1/v)

xy2<-data.frame(x, y2)
mod1<-lm(y2~x, data=xy2)
b21<-coef(mod1)[[1]]
b22<-coef(mod1)[[2]]
plot(x, y2, ylim=c(min(y2), max(y2)), xlim=c(min(x), max(x)), xlab="x", ylab="ge", type="p")
cov2=cov(x, y2)
theSecondB2<-(cov2/v)

xy3<-data.frame(x, y3)
mod1<-lm(y3~x, data=xy3)
b31<-coef(mod1)[[1]]
b32<-coef(mod1)[[2]]
plot(x, y3, ylim=c(min(y3), max(y3)), xlim=c(min(x), max(x)), xlab="x", ylab="ge", type="p")
cov3=cov(x, y3)
theSecondB3<-(cov3/v)

xy4<-data.frame(x, y4)
mod1<-lm(y4~x, data=xy4)
b41<-coef(mod1)[[1]]
b42<-coef(mod1)[[2]]
plot(x, y4, ylim=c(0, max(y4)), xlim=c(0, max(x))
     , xlab="x", ylab="Microsoft", pch=16)
abline(b41, b42)
cov4=cov(x, y4)
theSecondB4<-(cov4/v)

xy5<-data.frame(x, y5)
mod1<-lm(y5~x, data=xy5)
b51<-coef(mod1)[[1]]
b52<-coef(mod1)[[2]]
plot(x, y5, ylim=c(min(y5), max(y5)), xlim=c(min(x), max(x)), xlab="x", ylab="ge", type="p")
cov5=cov(x, y5)
theSecondB5<-(cov5/v)

xy6<-data.frame(x, y6)
mod1<-lm(y6~x, data=xy6)
b61<-coef(mod1)[[1]]
b62<-coef(mod1)[[2]]
plot(x, y6, ylim=c(min(y6), max(y6)), xlim=c(min(x), max(x)), xlab="x", ylab="ge", type="p")
cov6=cov(x, y6)
theSecondB6<-(cov6/v)

theB1<-c(b11, b21, b31, b41, b51, b61)
theB2<-c(b12, b22, b32, b42, b52, b62)
theSecondB<-c(theSecondB1, theSecondB2, theSecondB3, theSecondB4, theSecondB5, theSecondB6)
```