Chapter 5

Descriptive linear regression analysis

Exercise 22 *Answer* :

1. - $\dfrac{\partial^2 z}{\partial x^2} = \dfrac{\partial}{\partial x}\dfrac{\partial z}{\partial x} = \dfrac{\partial}{\partial x}(4x - 2y - 16) = 4.$
   - $\dfrac{\partial^2 z}{\partial y \partial x} = \dfrac{\partial}{\partial y}\dfrac{\partial z}{\partial x} = \dfrac{\partial}{\partial y}(4x - 2y - 16) = -2.$
   - $\dfrac{\partial^2 z}{\partial x \partial y} = \dfrac{\partial}{\partial x}\dfrac{\partial z}{\partial y} = \dfrac{\partial}{\partial x}(-2x + 10y + 2) = -2$ (lemme de Schwarz).
   - $\dfrac{\partial^2 z}{\partial y^2} = \dfrac{\partial}{\partial y}\dfrac{\partial z}{\partial y} = \dfrac{\partial}{\partial y}(-2x + 10y + 2) = 10.$

2. We get $H(x,y) = \begin{pmatrix} 4 & -2 \\ -2 & 10 \end{pmatrix}$ and $\det(H(x,y)) = 4 \times 10 - (-2 \times (-2)) = 36 > 0.$

3. Since H does not depend on x nor y, the Hessian matrix $H(x,y)$ equals to $\begin{pmatrix} 4 & -2 \\ -2 & 10 \end{pmatrix}$ at $(x,y) = (0,0)$ and $(x,y) = (-1,1)$.

   With `R` , we use the commands

```
> f <- function(x) 2*x[1]^2-2*x[1]*x[2]-16*x[1]+5*x[2]^2+2*x[2]+34
> x0=c(0,0)
> x1=c(-1,1)
> optim(x0,f,hessian=T)->solution0
> optim(x1,f,hessian=T)->solution1
> optimHess(solution0$par,f)
      [,1]  [,2]
 [1,]  4     -2
 [2,]  -2    10
> optimHess(solution1$par,f)
      [,1]  [,2]
 [1,]  4     -2
 [2,]  -2    10
```

Exercise 23 *Answer* :

1. (a) The revenue from selling Jordan shirts is equal to $R_1(x,y) = (40 - 50x + 40y)x$.
   (b) The revenue from selling O'Neal shirts is equal to $R_2(x,y) = (20 + 60x - 70y)y$.
   (c) The cost for shirts is equal to $C(x,y) = C_1(x,y) + C_2(x,y) = 2(60 + 10x - 30y)$ where $C_1(x,y) = 2(40 - 50x + 40y)$ and $C_2(x,y) = 2(20 + 60x - 70y)$.

(d) The overall profit equals to

$$P(x,y) = (40-50x+40y)(x-2)+(20+60x-70y)(y-2) = -50x^2-70y^2+100xy+20x+80y-120.$$

2. The necessary first order conditions are

$$\begin{cases} \dfrac{\partial P}{\partial x}(x,y)=0 \\ \dfrac{\partial P}{\partial y}(x,y)=0 \end{cases} \Leftrightarrow \begin{cases} -100x+100y+20=0 \\ 100x-140y+80=0 \end{cases} \Leftrightarrow \begin{cases} -40y+100=0 \\ 100x-140y-80=0 \end{cases} \Leftrightarrow \begin{cases} x=\frac{27}{10} \\ y=\frac{5}{2} \end{cases}.$$

The critical point of the profit function equals to $(x,y) = \left(\dfrac{27}{10}, \dfrac{5}{2}\right)$. We may use the command `optim` in `R` for minimizing a function depending on several variables. Since we have to maximize P, we apply `optim` to the opposite function $-P$:

```
> Profit<-function(x){50*x[1]^2+70*x[2]^2-100*x[1]*x[2]-20*x[1]-80*x[2]+120}
> solution<-optim(c(0,0),Profit)
> solution$par
[1] 2.700064 2.500025
```

3. It is straightforward from the previous question that Jordan T-shirts should be sold for \$2.70 and O'Neal shirts for \$2.50, provided this is a maximum

4. The Hessian matrix is

$$H = \begin{pmatrix} \dfrac{\partial P^2}{\partial x^2} & \dfrac{\partial P^2}{\partial x \partial y} \\ \dfrac{\partial P^2}{\partial y \partial x} & \dfrac{\partial P^2}{\partial y^2} \end{pmatrix} = \begin{pmatrix} -100 & 100 \\ 100 & -140 \end{pmatrix}$$

Indeed,

```
> solution$hessian
      [,1]  [,2]
 [1,]  100  -100
 [2,] -100  140
```

Its determinant equals to

$$\det(H) = (-100)(-140) - (100)^2 = 14000 - 10000 = 4000 > 0.$$

Since $\dfrac{\partial^2 P}{\partial x^2} = -100 < 0$ and $\det(H) > 0$, the solution found in 2. is a (relative) maximum.

Exercise 24

- We use the following datas :

	x	y	x^2	y^2	xy
	-2	0	4	0	0
	-1	0	1	0	0
	0	1	0	1	0
	1	1	1	1	1
	2	3	4	9	6
Totaux	0	5	10	11	7

Therefore, we obtain

- . $\overline{x} = \frac{1}{n}\sum x = 0,$
- . $s_X^2 = \frac{1}{n}\sum x^2 - \overline{x}^2 = \frac{10}{5} - 0^2 = 2$
- . $\overline{y} = \frac{1}{n}\sum y = \frac{5}{5} = 1$
- . $s_Y^2 = \frac{1}{n}\sum y^2 - \overline{y}^2 = \frac{11}{5} - 1^2 = \frac{6}{5}$
- . $s_{XY} = \frac{1}{n}\sum xy - \overline{x}.\overline{y} = \frac{7}{5} - 0 \times 2 = \frac{7}{5}$
- . $r = \frac{s_{XY}}{s_x s_Y} = \frac{\frac{7}{5}}{\sqrt{2} \times \sqrt{\frac{6}{5}}} = \frac{7}{5} \times \sqrt{\frac{5}{12}} \simeq 0,9037$

The formula (5.5) yields $b = \frac{\sqrt{\frac{6}{5}}}{\sqrt{2}} \times \frac{\frac{7}{5}}{\sqrt{2} \times \sqrt{\frac{6}{5}}} = \frac{7}{10}$ and $a = 1 - \frac{7}{10} \times 0 = 1.$

The formula (5.6) provides the resultant best-fit linear model $\hat{y} = 1 + \frac{7}{10}(x - 0) = \frac{7}{10}x + 1.$

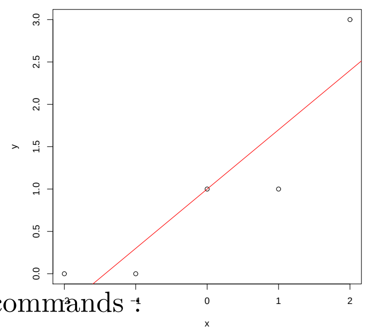
- The appropriate commands in R may be

```
> x<-c(-2,-1,0,1,2)
> y<-c(0,0,1,1,3)
> lm(y~x)

Call:
lm(formula = y ~ x)

Coefficients:
 (Intercept)     x
         1.0   0.7

> plot(x,y)
> abline(1.0,0.7,col='red')
```



Exercise 25 We use the following commands :

```
> xi<-c(10,12,9,27,47,112,36,241,59,167)
> yi<-c(9,14,7,29,45,109,40,238,60,170)
```

```
> lm(yi~xi)

Call:
lm(formula = yi   xi)

Coefficients:
 (Intercept)       xi
      0.7198   0.9914
```

1. The model fitted to the previous datas is $Y = 0.7198X + 0.9914$. The estimate for the expected change in audited value for a one-unit change in book value is 0.7198.
2. The best estimate is $a + b(100) \simeq (0.7198)(100) + 0.9914 \simeq 72.97$.

Exercise 26 We apply the `lm` function to a formula that describes the variable eruptions by the variable waiting, and save the linear regression model in a new variable eruption.lm.

```
> eruption.lm = lm(eruptions~waiting, data=faithful)
```

Then we extract the coefficient of determination from the `r.squared` attribute of its summary.

```
> summary(eruption.lm)$r.squared
[1] 0.81146
```

The coefficient of determination of the simple linear regression model for the data set faithful is 0.81146.