Ian Ma
s1743566
Group 12

Statistical Methodology
Hand-in 4

2019/11/12
Free Hong Kong
Revolution of Our Times

1

Given:

$$x : \text{time spent for a crypto-mining}$$
$$y : \text{no. of crypto-coins mined by a novel device}$$

Where:

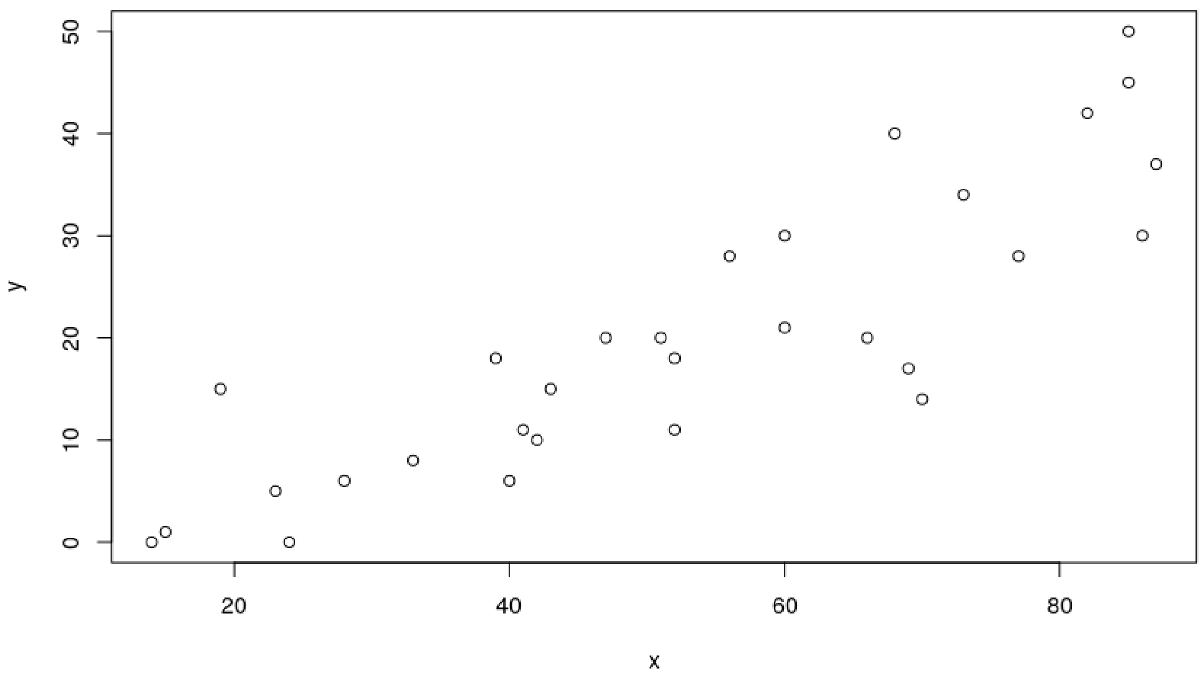
$$x : 73\ 82\ 14\ 56\ 43\ 47\ 85\ 86\ 24\ 33\ 19\ 66\ 68\ 51\ 40$$
$$60\ 70\ 39\ 28\ 69\ 87\ 52\ 52\ 15\ 85\ 41\ 23\ 77\ 42\ 60$$
$$y : 34\ 42\ 0\ 28\ 15\ 20\ 50\ 30\ 0\ 8\ 15\ 20\ 40\ 20\ 6$$
$$30\ 14\ 18\ 6\ 17\ 37\ 18\ 11\ 1\ 45\ 11\ 5\ 28\ 10\ 21$$

1.a

A diagram is plotted using R:
Code:

```
> x <- c(73,82,14,56,43,47,85,86,24,33,19,66,68,51,40,
60,70,39,28,69,87,52,52,15,85,41,23,77,42,60)
> y <- c(34,42,0,28,15,20,50,30,0,8,15,20,40,20,6,
30,14,18,6,17,37,18,11,1,45,11,5,28,10,21)
> plot(x,y)
```

Plot:



Trend: There is a positiv relation between the time and number of mined crypto-coins, therefore the more the time spent, the more the crypto-coins mined.

1.b

Assuming a simple linear regression is an appropriate model between x and y then:
Let: $y = \{y_1, y_2, ...\}$ where $y_i \sim Y_i \forall i \in \mathbb{N}$ then: $\mathbb{E}(Y_i|x_i) = \beta_0 + \beta_1 x_i$ Where:

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \dfrac{\sum_i (x_i - \bar{x}) y_i}{\sum_i (x_i - \bar{x}) x_i} = S_{xy}/S_{xx} \end{cases}$$

Where:

$$S_{xx} = \sum_i (x_i - \bar{x})^2$$

$$S_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y})$$

With Using R:

```
> xbar = mean(x)
> ybar = mean(y)
> Sxx <- sum((x - xbar)^2)
> Sxy <- sum((x-xbar)*(y-ybar))
> beta_1 = Sxy/Sxx
> beta_0 = ybar - beta_1*xbar
> round(c(beta_0,beta_1),5)
[1] -7.88777  0.52718
```

We now know $\hat{\beta}_0 = -7.88777 \quad \hat{\beta}_1 = 0.52718$ Therefore the least square estimate:

$$\mathbb{E}(Y_i|x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i = -7.88777 + 0.52718 x_i$$

1.c

As we know:

$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \dfrac{\sum_i (x_i - \bar{x}) y_i}{\sum_i (x_i - \bar{x}) x_i} = S_{xy}/S_{xx} \end{cases}$$

And Given: $\hat{\sigma}^2 = \frac{1}{n-2}(S_{yy} - S_{xy}^2/S_{xx}$
Using R:

```
> Syy <- sum((y - ybar)^2)
> n = 30
> sigmaSquare = (Syy - Sxy^2/Sxx)/(n-2)
> sigmaSquare
[1] 49.74041
```

We know that $\hat{\sigma^2} = 49.74041$
Estimated Standard Error of β_1:

$$\sigma^2_{\beta_0} = Var(\hat{\beta}_1|\mathbf{x}) = \frac{\sigma^2}{S_{xx}}$$

$$\sigma_{\beta_0} = \sqrt{\frac{\sigma^2}{S_{xx}}}$$

$$\mathbb{E}(\sigma_{\beta_0}) = \mathbb{E}\left(\sqrt{\frac{\sigma^2}{S_{xx}}}\right)$$

$$= 0.0579 \text{ (4 d.p.)}$$

Estimated Standard Error of β_0:

$$\sigma^2_{\beta_1} = Var(\hat{\beta}_0|\mathbf{x}) = \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\sigma^2$$

$$\sigma_{\beta_1} = \sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\sigma^2}$$

$$\begin{aligned}\mathbb{E}(\sigma_{\beta_0}) &= \mathbb{E}\left(\sqrt{\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)\sigma^2}\right)\\ &= 3.3247 \text{ (4 d.p.)}\end{aligned}$$

When fitting the model, we have assumed $\forall Y_i, i = 1, 2, ...$, all Y_i:

- are uncorrelated
- have commoned variance σ^2
- have expectation $\mathbb{E}(Y_i|x_i) = \beta_0 + \beta_1 x_i$

2

Prove, for $x_1, ..., x_n$ with mean $\bar{x}$ and $y_1, ..., y_n$ with mean $\bar{y}$:

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \equiv \sum_{i=1}^{n}(x_i - \bar{x})x_i \equiv \sum_{i=1}^{n} x_i^n + n^{-1}\left(\sum_{i=1}^{n} x_i\right)^2$$
$$S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \equiv \sum_{i=1}^{n}(x_i - \bar{x})y_i \equiv \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

Claim 1: $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \equiv \sum_{i=1}^{n}(x_i - \bar{x})x_i \equiv \sum_{i=1}^{n} x_i^n + n^{-1}\left(\sum_{i=1}^{n} x_i\right)^2$
Proof, direct proof:

$$\begin{aligned}
S_{xx} &= \sum_{i=1}^{n}(x_i - \bar{x})^2 \\
&= \sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x}) \\
&= \sum_{i=1}^{n}(x_i - \bar{x})x_i - \sum_{i=1}^{n}(x_i - \bar{x})\bar{x} \\
&= \sum_{i=1}^{n}(x_i - \bar{x})x_i - \bar{x}\sum_{i=1}^{n}(x_i - \bar{x}) \\
&= \sum_{i=1}^{n}(x_i - \bar{x})x_i - \bar{x} \times 0 \\
&= \sum_{i=1}^{n}(x_i - \bar{x})x_i
\end{aligned}$$

$$\therefore S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \equiv \sum_{i=1}^{n}(x_i - \bar{x})x_i$$

$$\begin{aligned}
S_{xx} &= \sum_{i=1}^{n}(x_i - \bar{x})x_i \\
&= \sum_{i=1}^{n}\left(x_i^2 - x_i\bar{x}\right) \\
&= \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i\bar{x} \\
&= \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i \\
&= \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2
\end{aligned}$$

$$\therefore S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \equiv \sum_{i=1}^{n}(x_i - \bar{x})x_i \equiv \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$$

$\therefore$ claim true

Claim 2: $S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y - \bar{y}) \equiv \sum_{i=1}^{n}(x_i - \bar{x})y_i \equiv \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$
Proof, direct proof:

$$\begin{aligned}
S_{xy} &= \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\
&= \sum_{i=1}^{n}(x_i - \bar{x})y_i - \sum_{i=1}^{n}(x_i - \bar{x})\bar{y} \\
&= \sum_{i=1}^{n}(x_i - \bar{x})y_i - \bar{y}\sum_{i=1}^{n}(x_i - \bar{x}) \\
&= \sum_{i=1}^{n}(x_i - \bar{x})y_i - \bar{y} \times 0 \\
&= \sum_{i=1}^{n}(x_i - \bar{x})y_i
\end{aligned}$$

$$\therefore S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \equiv \sum_{i=1}^{n}(x_i - \bar{x})y_i$$

$$\begin{aligned}
S_{xy} &= \sum_{i=1}^{n}(x_i - \bar{x})y_i \\
&= \sum_{i=1}^{n}(x_i y_i - \bar{x} y_i) \\
&= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \bar{x} y_i \\
&= \sum_{i=1}^{n} x_i y_i - n\bar{x}\frac{\sum_{i=1}^{n} y_i}{n} \\
&= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}
\end{aligned}$$

$$\therefore S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \equiv \sum_{i=1}^{n}(x_i - \bar{x})y_i \equiv \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

$\therefore$ claim true