Lab 3 STAT 151A

Data Table

X	Y	X_iY_i	X_i^2	$\hat{Y}_i$	$E_i = Y_i - \hat{Y}_i$	E_i^2
5	3.5	17.5	25	3.539048	-0.03904762	0.001524716
6	3.8	22.8	36	3.750476	0.04952381	0.002452607
3	3.1	9.3	9	3.116190	-0.01619048	0.0002621315
7	4	28	49	3.961905	0.03809524	0.001451247
4	3.2	12.8	16	3.327619	-0.12761905	0.016286621
2	3	6	4	2.904762	0.09523810	0.009070295

Step ①:

$$B = \frac{\sum_{i=1}^{6} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{6} x_i^2 - n\bar{x}^2}$$

where $x = [5, 6, 3, 7, 4, 2]$, $\bar{x} = \frac{5+6+\cdots+2}{6} = 4.5$

$y = [3.5, 3.8, 3.1, 4, 3.2, 3]$, $\bar{y} = \frac{3.5+3.8+\cdots+3}{6} = 3.433$

$$B = \frac{96.4 - 6 \cdot 4.5 \cdot 3.433}{139 - 6 \cdot (4.5)^2} = 0.2114286$$

$$\bar{Y} = A + B\bar{X} \Rightarrow A = \bar{Y} - B\bar{X} = 3.433 - 0.2114286 \cdot 4.5$$
$$= 2.481905$$

Step ②

$$\hat{Y}_i = A + BX_i \Rightarrow \hat{Y}_1 = 2.481905 + 5 \cdot 0.2114286 = 3.539048$$
$$\hat{Y}_2 = 2.481905 + 6 \cdot 0.2114286 = 3.750476$$
$$\hat{Y}_3 = 2.481905 + 3 \cdot 0.2114286 = 3.116190$$
$$\hat{Y}_4 = 2.481905 + 7 \cdot 0.2114286 = 3.961905$$
$$\hat{Y}_5 = 2.481905 + 4 \cdot 0.2114286 = 3.327619$$
$$\hat{Y}_6 = 2.481905 + 2 \cdot 0.2114286 = 2.904762$$

$$E_i = Y_i - \hat{Y}_i \Rightarrow E_1 = 3.5 - 3.539048 = -0.03904762$$
$$E_2 = 3.8 - 3.750476 = 0.04952381$$

Scanned by CamScanner

$E_3 = 3.1 - 3.116190 = -0.01619048$
$E_4 = 4 - 3.961905 = 0.03809524$
$E_5 = 3.2 - 3.327619 = -0.12761905$
$E_6 = 3 - 2.904762 = 0.09523810$

$E_i^2 = E_i \cdot E_i \Rightarrow E_1^2 = 0.0015247166 \quad E_2^2 = 0.0024326077$
$E_3^2 = 0.0002621315 \quad E_4^2 = 0.0014512472$
$E_5^2 = 0.0162866213 \quad E_6^2 = 0.0090702948$

$$SSR = \sum_{i=1}^{6} E_i^2 = E_1^2 + E_2^2 + E_3^2 + E_4^2 + E_5^2 + E_6^2$$
$$= 0.03104762$$

$$RMS = \frac{SSR}{n-2} = \frac{SSR}{4} = \frac{0.03104762}{4} = 0.007761905$$

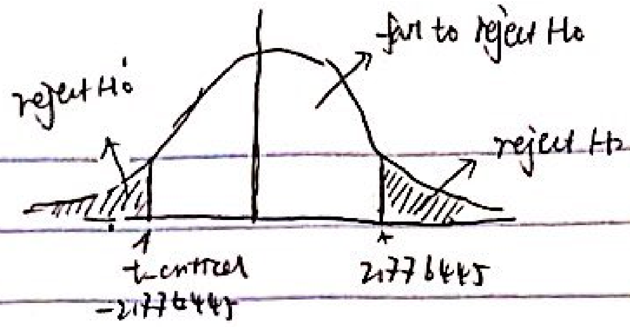
Step ③:

$$t = \frac{B}{\sqrt{\frac{RMS}{S_{xx}}}}, \quad S_{xx} = \sum_{i=1}^{6} x_i^2 - 6(\bar{x})^2$$
$$= \sum_{i=1}^{6} x_i^2 - 6 \cdot (4.5)^2 = 17.5$$

$$t = \frac{0.2114286}{\sqrt{\frac{0.007761905}{17.5}}} = 10.03919$$

t-test: Null Hypothesis (Ho): $\beta = 0$, there is no linear relationship between X and Y (# of hours of study in student life and his/her GPA)

Alternative Hypothesis (Ha): $\beta \neq 0$, there is a linear relationship between X and Y (# of hours of study and his/her GPA)



Degree of Freedom = $n-2 = 6-2 = 4$

$\alpha = 0.05$, $t\_critical_{\frac{0.05}{2}, df=4} = -2.776445$

$t\_critical_{0.975, df=4} = 2.776445$

Since 2 sided test, and $t > t\_critical_{0.975, df=4}$

$10.03919 > 2.776445$

We reject the Null Hypothesis that there is no linear relationship and conclude that there is a linear relationship between the # of hours of study in a student life and his/her GPA.

Scanned by CamScanner

STAT151A-Lab3

Xuanpei Ouyang

2/13/2017

```r
x = c(5,6,3,7,4,2)
y = c(3.5,3.8,3.1,4,3.2,3)
x_bar = mean(x)
x_bar
```

```
## [1] 4.5
```

```r
y_bar = mean(y)
y_bar
```

```
## [1] 3.433333
```

```r
X_iY_i = x * y
X_iY_i
```

```
## [1] 17.5 22.8  9.3 28.0 12.8  6.0
```

```r
X_sqr = x ^ 2
X_sqr
```

```
## [1] 25 36  9 49 16  4
```

```r
n = 6
sum(X_iY_i)
```

```
## [1] 96.4
```

```r
sum(x^2)
```

```
## [1] 139
```

```r
B = (sum(X_iY_i) - n*x_bar*y_bar) / (sum(x^2) - n*x_bar^2)
B
```

```
## [1] 0.2114286
```

```r
A = y_bar - B*x_bar
A
```

```
## [1] 2.481905
```

```r
y_hat <- A + B*x
y_hat
```

```
## [1] 3.539048 3.750476 3.116190 3.961905 3.327619 2.904762
```

```r
E <- y - y_hat
E
```

```
## [1] -0.03904762  0.04952381 -0.01619048  0.03809524 -0.12761905  0.09523810
```

```r
E_sqr = E^2
E_sqr
```

```
## [1] 0.0015247166 0.0024526077 0.0002621315 0.0014512472 0.0162866213
## [6] 0.0090702948
```

```r
SSR = sum(E_sqr)
SSR
```

```
## [1] 0.03104762
```

```r
RMS = SSR/4
RMS
```

```
## [1] 0.007761905
```

```r
S_xx = sum(x^2) - 6*x_bar^2
S_xx
```

```
## [1] 17.5
```

```r
t = B/(sqrt(RMS/S_xx))
t
```

```
## [1] 10.03919
```

```r
t_critical = qt(0.025, df = 4)
t_critical
```

```
## [1] -2.776445
```