Informatics Institute of Technology

B.Sc. (Hons) Artificial Intelligence and Data Science

Module: CM1606 Computational Mathematics

Module Coordinator: Ms. Ganesha Thondilege

Prof. Nimal Wickramasinghe

Coursework Report

Student Details: 20210537 2117526 Dineth Hasaranga

(Q1)

(1) $P(x) \geq 0$

$\sum P(x) = 1$

$\frac{1}{3} + \frac{1}{6} + P + 2P = 1$

$3P = \frac{1}{2}$

$P = \frac{1}{6}$

(2) $E(x) = \sum x \cdot P(x)$

$E(x) = (2 \times \frac{1}{3}) + (4 \times \frac{1}{6}) + (6 \times \frac{1}{6}) + (8 \times \frac{1}{3})$

$= \frac{2}{3} + \frac{2}{3} + 1 + \frac{8}{3}$

$= \frac{15}{3} = 5$

$Var(x) = E(x\text{^}2) - [E(x)]^2$

$E(x\text{^}2) \Rightarrow \sum x\text{^}2 \cdot P(x) \Rightarrow 4 \times \frac{1}{3} + 16 \times \frac{1}{6} + 36 \times \frac{1}{6} + 64 \times \frac{1}{3}$

$\Rightarrow \frac{4}{3} + \frac{8}{3} + 6 + \frac{64}{3}$

$\Rightarrow \frac{94}{3} = 31.3$

$[E(x)]^2 = 5^2 = 25$

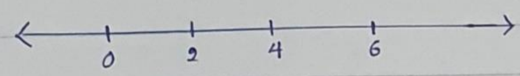
$Var(X) = 31.33 - 25$

$= 6.33$

(3) $Y = X - 2$

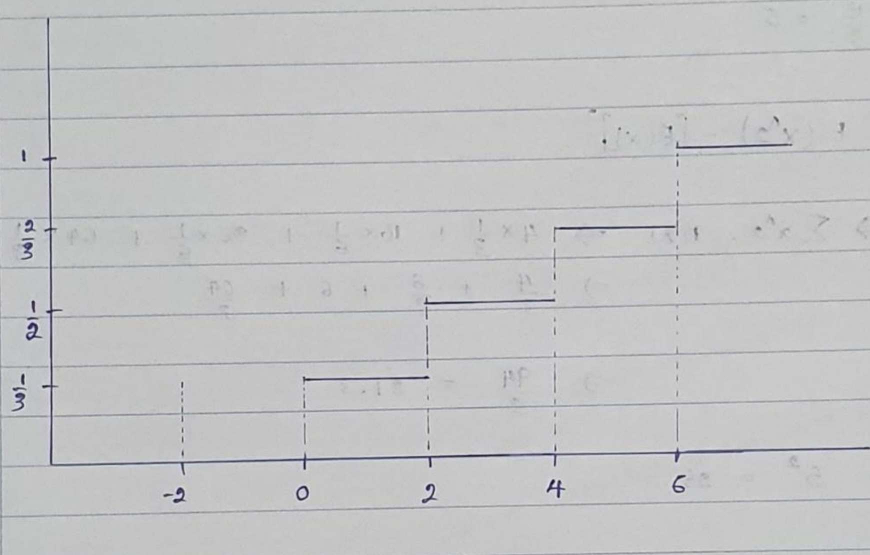
Y	0	2	4	6
$P(y)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
	0.33	0.166	0.166	0.33

(4)



Range(y)	$-\infty \leq y < 0$	$0 \leq y < 2$	$2 \leq y < 4$	$4 \leq y < 6$	$6 \leq y < +\infty$
F(y)	0	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	1

(5)



$$P(Y=4) = F(4+) - F(4-)$$
$$= \frac{2}{3} - \frac{1}{2}$$
$$= \frac{4-3}{6}$$
$$= \frac{1}{6}$$

(Q2)

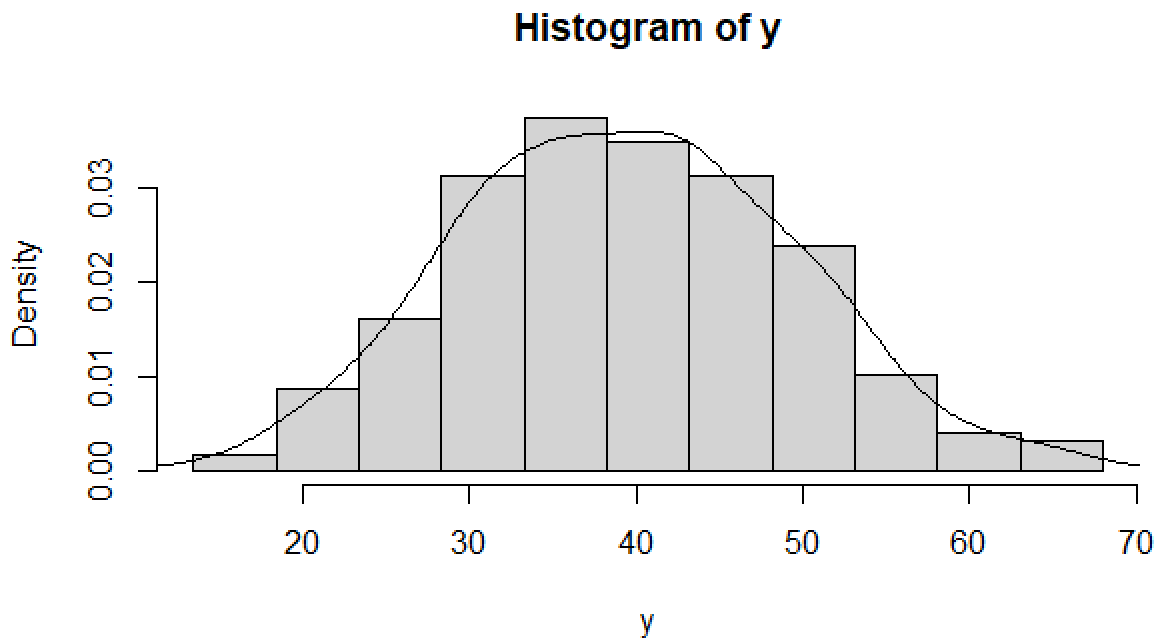
(q1)

```
y <- rnorm(400,mean=40,sd=10)
> y
  [1] 49.91523 44.80361 41.65612 46.78626 32.95662 30.83753 42.87218 26.19199
  [9] 48.13733 47.54115 45.73676 34.82109 35.29545 43.89861 35.06228 36.21921
 [17] 46.81194 32.25221 41.81426 26.89831 50.74884 30.65784 29.89844 29.17009
 [25] 38.22847 32.16749 46.13226 36.14053 22.75299 53.06842 54.86354 52.74251
 [33] 29.06027 28.70412 47.55704 49.75016 55.55318 19.78286 41.99641 40.62733
 [41] 27.97456 30.64086 53.04422 40.87563 42.15618 48.53994 47.53211 34.49275
 [49] 60.79381 40.77952 30.52326 33.62696 53.71616 51.63130 45.99981 45.75802
 [57] 47.69089 39.06473 23.93627 30.42712 49.54712 22.81402 30.11434 43.04450
 [65] 40.12524 37.81343 38.12783 39.95604 25.10668 43.43215 50.10041 19.26593
 [73] 36.67378 47.39760 44.13193 42.81101 34.07765 23.04493 35.19323 35.09411
 [81] 49.62107 26.32821 39.59088 51.48775 34.77150 42.65392 34.67518 32.03762
 [89] 38.45850 49.10560 31.63898 50.45910 54.18980 36.20109 27.48852 38.72379
 [97] 32.18739 49.42106 23.89721 31.13168 31.15359 40.19527 34.37440 53.94083
[105] 19.10842 33.62805 38.40101 33.93375 54.95925 37.95446 26.75786 40.22261
[113] 55.97031 35.38593 43.65937 42.45292 38.04685 44.45812 40.32561 26.80062
[121] 34.96401 51.78100 38.44529 39.28868 30.29623 29.30450 62.16273 42.54181
[129] 30.27531 65.09404 47.14243 25.04526 23.61277 30.30668 22.75649 30.70884
[137] 39.67676 45.96196 32.20882 48.29911 38.55732 31.58766 19.28771 45.75391
[145] 47.36452 43.46464 35.81201 42.96767 45.40160 49.75907 32.56551 40.91730
[153] 34.03293 26.20983 35.87608 41.06630 24.82921 32.76363 51.85455 33.06878
[161] 53.48301 42.31460 31.51301 48.40879 48.33006 35.66407 35.60584 31.92638
[169] 43.15121 44.40800 37.61400 41.56467 44.29070 43.99785 44.56865 58.53455
[177] 29.51391 26.40344 53.53391 31.96513 45.78191 24.55468 52.94571 20.69702
[185] 43.36925 64.84422 45.41369 35.26483 39.81617 41.13707 39.59445 37.20785
[193] 42.91069 42.16745 33.36140 48.76369 30.81039 27.36417 50.87825 24.34447
```

```
[201] 28.68608 35.40094 30.15856 38.67188 34.22787 37.37945 42.21004 39.32008
[209] 35.14320 51.90686 29.40840 32.35692 47.09854 54.22192 57.95322 54.21536
[217] 58.96598 55.85524 43.22311 36.37160 40.43539 26.54516 43.41345 30.37408
[225] 39.60324 25.70737 19.34796 20.20964 42.84219 38.76320 47.93261 48.43623
[233] 68.00435 42.02021 57.60832 27.48665 17.08240 38.09946 31.88318 48.10782
[241] 33.32940 50.29183 41.49530 29.93053 32.31349 51.00001 31.81888 58.68463
[249] 28.63040 40.41243 24.58124 39.42841 15.95774 48.50266 44.29831 29.70681
[257] 41.10969 48.74338 50.93699 53.35697 20.94005 26.58900 44.11488 57.66602
[265] 48.41098 44.03910 35.16475 43.27020 28.04708 37.14890 53.09787 42.62012
[273] 33.11927 40.07341 29.06515 45.73051 36.56490 39.75327 35.82372 29.89052
[281] 48.68019 30.98136 48.74827 41.94925 27.47436 34.48145 41.28178 44.59689
[289] 49.34044 44.02069 39.61900 45.00117 26.10496 25.07830 51.04855 33.48734
[297] 37.92994 53.08110 35.26896 36.77621 61.19483 51.78500 33.49528 43.42983
[305] 34.87746 36.48769 23.24508 58.12862 53.64806 35.41632 36.32915 23.54141
[313] 36.49063 32.99119 53.85957 44.93950 64.75086 41.45279 25.87177 29.08223
[321] 52.27515 35.49867 29.91569 22.20346 32.60837 53.95247 36.05669 28.22286
[329] 40.05009 39.74283 37.92924 28.43380 44.42682 38.91952 55.75599 30.45335
[337] 13.44876 33.96300 42.81002 52.95834 35.98150 28.47821 49.90559 42.13025
[345] 47.16990 50.35807 31.39632 45.26852 47.37571 26.66843 52.40869 44.21403
[353] 36.74364 30.35162 36.34014 30.49562 50.17731 40.12739 63.40199 49.49545
[361] 41.83014 36.93410 31.60533 41.13168 37.54010 47.10728 47.50149 46.03960
[369] 36.19113 43.74065 36.74542 46.99642 43.34532 50.95656 31.23203 39.43234
[377] 21.86242 38.19231 58.80332 43.60378 40.09958 32.93328 45.73336 37.82684
[385] 37.20074 33.63865 37.62912 43.14782 20.80450 43.35312 63.06485 30.35105
[393] 33.46528 33.95847 46.68913 18.84648 28.00102 41.31292 43.77266 36.98725
```

(q2)/(q3)

```
> breaks = seq(from=min(y),to=max(y),length=12)
> hist(y,breaks=breaks,freq = F)
> lines(density(y,))
```



(q4) When considering the shape of the density curve it shows approximately no skewness. Therefore, we can conclude that mean and median should have the same value.

(Q3)

(q1)

(1)

pmf of X $\Rightarrow P(X) = {}^{n}C_{x}\, p^{x} q^{(n-x)}$

$P(Y=0) \Rightarrow {}^{10}C_0\, p^0 q^{(10-0)}$

$\Rightarrow {}^{10}C_0\, (0.3)^0 (0.7)^{10}$

$\Rightarrow \dfrac{10!}{(0!)(10!)} (1) (0.7)^{10}$

$\Rightarrow (0.7)^{10} = \cancel{0.02824752}\ 0.02824752$

pmf satisfy $P(x) \geq 0$; and $\Sigma P(x) = 1$;

$P(Y=1) \Rightarrow {}^{10}C_1\, (0.3)^1 (0.7)^9$

$\Rightarrow \left(\dfrac{10!}{1!\ 9!}\right) (0.3)\ (0.7)^9$

$\Rightarrow 3 \times (0.7)^9 = 0.1210608$

$P(Y \leq 1) = P(Y=0) + P(Y=1) = \cancel{0.149336}\ 0.1493083$

(q2)

> pbinom(5,10,0.3)

[1] 0.952651

>1- pbinom(8,10,0.3)

[1] 0.00014368

(q3)

(3) Mean $\Rightarrow$ $E(Y) = (10) \times (0.3) \Rightarrow 3$

Variance $\Rightarrow$ $Var(Y) = npq$

$= (10)(0.3)(0.7)$

$= (3)(0.7)$

$= 2.1$

(q4)

```
> dbinom(x=0:3,10,0.3)
[1] 0.02824752 0.12106082 0.23347444
[4] 0.26682793
> sum(dbinom(x=0:3,10,0.3))
[1] 0.6496107
>
```

(Q4)

(q1)

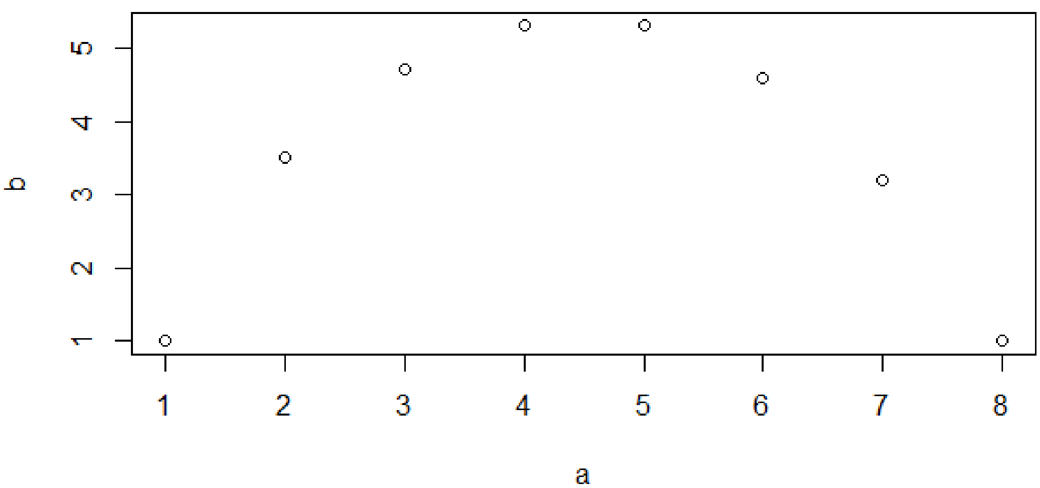
a<-c(1,2,3,4,5,6,7,8)

b<-c(1,3.5,4.7,5.3,5.3,4.6,3.2,1)

plot(a,b)

cor(a,b,method="pearson")

-0.02982897



(q2)

The correlation coefficient is a measure of linear relationship and thus a value of r = 0 or near to zero implies that there is no linear correlation however there is a quadratic relationship.

(q3)

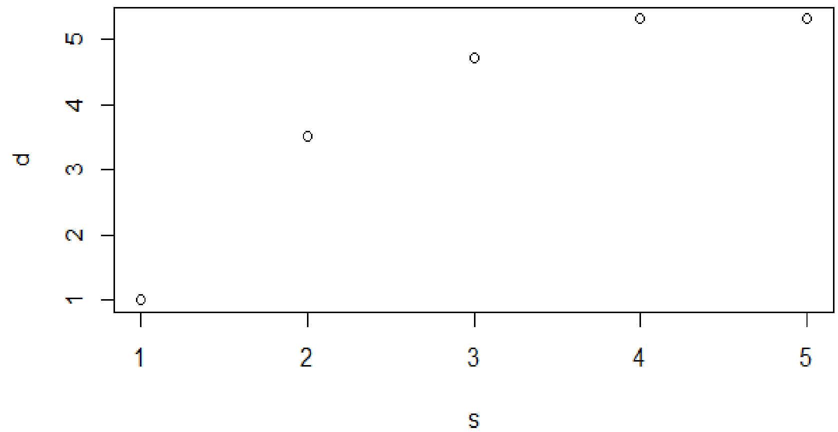
When considering the first five values, when X values increases the Y values also increases depicting a positive correlation. This indicates a strong correlation.

s<-c(1,2,3,4,5)

d<-c(1,3.5,4.7,5.3,5.3)

cor(s,d,method = "pearson")

0.9082363



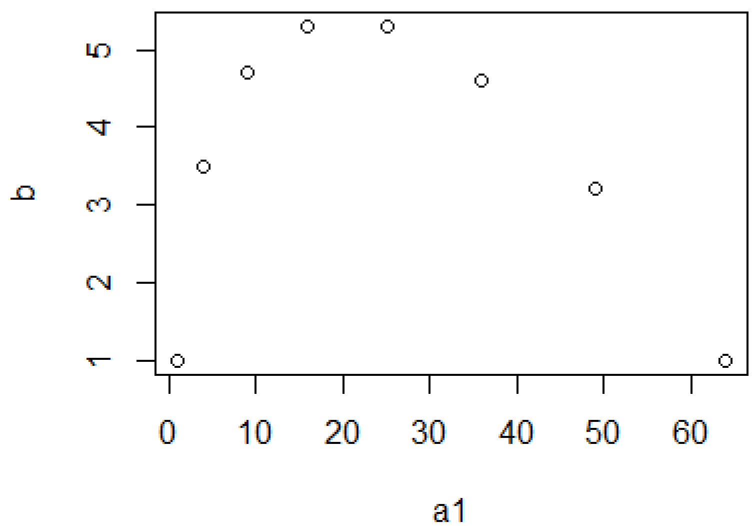
(q4)

>a1<-a^2

>a2<-2a+5

$r_{x,y} > r_{x1,y}$

$r_{x,y} = r_{x2,y}$



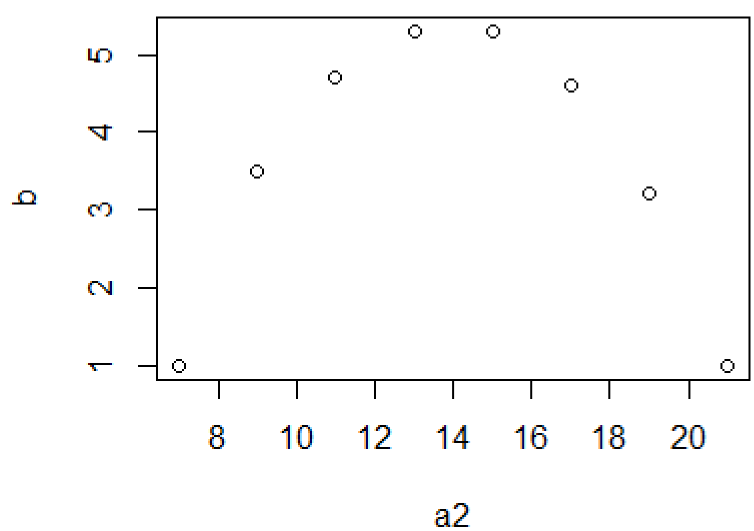
(q5) Verification:

cor(a1,b,method = "pearson")
[1] -0.2455314
> cor(a2,b,method = "pearson")
[1] -0.02982897

> plot(a1,b)

> plot(a2,b)

Graphs of ***b vs a1*** and ***b vs a2*** are also done for verification.

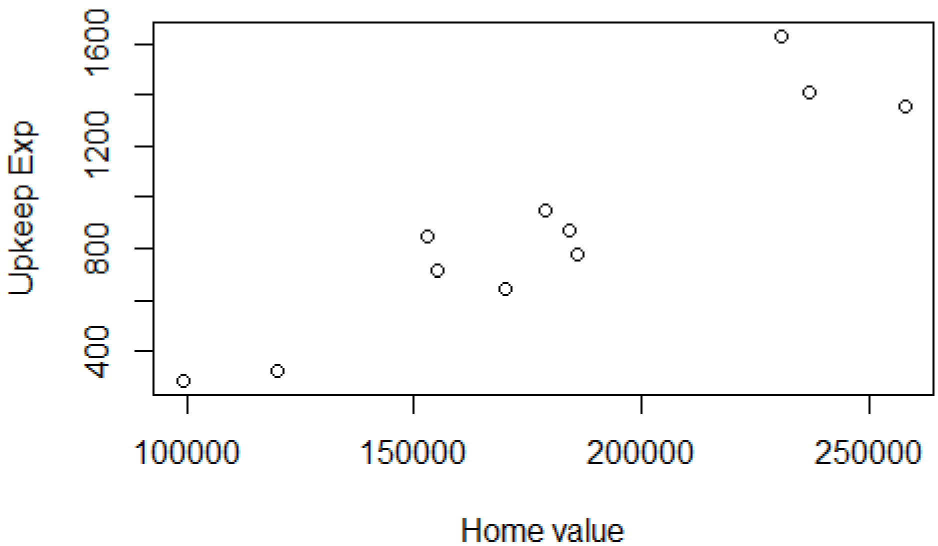


(Q5)

(q1)

```
x<-c(237000,184000,99000,258000,231000,186000,155000,153000,170000,120000,179000)
> y<-c(1412,872,288,1351,1627,775,711,849,642,324,950)
> data.frame<-c(x,y)
> plot(x,y,xlab="Home value",ylab="Upkeep Exp")
```

The both variables tend to increase together, therefore the association is positive.



(q2)

<u>Code</u>

```
fit<-lm(y~x)
> summary(fit)
```

<u>Output</u>

Call:

lm(formula = y ~ x)

Residuals:

Min 1Q Median 3Q Max

-184.83 -126.31   18.81   57.86 312.32

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -5.774e+02 1.970e+02 -2.931   0.0167 *

x   8.191e-03 1.064e-03   7.700   3e-05 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

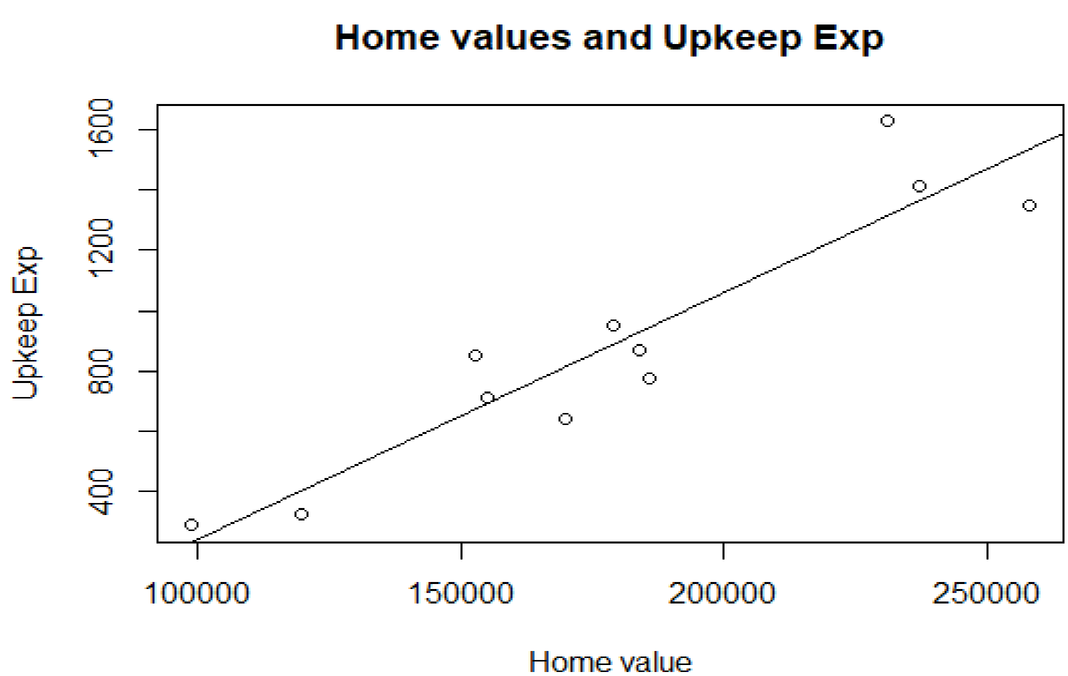
Residual standard error: 163.4 on 9 degrees of freedom

Multiple R-squared: 0.8682,   Adjusted R-squared: 0.8536

F-statistic: 59.29 on 1 and 9 DF, p-value: 3e-05

Code

```
plot(x,y,main = "Home values and Upkeep Exp",
+      abline(lm(y~x)),xlab = "Home value",ylab = "Upkeep Exp")
```



(q3)

8.191*10^-3   (*Using the graph*)

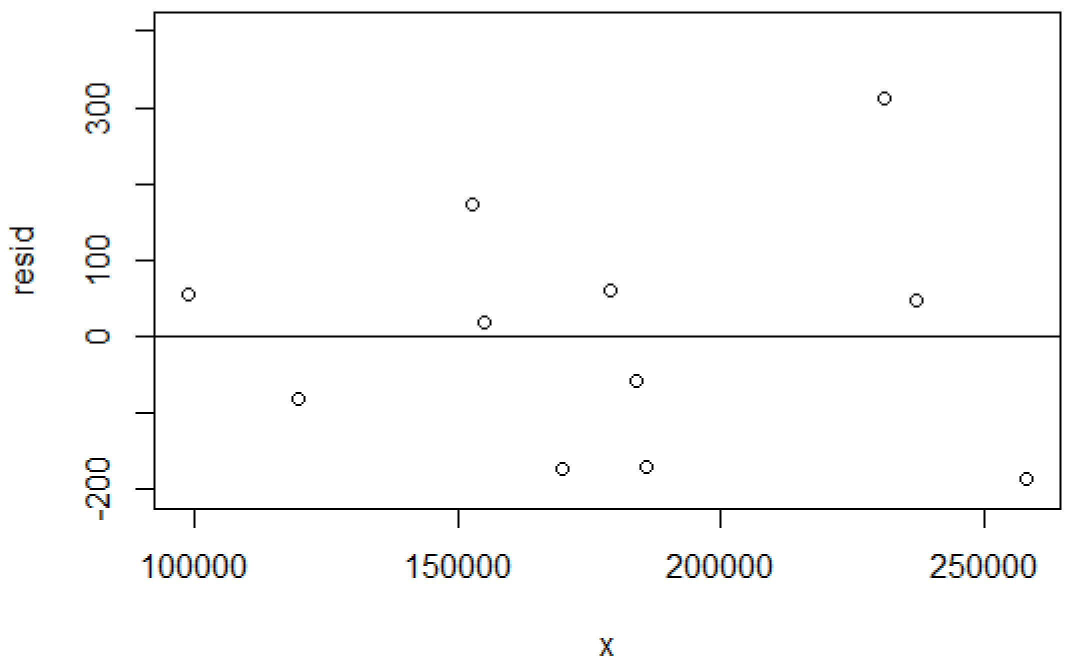
(q4)

fit<-lm(y~x)

> resid<-residuals(fit)

> plot(x,resid,ylim = c(-200,400))

> abline(0,0)



The fitted line is adequate, since the points on the plot are scattered randomly around the line y=0. The point between x = 200000 and x = 250000 with y(resid) = 300 can be considered as an outlier because of the large residual. So, we can conclude that the relationship between x and y is approximately linear.

(q5)

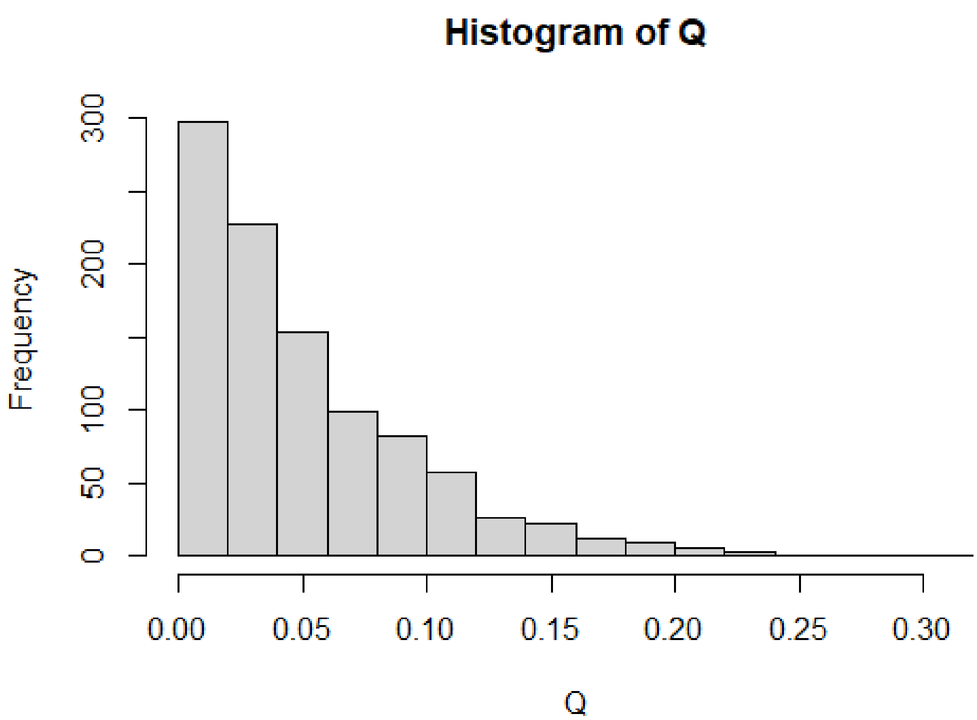
new<-data.frame(x=c(140000,225000))

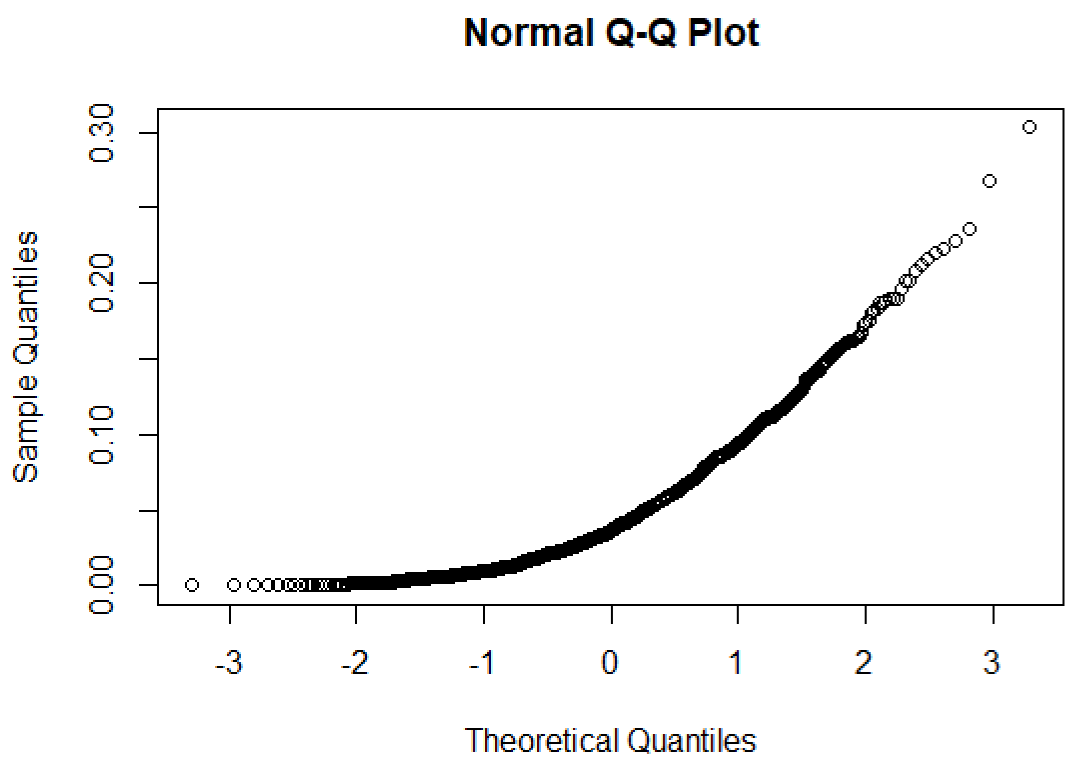
> predict(fit,new)

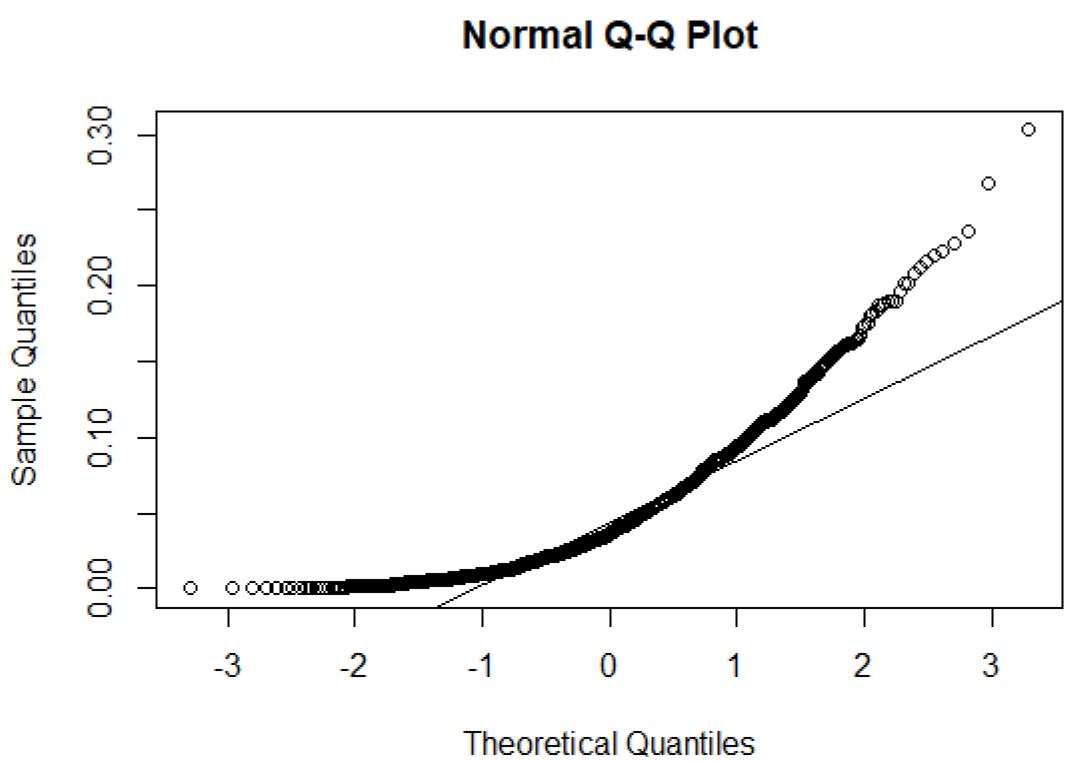
```
        1         2 
 569.3293 1265.5379 
```

(Q6)

```
Q<-numeric(1000)
> for(i in 1:1000){x<-runif(20)
+ Q[i]<-min(x)}
> hist(Q)
> qqnorm(Q)
> qqline(Q)
```




Normal Q-Q Plot
Sample Quantiles
0.00
0.10
0.20
0.30
-3
-2
-1
0
1
2
3
Theoretical Quantiles


Normal Q-Q Plot
Sample Quantiles
0.00
0.10
0.20
0.30
-3
-2
-1
0
1
2
3
Theoretical Quantiles