Assignment_01

Q1.

Q1.

(a) $X \begin{cases} A-C & \begin{cases} A-B \\ B-C \end{cases} \\ \text{Out of runway} \end{cases}$ So the distribution of X is discrete distribution. Binomial Distribution.

$P(A-C) = 0.9$ $P(\text{out of runway}) = 0.1$

$N=10$, $X=6$, $P=0.9$

$\therefore P(X=6) = 0.01116$

and $P(X \geq 6) = P(X=6) + P(X=7) + P(X=8) + P(X=9) + P(X=10)$

$= 0.99836$

(b) $P(X=AB) = P(X=AC) \times P(X=AB \mid \cancel{X}=AC) = 0.9 \times 0.4 = 0.36$

$\therefore P=0.34$, $N=10$, $X=6$

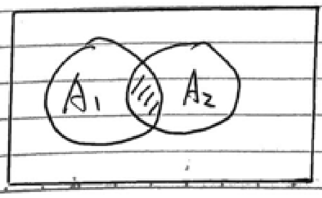
$P(X=6) = 0.06156$ for exactly 6 airplanes land on the AB section of the runway.

$1-P(X=6) = 0.93844$. for exactly 4 airplanes land at the out section of the runway.

and we can get $P(X \geq 6) = 0.0836$ and $1-P(X \geq 6) = 0.9164$.

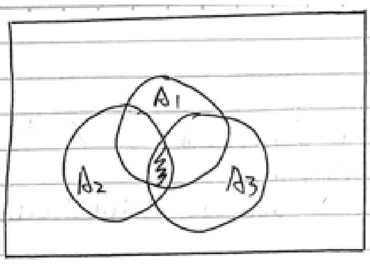
(c) 

$A \cap A^c = \emptyset$



$(A_1 \cap A_2) \in A_2$

$\therefore (A_1 \cap A_2) \cap A_2^c = \emptyset$



$A_1 \cap A_2 \cap A_3$

$= (A_1 \cap A_2) \cap A_3$

$= A_1 \cap (A_2 \cap A_3)$

$\therefore (A_1 \cap A_2 \cap A_3^c) \cap (A_1 \cap A_2^c \cap A_3) = \emptyset$

$(A_1 \cap A_1) \cap (A_2 \cap A_2^c) \cap (A_3 \cap A_3^c)$

$A_1 \cap \emptyset \cap \emptyset = \emptyset$

Probability Axioms

$(A_1 \cap A_2 \cap A_3^c)$ $\qquad$ $(A_1 \cap A_2^c \cap A_3)$

$= P(A_2) \cdot P(A_1 | A_2) \cdot P(A_3^c | A_1, A_2)$ $\quad = P(A_2^c) \cdot P(A_1 | A_2^c) \cdot P(A_3 | A_1, A_2^c)$

$(A_1 \cap A_2 \cap A_3^c) \in A_2$ $\qquad$ $(A_1 \cap A_2^c \cap A_3) \in A_2^c$

$A_2 \cap A_2^c = \emptyset$

$\therefore (A_1 \cap A_2 \cap A_3^c) \cap (A_1 \cap A_2^c \cap A_3) = \emptyset$

(d) $P = 1/6 = 0.1667$

$N = 5 \qquad X = 1$

$P(X=1) = 0.40188$

$P(X \geq 1) = 0.59812$

(e) $P = 5/6 = 0.8333$

$N = 6 - 1 = 5 \qquad X = 5$

$P(X=5) = 0.40188$

$0.40188 \times (0.9)^6 = 0.2136$

Q2.

(Q 2)

$$P(A|B) = \frac{P(A)\cdot P(B|A)}{P(B)}$$

$$P(X_i|Y=1) = \frac{P(X_i)\,P(Y=1|X_i)}{P(Y=1)} \quad \text{for } i = \cancel{1,2,3}\ 0,1,2$$

$$P(Y=1) = P(X=0,Y=1) + P(X=1,Y=1) + P(X=2,Y=1)$$

$$= 0.0679 + 0.0095 + 0.018$$

$$= 0.0954$$

$$P(X_0|Y=1) = \frac{P(X_0)\,P(Y=1|X=0)}{P(Y=1)} = \frac{P(X=0,Y=1)}{P(Y=1)} = 0.72$$

$$P(X_1|Y=1) = \frac{P(X=1,Y=1)}{P(Y=1)} \approx 0.0996$$

$$P(X_2|Y=1) = \frac{P(X=2,Y=1)}{P(Y=1)} \approx 0.1887$$

∴ When X-ray is positive:

- the probability of no cancer and disease is 0.72
- the probability of lung cancer is around 0.0996
- the probability of disease is around 0.1887

Q3.

Q3

(a) Sample space = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}

there are 8 possible outcomes.

$X = (X_1, X_2)$

$X = \{(0,3), (1,2), (2,1), (0,3)\}$

(b) and (c)

$X_1 \backslash X_2$	0	1	2	3	total P
0	0	0	0	1/8	1/8
1	0	0	3/8	0	3/8
2	0	3/8	0	0	3/8
3	1/8	0	0	0	1/8
total P	1/8	3/8	3/8	1/8	8/8

$$X \begin{cases} 1/8 & \text{if } (3,0) \\ 3/8 & \text{if } (2,1) \\ 3/8 & \text{if } (1,2) \\ 1/8 & \text{if } (0,3) \\ 0 & \text{if otherwise} \end{cases}$$

Deduce:

① the sum of X_1 and X_2 must be 3

② X_1 and X_2 are independent

PMF of X_1:

X_1	0	1	2	3
P	1/8	3/8	3/8	1/8

PMF of X_2:

X_2	0	1	2	3
P	1/8	3/8	3/8	1/8

(d) $E(X_1) = \sum X_1 P(X_1) = 1.5$

$E(X_2) = \sum X_2 P(X_2) = 1.5$

$E(X_1^2) = \sum X_1^2 P(X_1) = 3$

$E(X_2^2) = \sum X_2^2 P(X_2) = 3$

$V(X_1) = E(X_1^2) - [E(X_1)]^2 = 0.75$

$V(X_2) = E(X_2^2) - [E(X_2)]^2 = 0.75$

$E(X_1 X_2) = \sum_{X_1}\sum_{X_2} X_1 X_2 P(X_1, X_2) = 1.5$

$E(X_1 X_2) - E(X_1)E(X_2) = 1.5 - (1.5) \times (1.5) = -0.75$

$\frac{-0.75}{\sqrt{0.75} \times \sqrt{0.75}} = -1$

$\therefore$ ① X_1 and X_2 are dependent

and ② X_1 and X_2 are perfectly negatively Correlated.

Q4.

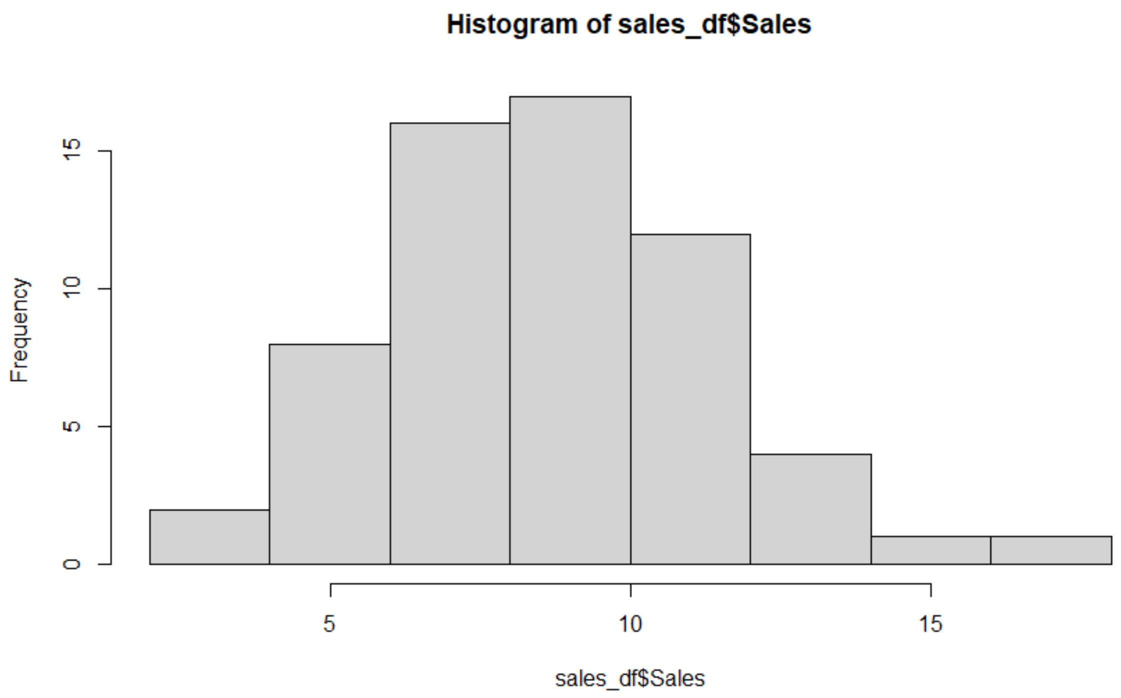
ready csv file：

```
library(readr)
sales_df <- read.csv("sales.csv")
```

1. **Identify the most appropriate distribution that models the variable, Sales**

```
hist(sales_df$Sales)
```

Resutl:



From the picture, it looks like a normal distribution.

2. **(2 marks) Use R to compute the mean of the Price ($) in the dataset over the time period**

```
mean(sales_df$Price....)
```

Resule:145.8862

3. **(3 marks) Use R to calculate the revenue and attach the daily revenue to the right of the sales table in Sales.csv. Please provide a screenshot of the daily revenue in your solution.**

```
sales_df$revenue <- sales_df$Price....*sales_df$Sales
```

Resutl:

	Day	Sales	Price....
1	1	11	127.48
2	2	9	145.39
3	3	11	127.48
4	4	9	145.39
5	5	8	154.91
6	6	7	164.46
7	7	7	164.46
8	8	7	164.46
9	9	7	164.46
10	10	9	145.39
11	11	12	119.50
12	12	11	127.48
13	13	11	127.48
14	14	8	154.91
15	15	10	136.17
16	16	9	145.39
17	17	9	145.39
18	18	15	100.66
19	19	10	136.17
20	20	5	182.34
21	21	4	189.98
22	22	12	119.50
23	23	12	119.50
24	24	8	154.91
25	25	10	136.17
26	26	10	136.17
27	27	14	106.06
28	28	9	145.39

29	29	10	136.17
30	30	11	127.48
31	31	5	182.34
32	32	14	106.06
33	33	12	119.50
34	34	6	173.72
35	35	9	145.39
36	36	7	164.46
37	37	13	112.34
38	38	17	92.38
39	39	11	127.48
40	40	5	182.34
41	41	9	145.39
42	42	8	154.91
43	43	8	154.91
44	44	10	136.17
45	45	7	164.46
46	46	6	173.72
47	47	7	164.46
48	48	7	164.46
49	49	6	173.72
50	50	2	201.06
51	51	7	164.46
52	52	11	127.48
53	53	10	136.17
54	54	10	136.17
55	55	9	145.39
56	56	5	182.34
57	57	8	154.91
58	58	6	173.72
59	59	11	127.48
60	60	14	106.06
61	61	8	154.91

4. **(2 marks) Use R to calculate the mean and variance of the revenue ($) over the 61 days**

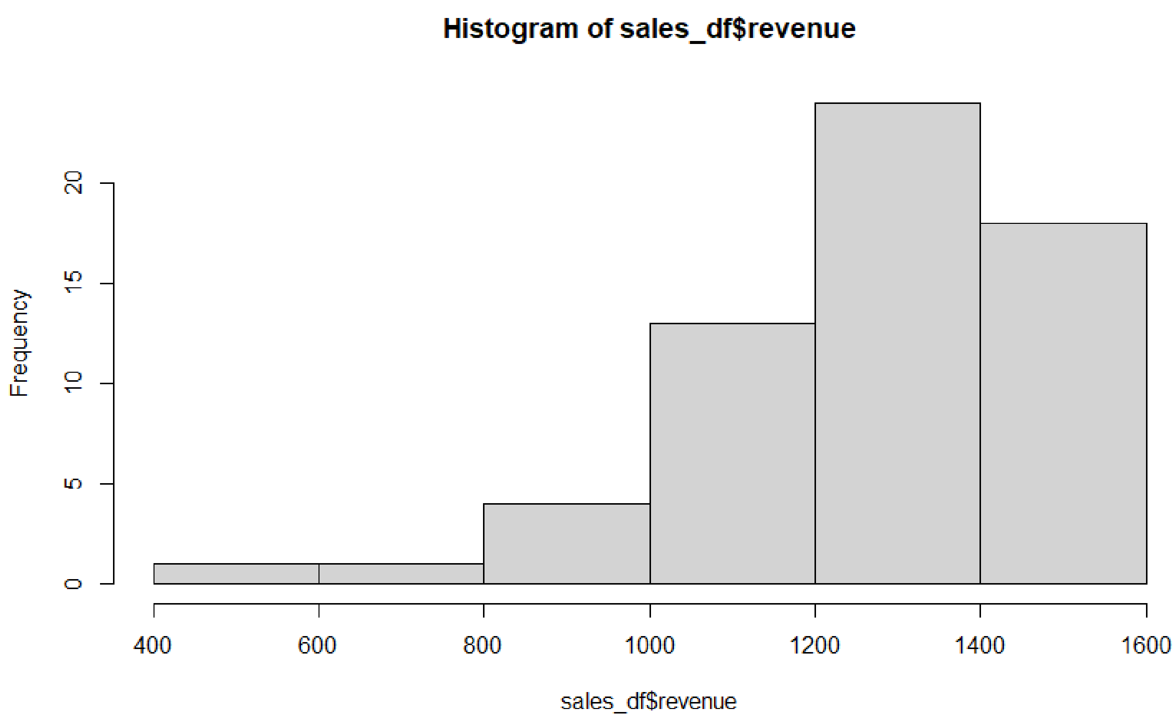
```
mean(sales_df$revenue)
var(sales_df$revenue)
```

Result:

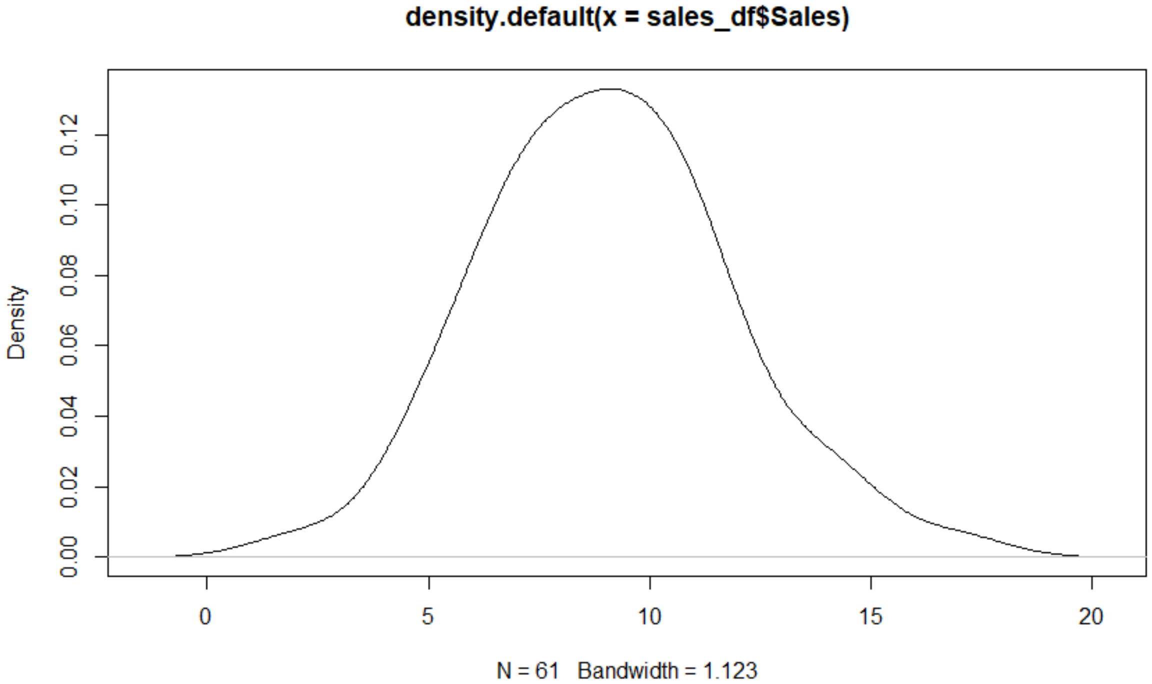
```
> mean(sales_df$revenue)
[1] 1256.292

> var(sales_df$revenue)
[1] 41785.5
```

5. **(3 marks) Use R to compute the probability density function for the number of Sales per day and use R to plot a histogram of the number of Sales and the revenue. Assume that the maximum sold per day is 20 units.**

```
hist(sales_df$revenue)

table(sales_df$Sales)/nrow(sales_df)

plot(density(sales_df$Sales))
```



```
> table(sales_df$Sales)/nrow(sales_df)

2          4          5          6          7          8          9          10
0.01639344 0.01639344 0.06557377 0.06557377 0.14754098 0.11475410 0.14754098 0.13114754

11         12         13
0.13114754 0.06557377 0.01639344

14         15         17
0.04918033 0.01639344 0.01639344
```



6. **(4 marks) The company chooses to fix the Price ($) per day to be the expected (mean) of Price ($) over 61 days. The company states that the lowest revenue amount is $1021 before there is a loss of money. What is the probability that the revenue is less than $1021?**

```
> sum(sales_df$Sales*mean(sales_df$Price....) <1021)/nrow(sales_df)
[1] 0.1639344
```

7. **(2 marks) Summarise the results in parts a) – f) and give a conclusion in relation to sales and revenue when the company fixes the price.**

The company has a very high probability of being profitable. The highest daily sales are between 1200~1400, followed by 1400~1600, and then 1000~1200.

Q5

Q3:

(a) Discrete distribution   Binamial Distribution

(b) $N=20$, $X \geq 8$, $P(X)=0.25$

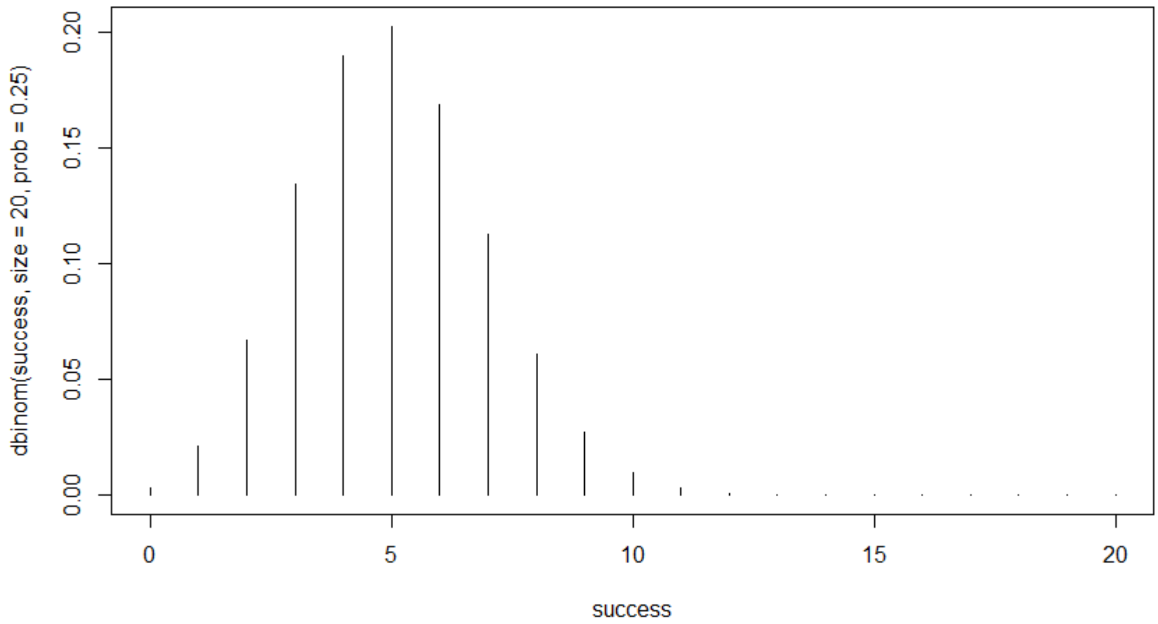
$$P(X \geq 8) = P(8)+P(9)+P(10)+ \sim +P(20)$$

$$= \cancel{0.108}\ 0.10181$$

$\mu = E(X) = 5$, $\sigma = SD(X) = 1.936$, $\sigma^2 = Var(X) = 3.75$

(c)(4 marks) Use R to plot a probability density function (p.d.f) for X, produce a table for the
p.d.f and calculate the probability of a student getting 8 or more correct answers.

```
success <- 0:20

plot(success, dbinom(success, size=20, prob=.25),type='h')
y<-dbinom(success, size=20, prob=.25)

pass.df <- data.frame(x=c(0:20), y=y)

sum(pass.df[pass.df$x>=8, ]$y)
```



and

```
> sum(pass.df[pass.df$x>=8, ]$y)
[1] 0.1018119
```

(d) $N=10$, $X \geq 3$, $P(x)=0.1081$

$P(X \geq 3) = P(3)+P(4)+ \cdots +P(10)$

$\approx \cancel{0.07334}$ 0.08482

Mean $= 1.018$ $\sigma = 0.9144$ $\sigma^2 = 0.9563$

e. (5 marks) Use R to simulate the number of correct answers someone gets for 50 students to populate a sample for the class. Compute and compare the mean, variance, plot the distribution and probability of passing to the distribution of X.

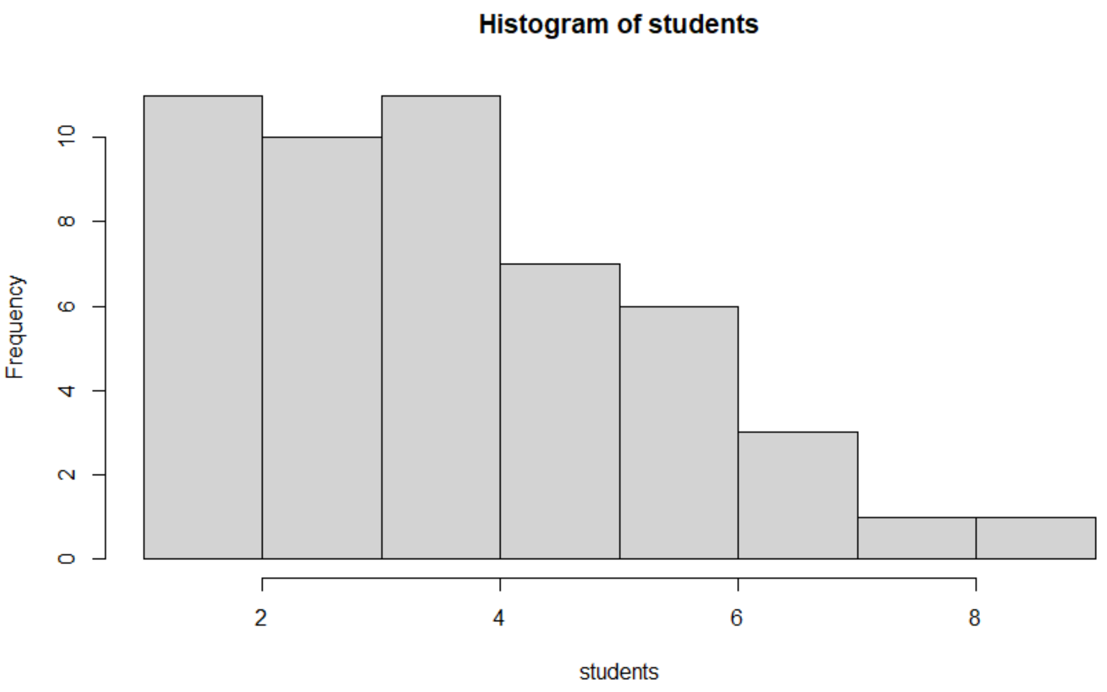
```
students <- rep(0, 50)

for(studentId in c(1:50)){
  correctNum <- 0
```

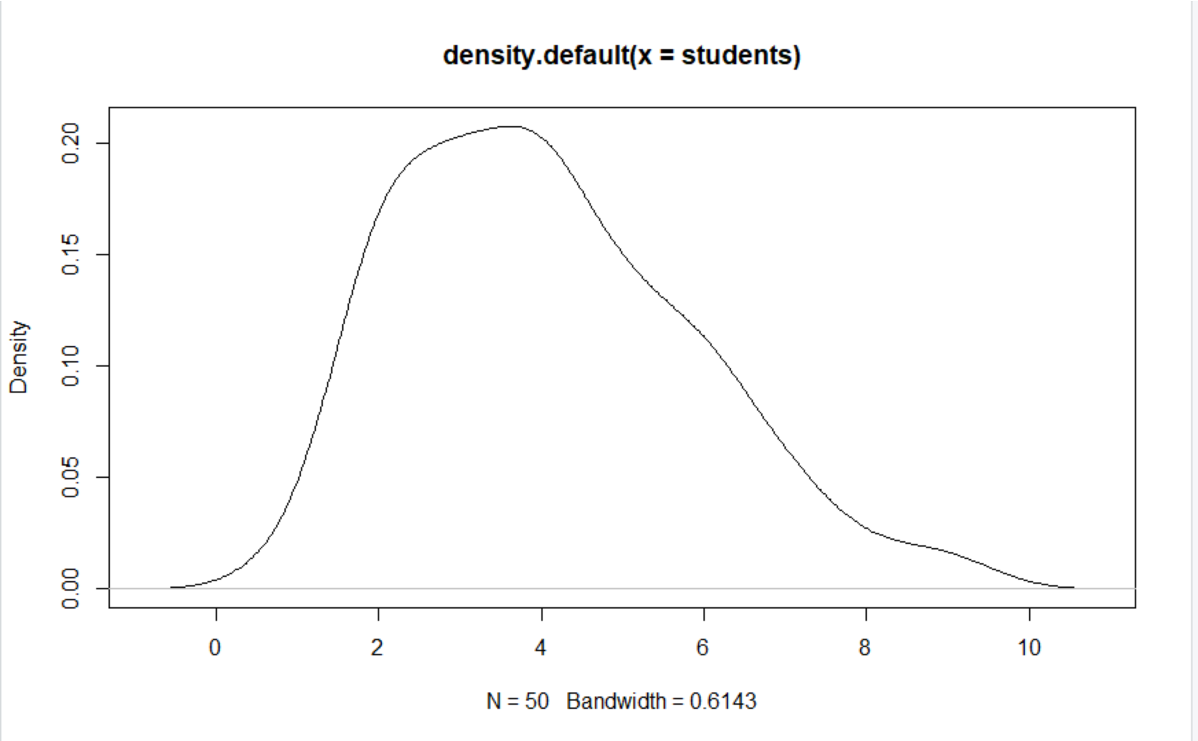
```
  for( i in c(1:20)){
    if(runif(1, 0, 1)<0.25){
      correctNum <- correctNum + 1
    }
  }
  print(correctNum)
  students[studentId] <- correctNum
}

mean(students)
var(students)

set.seed(0)
runif(3,0,1)

hist(students)
plot(density(students))

sum(students>=8)/length(students)
```

Result:

```
[1] 7
[1] 4
[1] 5
[1] 7
[1] 3
[1] 4
[1] 4
[1] 3
[1] 2
[1] 6
[1] 6
[1] 2
[1] 6
[1] 4
[1] 2
[1] 4
[1] 3
[1] 3
[1] 2
[1] 6
[1] 3
[1] 3
[1] 5
[1] 2
[1] 4
[1] 4
[1] 5
[1] 3
[1] 2
[1] 4
[1] 5
[1] 2
[1] 4
```

```
[1] 9
[1] 1
[1] 6
[1] 2
[1] 2
[1] 4
[1] 2
[1] 8
[1] 5
[1] 3
[1] 5
[1] 6
[1] 4
[1] 5
[1] 3
[1] 3
[1] 7


> mean(students)
[1] 4.08
> var(students)
[1] 3.217959
```






```
> sum(students>=8)/length(students)
[1] 0.04
```