S4686474-Yifan Zhang

1. Comparative Study

(a) The cake in different combination of baking time and temperature is the experiment unit.

(b) The factors in this experiment are baking times and temperatures.

(c) For baking time has two level: 25 mins and 30 mins.

For temperature has three level: 275◦F, 300◦F, and 325◦F.

(d) Baking Time: 25 mins, Temperature: 275◦F.

Baking Time: 25 mins, Temperature: 300◦F.

Baking Time: 25 mins, Temperature: 325◦F.

Baking Time: 30 mins, Temperature: 275◦F.

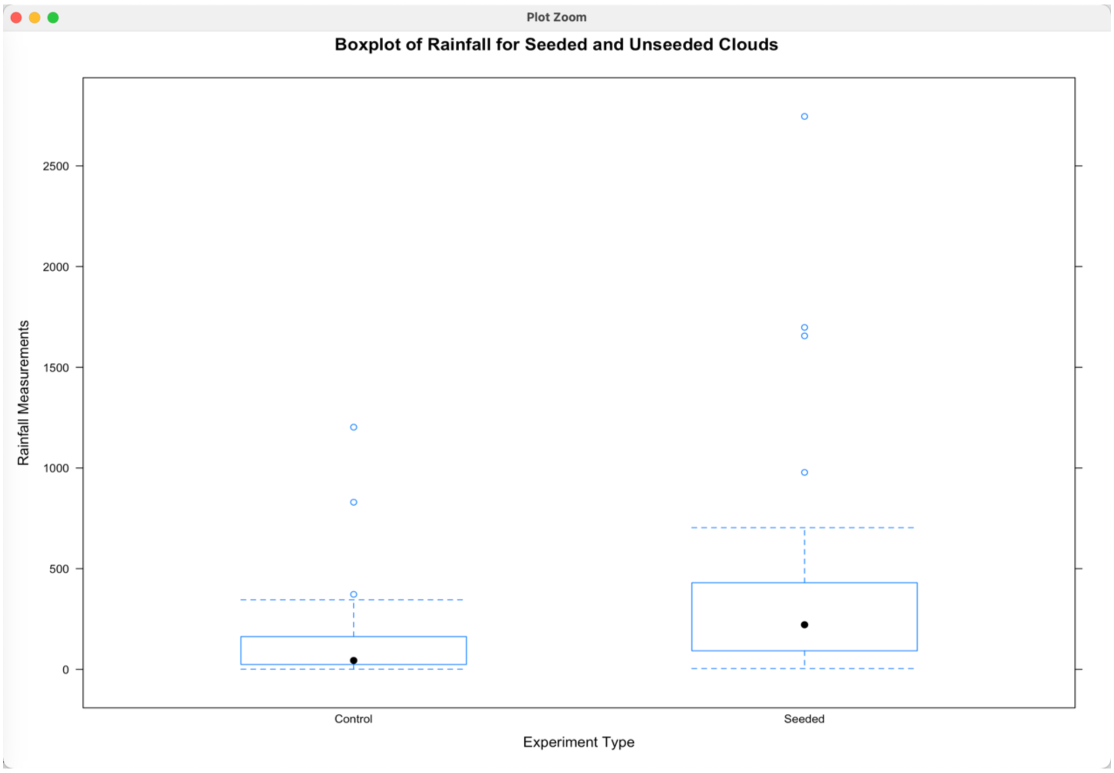
Baking Time: 30 mins, Temperature: 300◦F.

Baking Time: 30 mins, Temperature: 325◦F.

(e) In this experiment taste ("Great," "Mediocre," or "Terrible") is qualitative variable.

2. Visualization

```
library(lattice)
data <- read.csv("/Users/meviusz/UQ/sem2-23/STAT7203/ass1/CloudSeedingData.csv")
stacked <- stack(data)
stacked
bwplot(values ~ ind, data = stacked,
       main = "Boxplot of Rainfall for Seeded and Unseeded Clouds",
       xlab = "Experiment Type", ylab = "Rainfall Measurements")
```



We can observe that the median of the seeded group is higher than that of the control group through boxplot, and we can conclude that the rainfalls is higher in the seeded group than in the control. By IQR, we can find that the rainfall in the seeded group is higher, and the outliners of the boxplot of the seeded group show that there is an extreme number of rainfalls.

3. Counting(the raw code search from chatGPT, but this is modified)

(a) $P(X_1 = 2, X_2 = 5, X_3 = 3) = (10!/(2!*5!*3!))*(1/4)^2 * (1/2)^5 * (1/4)^3$

$= 0.0769$

(b) P(Box1 Empty) = (3/4)^10 = 0.0563

(c)

```
> count <- 0
>
> for (i in 1:100000) {
+     box_number <- sample(c(1, 2, 3), 10, replace = TRUE, prob = c(1/4, 1/2, 1/4))
+     if (sum(box_number == 1) == 2 && sum(box_number == 2) == 5 && sum(box_number == 3) == 3) {
+         count <- count + 1
+     }
+ }
>
> probability <- count / 100000
> cat("Simulated Probability:", probability, "\n")
Simulated Probability: 0.07538
```

```
> count <- 0
>
> for (i in 1:100000) {
+     box_number <- sample(c(1, 2, 3), 10, replace = TRUE, prob = c(1/4, 1/2, 1/4))
+     if (sum(box_number == 1) == 0) {
+         count <- count + 1
+     }
+ }
>
> probability <- count / 100000
> cat("Simulated Probability:", probability, "\n")
Simulated Probability: 0.05615
> |
```

4. Conditional Probability

We can assume:

$P(A) = 0.3$ (spam email)

$P(A^c) = 1-0.3 = 0.7$

$P(B)$ is the probability of filtering to spam email box

$P(B|A) = 0.9$ (the email is spam and correct filtered)

$P(B|A^c) = 0.05$ (the email is non-spam but filtered as spam email)

We need to calculate $P(A|B)$, the mail flagged as spam and correctly.

We can use law of total probability to calculate P(B).

$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$

$P(A|B) = P(B|A)P(A)/P(B) = 0.9*0.3/(0.9*0.3 + 0.05*0.7) = 0.885$

So, we can get the probability of the mail flagged as spam and correctly is 0.885

5. Discrete Random Variable

   (a) P(Win) = 0.139 + 0.014 = 0.153

   (b) The expectation is $E(X) = \sum x * f(x) = 0*0.417 + 1*0.43 + 2*0.139 + 3*0.014 = 0.75$

   The variance of the number matches is Var(X).

   $Var(X)=EX^2-(EX)^2 = (0*0.417+1*0.43+4*0.139+9*0.014)-0.75^2=0.5495$

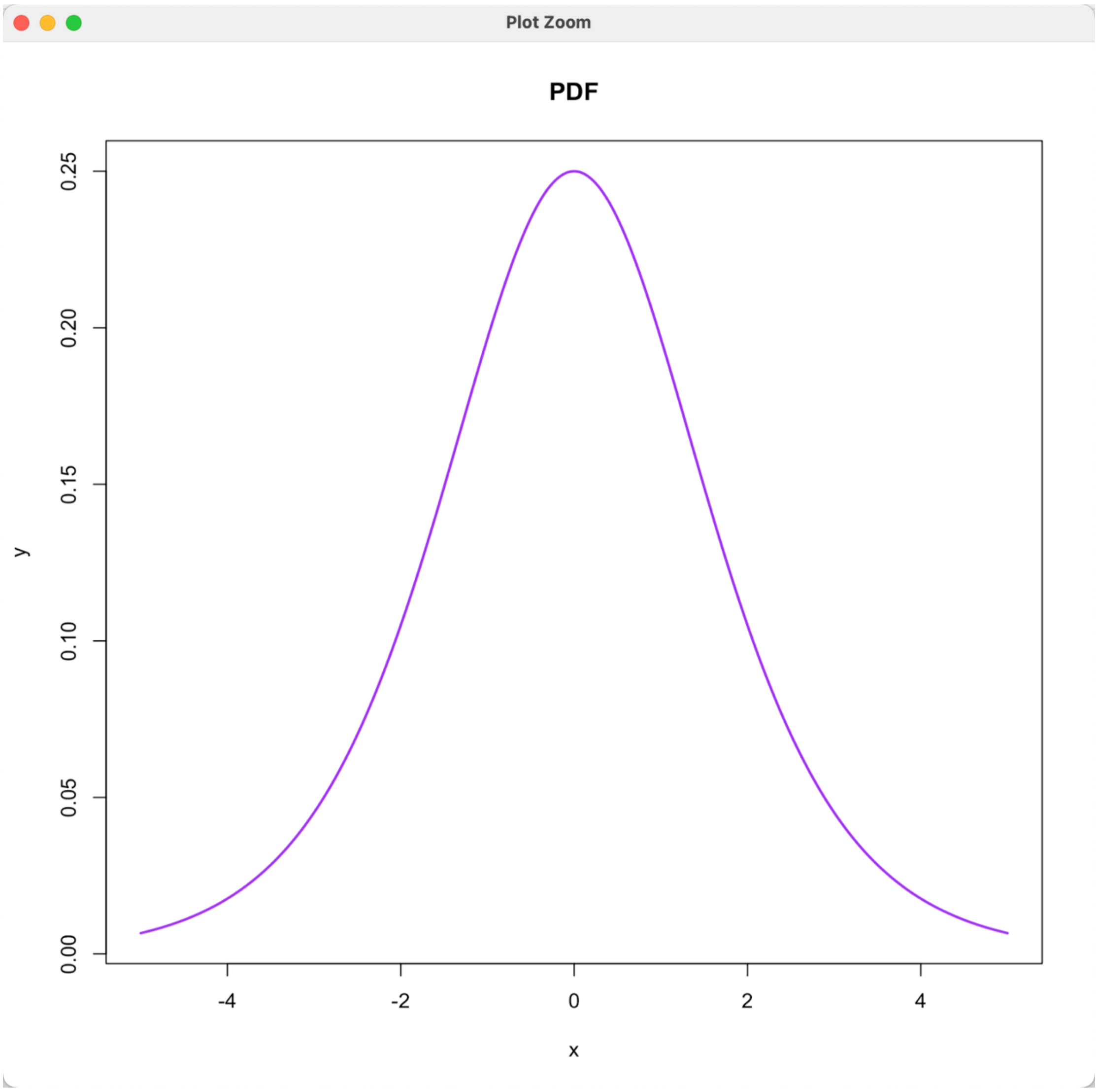
   $Sd(X) = \sqrt{0.5495} = 0.74$

6. Continuous Random Variable(the raw code search from chatGPT, but this is modified)

   (a)

```
pdf <- function(x) {
  exp(-x) / (1 + exp(-x))^2
}

x_values <- seq(-5, 5, length.out = 1000)
y_values <- pdf(x_values)
plot(x_values, y_values, type = "l", col = "purple", lwd = 2,
     main = "PDF", xlab = "x", ylab = "y")
```



(b)

$$\int_{x}^{+\infty} f(x)\,dx = -\int e_{x}^{+\infty} (1/(1+e^{-x})^{2})\, d(1+e^{-x}) = 1/1+e^{-x} \,\Big|_{x}^{+\infty}$$

$$=1 - 1/(1+e^{-x}) = 1/1+e^{x}$$

7. Independence

Proof:

Because A and B is independent, so $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

$P(A|B^c) = (P(A)P(B^c))/ P(B^c)$

Because B^c and B is opposite so $P(B^c) = 1-P(B)$

So we can get $P(A|B^c) = P(A)$

The A and B^c are also independent.

8. Expectation

(a) We have $f(x) + g(x) = \min\{f(x), g(x)\} + \max\{f(x), g(x)\}$

So, $X + Y = \min\{X, Y\} + \max\{X, Y\}$

$E(X, Y) = E(X) + E(Y)$

$E(X + Y) = E(\min\{X, Y\}) + E(\max\{X, Y\})$

$E(X) + E(Y) = E(\min\{X, Y\}) + E(\max\{X, Y\})$

$E(\max\{X, Y\}) = E(X) + E(Y) - E(\min\{X, Y\})$

(b) $M_X(s) = E[E^{sX}]$

Because this function is discrete uniform distribution in domain, so M(s)

$= \sum e^{sx} P_X(x)$

$= \sum e^{sx}(1/(b-a+1))$

$= (1/(b-a+1))(e^{sa}+e^{s(a+1)}+\cdots\cdots+e^{sb})$

$= (1/(b-a+1))(e^{sa}(1+e^{s}+\cdots\cdots+e^{(b-a)s}))$

$= (e^{sa}(1/(b-a+1))(1-e^{s(b-a+1)}))/(1-e^{s})$