ST117 E1

Homework Lab Group 003 Pod E

2025-01-23

This submission was created by:

1. Daniel Guo, Warwick ID: 5645242, Question: A: 1-5,6A,7-9, B:2
2. Zhijian Lin, Warwick ID: 5655296, Question: A: 1-4,6-9
3. Tom O'Connell, Warwick ID: 5628105 , Question: A: 1-3, 4A, 9C, B:2-3
4. Qinling Si, Warwick ID: 5614637, Question: A: 1-3,9, B: 1-3

Question A:

1. Cohort and lab groups

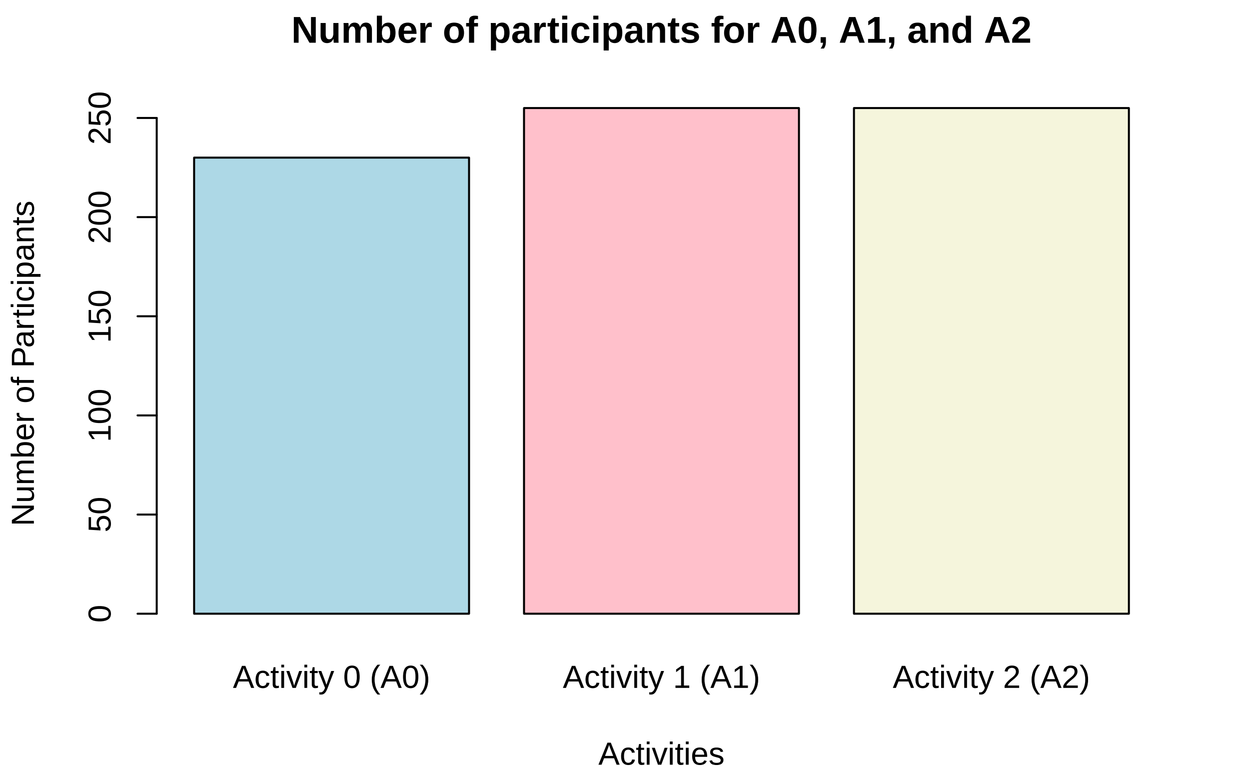
```r
library(randomNames)
set.seed(21012024)

first_names <- randomNames(273, which.names = "first")
last_names <- randomNames(273, which.names = "last")

lab_groups <- c(rep(1:13, each = 18), rep(14, 19), rep(15, 20))

grade_book <- data.frame(
  First_Name = first_names,
  Last_Name = last_names,
  Lab_Group = lab_groups
)

head(grade_book)
```

```
##   First_Name Last_Name Lab_Group
## 1     Karina        Vo         1
## 2     Aatika    Santio         1
## 3     Olivia     Vigil         1
## 4    Haleigh     Panya         1
## 5    Timothy   Sanchez         1
## 6      Taryn    Knight         1
```

2. Activities

```r
participants_A0 <- sample(1:273, 230)
grade_book$Participated_A0 <- ifelse(1:273 %in% participants_A0, "Yes", "No")
grade_book$Marks_A0 <- ifelse(grade_book$Participated_A0 == "Yes", 1, 0)
```

```r
additional_A1 <- sample(setdiff(1:273, participants_A0), 25)
participants_A1 <- unique(c(participants_A0, additional_A1))
grade_book$Participated_A1 <- ifelse(1:273 %in% participants_A1, "Yes", "No")
grade_book$Marks_A1 <- ifelse(grade_book$Participated_A1 == "Yes", 1, 0)

additional_A2 <- sample(setdiff(1:273, participants_A0), 25)
participants_A2 <- unique(c(participants_A0, additional_A2))
grade_book$Participated_A2 <- ifelse(1:273 %in% participants_A2, "Yes", "No")
grade_book$Marks_A2 <- ifelse(grade_book$Participated_A2 == "Yes", 1, 0)

participation_counts <- colSums(grade_book[, c("Marks_A0", "Marks_A1", "Marks_A2")])
barplot(participation_counts,
        main = "Number of participants for A0, A1, and A2",
        xlab = "Activities",
        ylab = "Number of Participants",
        col = c("lightblue", "pink", "beige"),
        names.arg = c("Activity 0 (A0)", "Activity 1 (A1)", "Activity 2 (A2)"))
```




```r
grade_book$Marks_A0 <- NULL
head(grade_book)
```

```
##   First_Name Last_Name Lab_Group Participated_A0 Participated_A1 Marks_A1
## 1     Karina        Vo         1             Yes             Yes        1
## 2     Aatika    Santio         1             Yes             Yes        1
## 3     Olivia     Vigil         1             Yes             Yes        1
## 4    Haleigh     Panya         1             Yes             Yes        1
## 5    Timothy   Sanchez         1             Yes             Yes        1
## 6      Taryn    Knight         1             Yes             Yes        1
##   Participated_A2 Marks_A2
## 1             Yes        1
```

```
## 2            Yes        1
## 3            Yes        1
## 4            Yes        1
## 5            Yes        1
## 6            Yes        1
```

3. Quizzes

3.(a)

```r
set.seed(129642169)
grade_book$Marks_Q1 <- ifelse(
  grade_book$Participated_A0 == "Yes",
  round(rbeta(sum(grade_book$Participated_A0 == "Yes"),4,2),2),
  round(rbeta(sum(grade_book$Participated_A0 == "No"),2,2),2))
grade_book$Marks_Q2 <- ifelse(
  grade_book$Participated_A2 == "Yes",
  round(rbeta(sum(grade_book$Participated_A2 == "Yes"),4,2),2),
  round(rbeta(sum(grade_book$Participated_A2 == "No"),2,6),2))
head(grade_book)
```

```
##   First_Name Last_Name Lab_Group Participated_A0 Participated_A1 Marks_A1
## 1     Karina        Vo         1             Yes             Yes        1
## 2     Aatika    Santio         1             Yes             Yes        1
## 3     Olivia     Vigil         1             Yes             Yes        1
## 4    Haleigh     Panya         1             Yes             Yes        1
## 5    Timothy   Sanchez         1             Yes             Yes        1
## 6      Taryn    Knight         1             Yes             Yes        1
##   Participated_A2 Marks_A2 Marks_Q1 Marks_Q2
## 1             Yes        1     0.75     0.79
## 2             Yes        1     0.43     0.88
## 3             Yes        1     0.85     0.81
## 4             Yes        1     0.81     0.85
## 5             Yes        1     0.72     0.80
## 6             Yes        1     0.87     0.73
```

3.(b)

Reason for choosing the beta distribution

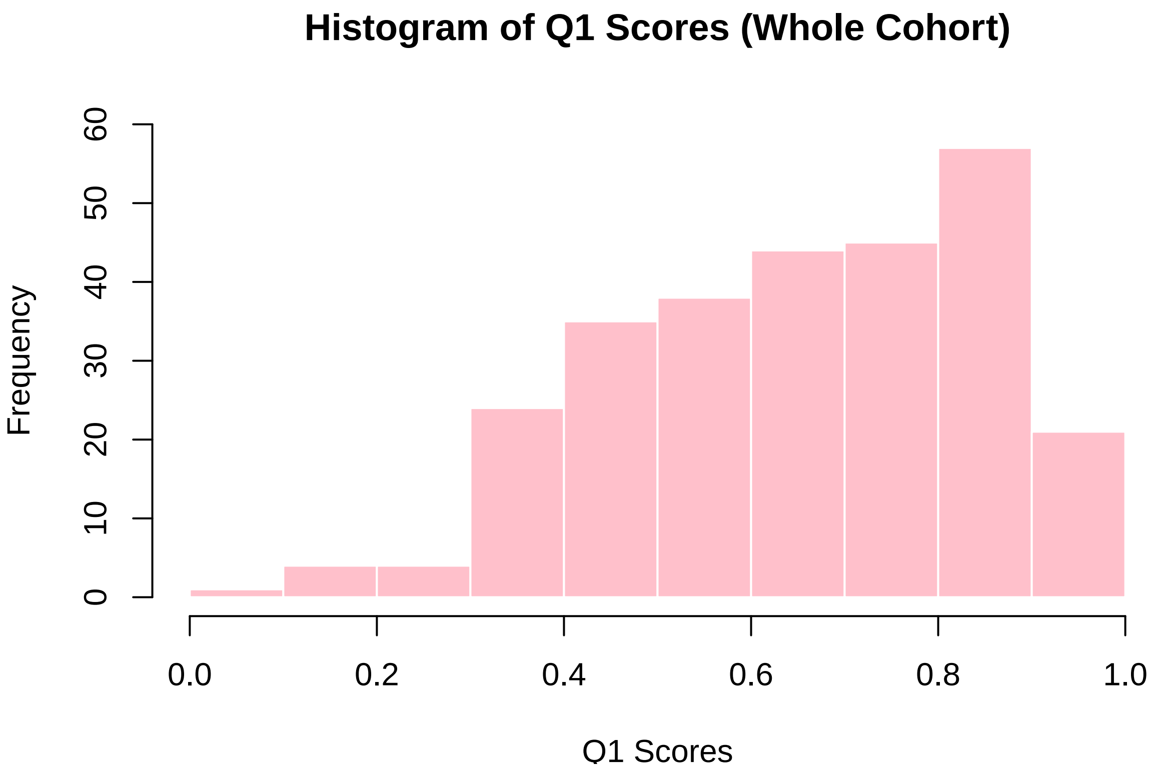
1. Beta distribution is defined on the interval [0,1], so that it is suitable to represent percentage of scores for a quiz or exercise.
2. Beta distribution is easy to adjust the skewness simply by changing the parameters, making it easier to reflect the performance on a quiz or exercise.
3. Parameters α and β can represent the overall performance of a quiz, where a larger α value indicates better overall performance and a larger β value indicates worse overall performance.

Rationale for using different parameters depending on the participation the activities.

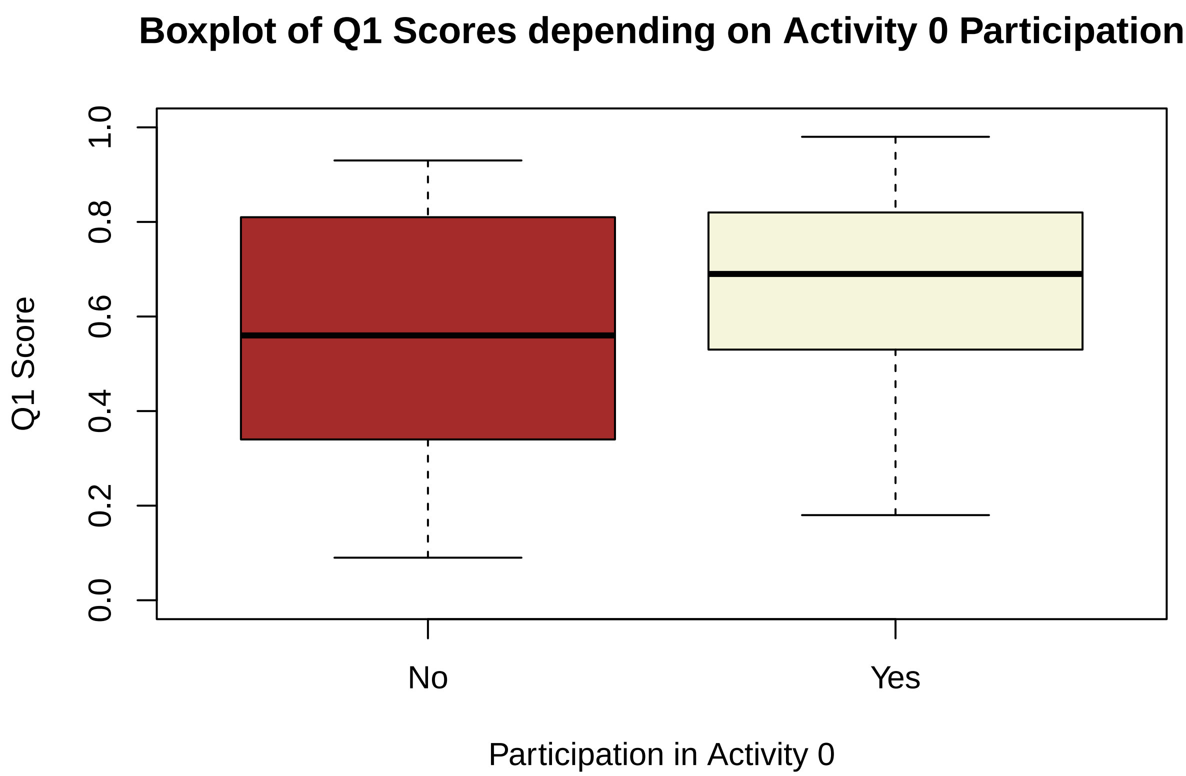
1. Different parameters can indicate different performance level as mentioned previously. Therefore, it can reflect the difference in preparedness of students who participated in the activities and those who did not. Since the one who participated would be more prepared for the quiz than those who did not, which would suggest a better overall performance, reflecting the skewness of the beta distribution due to different parameters (a larger α value indicates better overall performance and a larger β value indicates worse overall performance).

3.(c)

```r
hist(grade_book$Marks_Q1,
     main = "Histogram of Q1 Scores (Whole Cohort)",
     col = "pink",
     border = "white",
     xlab = "Q1 Scores",
     ylim=c(0,60))
```




```r
boxplot(
  Marks_Q1 ~ Participated_A0,
  data = grade_book,
  main = "Boxplot of Q1 Scores depending on Activity 0 Participation",
  xlab = "Participation in Activity 0",
  ylab = "Q1 Score",
  col = c("brown", "beige"),
  ylim = c(0, 1)
)
```



Observations from the Plots

1. The histogram shows the overall distribution of Q1 scores for the entire cohort, which is negatively skewed, with a mode between 0.8 to 0.9.
2. From the boxplots, students who participated in A0 have a higher median Q1 score, a smaller interquartile range, suggesting the ones who participated in A0 have more consistent performance.
3. From the boxplots, students who did not participate in A0 have a lower minimum value and median value, potentially due to lack of preparedness and experience compared the ones who attended A0

4. Assign Homework Pods

4.(a)

```
LG13<-rep(LETTERS[1:6], each = 3)
LG14<-rep(LETTERS[1:6],c(4,3,3,3,3,3))
LG15<-rep(LETTERS[1:6],c(4,4,3,3,3,3))

for (i in 1:13) {
  grade_book$Homework_Pods[grade_book$Lab_Group == i] <- sample(LG13, replace = FALSE)
}
grade_book$Homework_Pods[grade_book$Lab_Group == 14] <- sample(LG14,replace=FALSE)
grade_book$Homework_Pods[grade_book$LabG_roup == 15] <- sample(LG15,replace=FALSE)

grade_book$Pod_ID <-paste(grade_book$Lab_Group, grade_book$Homework_Pod)
head(grade_book)
```

```
##   First_Name Last_Name Lab_Group Participated_A0 Participated_A1 Marks_A1
## 1     Karina        Vo         1             Yes             Yes        1
## 2     Aatika    Santio         1             Yes             Yes        1
```

```
## 3     Olivia     Vigil          1             Yes             Yes         1
## 4    Haleigh     Panya          1             Yes             Yes         1
## 5    Timothy   Sanchez          1             Yes             Yes         1
## 6      Taryn    Knight          1             Yes             Yes         1
##   Participated_A2 Marks_A2 Marks_Q1 Marks_Q2 Homework_Pods Pod_ID
## 1             Yes        1     0.75     0.79             F    1 F
## 2             Yes        1     0.43     0.88             C    1 C
## 3             Yes        1     0.85     0.81             E    1 E
## 4             Yes        1     0.81     0.85             D    1 D
## 5             Yes        1     0.72     0.80             A    1 A
## 6             Yes        1     0.87     0.73             C    1 C
```

4.(b)

When $N mod 3 = 0$:

1. The total number of permutations of N students is $N!$, when $N \geq 1$ and $N mod 3 = 0$
2. Divide by the internal order within the pods is $(3!)^{\frac{N}{3}}$ where 3! is the order within each group and $\frac{N}{3}$ is the number of groups
3. Divide by the order of the pods is $(\frac{N}{3})!$

Therefore, the total number of ways $= \frac{N!}{(3!)^{\frac{N}{3}} \times (\frac{N}{3})!}$ when $N \geq 1$ and $N mod 3 = 0$

When $N mod 3 = 1$:

One Pod has 4 students, and the rest have 3:

1. Choose 4 students out of N to form the size-4 Pod: $\binom{N}{4}$
2. Similar to when $N mod 3 = 0$, we assign the remaining $N - 4$ students into $\frac{N-4}{3}$ pods with three students in each: $\frac{(N-4)!}{(3!)^{\frac{N-4}{3}} \times (\frac{N-1}{3})!}$

Therefore, the total number of ways $= \binom{N}{4} \times \frac{(N-4)!}{(3!)^{\frac{N-4}{3}} \times (\frac{N-1}{3})!}$ when $N \geq 1$ and $N mod 3 = 1$

When $N mod 3 = 2$:

Two Pods have 4 students, and the rest have 3:

1. Choose 4 students out of N for the first Pod with 4 students: $\binom{N}{4}$
2. Choose 4 students out of the remaining $N - 4$ for the second Pod with 4 students: $\binom{N-4}{4}$
3. Assign the remaining $N - 8$ students into $\frac{N-8}{3}$ pods with three students in each: $\frac{(N-8)!}{(3!)^{\frac{N-8}{3}} \times (\frac{N-2}{3})!}$
4. Ways to assign the order of the two 4 students pods: 2!

Therefore, the total number of ways $= \binom{N}{4} \times \binom{N-4}{4} \times \frac{(N-8)!}{(3!)^{\frac{N-8}{3}} \times (\frac{N-2}{3})! \times 2!}$ when $N \geq 1$ and $N mod 3 = 2$

When $N = 18$:

The total number of ways

$$= \frac{18!}{(3!)^{\frac{18}{3}} \times (\frac{18}{3})!}$$

$$= \frac{18!}{(3!)^6 \times 6!} == \frac{18!}{6^6 \times 6!} = 190590400$$

When $N = 19$:

The total number of ways

$$= \binom{19}{4} \times \frac{(19-4)!}{(3!)^{\frac{19-4}{3}} \times (\frac{19-1}{3})!}$$

$$= \binom{19}{4} \times \frac{15!}{6^5 \times 6!} = 905304400$$

When $N = 20$:

The total number of ways

$$= \binom{20}{4} \times \binom{20-4}{4} \times \frac{(20-8)!}{(3!)^{\frac{20-8}{3}} \times (\frac{20-1}{3})!}$$

$$= \binom{20}{4} \times \binom{16}{4} \times \frac{12!}{6^4 \times 6!} = 4526522000$$

4.(c)

```r
Possible_pods_allocation<-function(N) {
  if (N %% 3 == 0) {
    k <- N / 3
    return(factorial(N)/(factorial(3))^k/factorial(k))
  }

  if (N %% 3 == 1) {
    k <- (N - 4) / 3
    return(choose(N, 4)*factorial(N - 4)/factorial(3)^k/factorial(k+1))
  }

  if (N %% 3 == 2) {
    k <- (N - 8) / 3
    return(
      choose(N, 4)*choose(N - 4, 4)*factorial(N - 8)/factorial(3)^k/factorial(k+2))
  }
}

Possible_pods_allocation(18) #N = 18
```

```
## [1] 190590400
```

```r
Possible_pods_allocation(19) #N = 19
```

```
## [1] 905304400
```

```r
Possible_pods_allocation(20) #N = 20
```

```
## [1] 4526522000
```

5. Exercise Sets

5.(a)

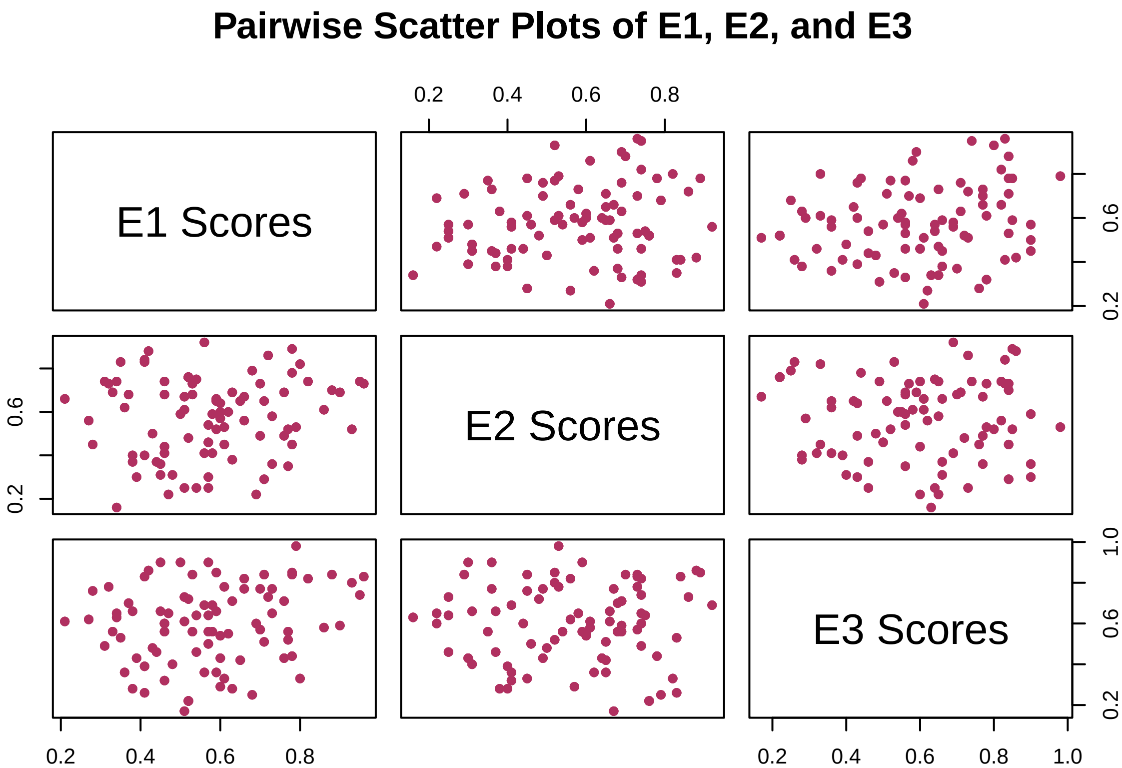
```r
set.seed(129642169)
pod_scores <- unique(grade_book$Pod_ID)
pod_scores_df <- data.frame(
  Pod_ID = pod_scores,
  E1_Score = round(rbeta(length(pod_scores), 4, 3), 2),
  E2_Score = round(rbeta(length(pod_scores), 4, 3), 2),
  E3_Score = round(rbeta(length(pod_scores), 4, 3), 2)
)

grade_book <- merge(grade_book, pod_scores_df, by = "Pod_ID")
head(grade_book)
```

```
##   Pod_ID First_Name Last_Name Lab_Group Participated_A0 Participated_A1
## 1    1 A     Aliyya   Borunda         1              No             Yes
## 2    1 A      Jerod el-Hashim         1             Yes             Yes
## 3    1 A    Timothy   Sanchez         1             Yes             Yes
## 4    1 B     Morgan  Arellano         1              No             Yes
## 5    1 B Jacqueline        Ky         1              No              No
## 6    1 B    Liberty    Torrez         1             Yes             Yes
##   Marks_A1 Participated_A2 Marks_A2 Marks_Q1 Marks_Q2 Homework_Pods E1_Score
## 1        1             Yes        1     0.60     0.79             A     0.62
## 2        1             Yes        1     0.51     0.85             A     0.62
## 3        1             Yes        1     0.72     0.80             A     0.62
## 4        1             Yes        1     0.32     0.94             B     0.79
## 5        0             Yes        1     0.49     0.92             B     0.79
## 6        1             Yes        1     0.45     0.36             B     0.79
##   E2_Score E3_Score
## 1     0.60     0.55
## 2     0.60     0.55
## 3     0.60     0.55
## 4     0.53     0.98
## 5     0.53     0.98
## 6     0.53     0.98
```

```r
hist(
  grade_book$E1_Score,
  breaks = 20,                 # Number of breaks
  col = "lightblue",           # Colour for the bars
  border = "white",         # Color for the border
  main = "Histogram of Exercise Set 1 Scores",  # Title
  xlab = "Exercise Ser 1 Scores",         # X-axis label
  ylab = "Frequency",          # Y-axis label
  xlim = c(0, 1),               # Limit for x-axis
  ylim = c(0,50)            # Limit for y-axis
)
```

5.(b)

```
pairs(
  grade_book[ , c("E1_Score", "E2_Score", "E3_Score")],
  main = "Pairwise Scatter Plots of E1, E2, and E3",
  pch = 16,
  col = "maroon",
  labels = c("E1 Scores", "E2 Scores", "E3 Scores")
)
```



Observation: 1. Since they are independently sampled from the Beta distributions, the scatter plots show no strong correlation between the scores of E1 E2 E3.

6. Assign Report Pods

6.(a)

```r
set.seed(129642169)
lab_20_students <- which(grade_book$Lab_Group == 15)  # Find students in the lab with 20 students
dropout_student <- sample(lab_20_students, 1)            # Randomly choose one student to drop
grade_book <- grade_book[-dropout_student, ]             # Remove the dropped student

# Randomly assign remaining students to Report Pods
remaining_students <- nrow(grade_book)
number_pods <- 68
pod_size <- 4

# Shuffle student indices and assign Report Pods
shuffle <- sample(1:remaining_students)
report_pods <- rep(1:number_pods, each = pod_size)  # Assign 4 students per pod
report_pods <- report_pods[1:remaining_students]
grade_book$Report_Pod <- paste("Pod", report_pods[shuffle])  # Assign pods to students

head(grade_book)
```

```
##   Pod_ID First_Name Last_Name Lab_Group Participated_A0 Participated_A1
## 1    1 A     Aliyya   Borunda         1              No             Yes
## 2    1 A      Jerod el-Hashim         1             Yes             Yes
## 3    1 A    Timothy   Sanchez         1             Yes             Yes
## 4    1 B     Morgan  Arellano         1              No             Yes
```

```
## 5      1 B Jacqueline         Ky         1             No              No
## 6      1 B    Liberty     Torrez         1            Yes             Yes
##   Marks_A1 Participated_A2 Marks_A2 Marks_Q1 Marks_Q2 Homework_Pods E1_Score
## 1        1             Yes        1     0.60     0.79             A     0.62
## 2        1             Yes        1     0.51     0.85             A     0.62
## 3        1             Yes        1     0.72     0.80             A     0.62
## 4        1             Yes        1     0.32     0.94             B     0.79
## 5        0             Yes        1     0.49     0.92             B     0.79
## 6        1             Yes        1     0.45     0.36             B     0.79
##   E2_Score E3_Score Report_Pod
## 1     0.60     0.55      Pod 9
## 2     0.60     0.55     Pod 13
## 3     0.60     0.55     Pod 49
## 4     0.53     0.98     Pod 10
## 5     0.53     0.98      Pod 2
## 6     0.53     0.98     Pod 21
```

6.(b)

2. Using the multinomial coefficient: with $273 - 1 = 272$ students divided into 68 groups of 4 students each: The number of ways of this assignments

$$= \frac{272!}{(4!)^{68} \times 68!}$$

- 272! represents the total number of ways to arrange all students.
- $(4!)^{68}$ represents the order arrangements of 4 students within each of the 68 pods.
- 68! represents the order arrangements of the 68 pods.

6.(c)

Consider one student, he is already in a report pod, so there are 3 remaining seats, so that $\binom{272}{3}$ represent all possibilities.

If no fellow student is from his homework pod, they come from other 269 students, so it is $\binom{269}{3}$

Probability that this student meet his fellow is $P = 1 - \binom{269}{3}/\binom{272}{3}$

There are 18 students with same situations, so expected number $= 18 \times P$

Expected_number$= 18 \times (1 - \binom{269}{3}/\binom{272}{3})$

7. Exercise Sets

7.(a)

```r
set.seed(129642169)
grade_book$Log_Participation <- rbinom(nrow(grade_book), 6, 0.8)
grade_book$Passed_Logs <- round(grade_book$Log_Participation/6, 2)
head(grade_book)
```

```
##   Pod_ID First_Name Last_Name Lab_Group Participated_A0 Participated_A1
## 1      1 A     Aliyya   Borunda         1              No             Yes
## 2      1 A      Jerod el-Hashim         1             Yes             Yes
## 3      1 A    Timothy   Sanchez         1             Yes             Yes
## 4      1 B     Morgan  Arellano         1              No             Yes
## 5      1 B Jacqueline         Ky         1              No              No
```

```
## 6     1 B    Liberty    Torrez         1             Yes            Yes
##   Marks_A1 Participated_A2 Marks_A2 Marks_Q1 Marks_Q2 Homework_Pods E1_Score
## 1        1             Yes        1     0.60     0.79             A     0.62
## 2        1             Yes        1     0.51     0.85             A     0.62
## 3        1             Yes        1     0.72     0.80             A     0.62
## 4        1             Yes        1     0.32     0.94             B     0.79
## 5        0             Yes        1     0.49     0.92             B     0.79
## 6        1             Yes        1     0.45     0.36             B     0.79
##   E2_Score E3_Score Report_Pod Log_Participation Passed_Logs
## 1     0.60     0.55      Pod 9                 5        0.83
## 2     0.60     0.55     Pod 13                 5        0.83
## 3     0.60     0.55     Pod 49                 4        0.67
## 4     0.53     0.98     Pod 10                 6        1.00
## 5     0.53     0.98      Pod 2                 6        1.00
## 6     0.53     0.98     Pod 21                 5        0.83
```

7.(b)

```r
grade_book$Participated_Draft <- ifelse(grade_book$Log_Participation >= 3, "Yes", "No")
grade_book$Mark_Draft <- ifelse(grade_book$Participated_Draft == "Yes", 1, 0)
grade_book$Participated_Draft <- NULL
head(grade_book)
```

```
##   Pod_ID First_Name Last_Name Lab_Group Participated_A0 Participated_A1
## 1     1 A     Aliyya   Borunda         1              No            Yes
## 2     1 A      Jerod el-Hashim         1             Yes            Yes
## 3     1 A    Timothy   Sanchez         1             Yes            Yes
## 4     1 B     Morgan  Arellano         1              No            Yes
## 5     1 B Jacqueline        Ky         1              No             No
## 6     1 B    Liberty    Torrez         1             Yes            Yes
##   Marks_A1 Participated_A2 Marks_A2 Marks_Q1 Marks_Q2 Homework_Pods E1_Score
## 1        1             Yes        1     0.60     0.79             A     0.62
## 2        1             Yes        1     0.51     0.85             A     0.62
## 3        1             Yes        1     0.72     0.80             A     0.62
## 4        1             Yes        1     0.32     0.94             B     0.79
## 5        0             Yes        1     0.49     0.92             B     0.79
## 6        1             Yes        1     0.45     0.36             B     0.79
##   E2_Score E3_Score Report_Pod Log_Participation Passed_Logs Mark_Draft
## 1     0.60     0.55      Pod 9                 5        0.83          1
## 2     0.60     0.55     Pod 13                 5        0.83          1
## 3     0.60     0.55     Pod 49                 4        0.67          1
## 4     0.53     0.98     Pod 10                 6        1.00          1
## 5     0.53     0.98      Pod 2                 6        1.00          1
## 6     0.53     0.98     Pod 21                 5        0.83          1
```

8. Final Report

```r
set.seed(129642169)
unique_pods <- unique(grade_book$Report_Pod)

# Vector to store WR marks for each pod
WR_pod_marks <- numeric(length(unique_pods))
```

```r
# Assign WR marks
for (i in seq_along(unique_pods)) {
  # Get the students in the current Report Pod
  pod_students <- grade_book$Report_Pod == unique_pods[i]

  # Check if all students in the pod submitted their draft reports
  all_submitted <- all(grade_book$Mark_Draft[pod_students] == 1)

  # Assign WR marks based on submission status
  WR_pod_marks[i] <- if (all_submitted) {
    round(rbeta(1, 5, 2),2)  # Beta(5, 2) if all members submitted
  } else {
    round(rbeta(1, 3, 2),2)  # Beta(3, 2) otherwise
  }
}

# Match the WR marks to each student in the Grade Book
grade_book$WR_Marks <- WR_pod_marks[match(grade_book$Report_Pod, unique_pods)]

# View the updated Grade Book
head(grade_book)
```

```
##   Pod_ID First_Name Last_Name Lab_Group Participated_A0 Participated_A1
## 1      1 A     Aliyya   Borunda         1              No             Yes
## 2      1 A      Jerod el-Hashim         1             Yes             Yes
## 3      1 A    Timothy   Sanchez         1             Yes             Yes
## 4      1 B     Morgan  Arellano         1              No             Yes
## 5      1 B Jacqueline        Ky         1              No              No
## 6      1 B    Liberty    Torrez         1             Yes             Yes
##   Marks_A1 Participated_A2 Marks_A2 Marks_Q1 Marks_Q2 Homework_Pods E1_Score
## 1        1             Yes        1     0.60     0.79             A     0.62
## 2        1             Yes        1     0.51     0.85             A     0.62
## 3        1             Yes        1     0.72     0.80             A     0.62
## 4        1             Yes        1     0.32     0.94             B     0.79
## 5        0             Yes        1     0.49     0.92             B     0.79
## 6        1             Yes        1     0.45     0.36             B     0.79
##   E2_Score E3_Score Report_Pod Log_Participation Passed_Logs Mark_Draft
## 1     0.60     0.55      Pod 9                 5        0.83          1
## 2     0.60     0.55     Pod 13                 5        0.83          1
## 3     0.60     0.55     Pod 49                 4        0.67          1
## 4     0.53     0.98     Pod 10                 6        1.00          1
## 5     0.53     0.98      Pod 2                 6        1.00          1
## 6     0.53     0.98     Pod 21                 5        0.83          1
##   WR_Marks
## 1     0.79
## 2     0.49
## 3     0.87
## 4     0.84
## 5     0.76
## 6     0.89
```

9. Module Mark

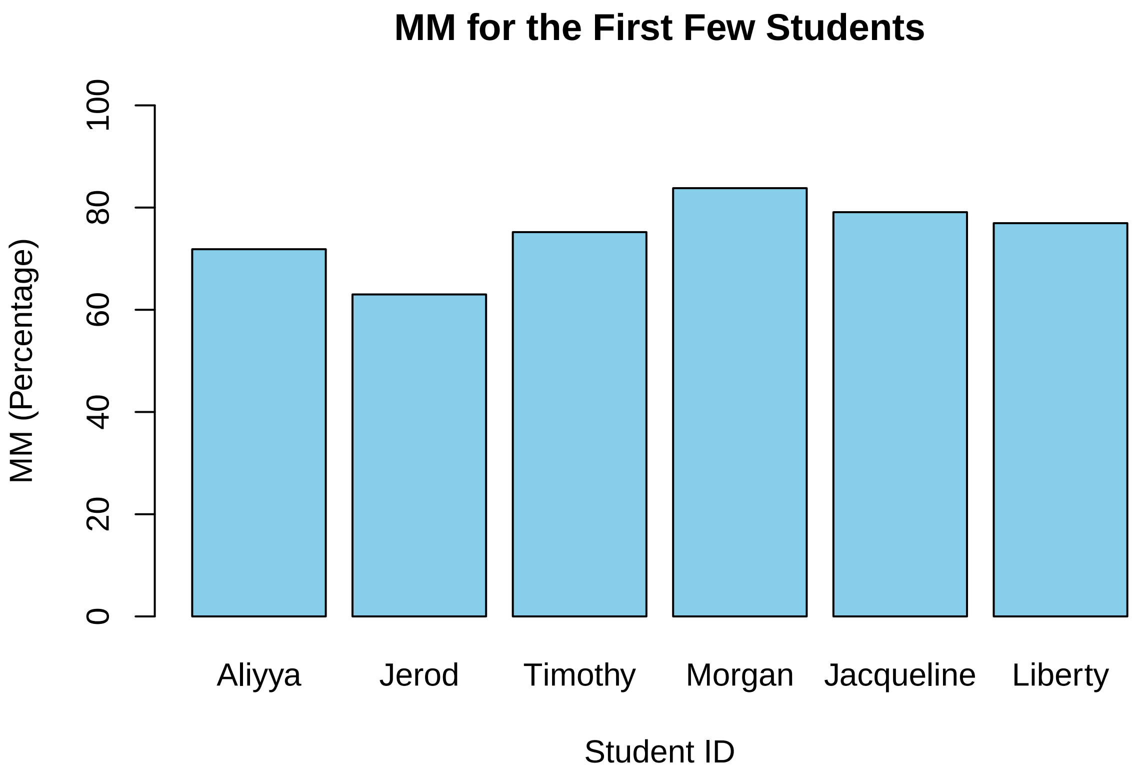
9.(a)

```r
grade_book$MM <- with(grade_book, {
  # Best 4 scores from Q1, Q2, E1, E2, E3 (60%)
  best_4_scores <- apply(cbind(Marks_Q1, Marks_Q2, E1_Score, E2_Score, E3_Score), 1, function(x)
    {mean(sort(x, decreasing = TRUE)[1:4])  # Average of the best 4 scores
  })

  # Final MM calculation
  final_mark <- 0.6*best_4_scores+0.02*(Marks_A1+Marks_A2)+0.01*Log_Participation+0.3*WR_Marks
  final_mark*100 #percentage
})

head(grade_book)
```

```
##   Pod_ID First_Name Last_Name Lab_Group Participated_A0 Participated_A1
## 1    1 A     Aliyya   Borunda         1              No             Yes
## 2    1 A      Jerod el-Hashim         1             Yes             Yes
## 3    1 A    Timothy   Sanchez         1             Yes             Yes
## 4    1 B     Morgan  Arellano         1              No             Yes
## 5    1 B Jacqueline        Ky         1              No              No
## 6    1 B    Liberty    Torrez         1             Yes             Yes
##   Marks_A1 Participated_A2 Marks_A2 Marks_Q1 Marks_Q2 Homework_Pods E1_Score
## 1        1             Yes        1     0.60     0.79             A     0.62
## 2        1             Yes        1     0.51     0.85             A     0.62
## 3        1             Yes        1     0.72     0.80             A     0.62
## 4        1             Yes        1     0.32     0.94             B     0.79
## 5        0             Yes        1     0.49     0.92             B     0.79
## 6        1             Yes        1     0.45     0.36             B     0.79
##   E2_Score E3_Score Report_Pod Log_Participation Passed_Logs Mark_Draft
## 1     0.60     0.55      Pod 9                 5        0.83          1
## 2     0.60     0.55     Pod 13                 5        0.83          1
## 3     0.60     0.55     Pod 49                 4        0.67          1
## 4     0.53     0.98     Pod 10                 6        1.00          1
## 5     0.53     0.98      Pod 2                 6        1.00          1
## 6     0.53     0.98     Pod 21                 5        0.83          1
##   WR_Marks    MM
## 1     0.79 71.85
## 2     0.49 63.00
## 3     0.87 75.20
## 4     0.84 83.80
## 5     0.76 79.10
## 6     0.89 76.95
```

```r
first_few_mm <- head(grade_book$MM)
student_names <- head(grade_book$First_Name)

# Plot MM for the first few students
barplot(
  first_few_mm,
  names.arg = student_names,
  main="MM for the First Few Students",
  xlab ="Student ID",
```

```r
  ylab = "MM (Percentage)",
  col= "skyblue",
  border ="black",
  ylim= c(0,100)
)
```



9.(b)

```r
mean(grade_book$MM)
```

```
## [1] 69.65257
```

```r
sd(grade_book$MM)
```

```
## [1] 7.692895
```

```r
summary(grade_book$MM)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   47.45   64.97   68.88   69.65   75.11   88.90
```

9.(c)

```r
set.seed(129642169)
simulate_grade_book <- function() {
  num_students <- 273
  grade_book <- data.frame(
    StudentID= 1:num_students,
    Marks_Q1= rbeta(num_students, 4, 2),
    Marks_Q2= rbeta(num_students, 4, 2),
```

```
    E1= rbeta(num_students, 4, 3),
    E2= rbeta(num_students, 4, 3),
    E3= rbeta(num_students, 4, 3),
    Marks_A1= rbinom(num_students, 1, 0.8),
    Marks_A2= rbinom(num_students, 1, 0.8),
    Log_Participation = rbinom(num_students, 6, 0.8),
    WR_Marks = ifelse(
      rbinom(num_students, 1, 0.9) == 1,
      rbeta(num_students, 5, 2),
      rbeta(num_students, 3, 2)
    )
  )
  grade_book$MM <- with(grade_book, {
  # calculate MM
  best_4_scores <- apply(cbind(Marks_Q1, Marks_Q2, E1, E2, E3), 1, function(x)
    {mean(sort(x, decreasing = TRUE)[1:4])
  })
  final_mark <- 0.6*best_4_scores+0.02*(Marks_A1+Marks_A2)+0.01*Log_Participation+0.3*WR_Marks
  final_mark*100
  })
  # Summary of MM
  summary_stats <- summary(grade_book$MM)
  return(list(mean = mean(grade_book$MM),
              sd= sd(grade_book$MM),
              min= summary_stats[1],
              q1 = summary_stats[2],
              median= summary_stats[3],
              q3 = summary_stats[4],
              max= summary_stats[5]))
}

#Run the simulation 10 times and store results
results <- data.frame(mean= numeric(10),
                      sd= numeric(10),
                      min= numeric(10),
                      q1 = numeric(10),
                      median= numeric(10),
                      q3= numeric(10),
                      max = numeric(10))

for (i in 1:10) {
  result <- simulate_grade_book()
  results$mean[i] <- result$mean
  results$sd[i] <- result$sd
  results$min[i] <- result$min
  results$q1[i] <- result$q1
  results$median[i] <- result$median
  results$q3[i] <- result$q3
  results$max[i] <- result$max
}

# Print the results of 10 simulations
print(results)
```

```
##        mean       sd      min       q1   median       q3      max
## 1  68.48325 7.360155 46.91509 63.61001 68.86744 68.48325 73.63671
## 2  69.71133 7.029426 50.12650 64.76191 69.92890 69.71133 74.75270
## 3  68.70847 7.452647 40.68269 64.09802 69.01744 68.70847 73.91819
## 4  68.60742 6.676521 50.45231 63.61818 68.93407 68.60742 73.54917
## 5  69.48858 6.551702 44.18072 64.84325 70.09015 69.48858 74.36311
## 6  68.98894 6.597471 51.38407 64.49951 68.64148 68.98894 73.94280
## 7  68.78190 7.414926 49.05056 63.74786 69.85558 68.78190 74.16132
## 8  69.00076 7.400844 47.24991 64.44857 69.16017 69.00076 74.17190
## 9  68.73888 7.859538 44.44570 63.93481 69.45774 68.73888 74.63328
## 10 69.55377 6.990255 50.54780 65.14649 69.53365 69.55377 75.02124
```

```
# Calculate the SD of the mean
sd_mean <- sd(results$mean)
print(paste("Standard Deviation of the 10 simulations Mean:", sd_mean))
```

```
## [1] "Standard Deviation of the 10 simulations Mean: 0.431143886656043"
```

Question B

1(a)

```
print("name of distribution:binomial distribution")
```

```
## [1] "name of distribution:binomial distribution"
```

```
print("X~Bin(13,0.5)")
```

```
## [1] "X~Bin(13,0.5)"
```

```
n<-13
p_0<-0.5
expected_value<-p_0*n
variance<-n*p_0*(1-p_0)
SD<-sqrt(variance)
cat("expected value is:",expected_value,"\n")
```

```
## expected value is: 6.5
```

```
cat("variance is:",variance,"\n")
```

```
## variance is: 3.25
```

```
cat("standard deviation is:",SD,"\n")
```

```
## standard deviation is: 1.802776
```

1(b)

```
n<-13
p_0<-0.5
probability_at_least_7 <- 1 - pbinom(6, size = n, prob = p_0)
cat("the probability that student i0 answers at least 7 is:",probability_at_least_7,"\n")
```

```
## the probability that student i0 answers at least 7 is: 0.5
```

1(c)

```r
# i)exact
n<-13
p_0<-0.5
k<-0:n
exact_probability<-dbinom(k,size = n,prob = p_0)
cat("exact value is:",exact_probability,"\n")
```

```
## exact value is: 0.0001220703 0.001586914 0.009521484 0.03491211 0.08728027 0.1571045 0.2094727 0.209
```

```r
# ii)normal distribution
sigma<-sqrt(variance)
normal_probability <- dnorm(k, mean =expected_value , sd = sigma)
cat("the normal approximation is:",normal_probability,"\n")
```

```
## the normal approximation is: 0.0003327011 0.002107797 0.009816893 0.03361169 0.0846014 0.1565437 0.2
```

```r
# iii)normal approximation with continuity correction
normal_probability_continuity_correction<-pnorm(k + 0.5, mean = expected_value, sd = sigma) - pnorm(k- 0
cat("the normal approximation with continuity correction is:",normal_probability_continuity_correction,
```

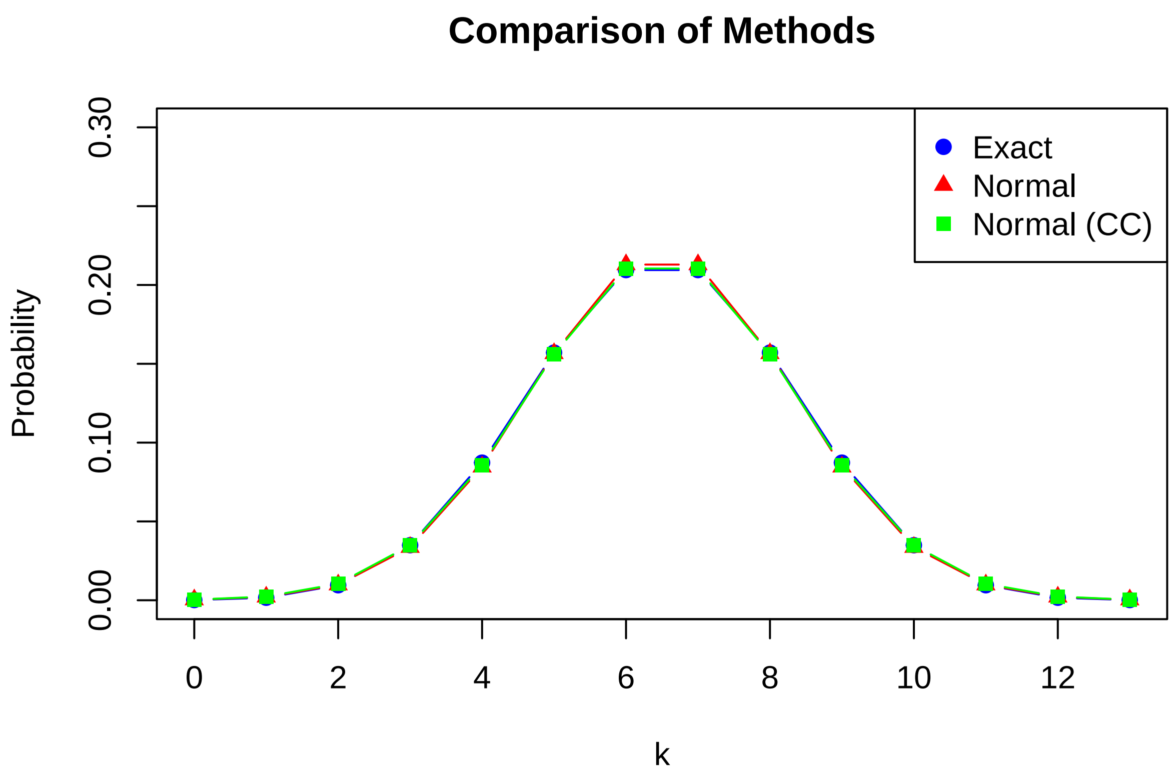
```
## the normal approximation with continuity correction is: 0.0003854349 0.00233579 0.01047731 0.0347960
```

```r
results <- data.frame(k = k,Exact = exact_probability,Normal = normal_probability,normal_probability_co
print(results)
```

```
##     k        Exact       Normal normal_probability_continuity_correction
## 1   0 0.0001220703 0.0003327011                             0.0003854349
## 2   1 0.0015869141 0.0021077965                             0.0023357901
## 3   2 0.0095214844 0.0098168926                             0.0104773066
## 4   3 0.0349121094 0.0336116880                             0.0347960244
## 5   4 0.0872802734 0.0846013976                             0.0855825818
## 6   5 0.1571044922 0.1565437309                             0.1559211244
## 7   6 0.2094726562 0.2129436782                             0.2104501290
## 8   7 0.2094726562 0.2129436782                             0.2104501290
## 9   8 0.1571044922 0.1565437309                             0.1559211244
## 10  9 0.0872802734 0.0846013976                             0.0855825818
## 11 10 0.0349121094 0.0336116880                             0.0347960244
## 12 11 0.0095214844 0.0098168926                             0.0104773066
## 13 12 0.0015869141 0.0021077965                             0.0023357901
## 14 13 0.0001220703 0.0003327011                             0.0003854349
```

1(d)

```r
plot(k, exact_probability, type = "b", col = "blue", pch = 19, xlab = "k", ylim=c(0,0.3),
     ylab = "Probability", main = "Comparison of Methods")
lines(k, normal_probability, type = "b", col = "red", pch = 17)
lines(k, normal_probability_continuity_correction, type = "b", col = "green", pch = 15)
legend("topright", legend = c("Exact", "Normal", "Normal (CC)"),
       col = c("blue", "red", "green"), pch = c(19, 17, 15))
```




```r
print("describe:`Exact values are standard.")
```

```
## [1] "describe:`Exact values are standard."
```

```r
print("`The normal distribution with continuity correction is closer to the exact value.")
```

```
## [1] "`The normal distribution with continuity correction is closer to the exact value."
```

```r
print("`The normal distribution deviation is more obvious")
```

```
## [1] "`The normal distribution deviation is more obvious"
```

2(a)

```r
n <- 13
p_0<- c(0.05, 0.1, 0.3)
k<-0:n
results_all <- data.frame()
for (p_0 in p_0) {
  exact_probability <- dbinom(k, size = n, prob = p_0)
  expected_value <- n * p_0
  sigma <- sqrt(n * p_0 * (1 - p_0))
  normal_probability <- dnorm(k, mean = expected_value, sd = sigma)
  normal_probability_continuity_correction <- pnorm(k + 0.5,mean = expected_value, sd = sigma) - pnorm(
  temp <- data.frame(
    k = k,
    p = rep(p_0, length(k)),
    Exact = exact_probability,
    Normal = normal_probability,
    normal_probability_continuity_correction = normal_probability_continuity_correction
  )
```

```r
  results_all <- rbind(results_all, temp)
}
results_all
```

```
##     k    p        Exact       Normal normal_probability_continuity_correction
## 1   0 0.05 5.133421e-01 3.605929e-01                             3.526362e-01
## 2   1 0.05 3.512341e-01 4.597416e-01                             4.359959e-01
## 3   2 0.05 1.109160e-01 1.160645e-01                             1.304165e-01
## 4   3 0.05 2.140485e-02 5.801955e-03                             9.136507e-03
## 5   4 0.05 2.816427e-03 5.742993e-05                             1.429842e-04
## 6   5 0.05 2.668194e-04 1.125619e-07                             4.803295e-07
## 7   6 0.05 1.872417e-05 4.368511e-11                             3.371895e-10
## 8   7 0.05 9.854826e-07 3.357103e-15                             4.862777e-14
## 9   8 0.05 3.890063e-08 5.108401e-20                             0.000000e+00
## 10  9 0.05 1.137445e-09 1.539198e-25                             0.000000e+00
## 11 10 0.05 2.394622e-11 9.183176e-32                             0.000000e+00
## 12 11 0.05 3.437256e-13 1.084877e-38                             0.000000e+00
## 13 12 0.05 3.015137e-15 2.537799e-46                             0.000000e+00
## 14 13 0.05 1.220703e-17 1.175501e-54                             0.000000e+00
## 15  0 0.10 2.541866e-01 1.791266e-01                             1.817251e-01
## 16  1 0.10 3.671584e-01 3.549061e-01                             3.435750e-01
## 17  2 0.10 2.447723e-01 2.991404e-01                             2.930250e-01
## 18  3 0.10 9.972203e-02 1.072618e-01                             1.126479e-01
## 19  4 0.10 2.770056e-02 1.636149e-02                             1.943461e-02
## 20  5 0.10 5.540113e-03 1.061717e-03                             1.494589e-03
## 21  6 0.10 8.207575e-04 2.930913e-05                             5.084420e-05
## 22  7 0.10 9.119528e-05 3.441956e-07                             7.595573e-07
## 23  8 0.10 7.599606e-06 1.719554e-09                             4.951501e-09
## 24  9 0.10 4.691115e-07 3.654553e-12                             1.401335e-11
## 25 10 0.10 2.084940e-08 3.304160e-15                             1.720846e-14
## 26 11 0.10 6.318000e-10 1.270855e-18                             0.000000e+00
## 27 12 0.10 1.170000e-11 2.079404e-22                             0.000000e+00
## 28 13 0.10 1.000000e-13 1.447405e-26                             0.000000e+00
## 29  0 0.30 9.688901e-03 1.489390e-02                             1.593358e-02
## 30  1 0.30 5.398102e-02 5.174733e-02                             5.336898e-02
## 31  2 0.30 1.388083e-01 1.246480e-01                             1.252339e-01
## 32  3 0.30 2.181274e-01 2.081615e-01                             2.059460e-01
## 33  4 0.30 2.337079e-01 2.410090e-01                             2.373937e-01
## 34  5 0.30 1.802890e-01 1.934568e-01                             1.918202e-01
## 35  6 0.30 1.030223e-01 1.076595e-01                             1.086403e-01
## 36  7 0.30 4.415240e-02 4.153734e-02                             4.311840e-02
## 37  8 0.30 1.419184e-02 1.111073e-02                             1.198829e-02
## 38  9 0.30 3.379010e-03 2.060461e-03                             2.333876e-03
## 39 10 0.30 5.792589e-04 2.649134e-04                             3.179743e-04
## 40 11 0.30 6.770558e-05 2.361354e-05                             3.030002e-05
## 41 12 0.30 4.836113e-06 1.459272e-06                             2.018181e-06
## 42 13 0.30 1.594323e-07 6.252146e-08                             9.390068e-08
```

2(b)

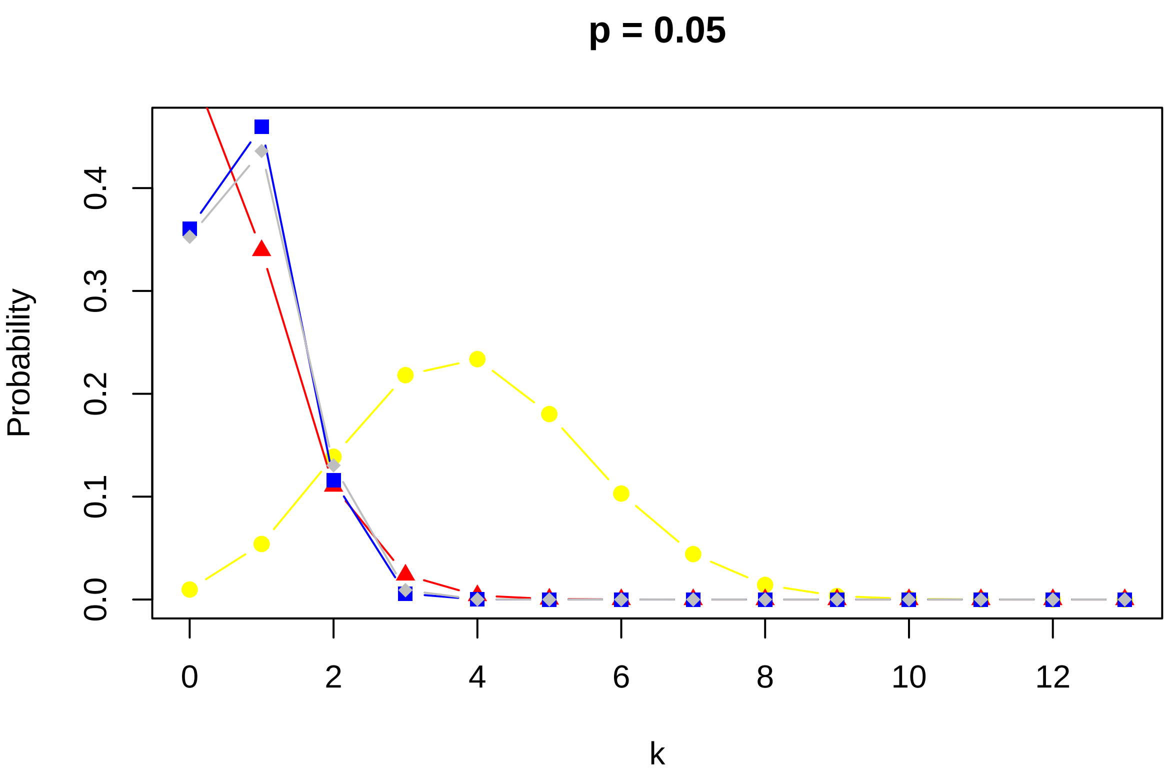
```r
n <- 13
p_0<- c(0.05, 0.1, 0.3)
k<-0:n
```

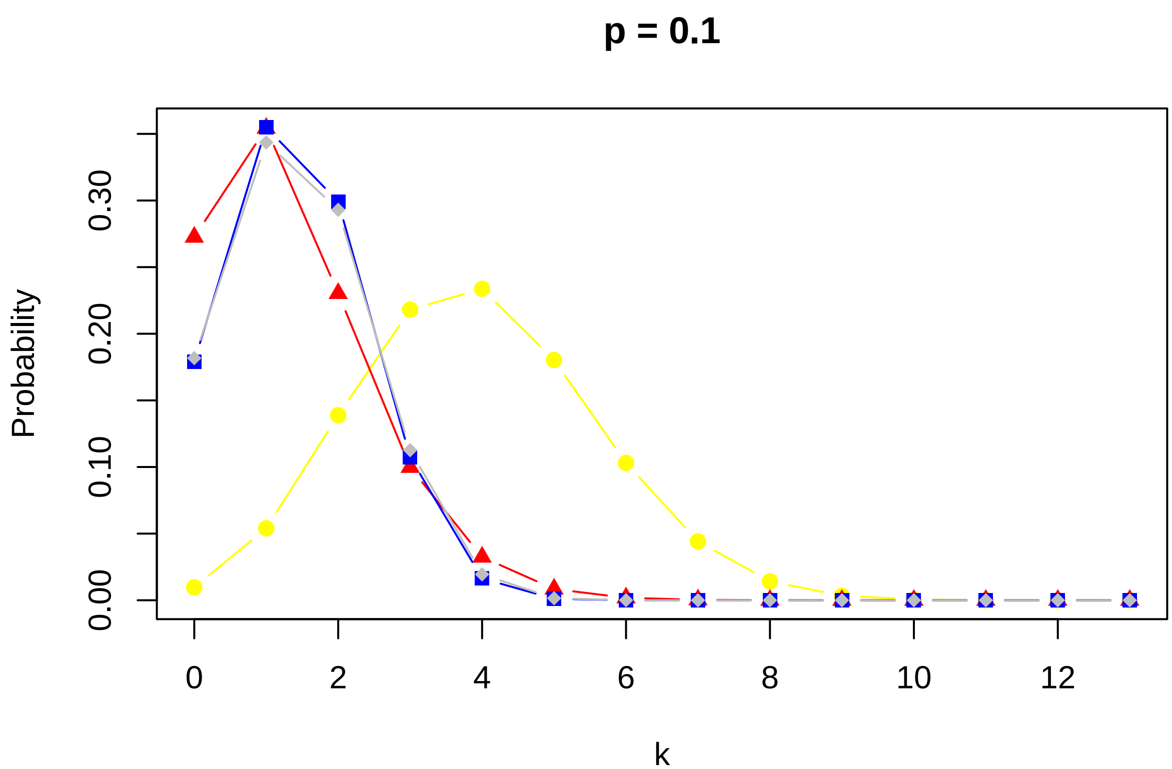
```r
results_another <- data.frame()
for (p_0 in p_0){
  lambda <- n * p_0
  poisson_probability <- dpois(k, lambda = lambda)
  expected_value <- n * p_0
  sigma <- sqrt(n * p_0 * (1 - p_0))
  normal_probability <- dnorm(k, mean = expected_value, sd = sigma)
  normal_probability_continuity_correction <- pnorm(k + 0.5,mean = expected_value, sd = sigma) - pnorm(
   temp <- data.frame(
    k = k,
    p = rep(p_0, length(k)),
    Exact = exact_probability,
    Normal = normal_probability,
    normal_probability_continuity_correction = normal_probability_continuity_correction
  )
  results_another<- rbind(results_another,temp)
}
results_another
```

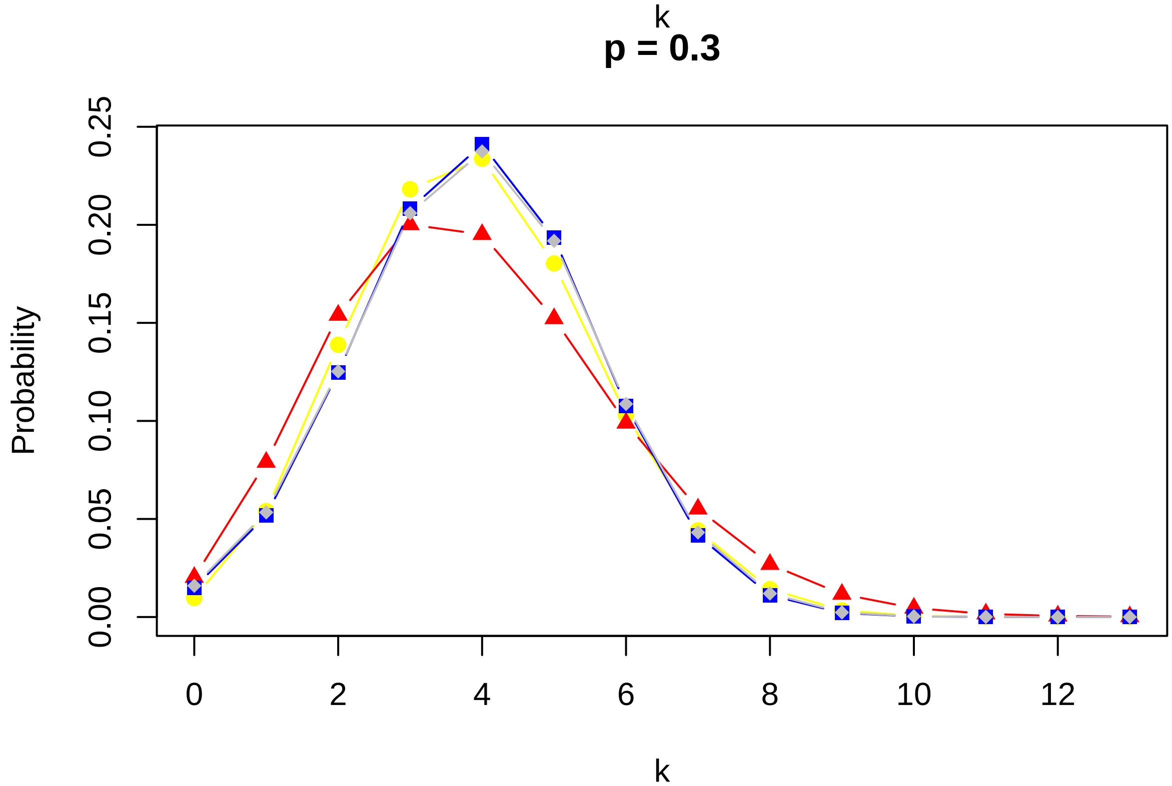
```
##     k    p        Exact       Normal normal_probability_continuity_correction
## 1   0 0.05 9.688901e-03 3.605929e-01                             3.526362e-01
## 2   1 0.05 5.398102e-02 4.597416e-01                             4.359959e-01
## 3   2 0.05 1.388083e-01 1.160645e-01                             1.304165e-01
## 4   3 0.05 2.181274e-01 5.801955e-03                             9.136507e-03
## 5   4 0.05 2.337079e-01 5.742993e-05                             1.429842e-04
## 6   5 0.05 1.802890e-01 1.125619e-07                             4.803295e-07
## 7   6 0.05 1.030223e-01 4.368511e-11                             3.371895e-10
## 8   7 0.05 4.415240e-02 3.357103e-15                             4.862777e-14
## 9   8 0.05 1.419184e-02 5.108401e-20                             0.000000e+00
## 10  9 0.05 3.379010e-03 1.539198e-25                             0.000000e+00
## 11 10 0.05 5.792589e-04 9.183176e-32                             0.000000e+00
## 12 11 0.05 6.770558e-05 1.084877e-38                             0.000000e+00
## 13 12 0.05 4.836113e-06 2.537799e-46                             0.000000e+00
## 14 13 0.05 1.594323e-07 1.175501e-54                             0.000000e+00
## 15  0 0.10 9.688901e-03 1.791266e-01                             1.817251e-01
## 16  1 0.10 5.398102e-02 3.549061e-01                             3.435750e-01
## 17  2 0.10 1.388083e-01 2.991404e-01                             2.930250e-01
## 18  3 0.10 2.181274e-01 1.072618e-01                             1.126479e-01
## 19  4 0.10 2.337079e-01 1.636149e-02                             1.943461e-02
## 20  5 0.10 1.802890e-01 1.061717e-03                             1.494589e-03
## 21  6 0.10 1.030223e-01 2.930913e-05                             5.084420e-05
## 22  7 0.10 4.415240e-02 3.441956e-07                             7.595573e-07
## 23  8 0.10 1.419184e-02 1.719554e-09                             4.951501e-09
## 24  9 0.10 3.379010e-03 3.654553e-12                             1.401335e-11
## 25 10 0.10 5.792589e-04 3.304160e-15                             1.720846e-14
## 26 11 0.10 6.770558e-05 1.270855e-18                             0.000000e+00
## 27 12 0.10 4.836113e-06 2.079404e-22                             0.000000e+00
## 28 13 0.10 1.594323e-07 1.447405e-26                             0.000000e+00
## 29  0 0.30 9.688901e-03 1.489390e-02                             1.593358e-02
## 30  1 0.30 5.398102e-02 5.174733e-02                             5.336898e-02
## 31  2 0.30 1.388083e-01 1.246480e-01                             1.252339e-01
## 32  3 0.30 2.181274e-01 2.081615e-01                             2.059460e-01
## 33  4 0.30 2.337079e-01 2.410090e-01                             2.373937e-01
## 34  5 0.30 1.802890e-01 1.934568e-01                             1.918202e-01
```

```
## 35   6 0.30 1.030223e-01 1.076595e-01                                 1.086403e-01
## 36   7 0.30 4.415240e-02 4.153734e-02                                 4.311840e-02
## 37   8 0.30 1.419184e-02 1.111073e-02                                 1.198829e-02
## 38   9 0.30 3.379010e-03 2.060461e-03                                 2.333876e-03
## 39  10 0.30 5.792589e-04 2.649134e-04                                 3.179743e-04
## 40  11 0.30 6.770558e-05 2.361354e-05                                 3.030002e-05
## 41  12 0.30 4.836113e-06 1.459272e-06                                 2.018181e-06
## 42  13 0.30 1.594323e-07 6.252146e-08                                 9.390068e-08
```

```
#graph
p_0<- c(0.05, 0.1, 0.3)
for (p_0 in p_0){
  lambda <- n * p_0
  poisson_probability <- dpois(k, lambda = lambda)
  expected_value <- n * p_0
  sigma <- sqrt(n * p_0 * (1 - p_0))
  normal_probability <- dnorm(k, mean = expected_value, sd = sigma)
  normal_probability_continuity_correction <- pnorm(k + 0.5,mean = expected_value, sd = sigma) - pnorm(
  plot(k, exact_probability, type = "b", pch = 19, col = "yellow", ylim = c(0, max(normal_probability))
  lines(k, poisson_probability, type = "b", pch = 17, col = "red")
  lines(k, normal_probability, type = "b", pch = 15, col = "blue")
  lines(k, normal_probability_continuity_correction, type = "b", pch = 18, col = "gray")
}
```








```
#compare
print("Poisson distribution is approximately suitable for cases where p is small (e.g. p = 0.05, 0.1)")
```

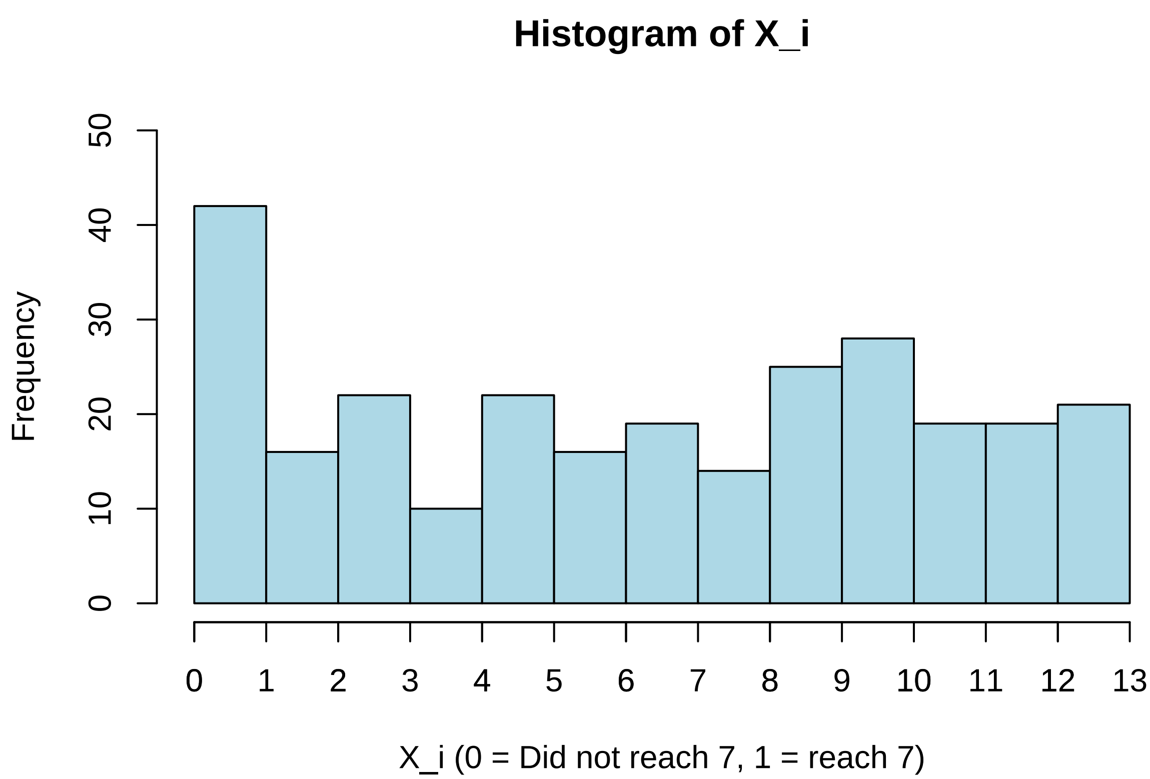
```
## [1] "Poisson distribution is approximately suitable for cases where p is small (e.g. p = 0.05, 0.1)"
```

3(a)

```r
#i)
n<-13
at_least<-7
P_Xi <- function(p) {
  sapply(p, function(prob) sum(dbinom(at_least:n, size = n, prob = prob)))
}
P_Zi <- integrate(function(p) P_Xi(p), lower = 0, upper = 1)$value
cat("P(Z_i = 1):",P_Zi,"\n")
```

```
## P(Z_i = 1): 0.5
```

```r
#ii)
set.seed(123321)
n_students <- 273
n <- 13
at_least <- 7
p_i <- runif(n_students, min = 0, max = 1)
X_i <- rbinom(n_students, size = n, prob = p_i)
Z_i <- ifelse(X_i >= at_least, 1, 0)
#histogram
hist(X_i, breaks = seq(0,13,by=1), col = "lightblue",
     xlab = "X_i (0 = Did not reach 7, 1 = reach 7)",
     main = "Histogram of X_i",
     ylim=c(0,50))
axis(1, at=0:13, labels=0:13)
```

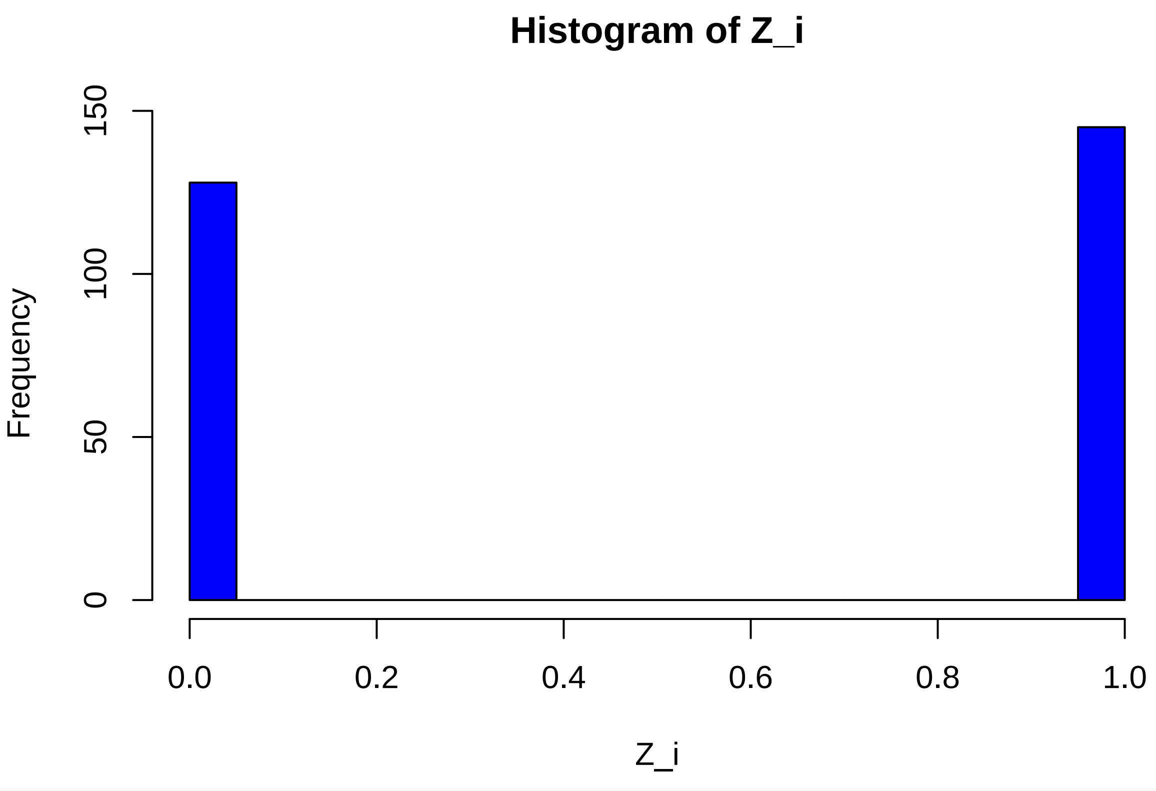



```r
#at least 8
num_students_correct_8 <- sum(X_i >= 8)
cat("Number of students who answered at least 8 questions correctly:", num_students_correct_8, "\n")
```

```
## Number of students who answered at least 8 questions correctly: 126
```

3(b)

```r
set.seed(123321)
n_students <- 273
n <- 13
at_least <- 7
simulate_Zi <- function(alpha, beta) {
  p_i <- rbeta(n_students, shape1 = alpha, shape2 = beta)
  X_i <- rbinom(n_students, size = n, prob = p_i)
  Z_i <- ifelse(X_i >= at_least, 1, 0)
  return(Z_i)
}
hist(Z_i, breaks=20,col="blue",
     xlab="Z_i",
     main="Histogram of Z_i")
```




```r
Z_beta_0_5_0_5 <- simulate_Zi(0.5, 0.5)
cat("Beta(0.5, 0.5):",mean(Z_beta_0_5_0_5),"\n")
```

```
## Beta(0.5, 0.5): 0.4652015
```

```r
Z_beta_5_5 <- simulate_Zi(5, 5)
cat("Beta(5, 5) :", mean(Z_beta_5_5), "\n")
```

```
## Beta(5, 5) : 0.5604396
```

```r
Z_beta_7_2 <- simulate_Zi(7, 2)
cat("Beta(7, 2):", mean(Z_beta_7_2), "\n")
```

```
## Beta(7, 2): 0.9267399
```

```r
#graph
par(mfrow = c(2, 2))
hist(Z_beta_0_5_0_5, breaks = 20 ,col = "lightyellow",
     main = "Beta(0.5, 0.5)", xlab = "Z_i")
hist(Z_beta_5_5, breaks = 20, col = "lightblue",
     main = "Beta(5, 5)", xlab = "Z_i")
hist(Z_beta_7_2, breaks = 20, col = "lightpink",
     main = "Beta(7, 2)", xlab = "Z_i")
print("come up with the idea: Because the graph is most similar to a when alpha is equal to beta is equa
```

```
## [1] "come up with the idea: Because the graph is most similar to a when alpha is equal to beta is eq
```