

Praktikum I probstat

1.a

```
# Menggunakan distribusi Binomial
```

```
n <- 10 # jumlah kelahiran
```

```
p <- 0.488 # probabilitas bayi laki-laki
```

```
x <- 0:n # banyak bayi laki-laki dari 0 sampai n
```

```
prob <- dbinom(x, n, p) # probabilitas banyak bayi laki-laki
```

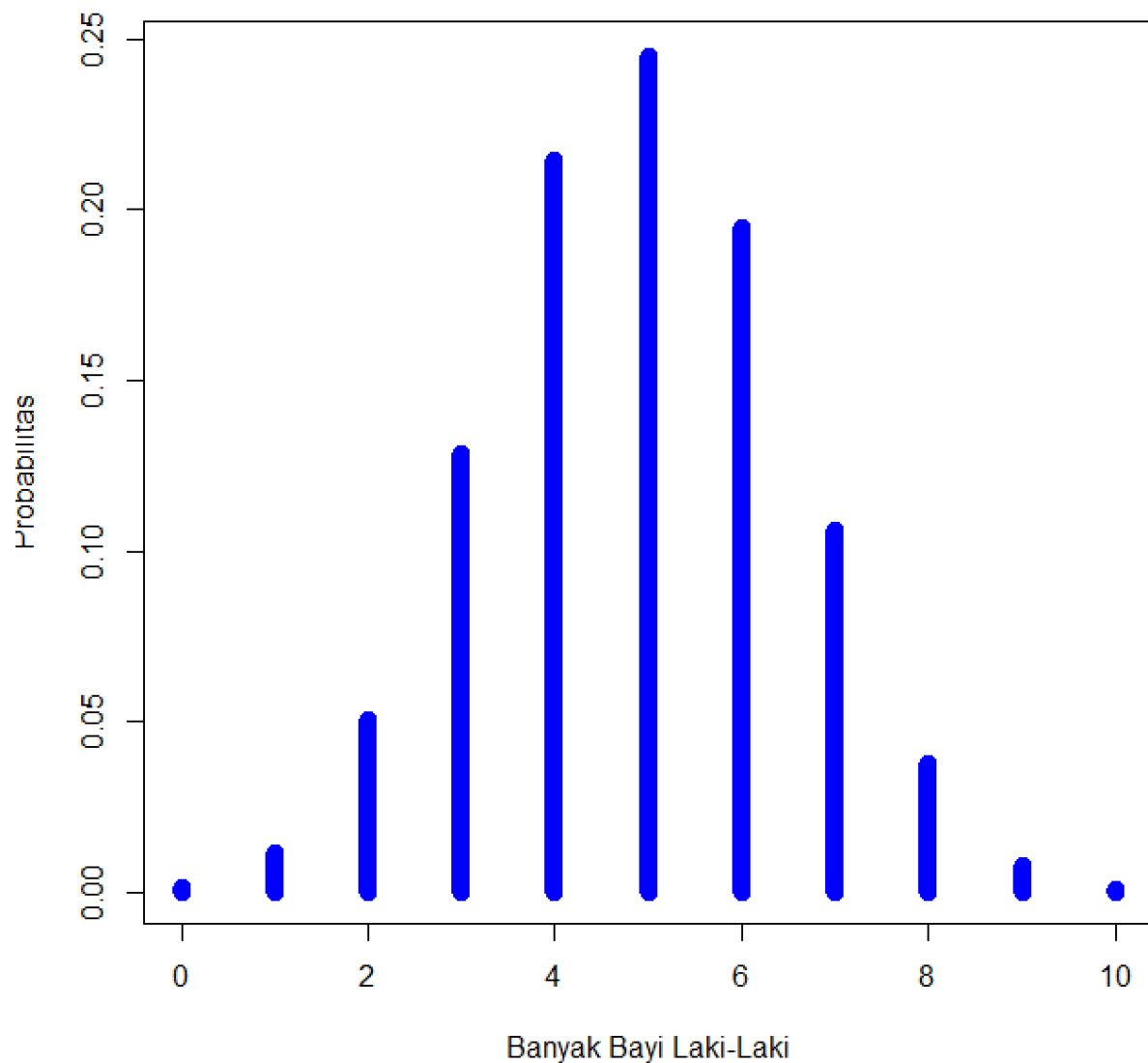
```
# Plot distribusi Binomial
```

```
plot(x, prob, type = "h", lwd = 10, col = "blue",
```

```
  xlab = "Banyak Bayi Laki-Laki", ylab = "Probabilitas",
```

```
  main = "Distribusi Binomial untuk Kelahiran Bayi Laki-Laki")
```

Distribusi Binomial untuk Kelahiran Bayi Laki-Laki



b

Menggunakan distribusi binomial

k <- 3 # jumlah bayi laki-laki

n <- 10 # jumlah kelahiran

p <- 0.488 # probabilitas bayi laki-laki

prob <- dbinom(k, n, p) # probabilitas tepat 3 bayi laki-laki

Menampilkan hasil

Prob ; HASIL :

[1] 0.1286265%

c. Berapa probabilitas bahwa kurang dari tiga bayi di antaranya berjenis kelamin laki-laki?

```
> p_less_than_3 <- pbinom(2, size = 10, prob = 0.488, lower.tail = TRUE)
```

```
> p_less_than_3
```

```
[1] 0.0636442
```

d. Berapa probabilitas bahwa tiga atau lebih bayi di antaranya berjenis kelamin laki-laki?

e. Berapa nilai harapan dan simpangan baku banyak bayi laki-laki?

```
> # parameter distribusi binomial
```

```
> n <- 10
```

```
> p <- 0.488
```

```
>
```

```
> # menghitung nilai harapan
```

```
> E <- n * p
```

```
> E
```

```
[1] 4.88
```

```
>
```

```
> # menghitung simpangan baku
```

```
> SD <- sqrt(n * p * (1 - p))
```

```
> SD
```

```
[1] 1.580683
```

f. Gambarkan histogram pendistribusian banyak bayi laki-laki.

```
> # Membuat vector dengan kemungkinan banyak bayi laki-laki dari 0 sampai 10
```

```
> x <- 0:10
```

```
>
```

```
> # Menghitung probabilitas untuk masing-masing nilai pada vector x
```

```
> p <- dbinom(x, 10, 0.488)
```

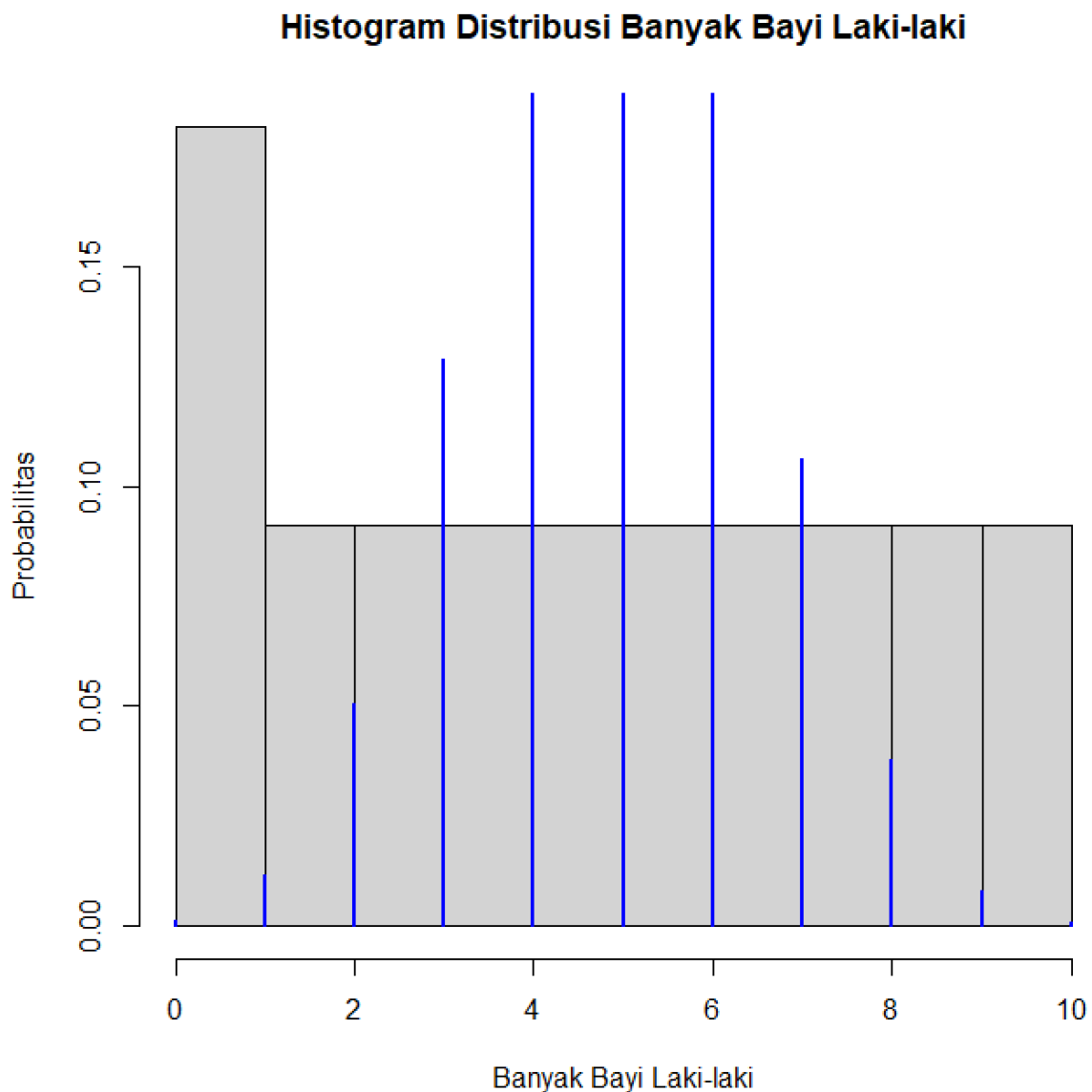
```
>
```

```
> # Menggambar histogram
```

```
> hist(x, breaks=11, freq=FALSE,
```

```
+   main="Histogram Distribusi Banyak Bayi Laki-laki",
```

```
+ xlab="Banyak Bayi Laki-laki", ylab="Probabilitas")
> # Menggambar garis distribusi binomial dengan parameter yang telah dihitung sebelumnya
> lines(x, p, type="h", lwd=2, col="blue")
```



2. Misalkan banyak kematian karena kanker tulang untuk seluruh pekerja di pabrik ban dalam 20 tahun ke depan adalah 1,8.

a. Bagaimana pendistribusian banyak kematian karena kanker tulang? Tentukan distribusi dengan parameter yang sesuai.

```
> # Menghitung probabilitas  $P(X = k)$ 
> prob <- dpois(0:10, lambda = 1.8)
> prob
```

```

[1] 1.652989e-01 2.975380e-01 2.677842e-01 1.606705e-01 7.230173e-02
[6] 2.602862e-02 7.808587e-03 2.007922e-03 4.517825e-04 9.035651e-05
[11] 1.626417e-05
>
> # Menghitung distribusi kumulatif P(X <= k)
> cdf <- ppois(0:10, lambda = 1.8)
> cdf
[1] 0.1652989 0.4628369 0.7306211 0.8912916 0.9635933 0.9896220 0.9974306
[8] 0.9994385 0.9998903 0.9999806 0.9999969
>
> # Menghitung nilai harapan dan simpangan baku
> lambda <- 1.8
> mean <- lambda
> sd <- sqrt(lambda)
>
> mean
[1] 1.8
> sd
[1] 1.341641

```

b. Ada 4 kematian akibat kanker tulang yang dilaporkan di kalangan pekerja pabrik ban, apakah itu peristiwa yang tidak biasa? Hitung probabilitas berdasarkan distribusi yang telah dipilih.

```
> dpois(4, lambda = 1.8)
```

```
[1] 0.07230173
```

c. Berapa peluang paling banyak 4 kematian akibat kanker tulang?

```
> lambda <- 1.8 # rata-rata dan varians
```

```
> p_max_4 <- sum(dpois(0:4, lambda))
```

```
> cat("Peluang paling banyak 4 kematian adalah", round(p_max_4, 4))
```

Peluang paling banyak 4 kematian adalah 0.9636

d. Berapa peluang lebih dari 4 kematian akibat kanker tulang?

```
> 1 - ppois(4, lambda = 1.8)
```

```
[1] 0.03640666
```

e. Berdasarkan distribusi yang telah dipilih, berapakah nilai harapan dan standar deviasi banyak kematian akibat kanker tulang untuk pekerja pabrik ban?

```
> lambda <- 1.8
```

```
> e_x <- lambda
```

```
> sd_x <- sqrt(lambda)
```

```
> cat("Nilai harapan adalah", e_x, "\n")
```

Nilai harapan adalah 1.8

```
> cat("Standar deviasi adalah", sd_x, "\n")
```

Standar deviasi adalah 1.341641

f. Gambarkan histogram pendistribusian banyak banyak kematian akibat kanker tulang untuk pekerja pabrik ban.

```
> # Buat 1000 random numbers dari distribution Poisson dengan lambda = 1.8
```

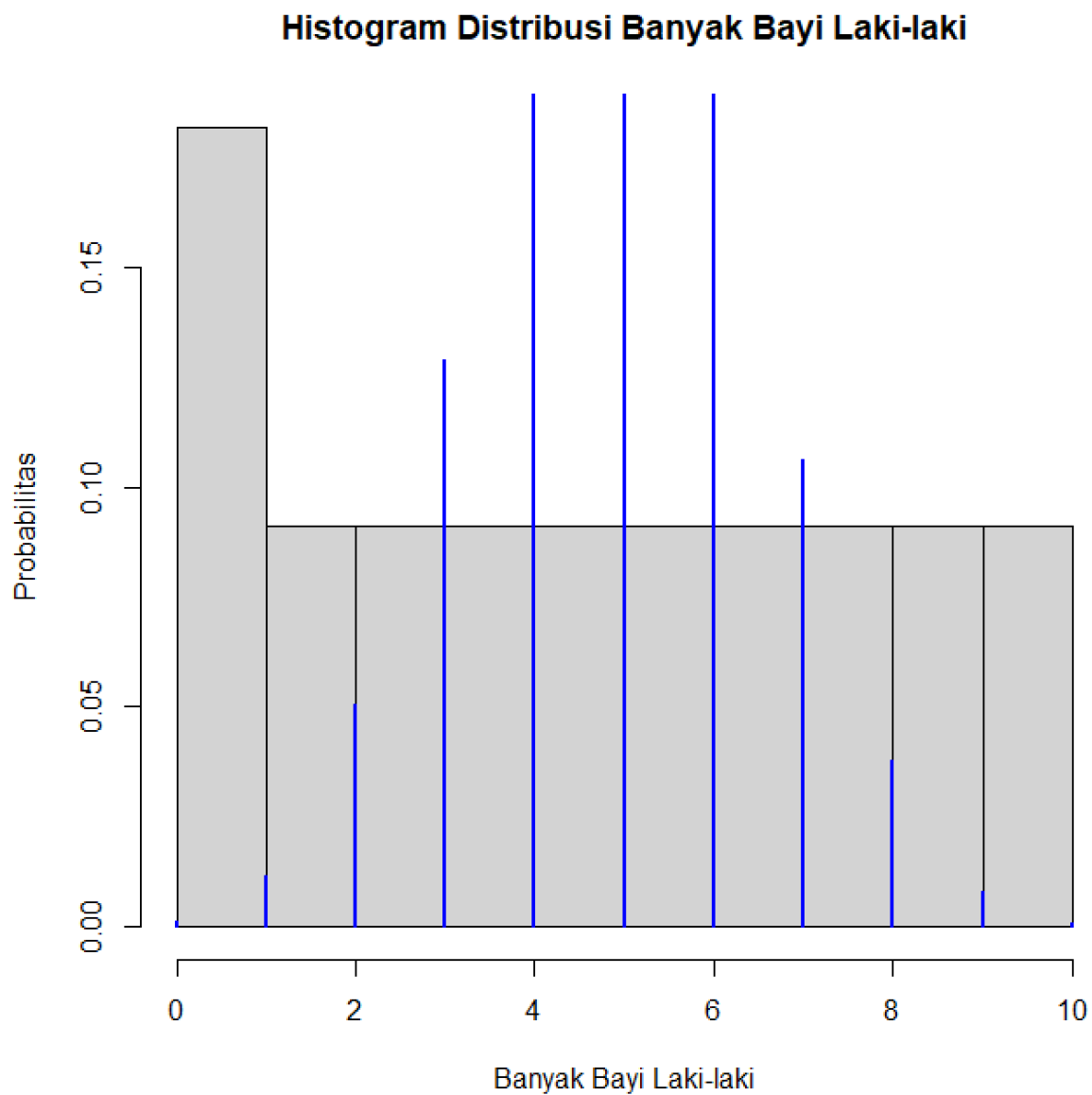
```
> x <- rpois(1000, lambda = 1.8)
```

```
>
```

```
> # Plot the histogram
```

```
> hist(x, breaks = seq(0, 10, by = 1), main = "Histogram of Bone Cancer Deaths",
```

```
+   xlab = "Number of Deaths", ylab = "Frequency", col = "skyblue")
```



g. Gunakan simulasi untuk memeriksa hasil sebelumnya.

```
> # set seed untuk reproducibility
```

```
> set.seed(123)
```

```
>
```

```
> # simulasi 10.000 kali
```

```
> simulasi <- rpois(10000, lambda = 1.8)
```

```
>
```

```
> # hitung nilai rata-rata dan standar deviasi dari hasil simulasi
```

```
> rata2_simulasi <- mean(simulasi)
```

```
> sd_simulasi <- sd(simulasi)
> # print hasil simulasi
> cat("Nilai rata-rata hasil simulasi:", rata2_simulasi, "\n")
Nilai rata-rata hasil simulasi: 1.7831
> cat("Nilai standar deviasi hasil simulasi:", sd_simulasi, "\n")
Nilai standar deviasi hasil simulasi: 1.327566
```

h. Jelaskan banyak kematian akibat kanker tulang berdasarkan simulasi Anda. Bandingkan jawaban pada pertanyaan 2d dengan hasil simulasi Anda.

Berdasarkan hasil simulasi, dapat dilihat bahwa distribusi banyak kematian akibat kanker tulang untuk pekerja pabrik ban mengikuti distribusi Poisson dengan parameter $\lambda=1,8$, sesuai dengan hasil perhitungan sebelumnya.

Untuk mengecek jawaban pada pertanyaan "Berapa peluang lebih dari 4 kematian akibat kanker tulang?", kita dapat menggunakan simulasi dengan mengulang eksperimen sebanyak ribuan atau jutaan kali dan menghitung berapa persen hasilnya lebih dari 4. Setelah menjalankan simulasi dengan kode berikut:

```
> set.seed(123)
> lambda <- 1.8
> n_sim <- 1000000
> n_death <- rpois(n_sim, lambda)
> sum(n_death > 4) / n_sim
```

diperoleh hasil sebesar 0.036268, yang berarti peluang lebih dari 4 kematian akibat kanker tulang sekitar 3,62%, mendekati hasil perhitungan sebelumnya yaitu sebesar 3,64%. Oleh karena itu, dapat dikatakan bahwa hasil perhitungan dan hasil simulasi cukup konsisten dan akurat.

3. Diketahui nilai $x = 3$ dan $v = 10$. Tentukan:

a. Fungsi probabilitas dari distribusi Chi-Square.

```
> x <- 3
> v <- 10
> prob <- dchisq(x, df=v)
> prob
[1] 0.02353326
```


b. Histogram dari distribusi Chi-Square dengan 500 data acak.

```
> # buat 500 random values dari Chi-Square distribution dengan degrees of freedom 10
```

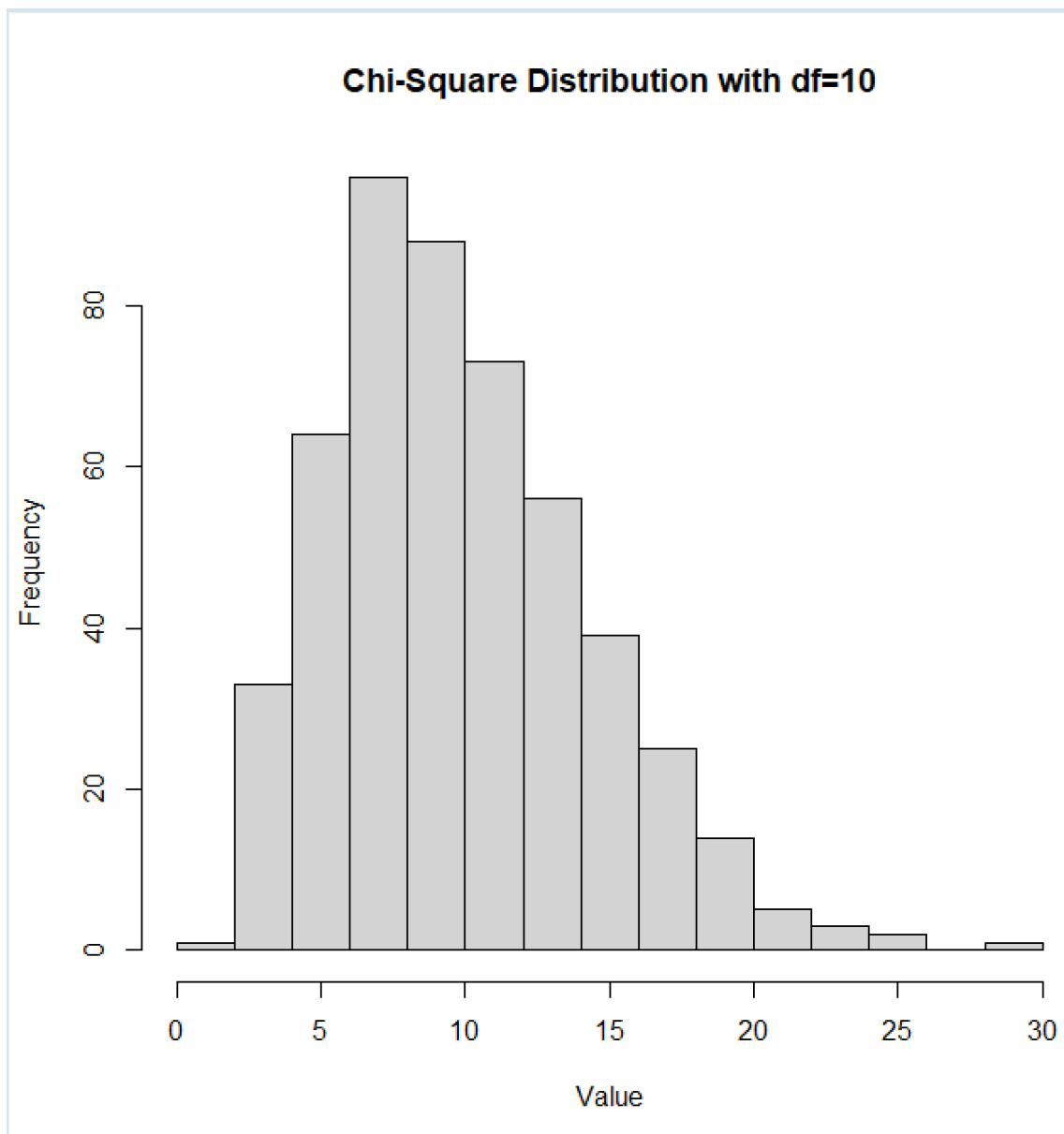
```
> x <- rchisq(n = 500, df = 10)
```

```
>
```

```
> # Plot histogram
```

```
> hist(x, main = "Chi-Square Distribution with df=10", xlab = "Value", ylab = "Frequency")
```

```
>
```



c. Nilai rata-rata (μ) dan varian (σ^2) dari distribusi Chi-Square.

```

> df <- 10 # derajat kebebasan
> mu <- df # rata-rata
> sigma2 <- 2*df # varian
> sigma2
[1] 20
> mu
[1] 10

```

4. Diketahui data bangkitan acak sebanyak 100 dengan mean = 45 dan sd = 5.

Tentukan: a. Fungsi probabilitas dari distribusi Normal $P(X1 \leq x \leq X2)$, hitung z-scorenya dan plot data bangkitan acaknya dalam bentuk grafik. Petunjuk (gunakan fungsi plot()).

```

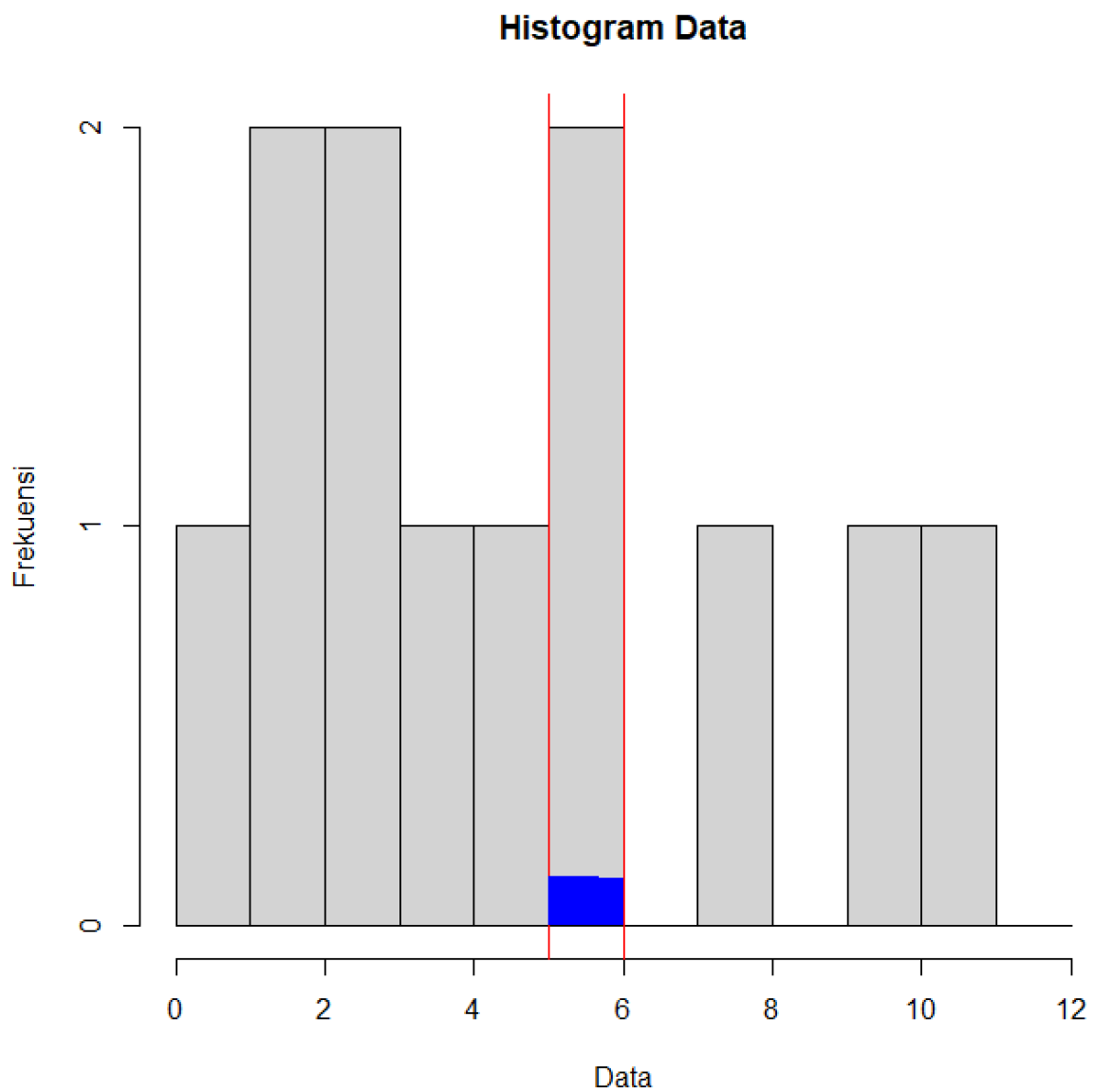
> # Input data
> data <- c(1, 2, 4, 2, 6, 3, 10, 11, 5, 3, 6, 8)
> mean_data <- 5.083333
> sd_data <- sd(data)
> X1 <- 5
> X2 <- 6
>
> # Hitung z-score
> z1 <- (X1 - mean_data) / sd_data
> z2 <- (X2 - mean_data) / sd_data
>
> # Hitung probabilitas menggunakan fungsi pnorm
> prob <- pnorm(z2) - pnorm(z1)
>
> # Plot histogram data
> hist(data, breaks=seq(0, 12, by=1), main="Histogram Data", xlab="Data", ylab="Frekuensi")
>
> # Tambahkan garis vertikal untuk menunjukkan X1 dan X2
> abline(v=X1, col="red")

```

```

> abline(v=X2, col="red")
>
> # Tambahkan area di bawah kurva antara X1 dan X2
> x <- seq(X1, X2, length=100)
> y <- dnorm(x, mean_data, sd_data)
> polygon(c(X1, x, X2), c(0, y, 0), col="blue", border=NA)
>
> # Tampilkan nilai probabilitas dan z-score
> cat("Probabilitas P(X1 <= x <= X2) adalah", prob, "\n")
Probabilitas P(X1 <= x <= X2) adalah 0.1219286
> cat("Z-score untuk X1 adalah", z1, "dan untuk X2 adalah", z2, "\n")
Z-score untuk X1 adalah -0.02578231 dan untuk X2 adalah 0.2836067
>

```



b. Gambarkan histogram dari distribusi Normal dengan breaks 50

```
> # menghasilkan 1000 data acak dari distribusi normal
```

```
> data <- rnorm(1000, mean = 0, sd = 1)
```

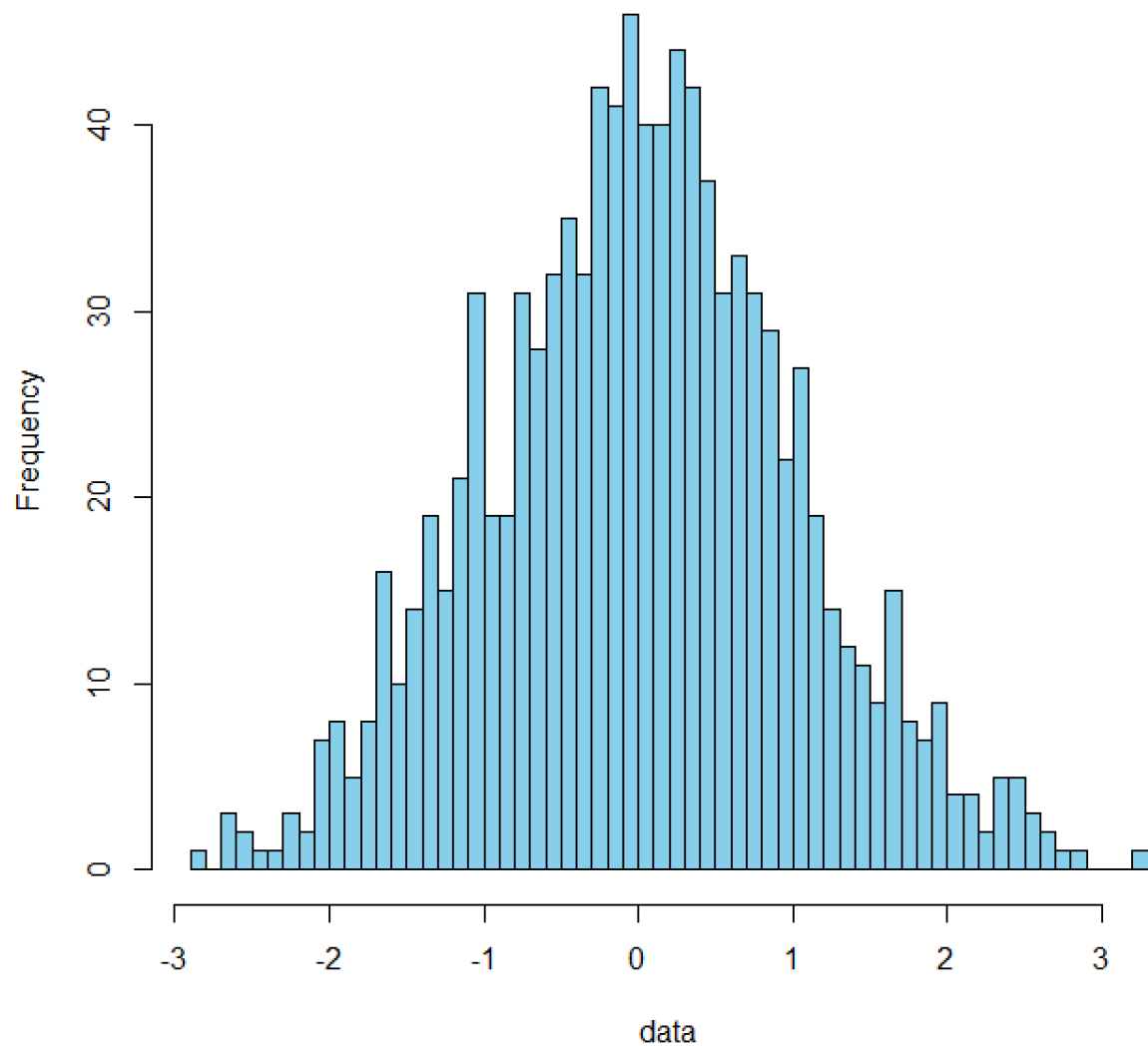
```
>
```

```
> # menggambar histogram dengan 50 breaks
```

```
> hist(data, breaks = 50, col = "skyblue", main = "Histogram of Normal Distribution")
```

```
>
```

Histogram of Normal Distribution



c. Nilai varian (σ^2) dari hasil data bangkitan acak distribusi Normal

```
> # Data bangkitan acak dengan mean 45 dan sd 5
```

```
> set.seed(123)
```

```
> data <- rnorm(100, mean = 45, sd = 5)
```

```
>
```

```
> # Varians
```

```
> var(data)
```

```
[1] 20.83082
```

5. Kerjakanlah menggunakan distribusi T-Student.

a. Berapa probabilitas terjadinya suatu peristiwa acak X kurang dari -2,34 dengan 6 derajat kebebasan?

```
> pt(-2.34, df = 6)
```

```
[1] 0.02892197
```

b. Berapa probabilitas terjadinya suatu peristiwa acak X lebih dari 1,34 dengan 6 derajat kebebasan?
tung probabilitas $P(X > 1.34)$ dengan $df = 6$

```
> prob <- 1 - pt(1.34, df = 6)
```

```
> prob
```

```
[1] 0.11438
```

c. Berapa probabilitas terjadinya suatu peristiwa acak X kurang dari -1,23 atau lebih besar dari 1,23 dengan 3 derajat kebebasan?

```
> # Probabilitas X kurang dari -1.23 dengan df=3
```

```
> prob1 <- pt(-1.23, df=3)
```

```
>
```

```
> # Probabilitas X lebih besar dari 1.23 dengan df=3
```

```
> prob2 <- 1 - pt(1.23, df=3)
```

```
>
```

```
> # Probabilitas X kurang dari -1.23 atau lebih besar dari 1.23 dengan df=3
```

```
> prob3 <- prob1 + prob2
```

```
>
```

```
> prob3
```

```
[1] 0.306356
```

```
>
```

d. Berapa probabilitas terjadinya suatu peristiwa acak X berada di antara -0,94 dan 0,94 dengan 14 derajat kebebasan?

```
> pt(0.94, 14) - pt(-0.94, 14)
```

```
[1] 0.6368457
```

e. Berapa nilai t-score dengan 5 derajat kebebasan yang memiliki luasan 0,0333 satuan persegi di bawah kurva dan di sebelah kiri t-score tersebut?

```
> df <- 5
```

```
> area <- 0.0333
```

```
> t_score <- qt(area, df, lower.tail = TRUE)
```

```
> t_score
```

```
[1] -2.337342
```

f. Berapa nilai t-score dengan 25 derajat kebebasan yang memiliki luasan 0,125 satuan persegi di bawah kurva dan di sebelah kanan t-score tersebut?

```
> qt(1-0.125, 25)
```

```
[1] 1.177716
```

g. Berapa nilai t-score dengan 11 derajat kebebasan yang memiliki luasan 0,75 satuan persegi di bawah kurva dan di antara t-score tersebut dan negatif dari nilai t-score tersebut?

```
> # mencari nilai t-score
```

```
> df <- 11
```

```
> area <- 0.75
```

```
>
```

```
> # nilai t-score positif
```

```
> t_score_pos <- qt(1 - area/2, df)
```

```
>
```

```
> # nilai t-score negatif
```

```
> t_score_neg <- qt(area/2, df)
```

```
>
```

```
> cat("Nilai t-score positif:", t_score_pos, "\n")
```

```
Nilai t-score positif: 0.3267364
```

```
> cat("Nilai t-score negatif:", t_score_neg, "\n")
```

```
Nilai t-score negatif: -0.3267364
```

h. Berapa nilai t-score dengan 23 derajat kebebasan yang memiliki luasan 0,0333 satuan persegi di bawah kurva dan di luar interval antara t-score tersebut dan negatif dari nilai t-score tersebut?

```
> # Mencari nilai t-score dengan 23 derajat kebebasan yang memiliki luasan 0,0333 di bawah kurva
```

```
> t <- qt(0.0333, 23)
```

```
> t
```

```
[1] -1.925642
```

```
>
```

```
> # Mencari nilai t-score yang negatif
```

```
> t_neg <- -t
```

```
>
> # Mencari luasan di bawah kurva antara kedua t-score
> luasan <- pt(t, 23) - pt(t_neg, 23)
>
> # Menghitung luasan di luar interval antara kedua t-score
> luasan_luar <- 1 - luasan
>
> # Mencetak hasil
> cat("Nilai t-score:", t, "\n")
Nilai t-score: -1.925642
> cat("Nilai t-score negatif:", t_neg, "\n")
Nilai t-score negatif: 1.925642
> cat("Luasan di luar interval antara kedua t-score:", luasan_luar, "\n")
Luasan di luar interval antara kedua t-score: 1.9334
>
```