Praktikum 2 SMA

Antonius Aditya Rizky Wijaya

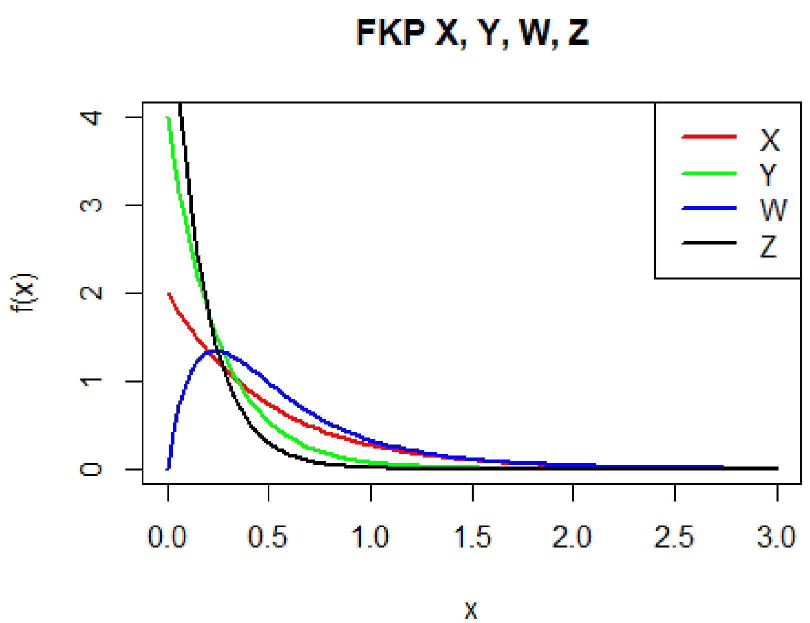
2025-08-27

NOMOR 1

> memahami bagaimana max/min mengubah bentuk distribusi

1a

```
x <- c(0:3000)/1000
fkp_X <- function(x) 2*exp(-2*x)
fkp_Y <- function(x) 4*exp(-4*x)
fkp_W <- function(x) 4*exp(-4*x) + 2*exp(-2*x) - 6*exp(-6*x)
fkp_Z <- function(x) 6*exp(-6*x)

plot(0, 0, type="n", xlim=c(0, 3), ylim=c(0, 4), xlab="x", ylab="f(x)",
main="FKP X, Y, W, Z")
curve(fkp_X, add=TRUE, col="red", lwd=2)
curve(fkp_Y, add=TRUE, col="green", lwd=2)
curve(fkp_W, add=TRUE, col="blue", lwd=2)
curve(fkp_Z, add=TRUE, col="black", lwd=2)
legend("topright", c("X", "Y", "W", "Z"), col=c("red", "green", "blue",
"black"), lwd=2)
```



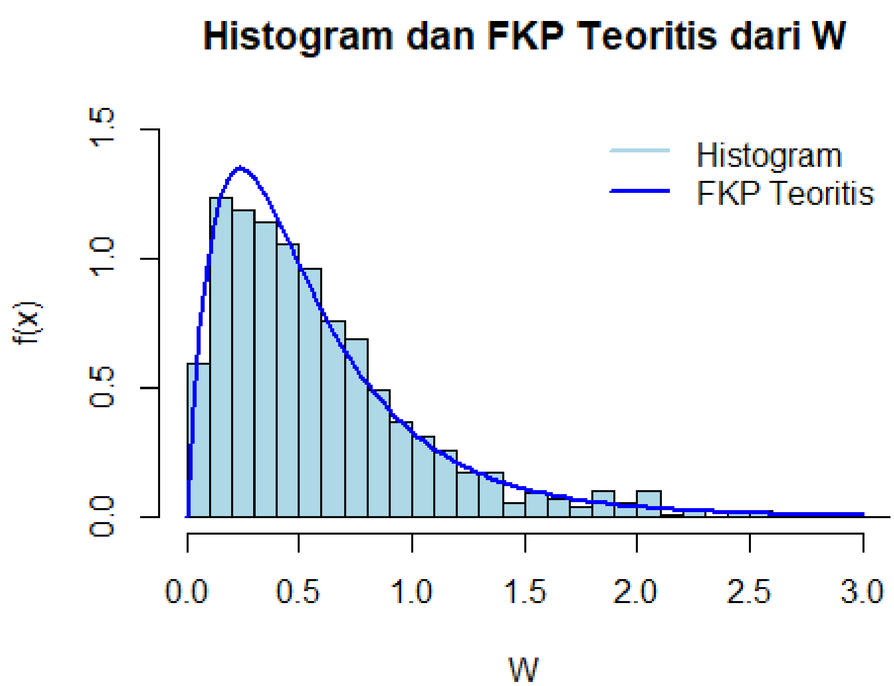
Kurva Y curam (menurun cepat) karena rate tinggi, kurva X lebih landai dari kurva Y, kurva Z paling curam, dan kurva W mulai dengan naik sedikit sebelum turun, karena max cenderung lebih besar dari min.

1b

```
rw <- function(n){
  x <- rexp(n, rate=2)
  y <- rexp(n, rate=4)
  pmax(x, y)
}
```

1c

```
set.seed(123)
w <- rw(1000)

hist(w, prob = TRUE, breaks = 30, xlab="W", ylab="f(x)", main="Histogram dan
FKP Teoritis dari W", xlim=c(0, 3), ylim=c(0, 1.5), col="lightblue")
curve(fkp_W, add=TRUE, col="blue", lwd=2)
legend("topright", legend = c("Histogram", "FKP Teoritis"), col =
c("lightblue", "blue"), lwd = 2, bty = "n")
```



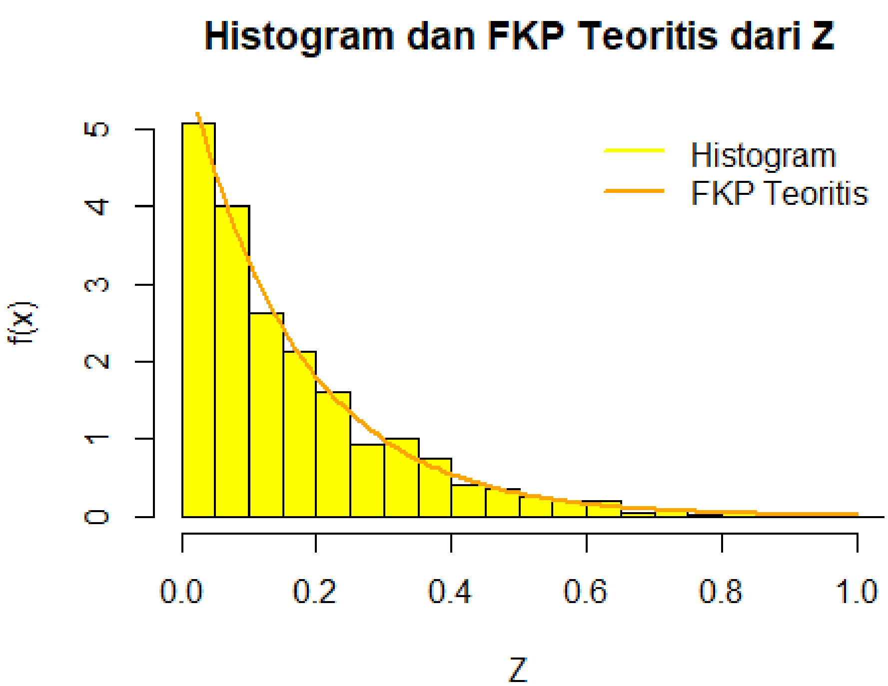
Histogram menunjukkan distribusi 1000 sampel W, dengan skala yang sama dengan FKP. Kurva biru akan overlay di atas histogram, dan terlihat cocok (histogram mendekati kurva karena $n = 1000$ besar). Nilai W cenderung antara $0 - 2$, dengan puncak sekitar $0.2 - 0.4$.

1d

```r
rz <- function(n) {
  x <- rexp(n, rate=2)
  y <- rexp(n, rate=4)
  pmin(x, y)
}
```

1e

```r
set.seed(123)
z <- rz(1000)

hist(z, prob = TRUE, breaks = 30, xlab="Z", ylab="f(x)", main="Histogram dan
FKP Teoritis dari Z", xlim=c(0, 1), ylim=c(0, 5), col="yellow")
curve(fkp_Z, add=TRUE, col="orange", lwd=2)
legend("topright", legend = c("Histogram", "FKP Teoritis"), col = c("yellow",
"orange"), lwd = 2, bty = "n")
```



Histogram menunjukkan distribusi 1000 sampel Z, yang sangat curam (banyak nilai dekat 0). Kurva jingga akan overlay di atas histogram, menurun eksponensial cepat. Nilai Z cenderung < 0.5, puncak di 0 dengan probabilitas kurang lebih 5.

NOMOR 2

menerapkan metode dugaan parameter yaitu metode momen dan metode kemungkinan maksimum

2a

Fungsi kepadatan peluang (FKP) diberikan sebagai berikut:

$$f(x;\theta) = (\theta + 1)x^{\theta}, \quad 0 < x \leq 1, \quad \theta > 1$$

Likelihood untuk n sampel independen:

$$L(\theta) = \prod_{i=1}^{n} (\theta + 1)\, x_i^{\theta} = (\theta + 1)^n \prod_{i=1}^{n} x_i^{\theta}$$

Log-likelihood:

$$l(\theta) = \ln L(\theta) = n\ln(\theta + 1) + \theta \sum_{i=1}^{n} \ln x_i$$

Turunan terhadap θ dan disamakan dengan 0 (untuk maksimum):

$$\frac{d}{d\theta} l(\theta) = \frac{n}{\theta + 1} + \sum_{i=1}^{n} \ln x_i = 0$$

Sehingga:

$$\frac{n}{\theta + 1} = -\sum_{i=1}^{n} \ln x_i$$

$$\theta + 1 = -\frac{n}{\sum_{i=1}^{n} \ln x_i}$$

Estimasi parameter $\hat{\theta}_1$ adalah:

$$\hat{\theta}_1 = -\frac{n}{\sum_{i=1}^{n} \ln x_i} - 1$$

2b

```r
x <- c(0.92, 0.79, 0.90, 0.65, 0.86, 0.47, 0.75, 0.92, 0.84)
n <- length(x)
theta_hat1 <- -(n/sum(log(x))) - 1
cat("θ1 =", theta_hat1)

## θ1 = 2.901999
```

2c

Fungsi kepekatan peluang (FKP) diberikan:

$$f(x;\theta) = (\theta + 1)x^{\theta}, \quad 0 < x \le 1, \quad \theta > 1$$

Nilai harapan teoretis:

$$E(X) = \int_0^1 x\,(\theta + 1)x^{\theta}\,dx$$

$$E(X) = (\theta + 1)\int_0^1 x^{\theta+1}\,dx = (\theta + 1)\left[\frac{x^{\theta+2}}{\theta + 2}\right]_0^1 = (\theta + 1)\cdot\frac{1}{\theta + 2} = \frac{\theta + 1}{\theta + 2}$$

Samakan dengan rata-rata sampel $\bar{x}$:

$$E(X) = \frac{\hat{\theta}_2 + 1}{\hat{\theta}_2 + 2} = \bar{x}$$

Selesaikan untuk $\hat{\theta}_2$:

$$\hat{\theta}_2 + 1 = \bar{x}(\hat{\theta}_2 + 2)$$

$$\hat{\theta}_2 + 1 = \bar{x}\hat{\theta}_2 + 2\bar{x}$$

$$\hat{\theta}_2 - \bar{x}\hat{\theta}_2 = 2\bar{x} - 1$$

$$\hat{\theta}_2(1 - \bar{x}) = 2\bar{x} - 1$$

$$\hat{\theta}_2 = \frac{2\bar{x} - 1}{1 - \bar{x}}$$

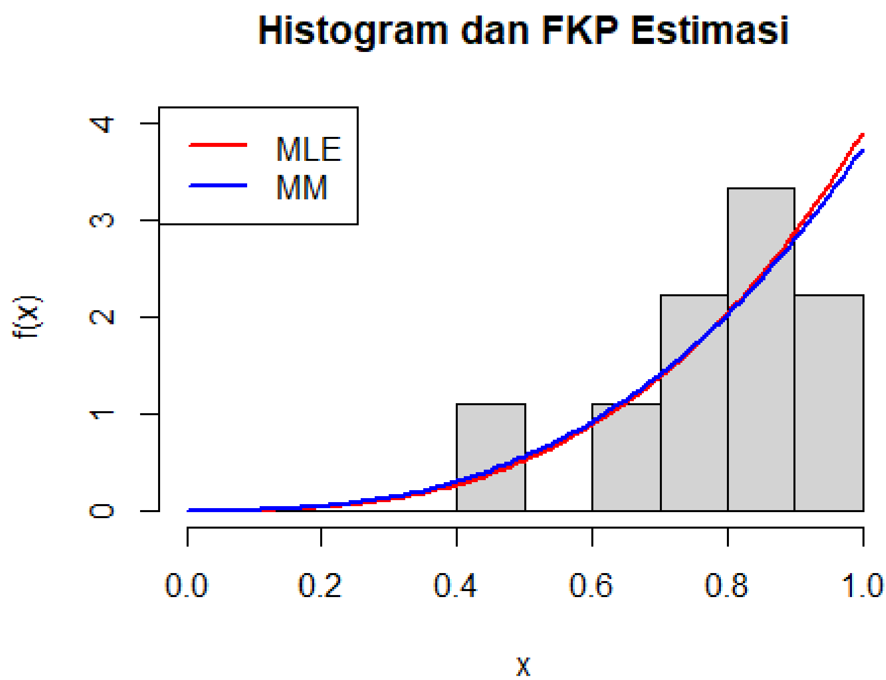
2d

```
xbar <- mean(x)

theta_hat2 <- (2*xbar - 1) / (1 - xbar)
cat("θ2 =", theta_hat2)

## θ2 = 2.736842
```

2e

```
fkp_theta <- function(x, theta) (theta + 1) * x^theta #0 ≤ x ≤ 1

hist(x, prob = TRUE, breaks = seq(0, 1, 0.1), xlab="x", ylab="f(x)",
main="Histogram dan FKP Estimasi", xlim=c(0,1), ylim=c(0,4))

curve(fkp_theta(x, theta_hat1), add=TRUE, col="red", lwd=2)
curve(fkp_theta(x, theta_hat2), add=TRUE, col="blue", lwd=2)
```

```r
legend("topleft", c("MLE", "MM"), col=c("red", "blue"), lwd=2)
```



Histogram menunjukkan distribusi empiris sampel (banyak nilai di sekitar $0.7 - 1$, menjulur ke kiri). Kedua kurva mendekati histogram, kurva merah (MLE) dan biru (MM) overlay, dengan MLE sedikit lebih curam karena θ lebih besar, membuat FKP lebih tinggi di x mendekati 1.

2f

```r
loglik <- function(x, theta){
  n <- length(x)
  return((n*log(theta + 1)) + (theta * sum(log(x))))
}

ll1 <- loglik(x, theta_hat1)
ll2 <- loglik(x, theta_hat2)
cat("Fungsi Loglikelihood dugaan θ1 =", ll1, "\n")

## Fungsi Loglikelihood dugaan θ1 = 5.559911

cat("Fungsi Loglikelihood dugaan θ2 =", ll2)

## Fungsi Loglikelihood dugaan θ2 = 5.551615
```

Log-likelihood MLE (5.559911) lebih tinggi daripada MM (5.551615), artinya MLE memberikan fit yang lebih baik karena secara definisi memaksimalkan likelihood, sehingga metode MLE lebih baik untuk data ini.

NOMOR 3

penerapan metode estimasi parameter MLE dan Metode Momen Fungsi kepadatan peluang (FKP) dari distribusi Laplace adalah:

$$f(x;\sigma) = \frac{1}{2\sigma}\exp\left(-\frac{|x|}{\sigma}\right), \quad x \in \mathbb{R}, \quad \sigma > 0$$

3a

Likelihood:

$$L(\sigma) = \prod_{i=1}^{n} \frac{1}{2\sigma}\exp\left(-\frac{|x_i|}{\sigma}\right) = \left(\frac{1}{2\sigma}\right)^n \exp\left(-\frac{\sum_{i=1}^{n}|x_i|}{\sigma}\right)$$

Loglikelihood:

$$l(\sigma) = n\log\left(\frac{1}{2\sigma}\right) - \frac{\sum_{i=1}^{n}|x_i|}{\sigma} = -n\log(2\sigma) - \frac{\sum_{i=1}^{n}|x_i|}{\sigma}$$

Turunan:

$$\frac{d}{d\sigma}l(\sigma) = -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}|x_i|}{\sigma^2} = 0$$

Persamaan:

$$\frac{\sum_{i=1}^{n}|x_i|}{\sigma^2} = \frac{n}{\sigma}$$

Hasil:

$$\sum_{i=1}^{n}|x_i| = n\sigma$$

Estimasi:

$$\hat{\sigma}_1 = \frac{1}{n}\sum_{i=1}^{n}|x_i|$$

3b

Bentuk umum FKP distribusi Laplace adalah:

$$f(x;\mu,\sigma) = \frac{1}{2\sigma}\exp\left(-\frac{|x-\mu|}{\sigma}\right), \quad x \in \mathbb{R}, \quad \sigma > 0$$

Dengan:

$$E[X] = \mu$$

$$Var[X] = 2\sigma^2$$

Sehingga, distribusi Laplace pada soal memiliki:

$$E[X] = 0$$

$$Var[X] = 2\sigma^2$$

Selanjutnya

$$Var[X] = E[X^2] - (E[X])^2 = E[X^2] = 2\sigma^2$$

Dan kita tahu, secara umum

$$E[(X-\mu)]^k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$$

Maka

$$E[X^2] = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = 2\sigma^2$$

$$\hat{\sigma}_2 = \sqrt{\frac{1}{2}\cdot\frac{1}{n}\sum_{i=1}^{n} x_i^2}$$

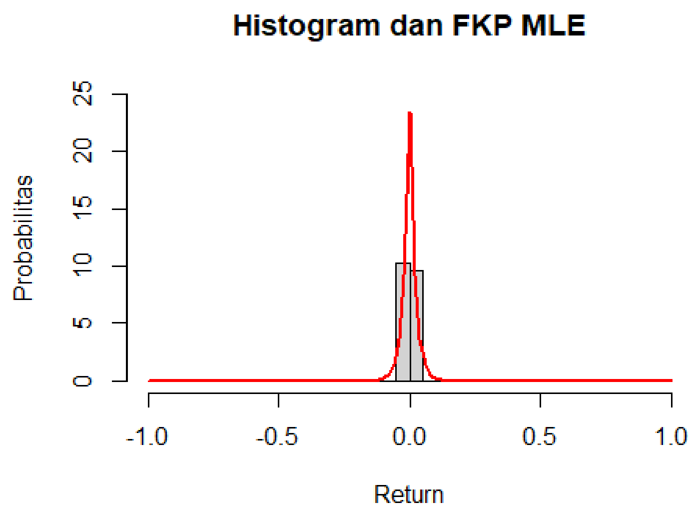
3c

```
return <- c(-1, -0.02, -0.015, 0.006, 0.015, 0.021, 0.75, 1)
freq <- c(1, 1650, 2, 1, 3, 1550, 1, 1)
x <- rep(return, freq)

sigma1 <- mean(abs(x))
sigma2 <- sqrt(sum((x - mean(x))^2)/(2*length(x)))

cat("MLE σ1 =", sigma1, "\n")

## MLE σ1 = 0.02130913

cat("MM σ2 =", sigma2)

## MM σ2 = 0.02467356
```

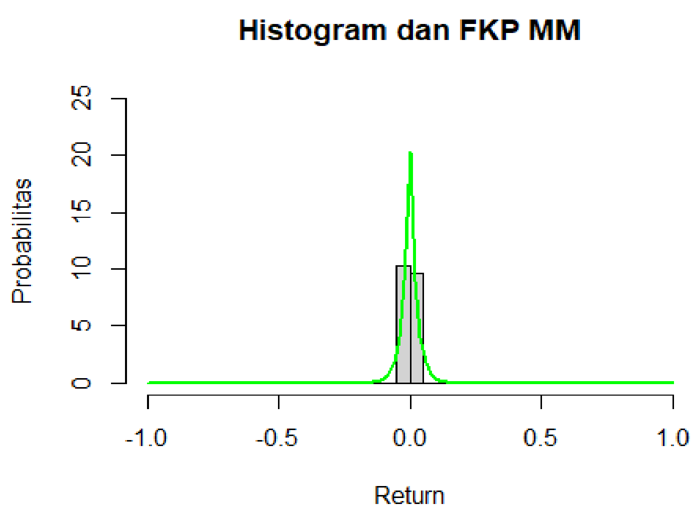
$\hat{\sigma}_1 \approx 0.02130913$ (MLE) lebih kecil karena berdasarkan rata-rata nilai absolut. $\hat{\sigma}_2 \approx 0.02467356$ (MM) lebih besar karena menggunakan momen kedua, mencerminkan variabilitas lebih luas.

3d

```r
fkp_laplace <- function(x, sigma) (1/(2*sigma)) * exp(-abs(x)/sigma)

hist(x, prob = TRUE, breaks = 30, main = "Histogram dan FKP MLE", xlab =
"Return", ylab = "Probabilitas", xlim=c(-1,1), ylim=c(0,25))
curve(fkp_laplace(x, sigma1), add = TRUE, col = "red", lwd = 2)
```




```r
hist(x, prob = TRUE, breaks = 30, main = "Histogram dan FKP MM", xlab =
"Return", ylab = "Probabilitas", xlim=c(-1,1), ylim=c(0,25))
curve(fkp_laplace(x, sigma2), add = TRUE, col = "green", lwd = 2)
```



FKP dengan MLE lebih tinggi di sekitar 0 dan ekor lebih tipis karena $\hat{\sigma}_1$ lebih kecil, menunjukkan konsentrasi data di nilai tengah.

FKP dengan MM lebih rendah di pusat dan ekor lebih tebal karena $\hat{\sigma}_2$ lebih besar, mencerminkan variabilitas lebih luas.

MLE lebih baik karena kurva merah lebih mengikuti histogram, sesuai kelebihan MLE dalam memaksimalkan likelihood.

3e

```r
loglikelihood <- function(x, s){
  n <- length(x)
  L <- -n*log(2*s) - (sum(abs(x))/s)
  return(L)
}

L1 <- loglikelihood(x, sigma1)
L2 <- loglikelihood(x, sigma2)

cat("Loglikelihood dugaan σ1 =", L1, "\n")

## Loglikelihood dugaan σ1 = 6916.911

cat("Loglikelihood dugaan σ2 =", L2)

## Loglikelihood dugaan σ2 = 6884.055
```

Loglikelihood MLE (6916.911) lebih tinggi daripada MM (6884.053), menunjukkan MLE lebih baik karena memaksimalkan likelihood secara langsung. MM kurang akurat untuk distribusi Laplace ini.

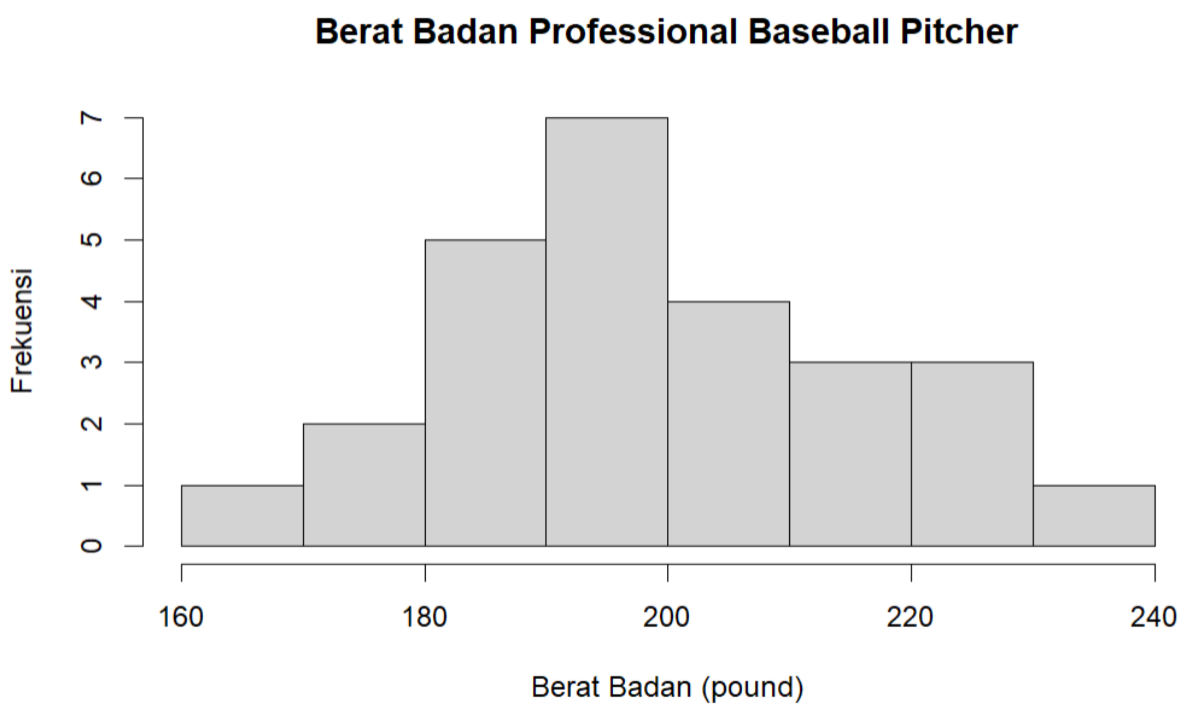
NOMOR 4

mengeksplorasi distribusi data A melalui histogram, memeriksa kecocokan dengan distribusi Normal dan Gamma menggunakan MLE (Maximum Likelihood Estimation), membandingkan log-likelihood untuk menentukan model terbaik, visualisasi, dan selang kepercayaan untuk selisih rata-rata antara dua gugus data independen.

```
dataA <- c(160, 175, 180, 185, 185, 185, 190, 190, 195, 195, 195, 200, 200,
200, 200, 205, 205, 210, 210, 218, 219, 220, 222, 225, 225, 232)

dataB <- c(155, 155, 160, 160, 160, 166, 170, 175, 175, 175, 180, 185, 185,
185, 185, 185, 185, 185, 190, 190, 190, 190, 190, 195, 195, 195, 195, 200,
205, 207, 210, 211, 230)
```

4a

```
hist(dataA, main = "Berat Badan Professional Baseball Pitcher", xlab = "Berat
Badan (pound)", ylab = "Frekuensi")
```



Bentuknya cukup simetris dengan puncak disekitar 200 pound, cukup sesuai sebaran Normal.

4b

```
library(MASS)

fit_normal <- fitdistr(dataA, "normal")
print(fit_normal)

##       mean          sd
##   201.000000    17.144185
##  (  3.362251) (  2.377471)

cat("\nNilai fungsi log-likelihood =", logLik(fit_normal))

##
## Nilai fungsi log-likelihood = -110.7755
```

Asumsi sebaran normal layak untuk memodelkan gugus data A karena histogram cukup simetri, dan nilai fungsi log-likelihood (-110.7755) lebih tinggi daripada sebaran gamma.

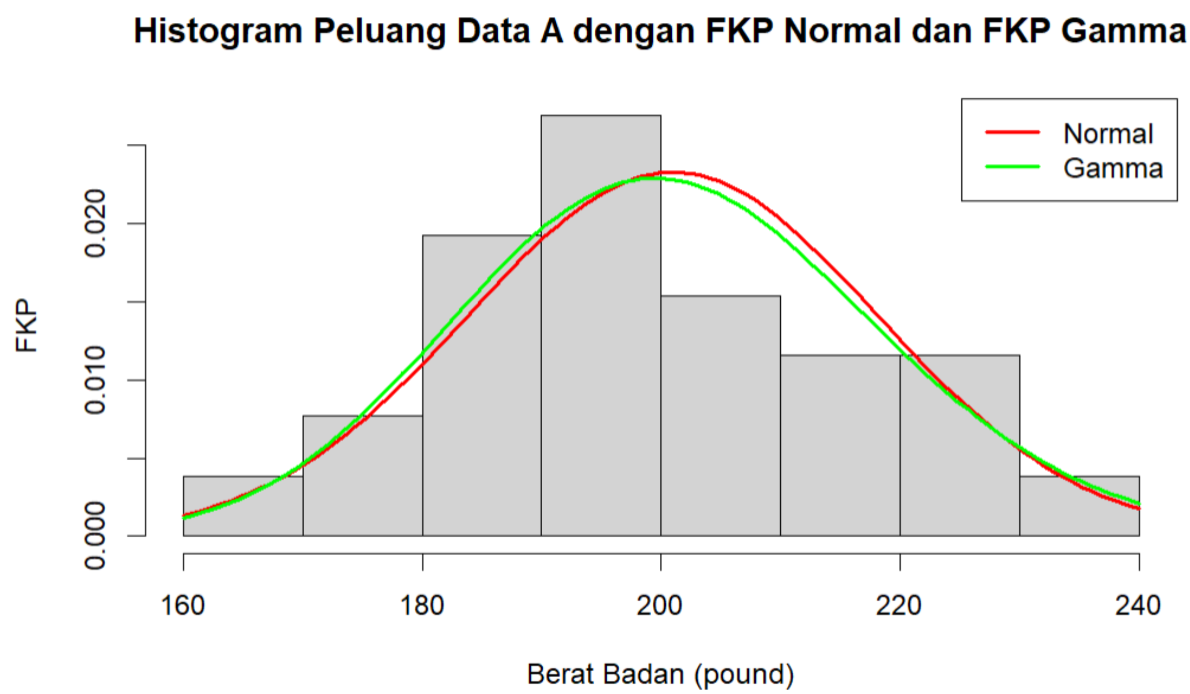
4c

```
fit_gamma <- fitdistr(dataA, "gamma")
print(fit_gamma)

##       shape         rate
##   132.1676293     0.6575516
##  ( 36.5999945) (  0.1824345)

cat("\nNilai fungsi log-likelihood =", logLik(fit_gamma))

##
## Nilai fungsi log-likelihood = -110.9703
```

Asumsi sebaran gamma tidak begitu layak untuk memodelkan gugus data A karena nilai fungsi log-likelihood (-110.9703) lebih rendah daripada sebaran normal.

4d

```
norm_dist <- function (x) dnorm(x, mean = fit_normal$estimate[1], sd =
fit_normal$estimate[2])
gamma_dist <- function (x) dgamma(x, shape = fit_gamma$estimate[1], rate =
fit_gamma$estimate[2])

hist(dataA, prob = TRUE, main = "Histogram Peluang Data A dengan FKP Normal
dan FKP Gamma", xlab = "Berat Badan (pound)", ylab = "FKP")
curve(norm_dist, add = TRUE, col = "red", lwd = 2)
curve(gamma_dist, add = TRUE, col = "green", lwd = 2)
legend("topright", c("Normal", "Gamma"), col = c("red", "green"), lwd = 2)
```



Histogram peluang menunjukkan distribusi data A, kurva normal lebih sesuai dengan bentuk simetris histogram, sedangkan kurva gamma sedikit miring dan kurang fit. Dengan demikian, sebaran normal lebih layak untuk memodelkan data A karena nilai fungsi loglikelihood lebih tinggi (-110.7755) dibanding nilai fungsi loglikelihood sebaran gamma (-110.9703).

4e

```
t.test(dataA, dataB, conf.level = 0.95, var.equal = FALSE, paired = FALSE)

##
##  Welch Two Sample t-test
##
## data:  dataA and dataB
## t = 3.4159, df = 53.502, p-value = 0.001221
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.432114 24.719401
## sample estimates:
## mean of x mean of y
##  201.0000  185.4242
```

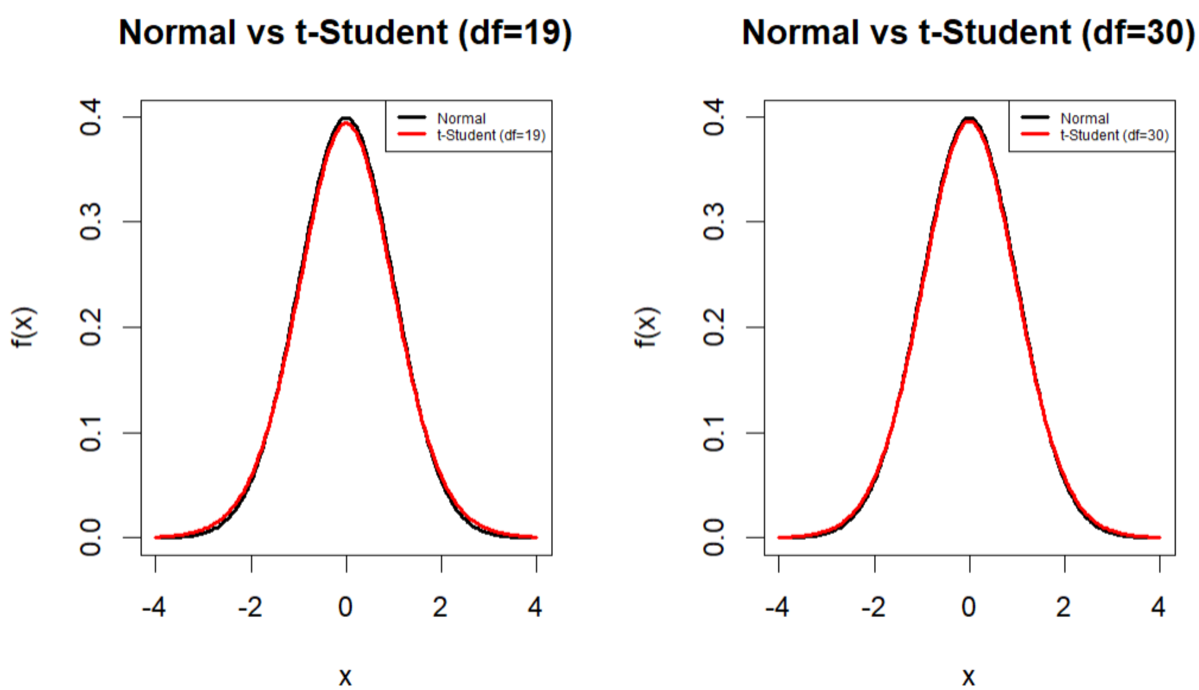
Selang kepercayaan 95% untuk selisih mean (A - B) adalah $[6.43, 24.72]$. Karena selang sepenuhnya positif, kita 95% yakin bahwa secara rata-rata, pitcher (A) lebih berat daripada hitter (B) secara signifikan, dengan perbedaan berat badan pitcher dan hitter berada antara 6.43 hingga 24.72 pound.

NOMOR 5

membuat selang kepercayaan 95% untuk rata-rata nilai perbedaan dari data pasangan (paired data),

Karena ada 20 berkas jawaban acak dari kelas A dan 20 berkas siswa pasangan dari kelas B, diasumsikan sampelnya berpasangan, sehingga perbedaan dihitung per pasangan mahasiswa. Karena sampel kita relatif kecil (`n = 20`) dan varians populasi tidak diketahui, distribusi t-Student digunakan untuk selang kepercayaan.

```
par(mfrow = c(1, 2))
x <- seq(-4, 4, length = 200)

normal <- dnorm(x, mean = 0, sd = 1)
t19 <- dt(x, df = 19)
t30 <- dt(x, df = 30)

plot(x, normal, type = "l", col = "black", lwd = 2, ylim = c(0, 0.4), xlab =
"x", ylab = "f(x)", main = "Normal vs t-Student (df=19)")
lines(x, t19, col = "red", lwd = 2)
legend("topright", legend = c("Normal", "t-Student (df=19)"), col =
c("black", "red"), lwd = 2, cex = 0.5)

plot(x, normal, type = "l", col = "black", lwd = 2, ylim = c(0, 0.4), xlab =
"x", ylab = "f(x)", main = "Normal vs t-Student (df=30)")
lines(x, t30, col = "red", lwd = 2)
legend("topright", legend = c("Normal", "t-Student (df=30)"), col =
c("black", "red"), lwd = 2, cex = 0.5)
```



```r
par(mfrow = c(1, 1))

kelas_A <- c(70, 65, 81, 85, 80, 60, 50, 83, 77, 81, 65, 45, 40, 52, 82, 80,
62, 46, 42, 61)
kelas_B <- c(70, 65, 81, 87, 80, 62, 50, 88, 78, 80, 65, 45, 42, 60, 83, 80,
63, 45, 45, 61)

perbedaan <- kelas_A - kelas_B
n <- length(perbedaan)

mean_beda <- mean(perbedaan)
sd_beda <- sd(perbedaan)

# Nilai kritis t untuk 95% kepercayaan, derajat bebas = n-1
# Gunakan qt() untuk quantile t, dengan alpha/2 = 0.025 (two-tailed)
c <- qt(0.975, df = n - 1)
error <- c * (sd_beda / sqrt(n))

# Selang kepercayaan 95%
lower_bound <- mean_beda - error
upper_bound <- mean_beda + error

cat("Rata-rata perbedaan:", mean_beda, "\n")

## Rata-rata perbedaan: -1.15

cat("Standar deviasi perbedaan:", sd_beda, "\n")

## Standar deviasi perbedaan: 2.158825

cat("Selang kepercayaan 95%:", lower_bound, "hingga", upper_bound, "\n")

## Selang kepercayaan 95%: -2.160361 hingga -0.1396387

t.test(kelas_A, kelas_B, conf.level = 0.95, var.equal = FALSE, paired = TRUE)

##
##  Paired t-test
##
## data:  kelas_A and kelas_B
## t = -2.3823, df = 19, p-value = 0.02781
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -2.1603613 -0.1396387
## sample estimates:
## mean difference
##           -1.15
```

Dengan tingkat kepercayaan 95%, dapat dinyatakan bahwa rata-rata perbedaan nilai UTS antara kelas A dan kelas B berada dalam interval -2.160361 hingga -0.1396387. Karena selang ini sepenuhnya negatif, berarti kita 95% yakin bahwa secara rata-rata, nilai kelas A sedikit lebih rendah daripada nilai kelas B, dengan perbedaan berkisar antara -2.160361 (kelas A lebih rendah 2.160361 poin) hingga -0.1396387 (kelas A lebih rendah 0.1396387 poin).

Namun, ujung atas selang (-0.1396387) cukup dekat dengan 0, menunjukkan bahwa perbedaan mungkin tidak cukup kuat untuk dikatakan “signifikan” secara statistik pada tingkat 5. Jadi, ada kemungkinan metode pengajaran di kelas B sedikit lebih efektif (karena nilai kelas A sedikit lebih rendah), tapi efeknya mungkin tidak besar.