HW 5

Q1) (1) $y \sim N(\mu, \sigma^2)$

pdf for normal : $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $[-\infty < y < \infty]$

(2) $Y_1, \ldots Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$ i.e IID'ed random sampled

$Y_1 = y_1, \ldots \ldots Y_n = y_n$.

data (sampled)

where each point (data) will have pdf $f(y_i) = \frac{1}{\sqrt{2\pi}(\sigma)} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}$
[where, $i = 1, \ldots n$]

$\because$ Sampled Randomly & Independently.

$f(y_1, y_2, \ldots y_n) = f(y_1)\, f(y_2) \times \ldots f(y_n)$

entire data.

Joint distribution.

(3) Joint dist = $f(y_1) \ldots f(y_n)$

$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_1-\mu)^2}{2\sigma^2}} \times \ldots \ldots \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_n-\mu)^2}{2\sigma^2}}$$

$$= L(\mu, \sigma^2 \mid y_1, \ldots y_n).$$

Likelihood function (a function of parameters μ & σ^2 here).

(14) for <u>log likelihood</u>.

$$l(\mu, \sigma^2) = \log\left[L(\mu, \sigma^2 \mid y_1, \ldots, y_n)\right]$$

$$= \log\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_1-\mu)^2}{2\sigma^2}} \times \ldots \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_n-\mu)^2}{2\sigma^2}}\right)$$

$$= \log\left[\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \left(e^{-\left(\frac{1}{2\sigma^2}\right)\sum_{i=1}^{n}(y_i-\mu)^2}\right)\right] \quad \text{— (I)}$$

$$\cancel{= n \log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2}$$

$$= \log\left[(2\pi\sigma^2)^{n/2} \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2}\right]$$

$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

or from (I)

$$= n \log \frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 .$$

$$-l(\mu, \sigma) = -\left(n \log \frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right).$$

(15) $(\hat{\mu}, \hat{\sigma})$ = optimum parameters = $\arg\max L(\mu, \sigma)$ = $\arg\max l(\mu, \sigma)$
= $\arg\min -l(\mu, \sigma)$.

$\because$ log is a monotonic function, maximizing (log) is same as minimizing (− log) & maximizing (L)

$\because$ ~~we~~ we can ~~reduce~~ improve the computations using log likelihood instead of likelihood, bec product calculations gets converted to Σ $\therefore$ we maximize loglikelihood instead. (To ~~solv~~ avoid underflow of float values)

(5) how is $\max \ell = \min -\ell$.

explanation.

$$\max \ell \Rightarrow \frac{d\ell}{d\mu}. \Rightarrow \hat{\mu} = \frac{\Sigma y_i}{n}.$$

$$\min -\ell. \Rightarrow \frac{d}{d\mu}\left(\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2 - n\log\frac{1}{\sigma\sqrt{2\pi}}\right).$$

$$= \frac{2}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)(+1) - 0 \equiv 0$$

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu) = 0$$

$$\therefore \Sigma y_i = \Sigma \mu$$

$$\therefore \boxed{\hat{\mu} = \frac{\Sigma y_i}{n}}.$$

$$\therefore (\hat{\mu}, \hat{\sigma}) = \arg\max(\mu,\sigma) = \arg\min -\ell(\mu,\sigma).$$
$$= \arg\max L(\mu,\sigma).$$

(6) to maximize $l(\mu, \sigma)$.

$$\frac{dl}{d\mu} = \frac{d}{d\mu}\left(n \log \frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right)$$

$$= 0 - \frac{1}{2\sigma^2} \, 2 \cdot \sum_{i=1}^{n}(y_i-\mu)(-1)$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i+\mu) \equiv 0$$

i.e $\sum_{i=1}^{n}(y_i-\mu) = 0$

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \mu$$

$$n \times \mu = \sum y_i$$

$\therefore$ $\boxed{\hat{\mu}_1 = \frac{\sum y_i}{n}}$ — (by max likelihood).

by least squares, error to minimize $= \sum_{i=1}^{n}(y_i-\mu)^2$

$$\frac{d}{d\mu}\sum_{i=1}^{n}(y_i-\mu)^2 = 0$$

$$2\sum_{i=1}^{n}(y_i-\mu)(-1) = 0$$

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n}\mu. \Rightarrow \boxed{\hat{\mu}_2 = \frac{\sum y_i}{n}}.$$

$\because$ $\hat{\mu}_1 = \hat{\mu}_2$, we can say that

maximizing likelihood & least squares gives the same answers.

S520_HW5_Q1

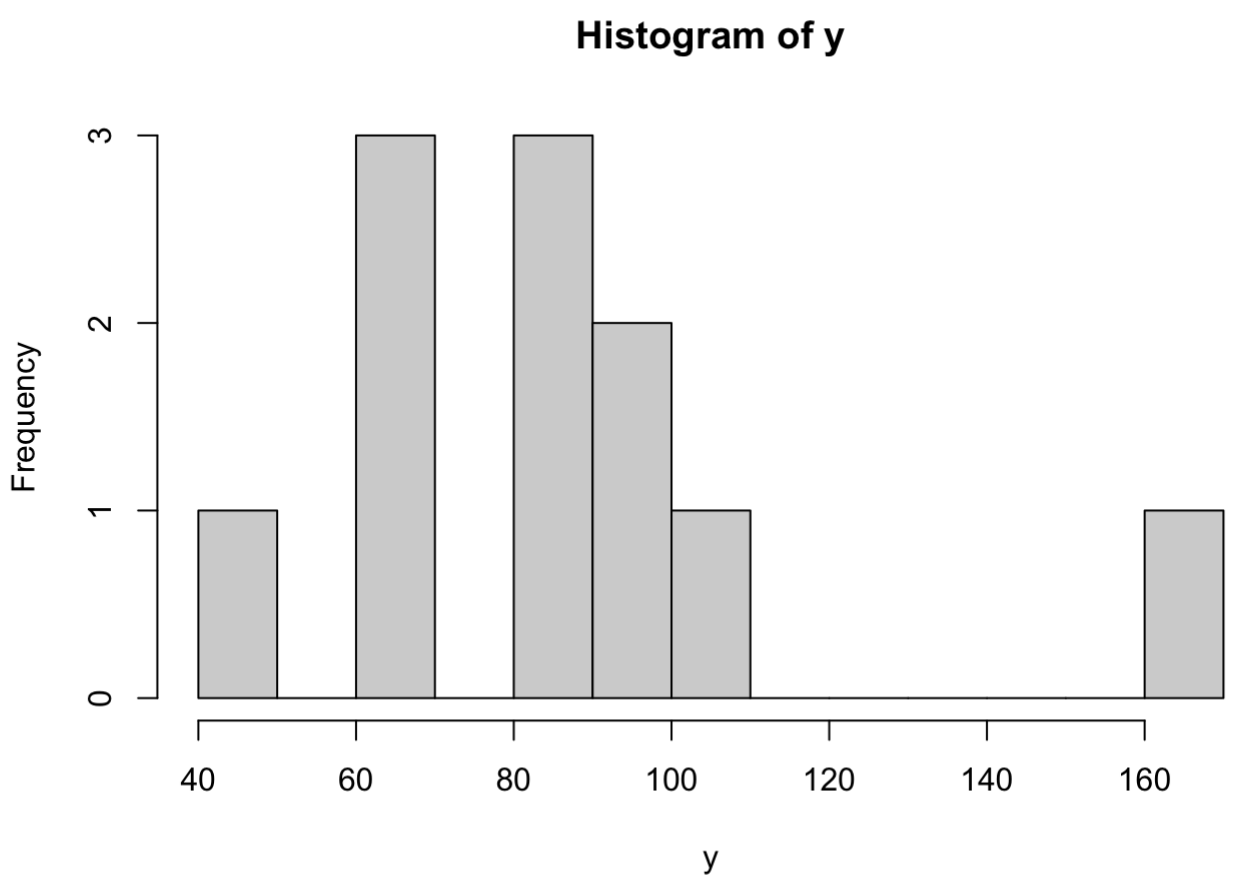
Rushank Ghanshyam Sheta

2022-11-11

7(a). Read data

```
data = read.csv('/Users/rushank/Downloads/q1_carprice.csv')
y = data$Price
```

```
hist(y, breaks=length(y))
```



7(b). estimates using OPTIM method

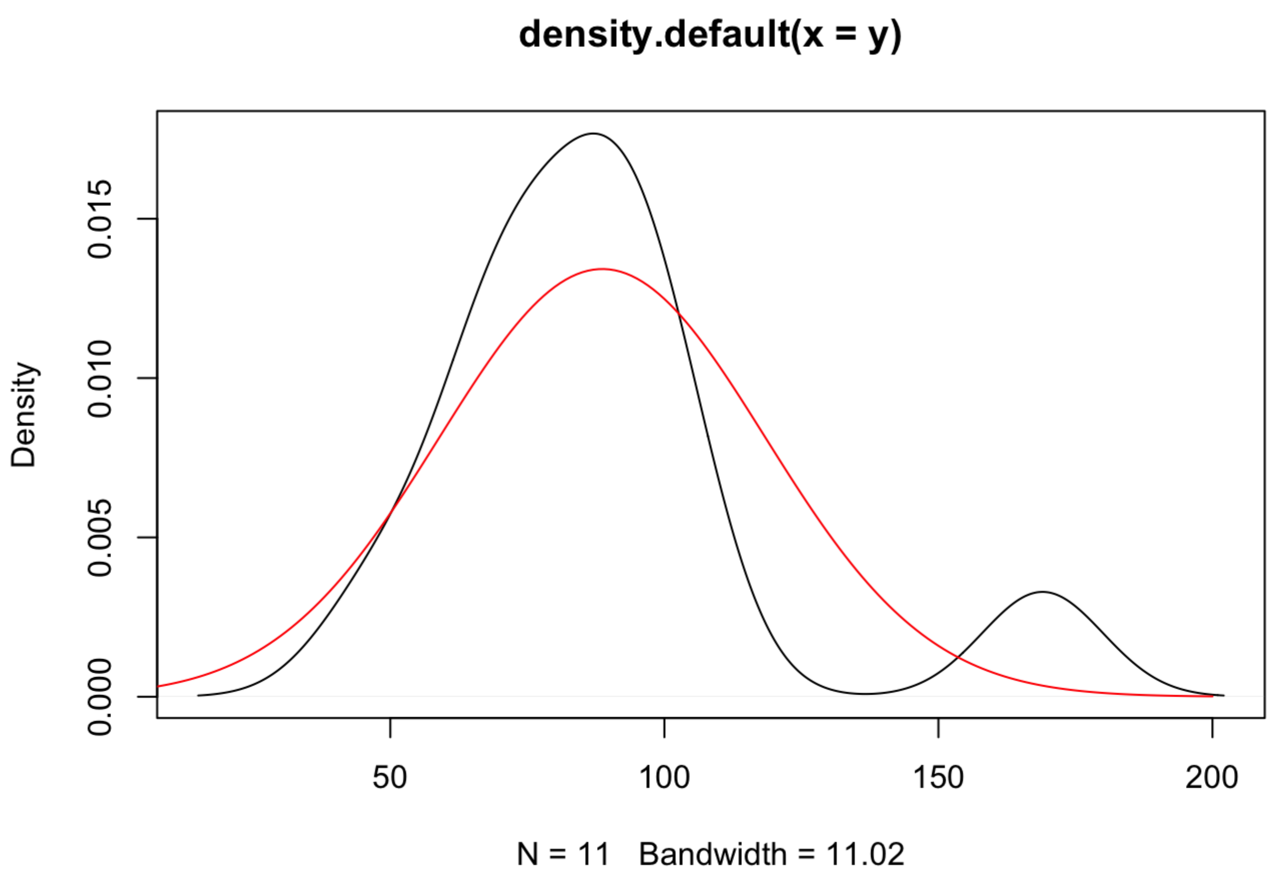
```
log_lik_norm = function(theta,y) {
  mu = theta[1]
  sd = exp(theta[2])
  vec_log_densities = dnorm(x=y,mean=mu,sd=sd,log = TRUE)
  log_lik = sum(vec_log_densities)
  return(log_lik)
}

estimates = optim(par=c(0,1),fn=log_lik_norm,y=y,control = list(fnscale=-1))$par
estimates[2] = exp(estimates[2])
sprintf('mu(optim): %f',estimates[1])
```

```
## [1] "mu(optim): 88.681715"
```

```
sprintf('sigma(optim): %f',estimates[2])
```

```
## [1] "sigma(optim): 29.722016"
```

```
plot(density(y))
x = seq(1,200,by=1)
lines(x,dnorm(x,mean=estimates[1],sd=estimates[2]),col="red")
```



8. estimates using MLE method

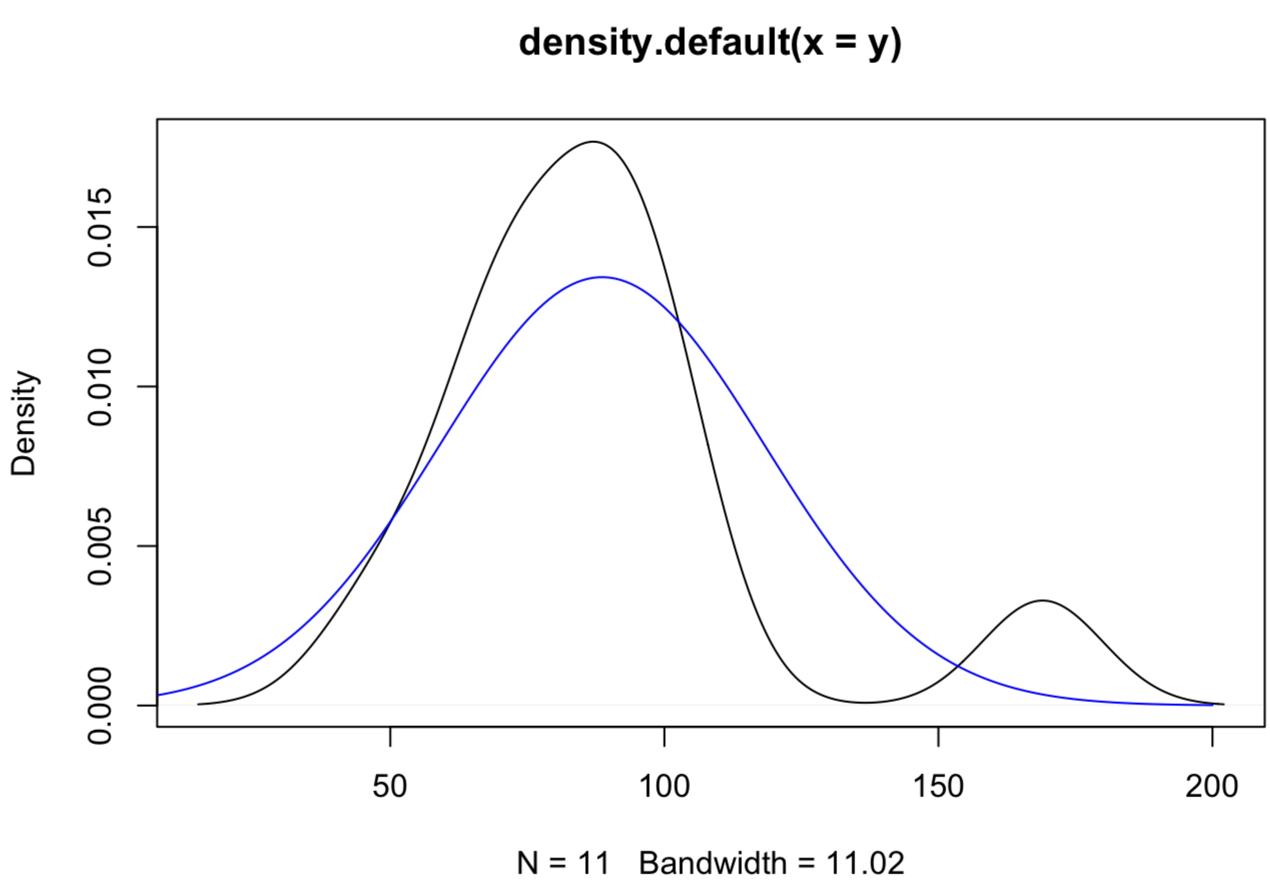
```
mu_mle = mean(y)
sprintf('mu(mle): %f',mu_mle)
```

```
## [1] "mu(mle): 88.636364"
```

```
sigma_mle = sqrt(sum((y-mu_mle)^2)/length(y))
sprintf('sigma(mle): %f',sigma_mle)
```

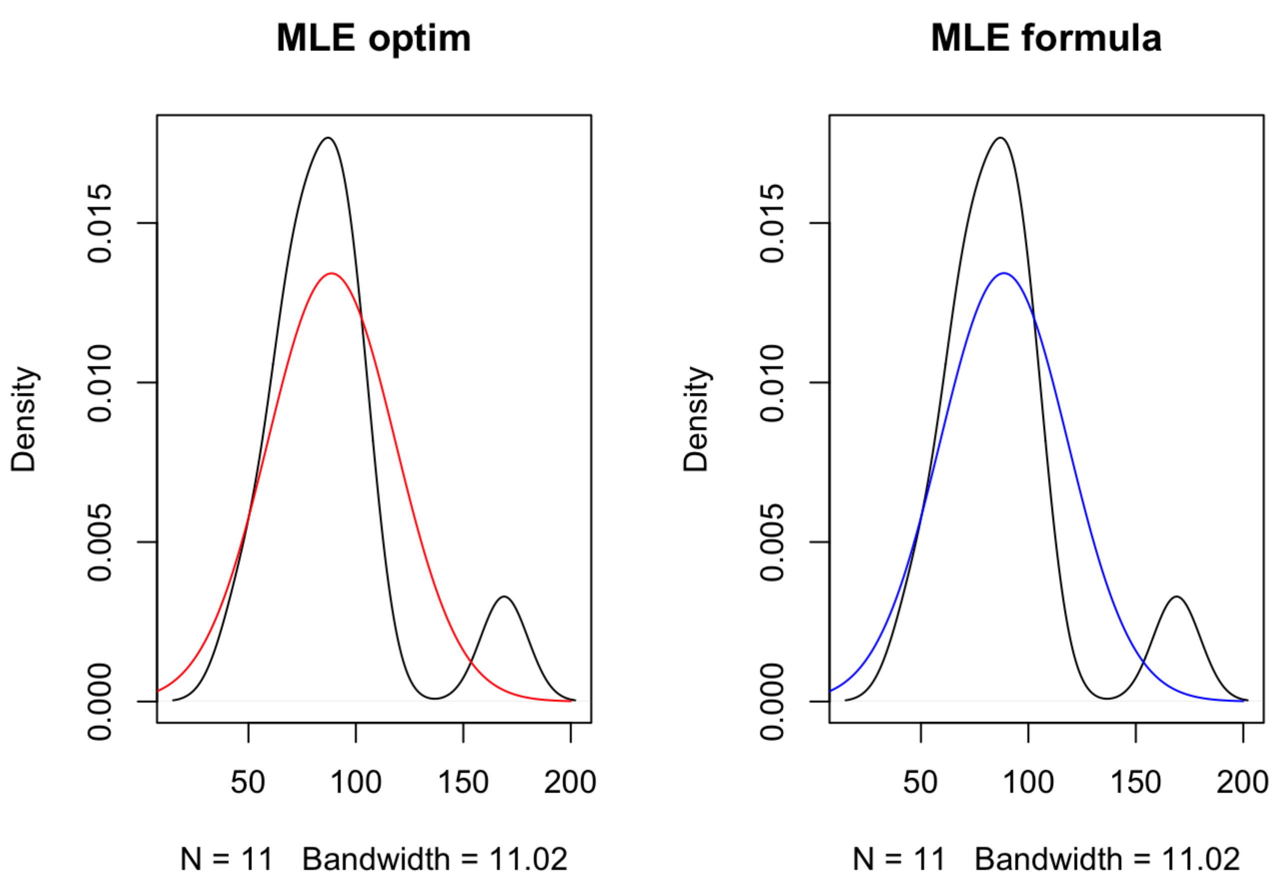
```
## [1] "sigma(mle): 29.708501"
```

```
plot(density(y))
x = seq(1,200,by=1)
lines(x,dnorm(x,mean=mu_mle,sd=sigma_mle),col="blue")
```

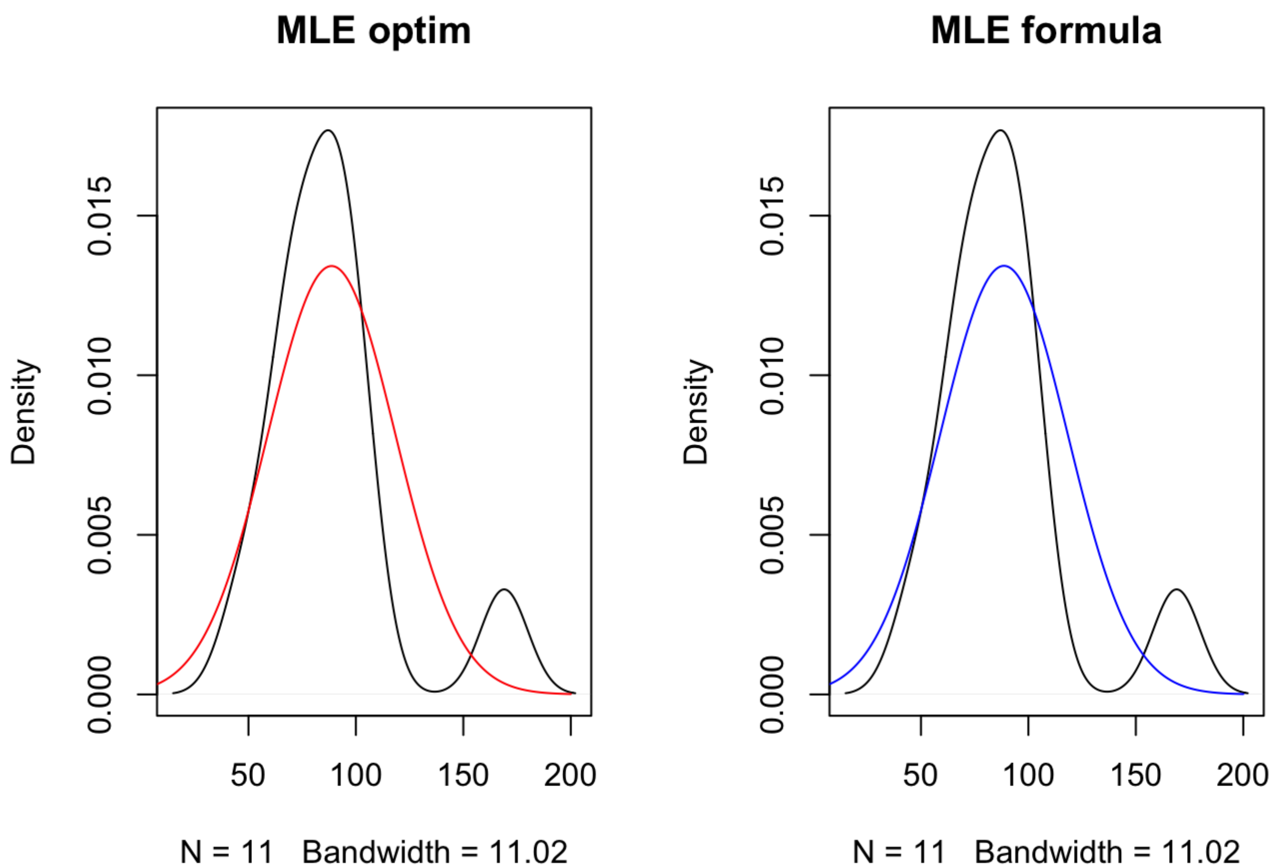


9. Subplots for estimates using both methods

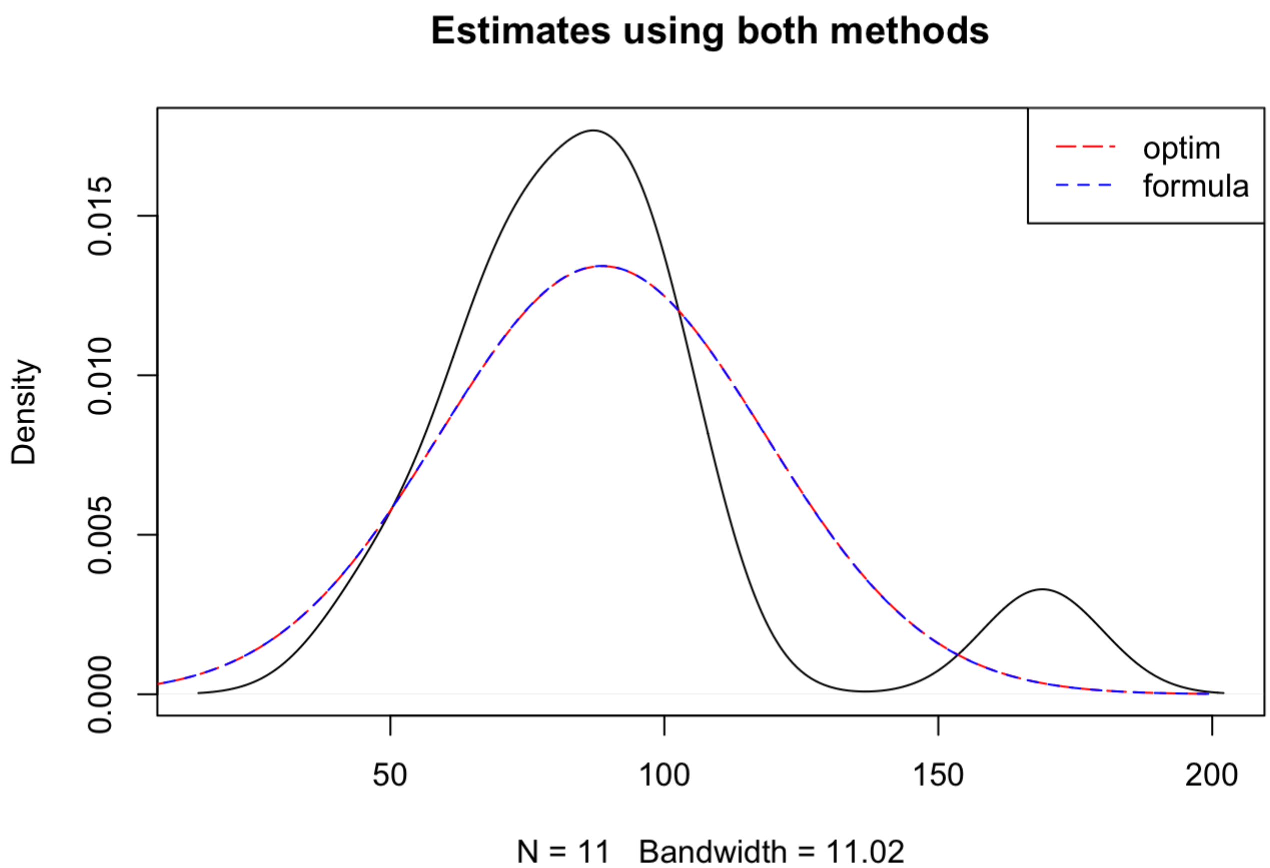
```
log_lik_norm = function(theta,y) {
  mu = theta[1]
  sd = exp(theta[2])
  vec_log_densities = dnorm(x=y,mean=mu,sd=sd,log = TRUE)
  log_lik = sum(vec_log_densities)
  return(log_lik)
}

par(mfrow=c(1,2))
x = seq(1,200,by=1)
plot(density(y),main="MLE optim")
lines(x,dnorm(x,mean=estimates[1],sd=estimates[2]),col="red")
plot(density(y),main="MLE formula")
lines(x,dnorm(x,mean=mu_mle,sd=sigma_mle),col="blue")
```




```
par(mfrow=c(1,2))
x = seq(1,200,by=1)
plot(density(y),main="MLE optim")
lines(x,dnorm(x,mean=estimates[1],sd=estimates[2]),col="red")
plot(density(y),main="MLE formula")
lines(x,dnorm(x,mean=mu_mle,sd=sigma_mle),col="blue")
```




```
par(mfrow=c(1,1))
x = seq(1,200,by=1)
plot(density(y),main="Estimates using both methods")
lines(x,dnorm(x,mean=estimates[1],sd=estimates[2]),col="red", lty='longdash')
lines(x,dnorm(x,mean=mu_mle,sd=sigma_mle),col="blue", lty='dashed')

legend(x = "topright",
       legend = c("optim", "formula"),
       lty = c('longdash', 'dashed'),
       col = c('red', 'blue'))
```



10. Comment on how well the chosen model (Normal) fits the data.

The estimates from both of the methods gives approximately the same results however these estimates(Normal) does not fit the given sampled data very well.