

# STAC67 A1

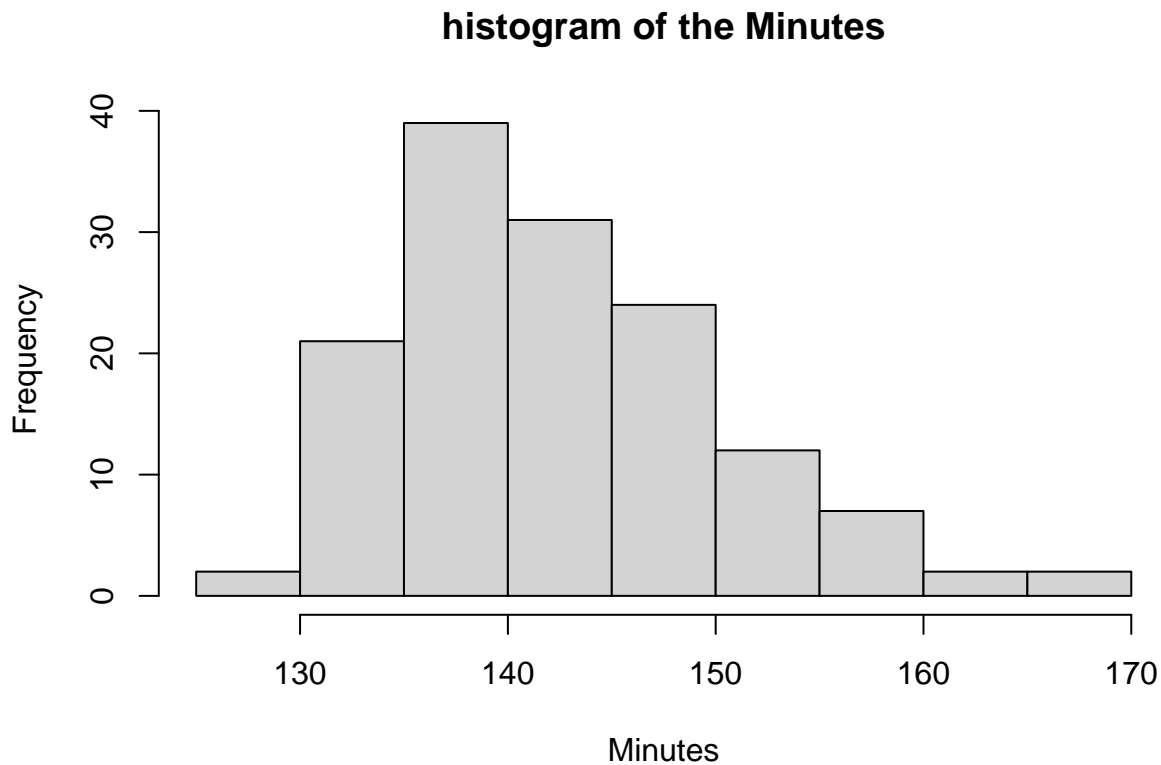
2026-01-14

Q1

(a)

```
csvFile = "OlympicMarathon2016.csv"
df <- read.csv(csvFile)
mu = mean(df[["Minutes"]])
sigma2 = var(df[["Minutes"]])
sprintf("Minutes has mean %f and variance %f", mu, sigma2)

## [1] "Minutes has mean 142.367143 and variance 59.638923"
hist(df[["Minutes"]], xlab="Minutes", main="histogram of the Minutes")
```



(b)

```
set.seed(1008494620)
population = df[["Minutes"]]
sampleCnt = 10000
n = 30
```

```

yBar = numeric(sampleCnt)
stats = numeric(sampleCnt)
for (i in 1:sampleCnt) {
  y = sample(population, size=n, replace=TRUE)
  yBar[i] = mean(y)
  s2 = var(y)
  stats[i] = (n-1) * s2 / sigma2
}

```

(b)(i)

```

mean_yBar = mean(yBar)
var_yBar = var(yBar)
mean_stats = mean(stats)
var_stats = var(stats)

percentiles = c(0.025, 0.25, 0.5, 0.75, 0.975)
percentiles_pop = mapply(
  function(p) {
    quantile(df[["Minutes"]], p)
  },
  percentiles
)

percentiles_yBar = mapply(
  function(p) {
    quantile(yBar, p)
  },
  percentiles
)

percentiles_stats = mapply(
  function(p) {
    quantile(stats, p)
  },
  percentiles
)

cat(
  sprintf("theoretical values\n"),
  sprintf("=====\n"),
  sprintf("mean = %f, var = %f\n", mu, sigma2),
  sprintf("percentiles:\n"),
  percentiles_pop
)

cat(
  sprintf("\n\nresult of sample means\n"),
  sprintf("=====\n"),
  sprintf("mean = %f, var = %f\n", mean_yBar, var_yBar),
  sprintf("percentiles:\n"),
  percentiles_yBar
)

cat(

```

```

    sprintf("\n\nresult of scaled sample var\n"),
    sprintf("=====\n"),
    sprintf("mean = %f, var = %f\n", mean_stats, var_stats),
    sprintf("percentiles:\n"),
    percentiles_stats
)

```

```

## theoretical values
## =====
## mean = 142.367143, var = 59.638923
## percentiles:
## 131.1555 137.1225 140.765 147.035 159.8852
##
## result of sample means
## =====
## mean = 142.354664, var = 1.984513
## percentiles:
## 139.7113 141.3853 142.3157 143.2964 145.1947
##
## result of scaled sample var
## =====
## mean = 28.747533, var = 69.091112
## percentiles:
## 14.39847 22.80234 28.02941 34.012 46.61924

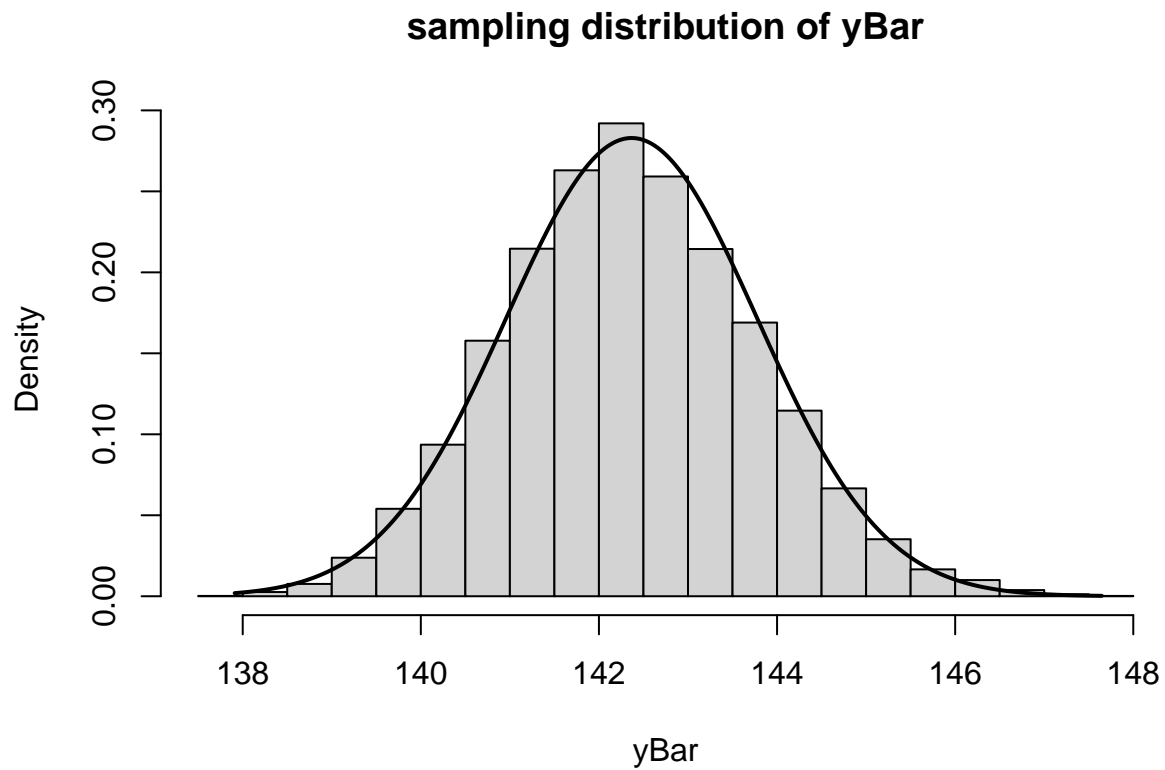
```

(b) (ii)

```

hist(yBar,
     probability = TRUE,
     main="sampling distribution of yBar",
     xlab="yBar",
)
x1 = seq(min(yBar), max(yBar), length.out = 500)
lines(x1, dnorm(x1, mean=mu, sd=sqrt(sigma2/n)), lwd=2)

```



```
hist(stats,  
      probability=TRUE,  
      main="sampling distribution of scaled sample variance",  
      xlab=expression(bar(y)),  
      )  
x2 = seq(0, max(stats), length.out = 500)  
lines(x2, dchisq(x2, df=n - 1), lwd=2)
```

# sampling distribution of scaled sample variance

