

Solutions to Worksheet 09: Sampling Distributions

Question 1: Sampling from a Normal Population

Let X denote the distribution of BMI of all adult men. We can approximate this distribution by $X \sim N(26, 4)$.

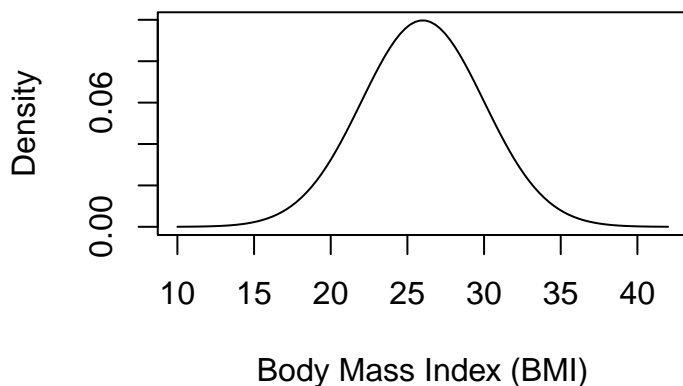
- a. Interpret and run the code below to create a graph of the pdf for the distribution of X , the BMI of all adult males.

```
# load possible BMI's going +/- 4 standard deviations from mean.
bmi <- seq(26-4*4, 26+4*4, length=100)

# Compute the value of f(x) at each bmi value if we assume
# BMI is normally distribution with mean 26 and standard dev 4
pdf.bmi <- dnorm(bmi, 26, 4)

# Plot bmi on x-axis and the value of pdf, f(x) on y-axis.
plot(bmi, pdf.bmi,
     type="l", lty=1,          # type="l" draws line lty=1 is solid line
     xlab="Body Mass Index (BMI)",
     ylab="Density", main="Distribution of Population")
```

Distribution of Population



- b. A sample of $n = 4$ adult men are randomly selected. The mean BMI of the sample is calculated:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4}{4}.$$

Then another random sample of $n = 4$ adult men are randomly selected, and again the mean BMI of this sample is computed. This is repeated 1000 times (each sample size $n = 4$), and the sampling distribution for the mean BMI can be constructed with the code below:

```
# creates an empty vector to store results
n4.bmi.bar <- numeric(1000)

# A for loop that generates 1000 random samples
# Each size n=4, and calculates the sample mean.
```

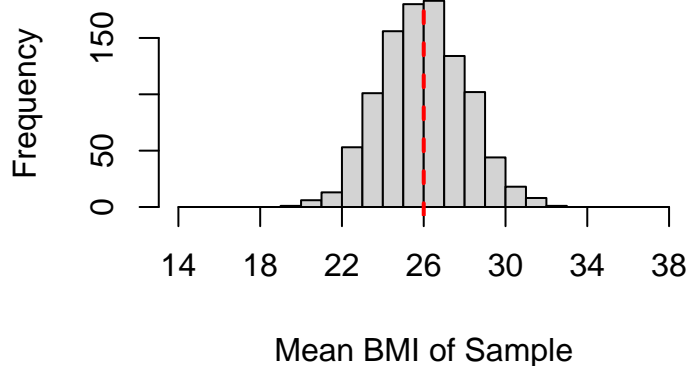
```

for (i in 1:1000)
{
  n4.bmi.sample <- rnorm(4, 26, 4) #Randomly picks 4 values from N(26,4)
  n4.bmi.bar[i] <- mean(n4.bmi.sample)
}

# Plot the sampling distribution
hist(n4.bmi.bar, xlim = c(14, 38),
     xlab = "Mean BMI of Sample",
     main = "Sampling Distribution of Mean BMI for n=4",
     xaxt='n')
axis(1, at=seq(14, 38, 4), pos=0)
abline(v = 26, col = "red", lwd = 2, lty = 2)

```

Sampling Distribution of Mean BMI for n=



- c. In the R code block below, enter commands to compute the center (as measured by the mean) and spread (as measured by the standard deviation) of the sampling distribution when $n = 4$.

```
mean(n4.bmi.bar)
```

```
## [1] 25.98417
```

```
sd(n4.bmi.bar)
```

```
## [1] 2.068444
```

- d. Repeat parts (b) and (c) if we increase the size of the samples from $n = 4$ to $n = 9$.

```

# creates an empty vector to store results
n9.bmi.bar <- numeric(1000)

# A for loop that generates 1000 random samples
# Each size n=9, and calculates the sample mean.
for (i in 1:1000)
{
  n9.bmi.sample <- rnorm(9, 26, 4) #Randomly picks 9 values from N(26,4)
  n9.bmi.bar[i] <- mean(n9.bmi.sample)
}

# Plot the sampling distribution
hist(n9.bmi.bar, xlim = c(14, 38),
     xlab = "Mean BMI of Sample",

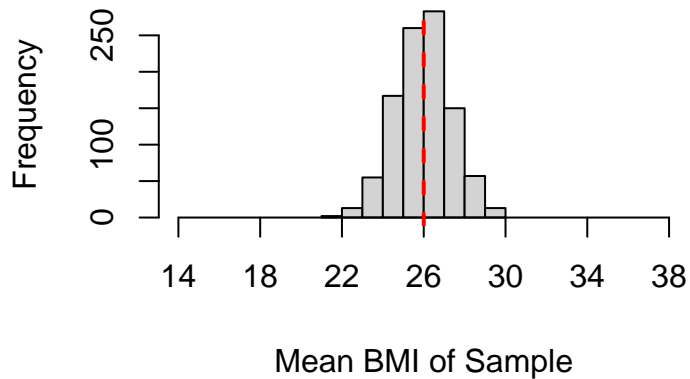
```

```

    main = "Sampling Distribution of Mean BMI for n=9",
    xaxt='n')
axis(1, at=seq(14, 38, 4), pos=0)
abline(v = 26, col = "red", lwd = 2, lty = 2)

```

Sampling Distribution of Mean BMI for n=



```
mean(n9.bmi.bar)
```

```
## [1] 25.97812
```

```
sd(n9.bmi.bar)
```

```
## [1] 1.350932
```

e. Repeat parts (b) and (c) if we increase the size of the samples to $n = 16$.

```

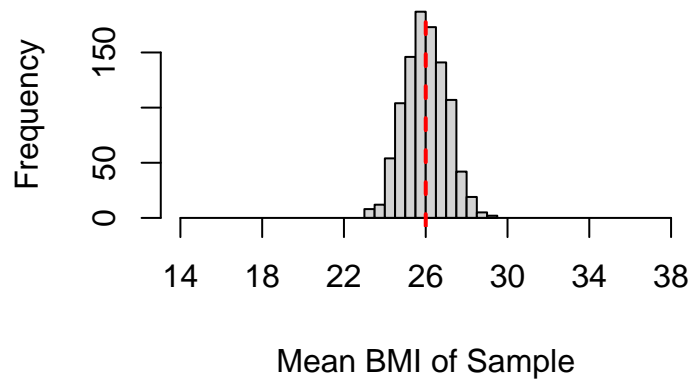
# creates an empty vector to store results
n16.bmi.bar <- numeric(1000)

# A for loop that generates 1000 random samples
# Each size n=16, and calculates the sample mean.
for (i in 1:1000)
{
  n16.bmi.sample <- rnorm(16, 26, 4) #Randomly picks 16 values from N(26,4)
  n16.bmi.bar[i] <- mean(n16.bmi.sample)
}

# Plot the sampling distribution
hist(n16.bmi.bar, xlim = c(14, 38),
     xlab = "Mean BMI of Sample",
     main = "Sampling Distribution of Mean BMI for n=16",
     xaxt='n')
axis(1, at=seq(14, 38, 4), pos=0)
abline(v = 26, col = "red", lwd = 2, lty = 2)

```

Sampling Distribution of Mean BMI for n=



```
mean(n16.bmi.bar)
```

```
## [1] 25.98596
```

```
sd(n16.bmi.bar)
```

```
## [1] 1.033291
```

Question 2: Sampling from a Skewed Population

Let X denote the distribution of the time (in minutes) between successive eruptions (called the wait time) of a certain geyser that is modeled by $X \sim \text{Exp}(\frac{1}{40})$.

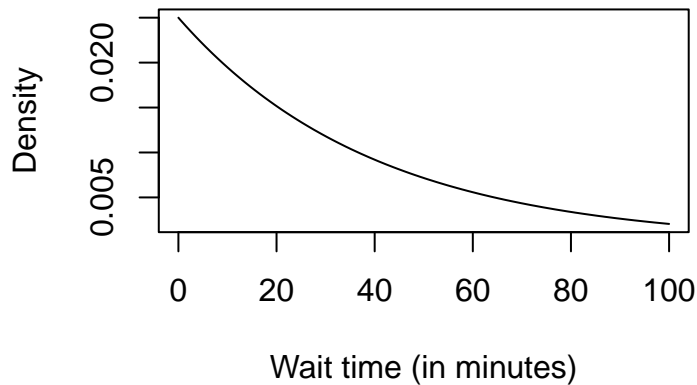
- a. Interpret and run the code below to create a graph of the pdf for the distribution of X , the wait time between all eruptions of a certain geyser.

```
# load possible wait times.
wait.time <- seq(0, 100, length=200)

# Compute the value of f(x) of each wait time x if we assume
# the times are exponentially distributed with mean 40 min.
pdf.wait.time <- dexp(wait.time, 1/40)

# Plot bmi on x-axis and the value of pdf, f(x) on y-axis.
plot(wait.time, pdf.wait.time,
     type="l", lty=1,          # type="l" draws line lty=1 is solid line
     xlab="Wait time (in minutes)",
     ylab="Density", main="Distribution of Population")
```

Distribution of Population



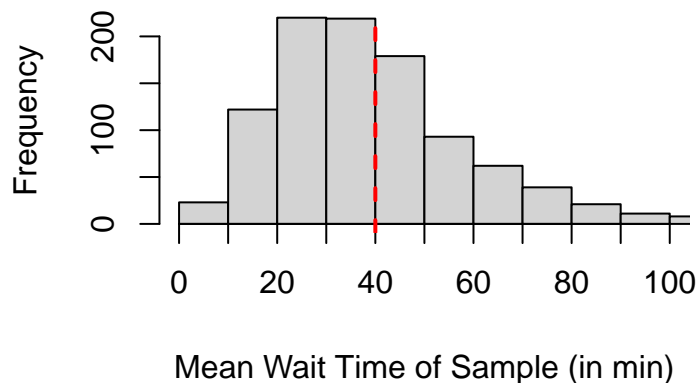
- b. A sample of $n = 4$ wait times are randomly selected and the sample mean wait mean is calculated. This is repeated until 1000 random samples each size $n = 4$ are selected.

```
# creates an empty vector to store results
n4.wait.bar <- numeric(1000)

# A for loop that generates 1000 random samples
# Each size n=4, and calculates the sample mean.
for (i in 1:1000)
{
  n4.wait.sample <- rexp(4, 1/40) #Randomly picks 4 values from Exp(1/40)
  n4.wait.bar[i] <- mean(n4.wait.sample)
}

# Plot the sampling distribution
hist(n4.wait.bar, xlim = c(0, 100),
     xlab = "Mean Wait Time of Sample (in min)",
     main = "Sampling Distribution of Mean Wait Time for n=4",
     xaxt='n')
axis(1, at=seq(0, 100, 10), pos=0)
abline(v = 40, col = "red", lwd = 2, lty = 2)
```

Sampling Distribution of Mean Wait Time for



- c. In the R code block below, enter commands to compute the center (as measured by the mean) and spread (as measured by the standard deviation) of the sampling distribution when $n = 4$.

```
mean(n4.wait.bar)
```

```
## [1] 39.22552
```

```
sd(n4.wait.bar)
```

```
## [1] 19.72777
```

d. Repeat parts (b) and (c) if we increase the size of the samples from $n = 4$ to $n = 9$.

```
# creates an empty vector to store results
```

```
n9.wait.bar <- numeric(1000)
```

```
# A for loop that generates 1000 random samples
```

```
# Each size n=9, and calculates the sample mean.
```

```
for (i in 1:1000)
```

```
{
```

```
  n9.wait.sample <- rexp(9, 1/40) #Randomly picks 9 values from Exp(1/40)
```

```
  n9.wait.bar[i] <- mean(n9.wait.sample)
```

```
}
```

```
# Plot the sampling distribution
```

```
hist(n9.wait.bar, xlim = c(0, 100),
```

```
      xlab = "Mean Wait Time of Sample (in min)",
```

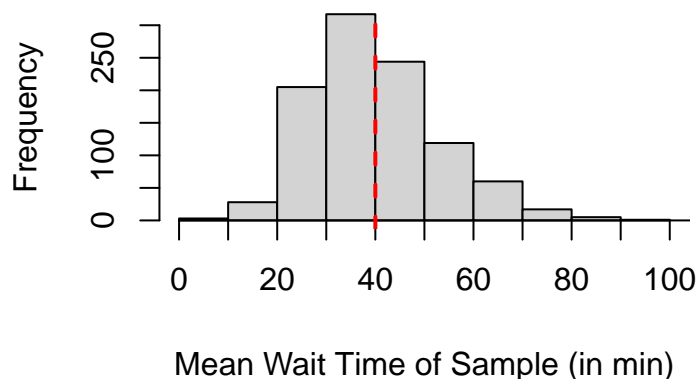
```
      main = "Sampling Distribution of Mean Wait Time for n=9",
```

```
      xaxt='n')
```

```
axis(1, at=seq(0, 100, 10), pos=0)
```

```
abline(v = 40, col = "red", lwd = 2, lty = 2)
```

Sampling Distribution of Mean Wait Time for



```
mean(n9.wait.bar)
```

```
## [1] 40.07623
```

```
sd(n9.wait.bar)
```

```
## [1] 13.34387
```

e. Repeat parts (b) and (c) if we increase the size of the samples to $n = 16$.

```
# creates an empty vector to store results
```

```
n16.wait.bar <- numeric(1000)
```

```
# A for loop that generates 1000 random samples
```

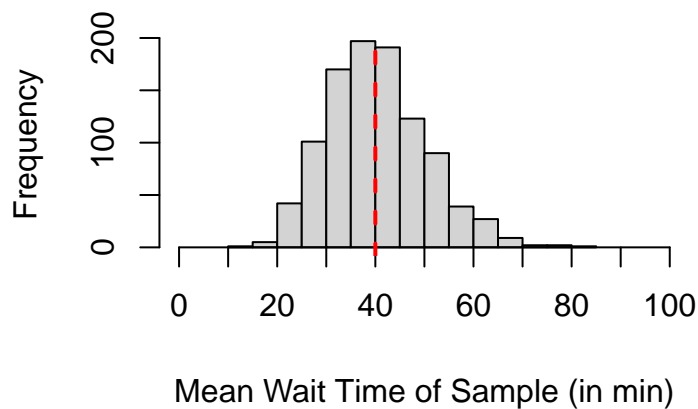
```

# Each size n=16, and calculates the sample mean.
for (i in 1:1000)
{
  n16.wait.sample <- rexp(16, 1/40) #Randomly picks 16 values from Exp(1/40)
  n16.wait.bar[i] <- mean(n16.wait.sample)
}

# Plot the sampling distribution
hist(n16.wait.bar, xlim = c(0, 100),
     xlab = "Mean Wait Time of Sample (in min)",
     main = "Sampling Distribution of Mean Wait Time for n=16",
     xaxt='n')
axis(1, at=seq(0, 100, 10), pos=0)
abline(v = 40, col = "red", lwd = 2, lty = 2)

```

Sampling Distribution of Mean Wait Time for



```
mean(n16.wait.bar)
```

```
## [1] 40.33276
```

```
sd(n16.wait.bar)
```

```
## [1] 10.04486
```

Question 3: Sampling from a Bimodal Population

The dataset **quakes** in R has the locations of 1000 seismic events that occurred near Fiji since 1964 with body wave magnitude (mb) > 4.0. Assume this data represents the population of all such earthquakes near Fiji since 1964. Let X denote the distribution of the depths (in km) where all such earthquakes occurred. Note this data is approximately **bimodal**.

- Calculate the mean and standard deviation of the depth of the population.

```
mean(quakes$depth)
```

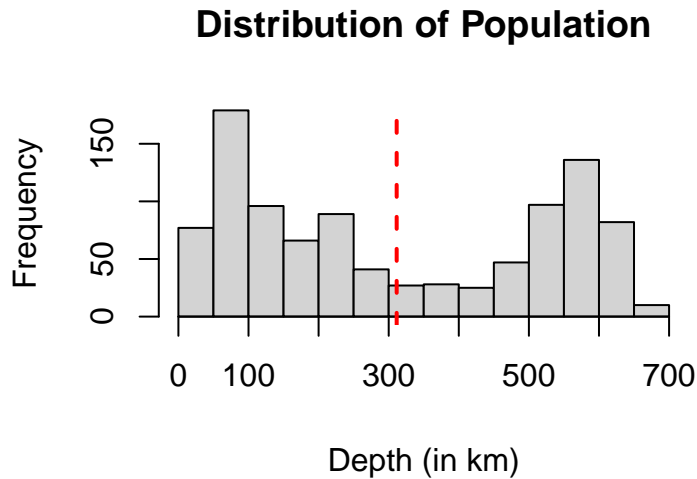
```
## [1] 311.371
```

```
sd(quakes$depth)
```

```
## [1] 215.5355
```

- b. Interpret and run the code below to create a histogram to show the distribution of the depths (variable is called **depth**) of all earthquakes in the dataset.

```
hist(quakes$depth,
     xlab = "Depth (in km)",
     main = "Distribution of Population",
     xaxt='n')
axis(1, at=seq(0, 700, 100), pos=0)
abline(v = mean(quakes$depth), col = "red", lwd = 2, lty = 2)
```



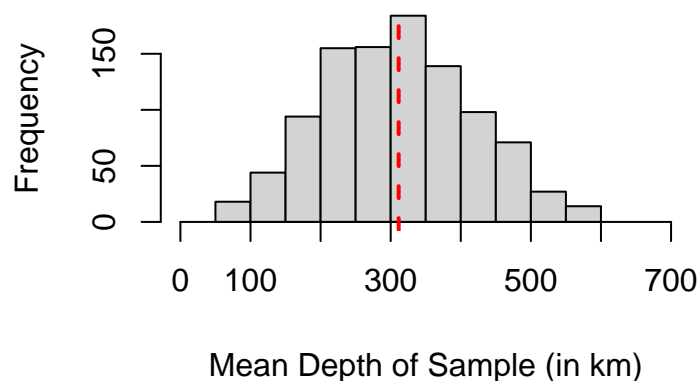
- c. A sample of $n = 4$ earthquakes are randomly selected and the mean depth of the quakes in the sample is calculated. This is repeated 1000 times (each sample size $n = 4$), and the sampling distribution for the mean depth can be constructed with the code below:

```
# creates an empty vector to store results
n4.quake.bar <- numeric(1000)

# A for loop that generates 1000 random samples
# Each size n=4, and calculates the sample mean.
for (i in 1:1000)
{
  n4.quake.sample <- sample(quakes$depth, 4, replace=FALSE) #Randomly picks 4 depths
  n4.quake.bar[i] <- mean(n4.quake.sample)
}

# Plot the sampling distribution
hist(n4.quake.bar, xlim = c(0, 700),
     xlab = "Mean Depth of Sample (in km)",
     main = "Sampling Distribution of Mean Depth for n=4",
     xaxt='n')
axis(1, at=seq(0, 700, 100), pos=0)
abline(v = mean(quakes$depth), col = "red", lwd = 2, lty = 2)
```


Sampling Distribution of Mean Depth for n



- d. In the R code block below, enter commands to compute the center (as measured by the mean) and spread (as measured by the standard deviation) of the sampling distribution when $n = 4$.

```
mean(n4.quake.bar)
```

```
## [1] 310.3752
```

```
sd(n4.quake.bar)
```

```
## [1] 107.1813
```

- e. Repeat parts (c) and (d) if we increase the size of the samples to $n = 9$.

```
# creates an empty vector to store results
```

```
n9.quake.bar <- numeric(1000)
```

```
# A for loop that generates 1000 random samples
```

```
# Each size n=9, and calculates the sample mean.
```

```
for (i in 1:1000)
```

```
{
```

```
  n9.quake.sample <- sample(quakes$depth, 9, replace=FALSE) #Randomly picks 9 depths
```

```
  n9.quake.bar[i] <- mean(n9.quake.sample)
```

```
}
```

```
# Plot the sampling distribution
```

```
hist(n9.quake.bar, xlim = c(0, 700),
```

```
  xlab = "Mean Depth of Sample (in km)",
```

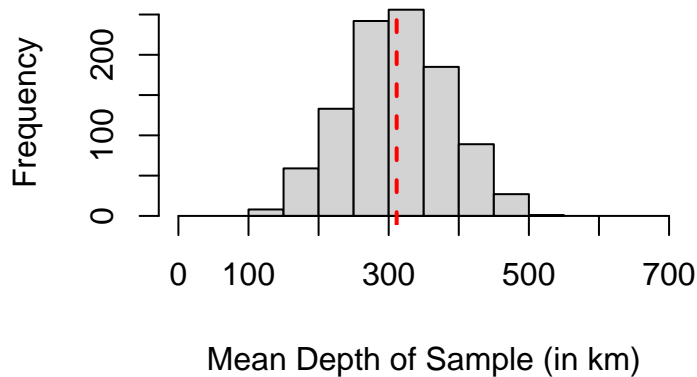
```
  main = "Sampling Distribution of Mean Depth for n=9",
```

```
  xaxt='n')
```

```
axis(1, at=seq(0, 700, 100), pos=0)
```

```
abline(v = mean(quakes$depth), col = "red", lwd = 2, lty = 2)
```

Sampling Distribution of Mean Depth for n



```
mean(n9.quake.bar)
```

```
## [1] 311.99
```

```
sd(n9.quake.bar)
```

```
## [1] 71.70165
```

f. Repeat parts (c) and (d) if we increase the size of the samples to $n = 16$.

```
# creates an empty vector to store results
```

```
n16.quake.bar <- numeric(1000)
```

```
# A for loop that generates 1000 random samples
```

```
# Each size n=16, and calculates the sample mean.
```

```
for (i in 1:1000)
```

```
{
```

```
  n16.quake.sample <- sample(quakes$depth, 16, replace=FALSE) #Randomly picks 16 depths
```

```
  n16.quake.bar[i] <- mean(n16.quake.sample)
```

```
}
```

```
# Plot the sampling distribution
```

```
hist(n16.quake.bar, xlim = c(0, 700),
```

```
      xlab = "Mean Depth of Sample (in km)",
```

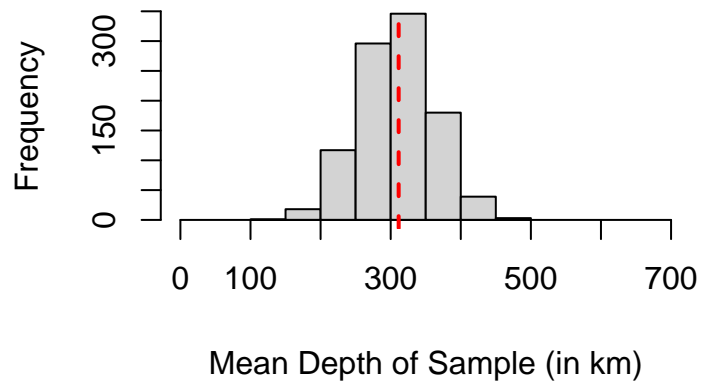
```
      main = "Sampling Distribution of Mean Depth for n=16",
```

```
      xaxt='n')
```

```
axis(1, at=seq(0, 700, 100), pos=0)
```

```
abline(v = mean(quakes$depth), col = "red", lwd = 2, lty = 2)
```

Sampling Distribution of Mean Depth for n:



```
mean(n16.quake.bar)
```

```
## [1] 308.6876
```

```
sd(n16.quake.bar)
```

```
## [1] 53.45576
```