Chapter 5

5.1

For a binomial with $n = 25$ and $p = 0.5$, determine:

a.) $P(\hat{p} \leq 15/25)$

```
pbinom(25 * 15/25, size = 25, prob = .5)
```

```
[1] 0.8852385
```

b.) $P(\hat{p} > 15/25)$

```
pbinom(15/25 * 25, size = 25, prob = .5, lower.tail = F)
```

```
[1] 0.1147615
```

c.) $P(10/25 \leq \hat{p} \leq 15/25)$

```
pbinom(25 * 15/25, size = 25, prob = .5) - pbinom(25 * 10/25, size = 25, prob = .5)
```

```
[1] 0.6730604
```

5.2

Many research teams intend to conduct a study regarding the proportion of people who have colon cancer. If a random sample of 10 individuals could be obtained, and if the probability of having colon cancer is 0.05, what is the probability that a research team will get $\hat{p} = 0.1$?

```
dbinom( 10 * .1, size = 10, prob = 0.05)
```

```
[1] 0.3151247
```

5.3

In the previous problem, what is the probability of $\hat{p} = 0.05$?

```
dbinom( 10 * .05, size = 10, prob = 0.05)
```

```
Warning in dbinom(10 * 0.05, size = 10, prob = 0.05): non-integer x = 0.500000
[1] 0
```

5.4

Someone claims that the probability of losing money, when using an investment strategy for buying and selling commodities, is 0.1. If this claim is correct, what is the probability of getting $\hat{p} \leq 0.05$ based on a random sample of 25 investments?

```
pbinom( 25 * .05, size = 25, prob = .1)
```

```
[1] 0.2712059
```

5.5

You interview a married couple and ask the wife whether she supports the current leader of their country. Her husband is asked the same question. Describe why it might be unreasonable to view these two responses as a random sample.

The liklihood of the two married people having the same political views is high.

5.6

Inagine that 1,000 research teams draw a random sample from a binomial distribution with $p = 0.4$, with each study based on a sample size of 30. So, this would result in 1,000 $\hat{p}$ values. If these 1,000 values were averages, what, approximately, would be the result?

```
p <- .4
n <- 30

N <- 1e3
result <- numeric(N)
for( i in 1:N )
{
   result[i] <- mean( rbinom(n, size = 1, prob = p) )
}

mean(result)
```

```
[1] 0.4034667
```

5.7

In the previous problem, if you computed the sample variance of the $\hat{p}$ values, what, approximatley, would be the result?

```
N <- 1e3
result <- numeric(N)
for( i in 1:N )
{
   result[i] <- mean( rbinom(n, size = 1, prob = p) )
```

```r
}

var(result) # estimated
```

```
[1] 0.008275821
```

```r
(p * (1 - p)) / n # exact
```

```
[1] 0.008
```

5.8

Suppose $n = 16, \sigma = 2, \mu = 30.$

Assume normality and determine:

a.) $P(\bar{X} \leq 29)$

```r
pnorm( 29, mean = 30, sd = sqrt(2^2 / 16))
```

```
[1] 0.02275013
```

b.) $P(\bar{X} \geq 30.5)$

```r
pnorm( 30.5, mean = 30, sd = sqrt(2^2 / 16), lower.tail = F)
```

```
[1] 0.1586553
```

c.) $P(29 \leq \bar{X} \leq 31)$

```r
pnorm(31, mean = 30, sd = sqrt(2^2 / 16)) - pnorm( 29, mean = 30, sd = sqrt(2^2 / 16))
```

```
[1] 0.9544997
```

5.9

Suppose $n = 25, \sigma = 5, \mu = 5$

Assume normality and determine:

a.) $P(\bar{X} \leq 4)$

```r
pnorm(4, mean = 5, sd = sqrt(5^2 / 25))
```

```
[1] 0.1586553
```

b.) $P(\bar{X} \leq 7)$

```r
pnorm(7, mean = 5, sd = sqrt(5^2 / 25), lower.tail = F)
```

```
[1] 0.02275013
```

c.) $P(3 \leq \bar{X} \leq 7)$

```
pnorm(7, mean = 5, sd = sqrt(5^2 / 25)) - pnorm(3, mean = 5, sd = sqrt(5^2 / 25))
```

```
[1] 0.9544997
```

5.10

Someone claims that within a certain neighborhood, the average cost of a house is $\mu = \$100,000$ with a standard deviation $\sigma = \$10,000$. Suppose that based on n = 16 homes, you find that the average cost of a house is $\bar{X} = \$95,000$. Assuming normality, what is the probability of getting a sample mean this low or lower if the claims about the mean and standard deviation are true?

```
pnorm(95e3, mean = 1e5, 1e4 / sqrt(16) )
```

```
[1] 0.02275013
```

5.11

In the previous problem, what is the probabiility of getting a sample mean between $97,500 and $102,500?

```
pnorm(1025e2, mean = 1e5, 1e4 / sqrt(16), lower.tail = T) - pnorm(975e2, mean = 1e5, 1e4 / sqrt
```

```
[1] 0.6826895
```

5.12

A company claims that the premiums paid by its clients for auto insurance has a normal distribution with mean $\mu = \$750$ dollars and standard deviation $\sigma = 100$.

Assuming normality, what is the probability that for $n = 9$ randomly sampled clients, the sample mean will have a value between $700 and $800?

```
s <- 100 / sqrt(9)

pnorm(800, mean = 750, sd = s) - pnorm(700, mean = 750, sd = s)
```

```
[1] 0.8663856
```

5.13

Imagine you are a health professional interested in the effects of medication on the diastolic blood pressure of adult women. For a particular drug being investigated, you find that for $n = 9$ women, the sample mean is $\bar{X} = 85$ and the sample variance is $s^2 = 160.78$.

Estimate the standard error of the sample mean, assuming random sampling.

```
sqrt(160.78 / 9)
```

```
[1] 4.226635
```

5.14

You randomly sample 16 observations from a discrete distribution with mean $\mu = 36$ and variance $\sigma^2 = 25$.

```
n <- 16; mu <- 36; variance <- 25

s <- sqrt(variance / n)
```

Use the central limit therom to determine:

a.) $P(\bar{X} \leq 34)$

```
pnorm(34, mean = mu, sd = s)
```

```
[1] 0.05479929
```

b.) $P(\bar{X} \leq 37)$

```
pnorm(37, mean = mu, sd = s)
```

```
[1] 0.7881446
```

c.) $P(\bar{X} \geq 33)$

```
pnorm(33, mean = mu, sd = s, lower.tail = F)
```

```
[1] 0.9918025
```

d.) $P(34 \leq \bar{X} \leq 37)$

```
pnorm(37, mean = mu, sd = s) - pnorm(34, mean = mu, sd = s)
```

```
[1] 0.7333453
```

5.15

You sample 25 observations from a nonnormal distribution with mean $\mu = 25$ and variance $\sigma^2 = 9$.

```
n <- 25; mu <- 25; var <- 9

s <- sqrt(var / n)
```

Use the central limit therom to determine:

a.) $P(\bar{X} \leq 24)$

```
pnorm(24, mean = mu, sd = s)
```

```
[1] 0.04779035
```

b.) $P(\bar{X} \leq 26)$

```
pnorm(26, mean = mu, sd = s)
```

```
[1] 0.9522096
```

c.) $P(\bar{X} \geq 24)$

```
pnorm(24, mean = mu, sd = s, lower.tail = F)
```

```
[1] 0.9522096
```

d.) $P(24 \leq \bar{X} \leq 26)$

```
pnorm(26, mean = mu, sd = s) - pnorm(24, mean = mu, sd = s)
```

```
[1] 0.9044193
```

5.16

Describe a situation where reliance on the central limit theorm to determine $P(\bar{X} \leq 24)$ might be unsatisfactory.

Skewed, heavy-tailed distributions (mixed normal).

5.17

Describe situations where a normal distribution provides a good approximation of the sampling distribution of the mean.

Symetric, light-tailed distributions.

5.18

For the values 4, 8, 23, 43, 12, 11, 32, 15, 6, 29, verify that the McKean-Schrader estimate of the standard error of the mean is 7.57.

```
x <- c(4, 8, 23, 43, 12, 11, 32, 15, 6, 29)

msmedse <- function(x) {
   n <- length(x)
   vals <- sort(x)

   k <- floor( (n + 1)/2 - 2.5758*sqrt(n/4) )
   m <- ((vals[n - k + 1] - vals[k]) / 5.1517)^2
```

```
  sqrt(m)
}

msmedse(x)
```

```
[1] 7.570317
```

5.19

In the previous exercise, how would you argue that the method used to estimate the standard error of the median is a reasonable approach?

There are not any tied values.

5.20

For the values 5, 7, 2, 3, 4, 5, 2, 6, 7, 3, 4, 6, 1, 7, 4, verify that the McKean-Schrader estimate of the standard error of the median is 0.97.

```
x <- c(5, 7, 2, 3, 4, 5, 2, 6, 7, 3, 4, 6, 1, 7, 4)
msmedse(x)
```

```
[1] 0.9705534
```

5.21

In the previous excersize, how would you argue that the method used to estimate the standard error of th median might be highly inaccurate?

Several tied values.

5.22

In Exercise 20, would it be advisable to approximate $P(M \leq 4)$, the probability that the sample median is less than or equal to 4, using Equation (5.7)?

No, there are ties.

5.23

For the values, 2, 3, 5, 6, 8, 12, 14, 18, 19, 22, 201, why would you suspect that the McKean-Schrader estimate of the standard error of the median will be smaller than the standard error of the mean?

There is an outlier.

```
x <- c(2, 3, 5, 6, 8, 12, 14, 18, 19, 22, 201)

msmedse(x)
```

```
[1] 38.62803
```

```
sqrt(var(x) / length(x))
```

```
[1] 17.4021
```

5.24

Summarize when it would and would not be reasonable to assume that the sampling distribution of M, the sample median, is normal.

When the distribution is not symetric the sampling distribution of M will not be normal.

5.25

For the values:

59, 106, 174, 207, 219, 237, 313, 365, 485, 497, 515, 529, 557, 615, 625, 645, 973, 1065, 3215

Estimate the standard error of the 20% trimmed mean.

```
x <- c(59, 106, 174, 207, 219, 237, 313, 365, 485, 497, 515, 529, 557, 615, 625, 645, 973, 1065

winsorize <- function( x, trim = .2) {
   n <- length(x)

   o <- sort(x)
   g <- floor(trim*n)

   o[1:(g+1)] <- o[(g+1)]
   o[(n-g):n] <- o[n-g]

   o
}

n <- length(x)

sqrt( var(winsorize(x)) / ( .36 * n ) )
```

```
[1] 69.48348
```

5.26

For the data in Exercise 25, why would you suspect that the standard error of the sample mean will be larger than the standard error of the 20% trimmed mean?

Because the mean doesn't drop any of the larger observations.

```r
sd(x) / sqrt(n)
```

```
[1] 157.8508
```

Verify this speculation.

5.27

The ideal estimator of location would have a smaller standard error than any other estimator we might use. Explain why such an estimator does not exist.

5.28

Under normality, the sample mean has a smaller standard error than the 20% trimmed mean or median. If observations are sampled from a distribution that appears to be normal, does this suggest that the mean should be preferred over the trimmed mean and median?

5.29

If the sample mean and 20% trimmed mean are nearly identical, it might be thought that for future studies, it will make little difference which measure of location is used. Comment on why this is not the case.

5.30

Observations can be generated from the distribution in Figure 5.4a using the R command rlnorm. for example, rlnorm(25) would generate 25 observations. Indicate the R commands for determining the sampling distribution of the sample median, M. In particular, how would you determine the probability that M will have a value less than 1.5?