

ISQA 8160 Exam II

Brian Detweiler

Thursday, July 28, 2016

Chapter 3.1

3.1 Problem 6.) A civic group reported to the town council that at least 60% of the town residents were in favor of a particular bond issue. Forty-eight said yes. Is the report of the civic group reasonable?

We have the following:

$$p = 0.60$$

$$p^* = 0.48$$

$$n = 100$$

$$\alpha = 0.05$$

The null hypothesis is that the sample probability is representative of a population probability of 60%. The alternative hypothesis is that they are not equal.

$$H_0 : p = p^*$$

$$H_a : p \neq p^*$$

```
library(binom)
binom.test(x=48, n=100, p=0.60, alternative=c('two.sided'), conf.level=0.95)
```

```
##
## Exact binomial test
##
## data: 48 and 100
## number of successes = 48, number of trials = 100, p-value =
## 0.01844
## alternative hypothesis: true probability of success is not equal to 0.6
## 95 percent confidence interval:
## 0.3790055 0.5822102
## sample estimates:
## probability of success
## 0.48
```

And thus we reject the null hypothesis with a p-value of 0.01844 and a 95% confidence interval of (0.3790055, 0.5822102). The population mean appears to be less than 60%.

■

3.1 Problem 7.) Out of 20 recent takeover attempts, 5 were successfully resisted by the companies being taken over. Assume these are independent events, and estimate the probability of a takeover attempt being successfully resisted. That is, find a 95% confidence interval.

Setting up this problem we have the following values:

$$n = 20$$

$$x = 5$$

We can find the confidence interval using the `binom.test` like before:

```
binom.test(x=5, n=20, alternative=c('two.sided'), conf.level=0.95)

##
## Exact binomial test
##
## data: 5 and 20
## number of successes = 5, number of trials = 20, p-value = 0.04139
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
## 0.08657147 0.49104587
## sample estimates:
## probability of success
## 0.25
```

(a) Use Table A4.

Table A4 confirms the 95% confidence interval as (0.087, 0.491).

(b) Use Table A1.

Although $n = 20$ is a smaller sample size than we generally require for using the normal distribution, we can nevertheless use it to obtain a slightly less accurate answer:

$$L = \frac{Y}{n} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{Y(n-Y)}{n^3}}$$

$$U = \frac{Y}{n} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{Y(n-Y)}{n^3}}$$

The z-value at $z_{1-\frac{0.05}{2}} = z_{0.975}$ is 1.96. Plugging in our values, we get

```
L <- (5/20) - 1.96 * sqrt((20 * (20 - 5)) / (20^3))
U <- (5/20) + 1.96 * sqrt((20 * (20 - 5)) / (20^3))
print(paste(paste(paste(paste('(', L), ', ', U), ')'))
```

```
## [1] "( -0.129552367928327 , 0.629552367928327 )"
```

Clearly, with such a small sample size, the normal approximation will not return accurate results.



Chapter 3.2

3.2 Problem 1.) A random sample of tenth-grade boys resulted in the following 20 observed weights.

142	134	98	119	131
103	154	122	93	137
86	119	161	144	158
165	81	117	128	103

Test the hypothesis that the median weight is 103.

Let's first order the data.

```
weights <- sort(c(142, 134, 98, 119, 131,
                 103, 154, 122, 93, 137,
                 86, 119, 161, 144, 158,
                 165, 81, 117, 128, 103))
weights.df <- as.data.frame(weights)
weights.df
```

```
##      weights
## 1         81
## 2         86
## 3         93
## 4         98
## 5        103
## 6        103
## 7        117
## 8        119
## 9        119
## 10       122
## 11       128
## 12       131
## 13       134
## 14       137
## 15       142
## 16       144
## 17       154
## 18       158
## 19       161
## 20       165
```

We will test the hypothesis that $x_{0.50} = 103$ with $\alpha = 0.05$.

$$H_0 : p = p^*$$

$$H_a : p \neq p^*$$

$$x^* = 103$$

$$T_1 = \text{number of } x_i \leq x^*$$

$$T_2 = \text{number of } x_i < x^*$$

```
T1 <- length(weights.df[weights.df$weights <= 103, ])
T1
```

```
## [1] 6
```

```
T2 <- length(weights.df[weights.df$weights < 103, ])
T2
```

```
## [1] 4
```

Since $n \leq 20$, the critical region is $c = \{T_1 \leq t_1, T_2 \geq t_2\}$, such that $P(y \leq t_1) = \frac{\alpha}{2}$ and $P(y \leq t_2) = 1 - \frac{\alpha}{2}$.

$T_1 = \text{number of } x_i \leq 103 = 6$

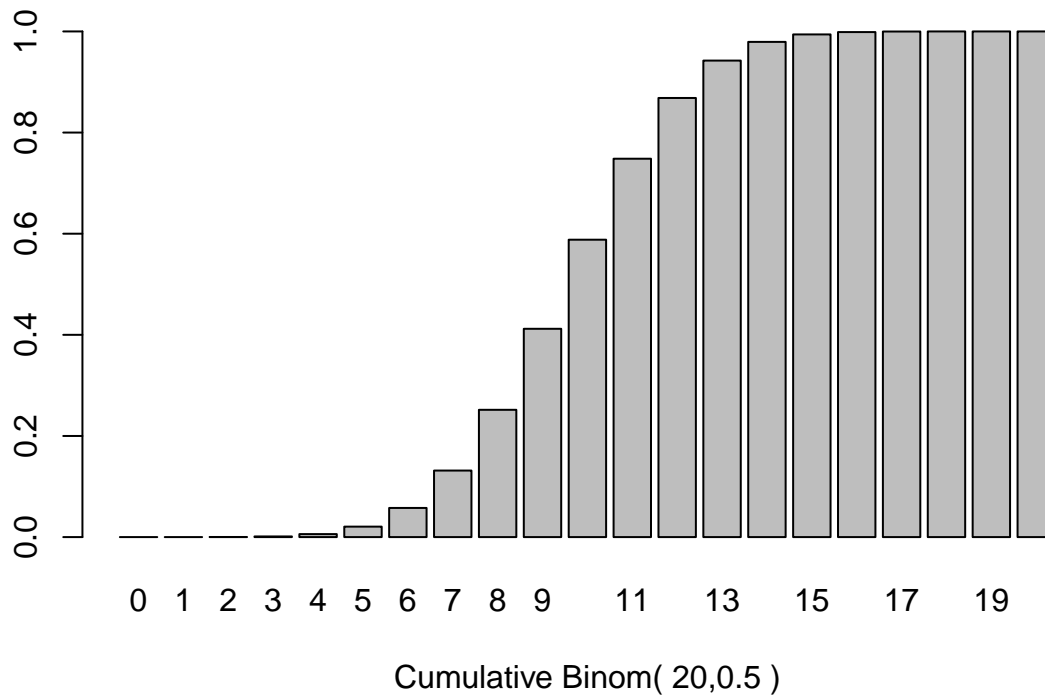
$T_2 = \text{number of } x_i < 103 = 4$

$c = \{6 \leq t_1 \text{ or } 4 > t_2\}$

$P(y \leq t_1) = 0.025$

$P(y \leq t_2) = 0.975$

```
trials <- 20
probability <- 0.50
alpha <- 0.05
left.crit.region <- alpha / 2
right.crit.region <- 1 - (alpha / 2)
```



```

# binomsum(start, end, n, p)
# Gives a cumulative sum of the binomial from start to end, for n trials with p probability
binomsum <- function(start, end, n, p) sum(dbinom(x=c(start:end), size=n, prob=p))

# cumsums is a list of the cumulative probability at each point
cumsums <- list(mapply(binomsum, 0, c(0:trials), trials, probability))

# gives a TRUE/FALSE list where values are <= alpha/2
left.tail <- lapply(cumsums, function(x) x <= left.crit.region)

# gives a TRUE/FALSE list where values are >= 1 - (alpha / 2)
right.tail <- lapply(cumsums, function(x) x >= right.crit.region)

# Table sums up values
left.tail.table <- table(left.tail)
right.tail.table <- table(right.tail)

# Note that zero is counted in here, so we must subtract 1, otherwise we get
# the value that sits outside outside of the range
t1 <- as.data.frame(left.tail.table[names(left.tail.table) == T])[1, 1] - 1
t1

```

```
## [1] 5
```

```

# Zero is counted in here, so we need to add one to get the value we're looking for
# Also, recall we want t2 s.t.  $P(y \leq t2) = 1 - \alpha/2$ , or  $P(y \leq t2) = 0.975$ 
# In other words, this will be the majority of the distribution
t2 <- trials - as.data.frame(right.tail.table[names(right.tail.table) == T])[1, 1] + 1
t2

```

```
## [1] 14
```

```

T1.prob <- pbinom(q=T1, size=trials, prob=probability)
T1.prob

```

```
## [1] 0.05765915
```

```

T2.prob <- 1 - pbinom(q=T2, size=trials, prob=probability)
T2.prob

```

```
## [1] 0.994091
```

Now it is clear,

$$\begin{aligned}
 T_1 &= 6 \\
 T_2 &= 4 \\
 t_1 &= 5 \\
 t_2 &= 14 \\
 c &= \{6 \not\leq 5 \text{ or } 4 \not\geq 14\}
 \end{aligned}$$

So we cannot reject the null hypothesis with a p-value of

```
min(2*T1.prob, 2*T2.prob)
```

```
## [1] 0.1153183
```



2.) In Exercise 1 test the hypothesis that the upper quartile is at least 150.

We will test the hypothesis that $x_{0.75} = 150$ with $\alpha = 0.05$ using a left-tailed test.

$$\begin{aligned}H_0 &: p \geq p^* \\H_a &: p < p^* \\c &= \{T_1 : T_1 \leq t_1\}\end{aligned}$$

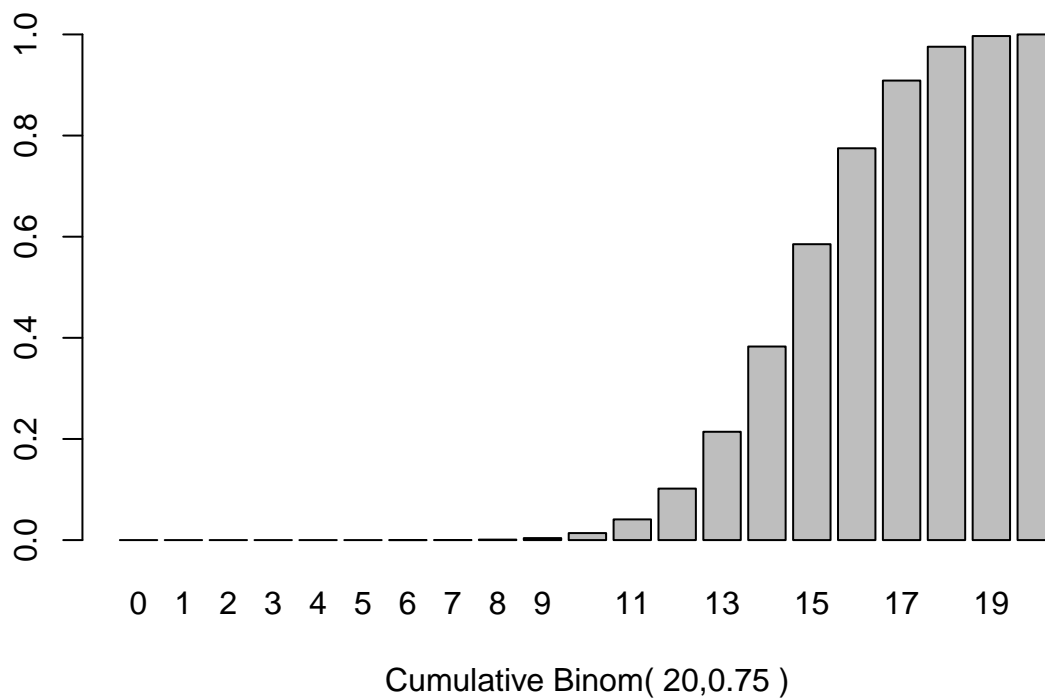
```
T1 <- length(weights.df[weights.df$weights <= 150, ])  
T1
```

```
## [1] 16
```

Since $n \leq 20$, the critical region is $c = \{T_1 \leq t_1\}$, such that $P(y \leq t_1) = \alpha$.

$$\begin{aligned}T_1 &= \text{number of } x_i \leq 150 = 16 \\c &= \{16 \leq t_1\} \\P(y \leq t_1) &= 0.05\end{aligned}$$

```
trials <- 20  
probability <- 0.75  
alpha <- 0.05  
left.crit.region <- alpha
```



```

# binomsum(start, end, n, p)
# Gives a cumulative sum of the binomial from start to end, for n trials with p probability
binomsum <- function(start, end, n, p) sum(dbinom(x=c(start:end), size=n, prob=p))

# cumsums is a list of the cumulative probability at each point
cumsums <- list(mapply(binomsum, 0, c(0:trials), trials, probability))

# gives a TRUE/FALSE list where values are <= alpha/2
left.tail <- lapply(cumsums, function(x) x <= left.crit.region)

# Table sums up values
left.tail.table <- table(left.tail)

# Note that zero is counted in here, so we must subtract 1, otherwise we get
# the value that sits outside outside of the range
t1 <- as.data.frame(left.tail.table[names(left.tail.table) == T])[1, 1] - 1
t1

```

```
## [1] 11
```

```

T1.prob <- pbinom(q=T1, size=trials, prob=probability)
T1.prob

```

```
## [1] 0.774844
```

Now it is clear,

$$\begin{aligned}
 T_1 &= 16 \\
 t_1 &= 11 \\
 c &= \{16 \not\leq 11\}
 \end{aligned}$$

So we cannot reject the null hypothesis with a p-value of

```
T1.prob
```

```
## [1] 0.774844
```



3.) In Exercise 1 test the hypothesis that the third decile is no greater than 100.

We will test the hypothesis that $x_{0.30} \leq 100$ with $\alpha = 0.05$ using a right-tailed test.

$$\begin{aligned}H_0 &: p \leq p^* \\H_a &: p > p^* \\c &= \{T_2 : T_2 \leq t_2\}\end{aligned}$$

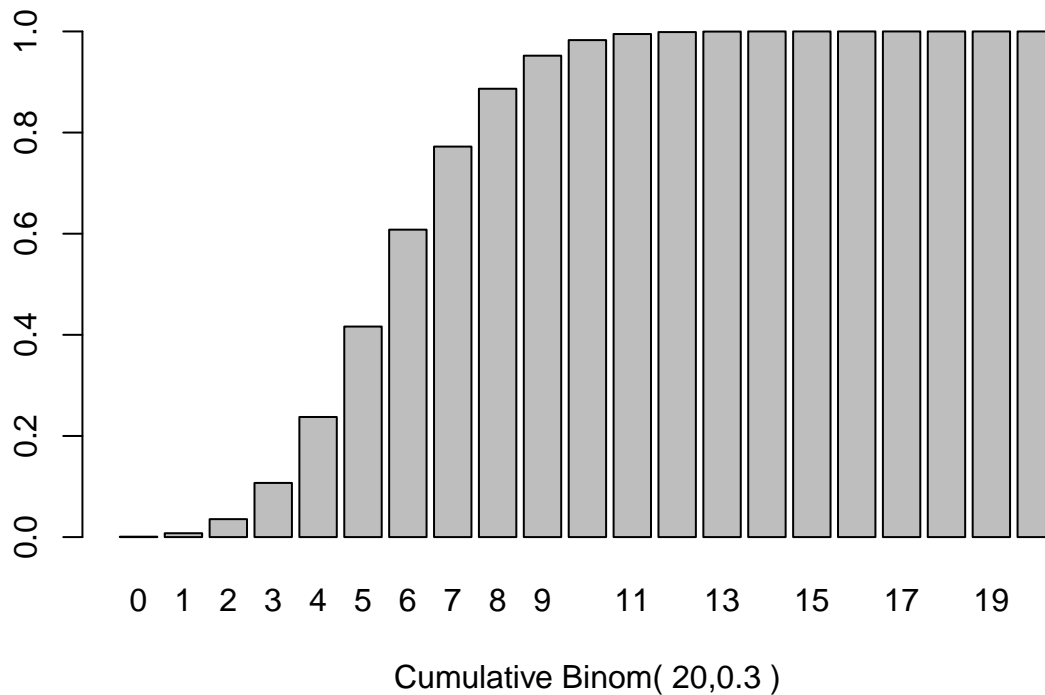
```
T2 <- length(weights.df[weights.df$weights < 100, ])  
T2
```

```
## [1] 4
```

Since $n \leq 20$, the critical region is $c = \{T_2 \geq t_2\}$, such that $P(y \geq t_2) = 1 - \alpha$.

$$\begin{aligned}T_2 &= \text{number of } x_i < 100 = 4 \\c &= \{4 > t_2\} \\P(y \leq t_2) &= 0.95\end{aligned}$$

```
trials <- 20  
probability <- 0.30  
alpha <- 0.05  
right.crit.region <- 1 - alpha
```



```

# binomsum(start, end, n, p)
# Gives a cumulative sum of the binomial from start to end, for n trials with p probability
binomsum <- function(start, end, n, p) sum(dbinom(x=c(start:end), size=n, prob=p))

# cumsums is a list of the cumulative probability at each point
cumsums <- list(mapply(binomsum, 0, c(0:trials), trials, probability))

# gives a TRUE/FALSE list where values are >= 1 - (alpha / 2)
right.tail <- lapply(cumsums, function(x) x >= right.crit.region)

# Table sums up values
right.tail.table <- table(right.tail)

# Zero is counted in here, so we need to add one to get the value we're looking for
# Also, recall we want t2 s.t. P(y <= t2) = 1 - alpha/2, or P(y <= t2) = 0.975
# In other words, this will be the majority of the distribution
t2 <- trials - as.data.frame(right.tail.table[names(right.tail.table) == T])[1, 1] + 1
t2

```

```
## [1] 9
```

```

T2.prob <- 1 - pbinom(q=T2, size=trials, prob=probability)
T2.prob

```

```
## [1] 0.7624922
```

$$\begin{aligned}
 T_2 &= 9 \\
 t_2 &= 10 \\
 c &= \{9 \not\geq 10\}
 \end{aligned}$$

So we cannot reject the null hypothesis with a p-value of

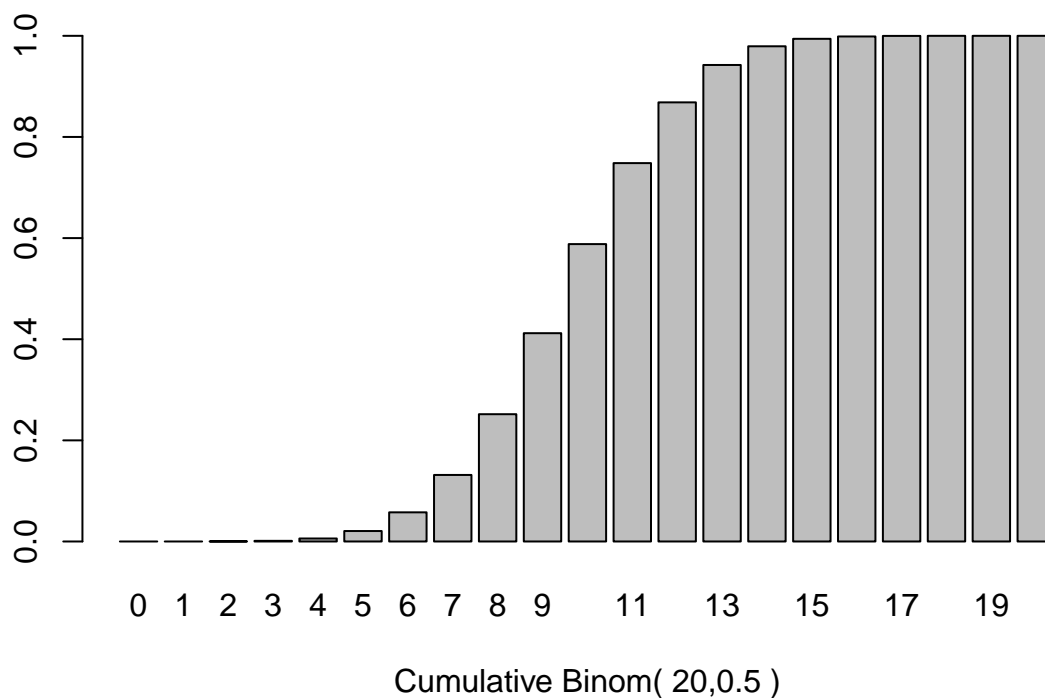
```
T2.prob
```

```
## [1] 0.7624922
```

■

4.) In Exercise 1 find an approximate 90% confidence interval for the median. What is the exact confidence coefficient? Also compare the results using the exact method with the results obtained using the large sample approximation.

```
trials <- 20
probability <- 0.50
alpha <- 0.10
left.crit.region <- alpha / 2
right.crit.region <- 1 - (alpha / 2)
```



```
# binomsum(start, end, n, p)
# Gives a cumulative sum of the binomial from start to end, for n trials with p probability
binomsum <- function(start, end, n, p) sum(dbinom(x=c(start:end), size=n, prob=p))

# cumsums is a list of the cumulative probability at each point
cumsums <- list(mapply(binomsum, 0, c(0:trials), trials, probability))

# gives a TRUE/FALSE list where values are <= alpha/2
left.tail <- lapply(cumsums, function(x) x <= left.crit.region)

# gives a TRUE/FALSE list where values are >= 1 - (alpha / 2)
right.tail <- lapply(cumsums, function(x) x >= right.crit.region)

# Table sums up values
```

```
left.tail.table <- table(left.tail)
right.tail.table <- table(right.tail)
```

```
# Note that zero is counted in here, so we must subtract 1, otherwise we get
# the value that sits outside outside of the range
r <- as.data.frame(left.tail.table[names(left.tail.table) == T])[1, 1] - 1
r
```

```
## [1] 5
```

```
# Zero is counted in here, so we need to add one to get the value we're looking for
# Also, recall we want t2 s.t.  $P(y \leq t2) = 1 - \alpha/2$ , or  $P(y \leq t2) = 0.975$ 
# In other words, this will be the majority of the distribution
s <- trials - as.data.frame(right.tail.table[names(right.tail.table) == T])[1, 1] + 1
s
```

```
## [1] 14
```

```
conf.coef <- pbinom(q=r, size=trials, p=probability) + (1 - pbinom(q=s, size=trials, p=probability))
conf.coef
```

```
## [1] 0.04138947
```

But r and s are the values *below* 0.05 and above 0.95 respectively. They may not be the closest values. We'll test for these by hand (or we could look them up in the table, by why bother when we have R?)

```
abs(0.05 - pbinom(q=r, size=trials, p=probability))
```

```
## [1] 0.02930527
```

```
abs(0.05 - pbinom(q=r+1, size=trials, p=probability))
```

```
## [1] 0.007659149
```

```
abs(0.95 - pbinom(q=s, size=trials, p=probability))
```

```
## [1] 0.02930527
```

```
abs(0.95 - pbinom(q=s-1, size=trials, p=probability))
```

```
## [1] 0.007659149
```

So we can see that $r + 1 = 6$ and $s - 1 = 13$ are the closest values to a 90% confidence coefficient.

The actual value of the confidence coefficient is

```
conf.coef <- pbinom(q=r+1, size=trials, p=probability) + (1 - pbinom(q=s-1, size=trials, p=probability))
conf.coef
```

```
## [1] 0.1153183
```

This gives us values of

```
weights[r + 1]
```

```
## [1] 103
```

```
weights[s - 1]
```

```
## [1] 134
```

We can see that $r + 1$ and $s - 1$ are the values of the order statistics, which we found in the *weights* list.

So our approximate 90% confidence interval for $x_{0.50}$ is (103, 134).

Finding the large sample approximation, we get

$$\begin{aligned} r^* &= np^* + z_{\frac{\alpha}{2}} \sqrt{np^*(1-p^*)} \\ &= 20(0.50) + 1.96 \sqrt{20(.5)(1-.5)} \\ &= 10 + 4.38 \\ &= \lceil 14.4 \rceil \\ &= 15 \end{aligned}$$

$$\begin{aligned} s^* &= np^* + z_{1-\frac{\alpha}{2}} \sqrt{np^*(1-p^*)} \\ &= 20(0.50) - 1.96 \sqrt{20(.5)(1-.5)} \\ &= 20(0.50) - 1.96 \sqrt{20(.5)(1-.5)} \\ &= 10 - 4.38 \\ &= \lceil 5.62 \rceil \\ &= 6 \end{aligned}$$

```
weights[r]
```

```
## [1] 103
```

```
weights[s]
```

```
## [1] 137
```

And so our large sample approximation for a 90% confidence interval is (103, 137)

■