

Interval Estimation and Coverage Probabilities

Contents

4.3 Interval Estimation and Coverage Probabilities	1
4.3.1 Revisiting Sampling Distributions	1
4.3.2 Random vs. Observed Intervals	4
Random Intervals	6
Observed Intervals	7
Example: Coverage Calculation via Simulation	8
4.3.3 Student t Based Intervals	10
Standard Error vs. Standard Deviation	10
Example: Coverage Calculation via Simulation	14
4.3.4 Connection to Hypothesis Testing	17

4.3 Interval Estimation and Coverage Probabilities

4.3.1 Revisiting Sampling Distributions

- Recall that when we look at $a(\mathcal{S})$ for all possible samples \mathcal{S} of some size n from a population \mathcal{P} that the values of $a(\mathcal{S})$ have a distribution.
 - We called this the sampling distribution of $\tilde{a}(\mathcal{S})$.
- For example, we looked at the Australian shark encounter population ($N = 28$), and we considered all possible samples of size $n = 5$.
 - We calculated $a(\mathcal{S}) = \bar{y}$, the average shark length, for each sample and then constructed a histogram.

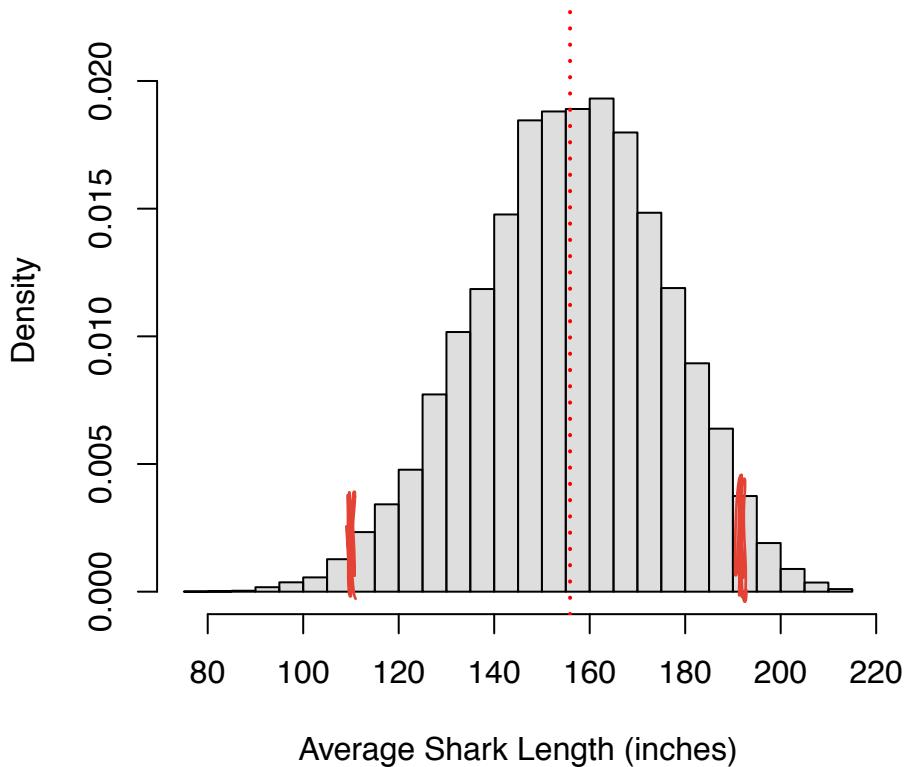
```
sharks <- read.csv("/Users/nstevens/Dropbox/Teaching/STAT_341/Lectures//Data/sharks.csv")
popSharks <- rownames(sharks)
popSharksAustralia <- popSharks[sharks$Australia == 1]

samples <- combn(popSharksAustralia, 5)
N_s <- ncol(samples)

avePop <- mean(sharks[popSharksAustralia, "Length"])
avesSamp <- apply(samples, MARGIN = 2,
                  FUN = function(s){mean(sharks[s, "Length"])})

## The plot
hist(avesSamp, col=adjustcolor("grey", alpha = 0.5),
      freq = FALSE,
      main="Sampling Distribution (n=5)",
      xlab="Average Shark Length (inches)",
      ylim=c(0, 0.022),
      breaks=25)
```

Sampling Distribution (n=5)



- The histogram is fairly symmetric and centered about the population average $a(\mathcal{P}) = 155.89$
 - The 2.5th and 97.5th quantiles of this distribution were 115.4 and 192.8
 - Thus we know that the interval (115.4, 192.8) contains the 95% of the possible sample averages.
- Quantifying the spread in this way is very useful – but in order to do so exactly we require having observed **all possible samples**
 - Alternatively, we could instead consider a large number of possible samples to approximate the sampling distribution
 - Or, we could approximate the sampling distribution via **resampling techniques** such as the **bootstrap** (see Sections 4.4 and 4.5)
 - Or, we could approximate the sampling distribution via the **normal distribution** to simplify calculations
- The normal distribution that best approximates the sampling distribution is the one with mean and standard deviation equal to the mean, `mean(avesSamp) = 155.89` and standard deviation, `sd(avesSamp) = 19.8`, from all possible $a(\mathcal{S})$ values.
 - The plots below evaluate the adequacy of this approximation.

```
par(mfrow=c(1,2))
hist(avesSamp, col=adjustcolor("grey", alpha = 0.5), freq = FALSE,
     main="Sampling Distribution (n=5)",
     xlab="Average Shark Length (inches)",
```

```

    ylim=c(0, 0.022),
    breaks=25
)
### Mark the population attribute in red
abline(v=avePop, col="red", lty=3, lwd=2)

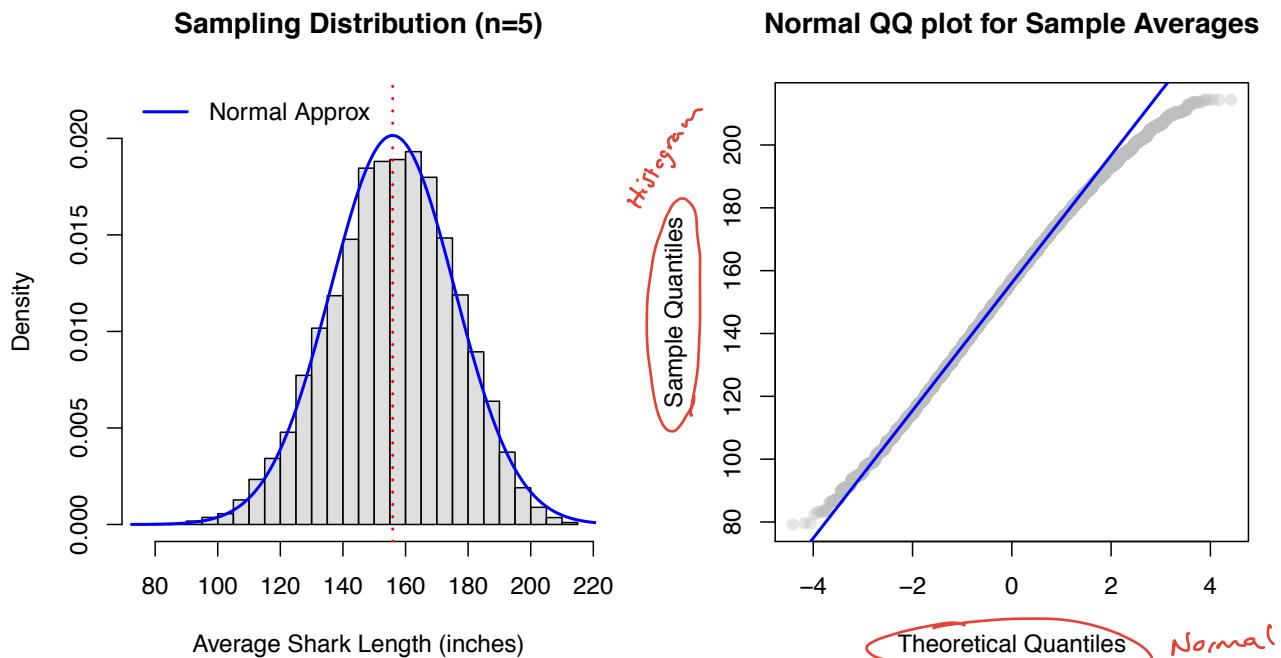
sdN = function(x){sqrt(var(x)*(length(x)-1)/length(x))}

### Add a Gaussian density
tmpAve <- mean(avesSamp)
tmpSD <- sdN(avesSamp)
tmpX <- extendrange(avesSamp)
tmpX <- seq(tmpX[1], tmpX[2], length.out = 200)
lines(tmpX, dnorm(tmpX, mean = tmpAve, sd = tmpSD), lwd = 2, col = "blue")

legend("topleft", legend = c("Normal Approx"), lwd = 2,
       col = "blue", lty = 1, bty = "n")

###Drawing the qqplot
qqnorm(avesSamp, main='Normal QQ plot for Sample Averages', col=adjustcolor("Grey", 0.4), pch=19)
qqline(avesSamp, lwd = 2, col = "blue", xlab = "Normal Quantiles")

```



- The plots suggest that the normal approximation is pretty good, except for in the tails.
 - This is apparent in both the histogram and the QQ-plot.
 - The closeness of this approximation is unsurprising – recall the **central limit theorem (CLT)** and its implications for distributions of averages

* The normal approximation provides a model for the sampling distribution and can be used as a basis to construct confidence intervals for population averages.

4.3.2 Random vs. Observed Intervals

- Suppose the attribute of interest is the population average $a(\mathcal{P}) = \mu = \frac{1}{N} \sum_{u \in \mathcal{P}} y_u = \bar{y}$
- Recall that the estimator $\tilde{a}(\mathcal{P}) = \tilde{\mu} = \bar{Y}$ (a random variable) has the following properties:

$$\underbrace{E[\bar{Y}] = \mu}_{\text{and}} \quad \underbrace{Var[\bar{Y}] = \frac{\sigma^2}{n}}$$

where

$$\sigma^2 = \frac{1}{N} \sum_{u \in \mathcal{P}} (y_u - \mu)^2$$

is the population variance.

- If the normality assumption holds (this may be appropriate due to the CLT) then the estimator $\tilde{a}(\mathcal{P}) = \bar{Y}$ has the following distribution:

$$\bar{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- In the case of the Australian shark data

- $N = 28$ and $n = 5$
- $\mu = a(\mathcal{P}) = 155.89$ is the population average
- $\sigma = 49.74$ is the population standard deviation
- $SD[\bar{Y}] = \frac{\sigma}{\sqrt{n}} = 19.8$.

- Standardizing this random variable yields

$$Z = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

- Note:** Strictly speaking we are dealing with a finite population and so our variance term should include the **finite population correction**:

$$Var[\bar{Y}] = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)$$

(Recall that we derived this in the context of the variance of the Horvitz-Thompson estimator under SRSWOR)

- However, when $N \gg n$ then $\frac{N-n}{N-1} \rightarrow 1$ in which case this term can be ignored
- This is why most texts omit the correction factor, giving rise to the more familiar formulation $Var[\bar{Y}] = \frac{\sigma^2}{n}$
- For completeness we will henceforth include the finite population correction, in which case

$$\bar{Y} \sim N\left(\mu, \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)\right)$$

and

$$Z = \frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}} \sim N(0, 1)$$

Pivotal Quantity

- A histogram and corresponding $N(0, 1)$ curve for this standardized quantity is visualized below:

```
Z = (avesSamp - mean(avesSamp))/sdN(avesSamp)
```

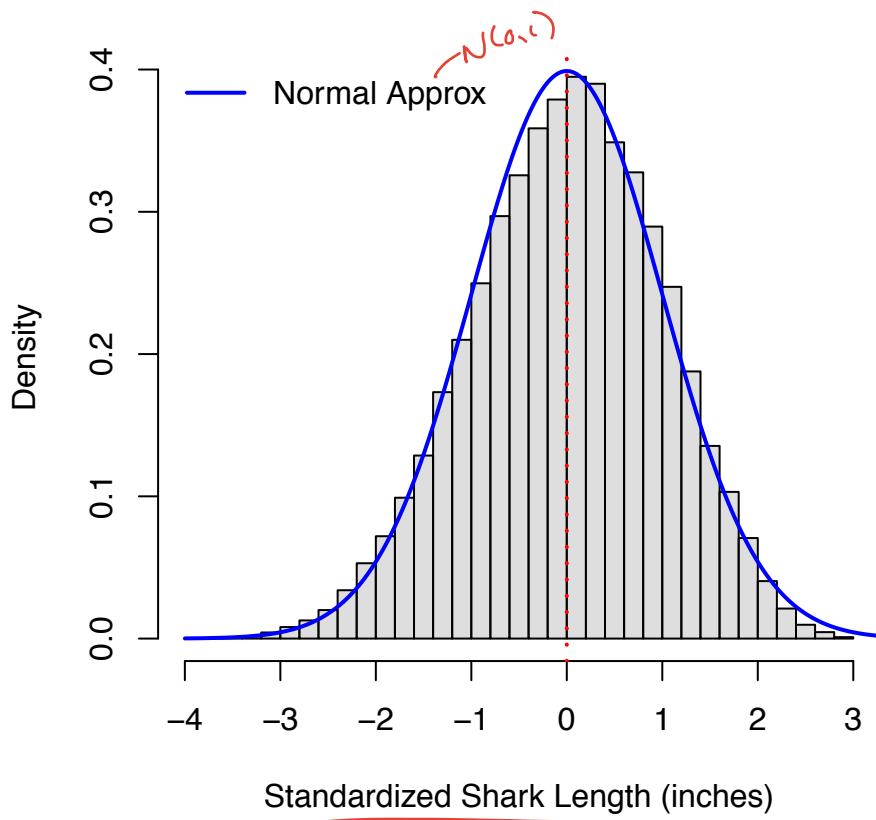
```
hist(Z, col=adjustcolor("grey", alpha = 0.5),
      freq = FALSE,
      main="Sampling Distribution (n = 5)",
      xlab="Standardized Shark Length (inches)",
      breaks=25 )

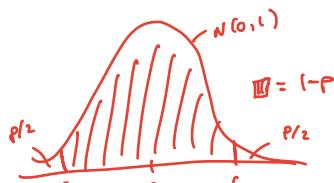
### Mark the population attribute in red
abline(v=0, col="red", lty=3, lwd=2)

### Add a Gaussian density
x = seq(-4, 4, length.out=1000)
lines( x, dnorm(x) , col = "blue", lwd = 2)

legend("topleft", legend = c("Normal Approx"), lwd = 2,
       col = "blue", lty = 1, bty = "n")
```

Sampling Distribution (n = 5)





Random Intervals

- Using the standardized random variable and specified $p \in (0, 1)$ we can find a constant $c > 0$ such that

$$\begin{aligned} 1 - p &= \Pr(-c \leq Z \leq c) \\ 1 - p &= \Pr\left(-c \leq \frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}} \leq c\right) \quad \text{Going from here to here requires a pivot} \\ &= \Pr\left(\bar{Y} - c \times \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \leq \mu \leq \bar{Y} + c \times \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}\right). \end{aligned}$$

- Rearranging this statement yields a random interval

$$\left[\underbrace{\bar{Y} - c \times \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}}_{\text{---}}, \underbrace{\bar{Y} + c \times \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}}_{\text{---}} \right]$$

which contains μ with probability $1 - p$.

- This interval is considered random because it's never actually observed – it is defined in terms of random variables

* However, observed intervals calculated by substituting \bar{Y} with \bar{y} are guaranteed to contain μ , $100(1-p)\%$ of the time.

- $1 - p$ is therefore called the coverage probability.
- μ is contained in (or covered by) such an interval $100(1 - p)\%$ of the time.
- All such intervals have the same width, they just different (random) centres.

A Note on determining c :

- The normal distribution is symmetric about its mean μ , so p and c are related through

$$1 - p = \Pr(-c \leq Z \leq c)$$

or, equivalently,

$$1 - p/2 = \Pr(Z \leq c)$$

where $Z \sim N(0, 1)$ is a standard normal random variable.

- Therefore, given any $p \in (0, 1)$ the value of c can determine from the quantile function of a standard normal random variable:

$$c = Q_Z\left(1 - \frac{p}{2}\right)$$

which in R is calculated as qnorm(1 - p/2).

- e.g., $c \approx 1.96$ when $1 - p = 0.95$ for a standard normal random variable.

* $\Phi(\cdot)$ is the standard normal CDF.

Observed Intervals

- In practice, we will have only one sample
 - And thus a single numerical average \bar{y}
 - And thus a single instance of these randomly generated intervals:

$$\left[\bar{Y} - c \times \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}, \bar{Y} + c \times \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right] \quad \text{←}$$

- In particular, we observe

$$\left[\bar{y} - c \times \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}, \bar{y} + c \times \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right] \quad \text{←}$$

- Such observed intervals are referred to as **confidence intervals** and you must take care to distinguish them conceptually from a random interval
 - In the context of random intervals, probabilistic statements are sensible.
 - However, an observed confidence interval is not random – it either contains μ or it doesn't.
 - Probability statements are in reference to the **method used to generate the intervals**, NOT to the particular interval we have observed.
- If the normality assumption holds up, $100(1-p)\%$ of such intervals will contain μ ,
 - We thus have some **confidence** that our particular observed interval will contain μ as well, but unfortunately we'll never know if it does

Probability that the observed interval contains μ is either 0 or 1. There's no randomness.

- The larger $1-p$, the *more* confident we are that the interval will contain μ .

** Changing our confidence level necessarily changes the width of the interval (more confident \Rightarrow wider less confident \Rightarrow narrower)*

Example: Coverage Calculation via Simulation

- Given that $\mu = a(\mathcal{P}) = 155.89$ and $SD(\bar{Y}) = \text{sdN}(\text{avesSamp}) = 19.8$.
 - We randomly select 100 samples of size n
 - For each sample we calculate the 95% confidence interval (i.e., $(1-p) = 0.95$ and $c \approx 1.96$)
 - We should hope to find that approximately $100(1-p)\% = 95\%$ of the intervals contain (i.e., cover) $\mu = 155.89$.

```
## R code to simulate the GAUSSIAN approximation
```

```
ylim <- c(0, 0.022)

### Add a Gaussian density
tmpAve <- mean(avesSamp)
tmpSD <- sdN(avesSamp)
```

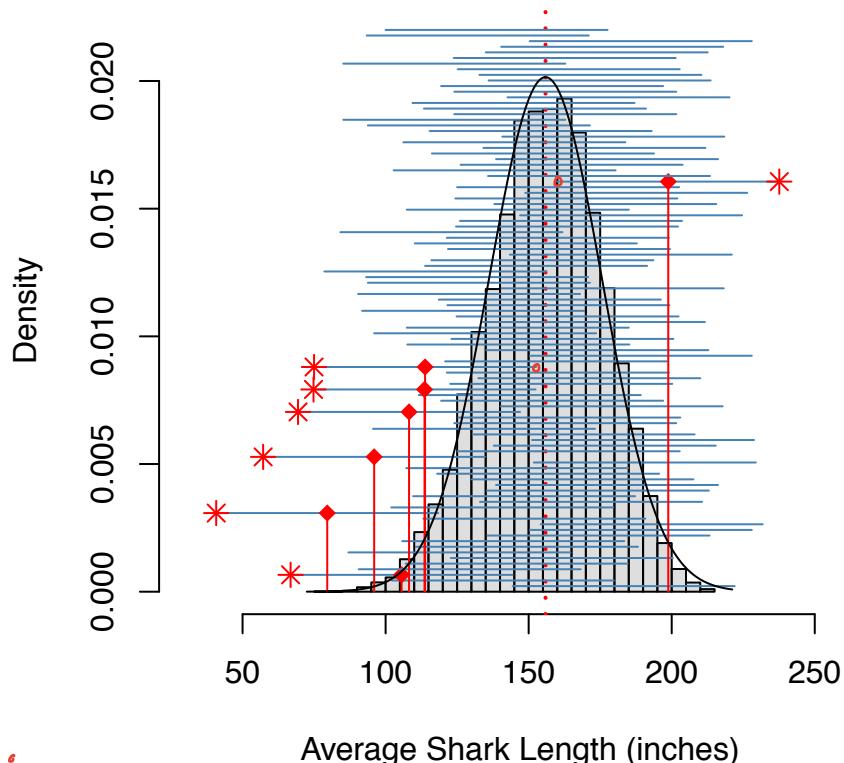
```

confidence <- 0.95
p <- 1 - confidence
numIntervals <- 100
cValue <- qnorm(1 - p/2) # or qnorm((confidence + 1)/2)
set.seed(34781453) # comment out this line to get different samples every time you run the code.
ybarSampled <- sample(avesSamp, numIntervals)
heights <- seq(diff(ylim)/numIntervals, max(ylim), length.out = numIntervals)
xlim <- extendrange(avesSamp + cValue * c(-tmpSD, tmpSD))
hist(avesSamp, col=adjustcolor("grey", alpha = 0.5), freq = FALSE,
      main=paste0(numIntervals, " Individual ", round(100 * confidence), "% Confidence Intervals"),
      xlab="Average Shark Length (inches)",
      ylim=ylim, xlim = xlim,
      breaks=25
)
### Mark the population attribute in red
abline(v=avePop, col="red", lty=3, lwd=2)
numIntervalsMissed <- 0
for(i in 1:numIntervals) {
  lines(ybarSampled[i] + cValue * c(-tmpSD, tmpSD), rep(heights[i],2),
        col = "steelblue")
  if (tmpAve > ybarSampled[i] + cValue*tmpSD) {
    points(ybarSampled[i] - cValue*tmpSD, heights[i], pch=8, cex=1.2, col="red")
    points(ybarSampled[i], heights[i], pch=18, cex=1.2, col="red")
    lines(rep(ybarSampled[i], 2), c(0, heights[i]), col = "red")
  } else if (tmpAve < ybarSampled[i] - cValue*tmpSD) {
    points(ybarSampled[i] + cValue*tmpSD, heights[i], pch=8, cex=1.2, col="red")
    points(ybarSampled[i], heights[i], pch=18, cex=1.2, col="red")
    lines(rep(ybarSampled[i], 2), c(0, heights[i]), col = "red")
  } else numIntervalsMissed <- numIntervalsMissed + 1
}

tmpX <- extendrange(avesSamp)
tmpX <- seq(tmpX[1], tmpX[2], length.out = 200)
lines(tmpX, dnorm(tmpX, mean = tmpAve, sd = tmpSD))

```

100 Individual 95% Confidence Intervals



- 93 of these 100 intervals cover the value μ (which is indicated by the vertical dashed red line)
 - Those which *do not* cover μ are marked with an asterisk and the corresponding sample means are indicated by diamonds and vertical solid red lines.
 - What do we notice?

All problematic samples have averages coming from the tails of the sampling distribution.

4.3.3 Student t Based Intervals

Standard Error vs. Standard Deviation

- In the previous section the confidence intervals we calculated assumed $SD[\bar{Y}]$ was known
 - This is an unrealistic assumption.
 - Only very rarely would we have this value.
- However, for many sample attributes $a(\mathcal{S})$ (e.g., Horvitz-Thompson estimators), we can estimate the standard deviation $SD[\tilde{a}(\mathcal{S})]$.
 - The **standard error** is an estimate of the standard deviation of the corresponding estimator:

$$SE[\tilde{a}(\mathcal{S})] = \widehat{SD}[\tilde{a}(\mathcal{S})]$$

$$\sqrt{\frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)}$$

Using this requires
 that we know σ

- We could use this instead. The estimate now becomes

$$\frac{a(\mathcal{S}) - a(\mathcal{P})}{SE[\tilde{a}(\mathcal{S})]} \quad \text{Rather than} \quad \frac{a(\mathcal{S}) - a(\mathcal{P})}{SD[\tilde{a}(\mathcal{S})]}$$

>Note that using the estimate SE in place of SD will increase the variability of the random intervals.

The corresponding estimator has much more variability – since we have to estimate \widehat{SD} now as well.

- **Example:** Let us illustrate this phenomenon using Australia sharks lengths aata

- Because we've taken all possible samples, we know $SD[\tilde{a}(\mathcal{S})]$ but let's pretend like we don't and estimate it.
- Thus let's calculate both

$$\textcircled{X} \quad \frac{a(\mathcal{S}) - a(\mathcal{P})}{SD[\tilde{a}(\mathcal{S})]} \quad \text{and} \quad \frac{a(\mathcal{S}) - a(\mathcal{P})}{SE[\tilde{a}(\mathcal{S})]} \quad \textcircled{X} \textcircled{X}$$

```
par(mfrow=c(1,2))

Z = (avesSamp - mean(avesSamp))/sdN(avesSamp)

sdsSamp <- apply(samples, MARGIN = 2,
                  FUN = function(s){sdN(sharks[s, "Length"])})
```

```
n = 5
N = 28
se = sdsSamp/sqrt(n)*sqrt((N-n)/(N-1))
t = (avesSamp - mean(avesSamp))/se
```

```
## So the two histograms have the same bins
delta = 0.2
brk = seq(min(t)-delta, max(t)+delta, delta)
```

```
hist(Z, col=adjustcolor("grey", alpha = 0.5),
      freq = FALSE,
      main="Standardized averages \n with known standard deviation",
      xlab="Standardized Shark Length (inches)",
      breaks=brk, xlim=c(-4,4), ylim=c(0,0.4) )
```

```
### Mark the population attribute in red
abline(v=0, col="red", lty=3, lwd=2)
```

```
### Add a Gaussian density
x = seq(-4, 4, length.out=1000)
lines(x, dnorm(x) , lwd = 2, col = "blue")
```

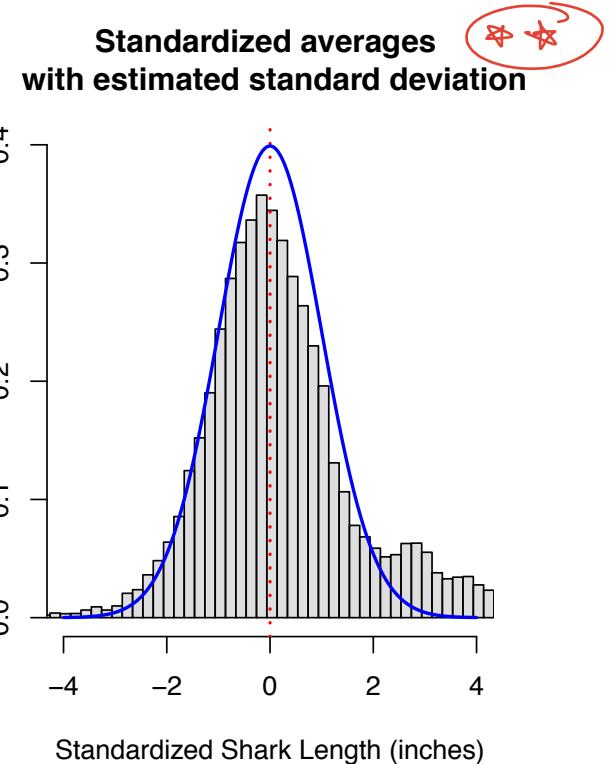
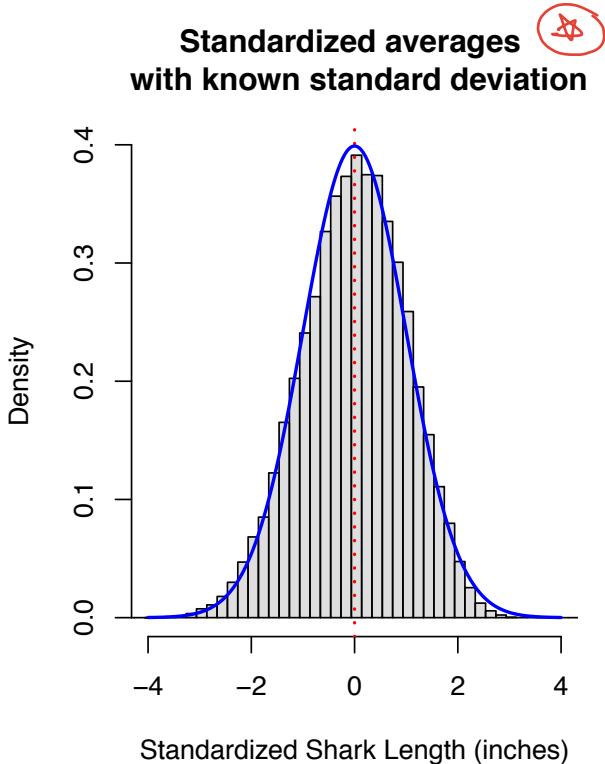
```
hist(t, col=adjustcolor("grey", alpha = 0.5),
      freq = FALSE,
      main="Standardized averages \n with estimated standard deviation",
      xlab="Standardized Shark Length (inches)",
      breaks=brk, xlim=c(-4,4), ylim=c(0,0.4) )
```

```

### Mark the population attribute in red
abline(v=0, col="red", lty=3, lwd=2)

### Add a Gaussian density
x = seq(-4, 4, length.out=1000)
lines( x, dnorm(x) , lwd = 2, col = "blue")

```



- The blue lines correspond to the $N(0,1)$ density curve
 - the normal distribution does not seem to be a good approximation when we have to estimate the standard deviation.
- Under the normality assumption we have the following distributional result

$$\frac{\bar{Y} - \mu}{SD[\bar{Y}]} = \frac{\bar{Y} - \mu}{\tilde{\sigma}/\sqrt{n}} \sim t_{n-1}$$

- This statistic is called a pivotal quantity because
 - it is a function of the sample data ($Y_u, u \in \mathcal{S}$) and unknown parameter (μ) and its sampling distribution is completely known. distribution doesn't depend on unknown parameters
- Let's try the t_{n-1} distribution as an approximation instead of the $N(0,1)$:

```

n = 5
N = 28
se = sdsSamp/sqrt(n)*sqrt((N-n)/(N-1))
t = (avesSamp - mean(avesSamp))/se

## So the two histograms have the same bins

```

```

delta = 0.2
brk = seq(min(t)-delta, max(t)+delta, delta)

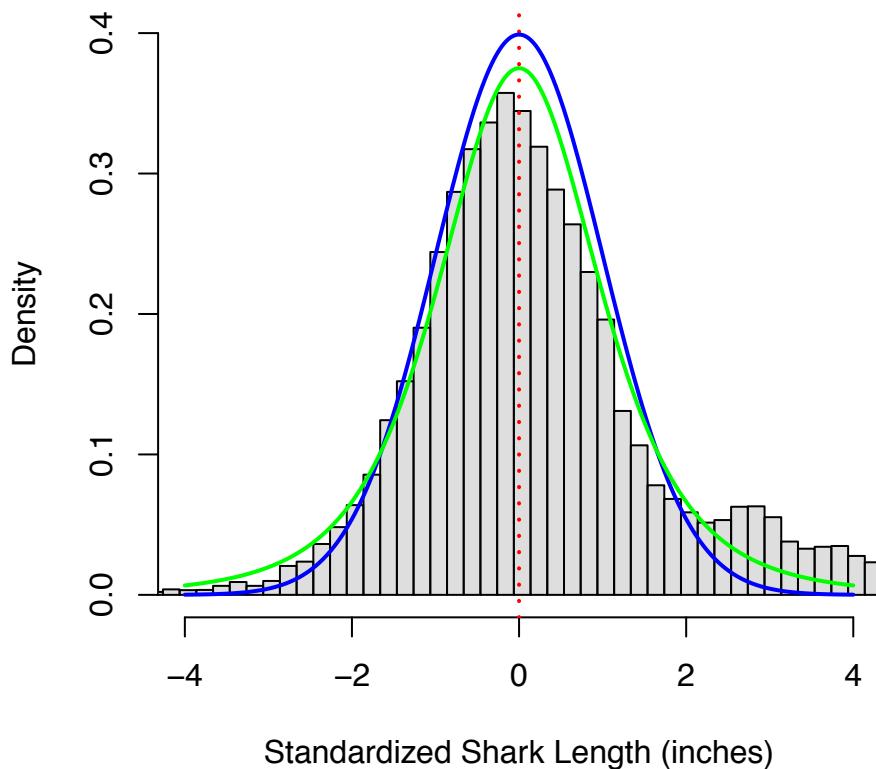
hist(t, col=adjustcolor("grey", alpha = 0.5),
      freq = FALSE, ylim = c(0,.4),
      main="Standardized averages with \n estimated standard deviation",
      xlab="Standardized Shark Length (inches)",
      breaks=brk, xlim=c(-4,4) )

### Mark the population attribute in red
abline(v=0, col="red", lty=3, lwd=2)

### Add a Gaussian density
x = seq(-4, 4, length.out=1000)
lines( x, dnorm(x), col = "blue", lwd=2 )
lines( x, dt(x, n-1), col= "green", lwd=2)

```

Standardized averages with estimated standard deviation



- The green and blue curves represent the t -distribution and normal distribution density curves, respectively
 - The t -distribution with $n - 1 = 4$ is a better approximation, but still not great.
- Now if we suppose that

$$\frac{\tilde{a}(\mathcal{S}) - a(\mathcal{P})}{\widetilde{SD}[\tilde{a}(\mathcal{S})]} \sim t_{n-1}$$

- Then we choose a $p \in (0, 1)$ and a corresponding $c > 0$ with

$$\begin{aligned} 1 - p &= Pr\left(-c \leq \frac{\tilde{a}(\mathcal{S}) - a(\mathcal{P})}{[SD](\tilde{a}(\mathcal{S}))} \leq c\right) \\ &= Pr\left(\tilde{a}(\mathcal{S}) - c \times \widetilde{SD}[\tilde{a}(\mathcal{S})] \leq \mu \leq \tilde{a}(\mathcal{S}) + c \times \widetilde{SD}[\tilde{a}(\mathcal{S})]\right) \end{aligned}$$

- This random interval has both a random center and a random length.

- The value of c is determined using the t distribution with $n - 1$ degrees of freedom.

- In R use `qt(1-p/2, df = n-1)` to get the value of c
- You can check that $c \approx 2.78$ when $1 - p = 0.95$ for a t_4 random variable.

- In the special case that $a(\mathcal{P}) = \bar{Y}$ is the population average then the standard deviation is

$$SD[\tilde{a}(\mathcal{S})] = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

and the standard deviation estimator is

$$\widetilde{SD}[\tilde{a}(\mathcal{S})] = \frac{\tilde{\sigma}}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

and the standard deviation estimate (the **standard error**) is

$$\widehat{SD}[\tilde{a}(\mathcal{S})] = \frac{\widehat{\sigma}}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$$

- Thus the random interval in this case is

$$\left[\bar{Y} - c \times \frac{\tilde{\sigma}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}, \bar{Y} + c \times \frac{\tilde{\sigma}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right]$$

- And the corresponding confidence interval is

$$\left[\bar{y} - c \times \frac{\widehat{\sigma}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}, \bar{y} + c \times \frac{\widehat{\sigma}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right]$$

- **Note:** In the formula above σ denotes the population standard deviation, which may be estimated in a sample by

$$\widehat{\sigma} = \sqrt{\frac{\sum_{u \in \mathcal{S}} (y_u - \bar{y})^2}{n}}$$

Example: Coverage Calculation via Simulation

- We randomly select 100 samples of size $n = 5$, and
- For each sample we calculate the 95% confidence interval (defined above)
 - using $c \approx 2.78$ from the t_4 distribution
- We should hope to find that approximately 95% of the intervals contain (or cover) $\mu = 155.89$.

```
## R code to simulate the t-STUDENT approximation
```

```
ylim <- c(0, 0.022)
n <- 5
N <- 28
### Add a Gaussian density
tmpAve      <- mean(avesSamp)
tmpSD       <- sdN(avesSamp)
confidence   <- 0.95
numIntervals <- 100
cValue       <- qt((confidence +1)/2, df = n-1)

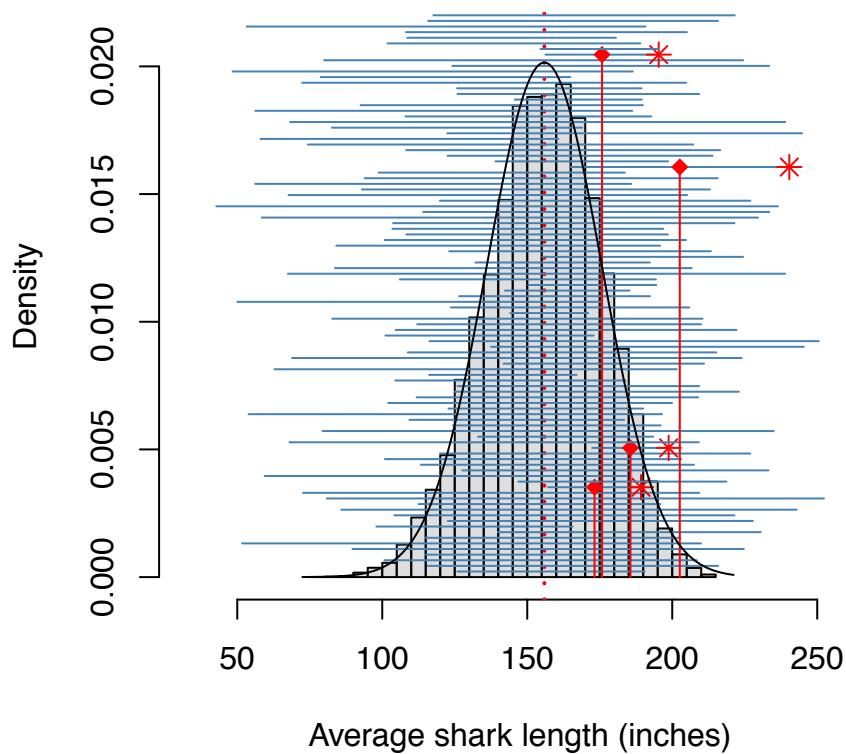
set.seed(341)
tmpSamp      <- sample(1:length(avesSamp), numIntervals)
ybarSampled <- avesSamp[tmpSamp]
sdSampled   <- sdsSamp[tmpSamp]
maxSD       <- max(sdSampled)/sqrt(n)

heights <- seq(diff(ylim)/numIntervals, max(ylim), length.out = numIntervals)
xlim <- extendrange(c(ybarSampled - cValue * sqrt((N-n)/(N-1)) * sdSampled/sqrt(n),
                      ybarSampled + cValue * sqrt((N-n)/(N-1)) * sdSampled/sqrt(n)))
hist(avesSamp, col=adjustcolor("grey", alpha = 0.5), freq = FALSE,
     main=paste0(numIntervals, " Individual ", round(100 * confidence), "% Confidence Intervals"),
     xlab="Average shark length (inches)",
     ylim=ylim, xlim = xlim,
     breaks=25
)
### Mark the population attribute in red
abline(v=avePop, col="red", lty=3, lwd=2)
numIntervalsMissed <- 0
for(i in 1:numIntervals) {
  tmpSampSD <- sdSampled[i]/sqrt(n)*sqrt((N-n)/(N-1))

  lines(ybarSampled[i] + cValue * c(-tmpSampSD, tmpSampSD), rep(heights[i],2),
        col = "steelblue")
  if (tmpAve > ybarSampled[i] + cValue*tmpSampSD) {
    points(ybarSampled[i] - cValue*tmpSampSD, heights[i], pch=8, cex=1.2, col="red")
    points(ybarSampled[i], heights[i], pch=18, cex=1.2, col="red")
    lines(rep(ybarSampled[i], 2), c(0, heights[i]), col = "red")
  } else if (tmpAve < ybarSampled[i] - cValue*tmpSampSD) {
    points(ybarSampled[i] + cValue*tmpSampSD, heights[i], pch=8, cex=1.2, col="red")
    points(ybarSampled[i], heights[i], pch=18, cex=1.2, col="red")
    lines(rep(ybarSampled[i], 2), c(0, heights[i]), col = "red")
  } else numIntervalsMissed <- numIntervalsMissed + 1
}
tmpX <- extendrange(avesSamp)
```

```
tmpX <- seq(tmpX[1], tmpX[2], length.out = 200)
lines(tmpX, dnorm(tmpX, mean = tmpAve, sd = tmpSD))
```

100 Individual 95% Confidence Intervals



- 96 of these 100 intervals cover the value μ (which is indicated by the vertical dashed red line)
 - Those which *do not* cover μ are marked with an asterisk and the corresponding sample means are indicated by diamonds and vertical solid red lines.
 - What do we notice?

Interval centers and widths are both random (i.e., they change from one sample to another). Also the actual coverage is slightly higher than 95%.

* Note that we can find the exact coverage probability by calculating all possible confidence intervals.

A Note on Pivotal Quantities:

- Pivotal quantities are the basis for constructing random intervals.
- Many random intervals are constructed via pivotal quantities such as

$$\frac{\tilde{a}(\mathcal{S}) - a(\mathcal{P})}{\widetilde{SD}[\tilde{a}(\mathcal{S})]}$$

- The term *pivot* comes from the fact that with this quantity (which is a function of both \mathcal{S} and \mathcal{P}) we are able to pivot and isolate for $a(\mathcal{P})$
- This is the general prescription for constructing random intervals.

- This particular pivotal is a t -like pivotal, but it's not the *only* form of pivotal quantity.
 - One common pivotal quantity used for scale attributes $s(\cdot)$, are of the form

$$\frac{\tilde{s}(\mathcal{S})}{s(\mathcal{P})}$$

- e.g., this gives rise to χ^2 -based intervals for population variances.

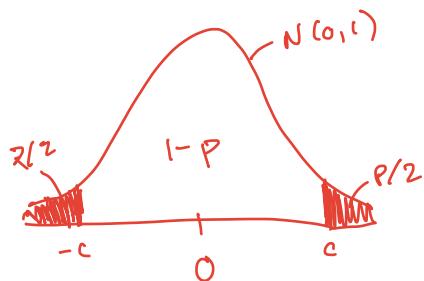
HOWEVER all of this assumes we know the distribution of the pivotal quantity...

So what have we learned?

- Random intervals for population attributes can be constructed by picking an appropriate pivotal quantity, and then pivoting
- Corresponding confidence intervals are obtained by plugging in sample estimates for estimators.

4.3.4 Connection to Hypothesis Testing

- Note that there is a 1:1 correspondence between confidence intervals and hypothesis tests.
- Thus, if one wished to test a hypothesis (like $H_0 : a(\mathcal{P}) = a$) about a population attribute, they could do so with an appropriately defined confidence interval.
- The reasoning is as follows:
 - Suppose that there exists a threshold below which a p -value would be sufficiently small so as to disbelieve the null hypothesis. This is often referred to as the **significance level** of the test.
 - If the same pivotal quantity is used as both a discrepancy measure for the test and the basis for a confidence interval, the $100(1 - p)\%$ confidence interval for $a(\mathcal{P})$ contains all values of a for which the test of $H_0 : a(\mathcal{P}) = a$ would yield a p -value $> p$.
- * Stated more usefully: Do not reject $H_0 : a(\mathcal{P}) = a$ at a $100p\%$ significance level if and only if a is contained in the $100(1 - p)\%$ confidence interval for $a(\mathcal{P})$.



If the discrepancy falls in \mathbb{I} , then the corresponding p -value is $\leq p$. If p is your significance threshold, then H_0 is rejected. Values of a outside the $(1-p) \times 100\%$ confidence are those values that push the discrepancy into \mathbb{I} .