4 Resampling Methods: The Bootstrap

- Situation: Let $x_1, x_2, \ldots, x_n$ be a SRS of size n taken from a distribution that is unknown. Let θ be a parameter of interest associated with this distribution and let $\widehat{\theta} = S(x_1, x_2, \ldots, x_n)$ be a statistic used to estimate θ.
  - For example, the sample mean $\widehat{\theta} = S(x_1, x_2, \ldots, x_n) = \overline{x}$ is a statistic used to estimate the true mean.
- Goals: (i) provide a standard error $\text{se}_B(\widehat{\theta})$ estimate for $\widehat{\theta}$, (ii) estimate the bias of $\widehat{\theta}$, and (iii) generate a confidence interval for the parameter θ.
- **Bootstrap methods** are computer-intensive methods of providing these estimates and depend on *bootstrap samples*.
- An (independent) **bootstrap sample** is a SRS of size n taken with replacement from the data $x_1, x_2, \ldots, x_n$.
- We denote a bootstrap sample as $x_1^*, x_2^*, \ldots, x_n^*$ which consists of members of the original data set $x_1, x_2, \ldots, x_n$ with some members appearing zero times, some appearing only once, some appearing twice, and so on.
- A **bootstrap sample replication** of $\widehat{\theta}$, denoted $\widehat{\theta}^*$, is the value of $\widehat{\theta}$ evaluated using the bootstrap sample $x_1^*, x_2^*, \ldots, x_n^*$.
- The bootstrap algorithm requires that a large number (B) of bootstrap samples be taken. The bootstrap sample replication $\widehat{\theta}^*$ is then calculated for each of the B bootstrap samples. We will denote the **b$^{\text{th}}$ bootstrap replication** as $\widehat{\theta}^*(b)$ for $b = 1, 2, \ldots, B$.
- The notes in this section are based on Efron and Tibshirani (1993) and Manly (2007).

4.1 The Bootstrap Estimate of the Standard Error

- The **bootstrap estimate of the standard error** of $\widehat{\theta}$ is

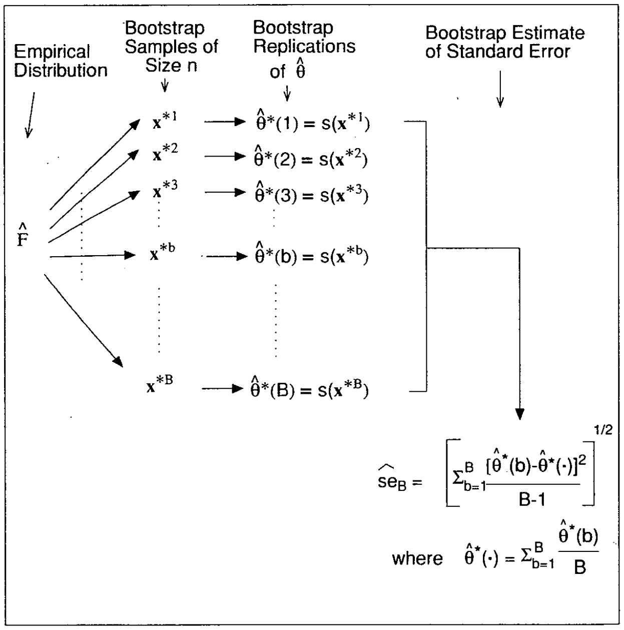
$$\text{se}_B(\widehat{\theta}) = \sqrt{\frac{\sum_{b=1}^{B} \left[\widehat{\theta}^*(b) - \widehat{\theta}^*(\cdot) \right]^2}{B-1}} \qquad (7)$$

where $\widehat{\theta}^*(\cdot) = \dfrac{\sum_{b=1}^{B} \widehat{\theta}^*(b)}{B}$ is the sample mean of the B bootstrap replications.

- Note that $\text{se}_B(\widehat{\theta})$ is just the sample standard deviation of the B bootstrap replications.
- The limit of $\text{se}_B(\widehat{\theta})$ as $B \longrightarrow \infty$ is the **ideal bootstrap estimate** of the standard error.
- Under most circumstances, as the sample size n increases, the sampling distribution of $\widehat{\theta}$ becomes more normally distributed. Under this assumption, an approximate ***t*-based bootstrap confidence interval** can be generated using $\text{se}_B(\widehat{\theta})$ and a t-distribution:

$$\widehat{\theta} \pm t^* \, \text{se}_B(\widehat{\theta})$$

where t^* has $n-1$ degrees of freedom.



4.2 The Bootstrap Estimate of Bias

- The bias of $\widehat{\theta} = S(X_1, X_2, \ldots, X_n)$ as an estimator of θ is defined to be

$$\text{bias}(\widehat{\theta}) \;=\; E_F(\widehat{\theta}) \;-\; \theta$$

  with the expectation taken with respect to distribution F.

- The **bootstrap estimate of the bias of** $\widehat{\theta}$ as an estimate of θ is calculated by replacing the distribution F with the empirical cumulative distribution function $\widehat{F}$. This yields

$$\widehat{\text{bias}}_B(\widehat{\theta}) \;=\; \widehat{\theta}^*(\cdot) \;-\; \widehat{\theta} \qquad \text{where } \widehat{\theta}^*(\cdot) = \frac{1}{B}\sum_{b=1}^{B}\widehat{\theta}^*(b).$$

- Then, the **bias-corrected estimate** of θ is

$$\tilde{\theta}_B \;=\; \widehat{\theta} - \widehat{\text{bias}}_B(\widehat{\theta}) \;=\; 2\widehat{\theta} - \widehat{\theta}^*(\cdot).$$

- One problem with estimating the bias is that the variance of $\widehat{\text{bias}}_B(\widehat{\theta})$ is often large. Efron and Tibshirani (1993) recommend using:

  1. $\widehat{\theta}$ if $\widehat{\text{bias}}_B(\widehat{\theta})$ is small relative to the $\text{se}_B(\widehat{\theta})$.
  2. $\tilde{\theta}$ if $\widehat{\text{bias}}_B(\widehat{\theta})$ is large relative to the $\text{se}_B(\widehat{\theta})$.

  They suggest and justify that if $\widehat{\text{bias}}_B(\widehat{\theta}) < .25\text{se}_B(\widehat{\theta})$ then the bias can be ignored (unless the goal is precise confidence interval estimation using this standard error).

- Manly (2007) suggests that when using bias correction, it is better to center the confidence interval limits using $\tilde{\theta}$. This would yield the approximate **bias-corrected t-based confidence interval**:

$$\tilde{\theta} \;\pm\; t^*\,\text{se}_B(\widehat{\theta})$$

4.3 Introductory Bootstrap Example

- Consider a SRS with $n = 10$ having y-values   0 1 2 3 4 8 8 9 10 11
- The following output is based on $B = 40$ bootstrap replications of the sample mean $\overline{x}$, the sample standard deviation s, the sample variance s^2, and the sample median.
- The terms in the output are equivalent to the following:

theta(hat)	$= \widehat{\theta} =$ the sample estimate of a parameter
mean	= the sample mean $\overline{x}$
s	= the sample standard deviation s
variance	= the sample variance s^2
median	= the sample median m

- These statistics are used as estimates of the population parameters μ, σ, σ^2, and M.

```
theta(hat) values for mean, standard deviation, variance, median

      mean          s   variance     median
    5.6000     4.0332    16.2667     6.0000

The number of bootstrap samples B = 40

The bootstrap samples
     8     3     2     0     2    11    11    10     8     2
     2     9     9     1     9     1    10     4    11     4
     8     9     3     8     2     8     3     9     2     2
    10     2    11     9     4     8     2     8     3     9
     2     8     3     9     8     2     1     8     9     9
    10    10     2    11     4     8     8     3     2     3
    10     8     0     8     9    10     9     1     1    10
     3    10     8     3     2     3    11     9     3    10
    10     1    10     8     1     9    10     0     3     8
     3     3     3     4     3     0     1    11    11     4
     2     3     8     2    10     1     4     8     8     2
     2    10     1     3     2     0    10     3     8     4
     8     2    11    11     1     9    11     3     8     4
     9    11    10     2     8     1     9    10     9     9
     8     4     8     4     0     1     3     1     9    11
     4     9     8     0    10    10     4     1     8     8
     9    11     8    10     3     0     8    10     2     4
     8    11     0     3     2    10     1     4     8     8
     8     8     3     4     8     0     3    10    11     2
     2     1     8     0     2     3    11    11     9     8
     2    11     4     2     8     8     8     0     1    10
     4     2     0     4    11     0     0     9     1    10
    10     2     4     1     4     3     3    10     3     2
     1     8     1     8     4     8     3     0     9    10
     8     3     9     1     1     8    10     3    10     0
     8    11     8     2     1     2     0    11     3    11
     0     8     3     8    10    10     8     1     8     9
     2     0     3     3     9    11     2    11     3    11
    11     1     0    10     9     9     8     2    11     1
     2     8    10     1    11     4     8     4     3     2
    10     9     1     0     3     2     2     0     2    10
     2     8    11     3     0     4     4     4     3     3
     3     0     8     2     4     8     1     3    10     4
    10     9     8    10     8    10     0     0    11     8
     8    11     1     3     3    11     3     0    10     8
     8     3     1     4     8     3     8     9     1     1
     0     9    11     2     4    10     2     8    10     4
     8    11     0     8     0     8     1     3     0     2
     9     9     8    11    10     2    11    11     9     1
     1    11    11    11     3     0     0     3     0     9
```

Next, we calculate the sample mean, sample standard deviation, sample variance, and sample median for each of the bootstrap samples. These represent $\widehat{\theta}_b^*$ values.

The column labels below represent bootstrap labels for four different $\widehat{\theta}^*_b$ cases:

mean = **a bootstrap sample mean** $\overline{x}^*_b$
std dev = **a bootstrap sample standard deviation** s^*_b
variance = **a bootstrap sample variance** s^{2*}_b
median = **a bootstrap sample median** m^*_b

```
Bootstrap replications: theta(hat)^*_b
      mean   std dev  variance    median
    5.7000    4.2960   18.4556    5.5000
    6.0000    3.9721   15.7778    6.5000
    5.4000    3.2042   10.2667    5.5000
    6.6000    3.4705   12.0444    8.0000
    5.9000    3.4140   11.6556    8.0000
    6.1000    3.6347   13.2111    6.0000
    6.6000    4.1687   17.3778    8.5000
    6.2000    3.6757   13.5111    5.5000
    6.0000    4.2164   17.7778    8.0000
    4.3000    3.7431   14.0111    3.0000
    4.8000    3.3267   11.0667    3.5000
    4.3000    3.6833   13.5667    3.0000
    6.8000    3.9384   15.5111    8.0000
    7.8000    3.4254   11.7333    9.0000
    4.9000    3.8427   14.7667    4.0000
    6.2000    3.6757   13.5111    8.0000
    6.5000    3.8944   15.1667    8.0000
    5.5000    3.9511   15.6111    6.0000
    5.7000    3.7431   14.0111    6.0000
    5.5000    4.3012   18.5000    5.5000
    5.4000    4.0332   16.2667    6.0000
    4.1000    4.3576   18.9889    3.0000
    4.2000    3.1903   10.1778    3.0000
    5.2000    3.7947   14.4000    6.0000
    5.3000    4.0565   16.4556    5.5000
    5.7000    4.5228   20.4556    5.5000
    6.5000    3.7193   13.8333    8.0000
    5.5000    4.4284   19.6111    3.0000
    6.2000    4.5898   21.0667    8.5000
    5.3000    3.6225   13.1222    4.0000
    3.9000    4.0947   16.7667    2.0000
    4.2000    3.1198    9.7333    3.5000
    4.3000    3.3015   10.9000    3.5000
    7.4000    4.0332   16.2667    8.5000
    5.8000    4.2374   17.9556    5.5000
    4.6000    3.3066   10.9333    3.5000
    6.0000    4.0277   16.2222    6.0000
    4.1000    4.2019   17.6556    2.5000
    8.1000    3.6347   13.2111    9.0000
    4.9000    4.9766   24.7667    3.0000
```

Take the mean of each column ($\widehat{\theta}^*_{(\cdot)}$) yielding $\overline{x}^*_{(\cdot)}$, $s^*_{(\cdot)}$, $s^{2*}_{(\cdot)}$, and $m^*_{(\cdot)}$.

```
Mean of the B bootstrap replications: theta(hat)^*_(.)
      mean         s   variance    median
    5.5875    3.8707    15.1581    5.6250
```

Take the standard deviation of each column ($se_B(\widehat{\theta})$) yielding $se_B(\overline{x}^*)$, $se_B(s^*)$, $se_B(s^{2*})$, and $se_B(m^*)$.

```
Bootstrap standard error:  s.e._B(theta(hat))
      mean         s   variance    median
    1.0229    0.4248     3.3422    2.1266
```

Finally, calculate the estimates of bias $\widehat{\text{Bias}}_B(\widehat{\theta}) = \widehat{\theta}^*_{(\cdot)} - \widehat{\theta}$ for the four cases.

```
Bootstrap bias estimate:  bias(hat)_B(theta(hat))
      mean         s   variance    median
   -0.0125   -0.1625    -1.1086   -0.3750
```

4.4 Bootstrap Confidence Intervals

- Several methods for generating confidence intervals based on the bootstrap replications will now be presented.

4.4.1 Bootstrap CIs Assuming Approximate Normality

- An approximate $100(1-\alpha)\%$ confidence interval for θ is

$$\widehat{\theta} \, \pm \, t^* \mathrm{se}_B(\widehat{\theta}) \qquad \text{or} \qquad \widehat{\theta} \, \pm \, z^* \mathrm{se}_B(\widehat{\theta}) \tag{8}$$

  where t^* is the upper $\alpha/2$ critical value from a t-distribution having $n-1$ degrees of freedom and z^* is the upper $\alpha/2$ critical value from a standard normal (z) distribution.

- For an approximate 90%, 95%, or 99% confidence intervals for θ to be useful, we would expect that approximately 90%, 95%, or 99% of confidence intervals generated using this method will contain θ.

- If the n is not large enough and the distribution sampled from is highly skewed (or, in general, is not close in shape to a normal distribution), then the confidence interval given in (8) will not be very reliable. That is, the nominal (stated) confidence level is not close to the true (actual) confidence level.

4.4.2 Confidence Intervals Using Bootstrap Percentiles

- If the sample size is relatively small or it is suspected that the sampling distribution of $\widehat{\theta}$ is skewed or non-normal, we want an alternative to (8) for generating a confidence interval.

- The simplest alternative is to use percentiles of the B bootstrap replications of $\widehat{\theta}^*$.

- The reliability of the percentile confidence interval method depends on one assumption. It is assumed that there exists a monotonic increasing function f such that the transformed values $f(\widehat{\theta})$ are normally distributed with mean $f(\theta)$ and standard deviation 1.

- Thus, with probability $1-\alpha$, the following statement is true:

$$f(\theta) \, - \, z_{\alpha/2} \; < \; f(\widehat{\theta}) \; < \; f(\theta) \, + \, z_{\alpha/2}.$$

  After rearranging the terms we have:

$$f(\widehat{\theta}) \, - \, z_{\alpha/2} \; < \; f(\theta) \; < \; f(\widehat{\theta}) \, + \, z_{\alpha/2} \tag{9}$$

- If the transformation f was known, then by applying a back-transformation f^{-1} to the confidence limits for $f(\theta)$ in (9), we have the confidence limits for θ.

- It is not necessary, however, to know the form of the function f. We only need to assume its existence. Because of the assumption that f is a monotonic and increasing function, then the ordering of the B transformed bootstrap estimates from smallest to largest must correspond to the ordering of the original B untransformed bootstrap replicates $\widehat{\theta}^*(b)$ from smallest to largest.

- Thus, the confidence limits for $f(\theta)$ in (9) are those values that exceed the $\alpha/2$ percentiles in the left and right tails of the distribution of the B bootstrap replicates.

- That is, the approximate **bootstrap percentile-based confidence interval** for θ is

$$\widehat{\theta}^*_L \; < \; \theta \; < \; \widehat{\theta}^*_U \tag{10}$$

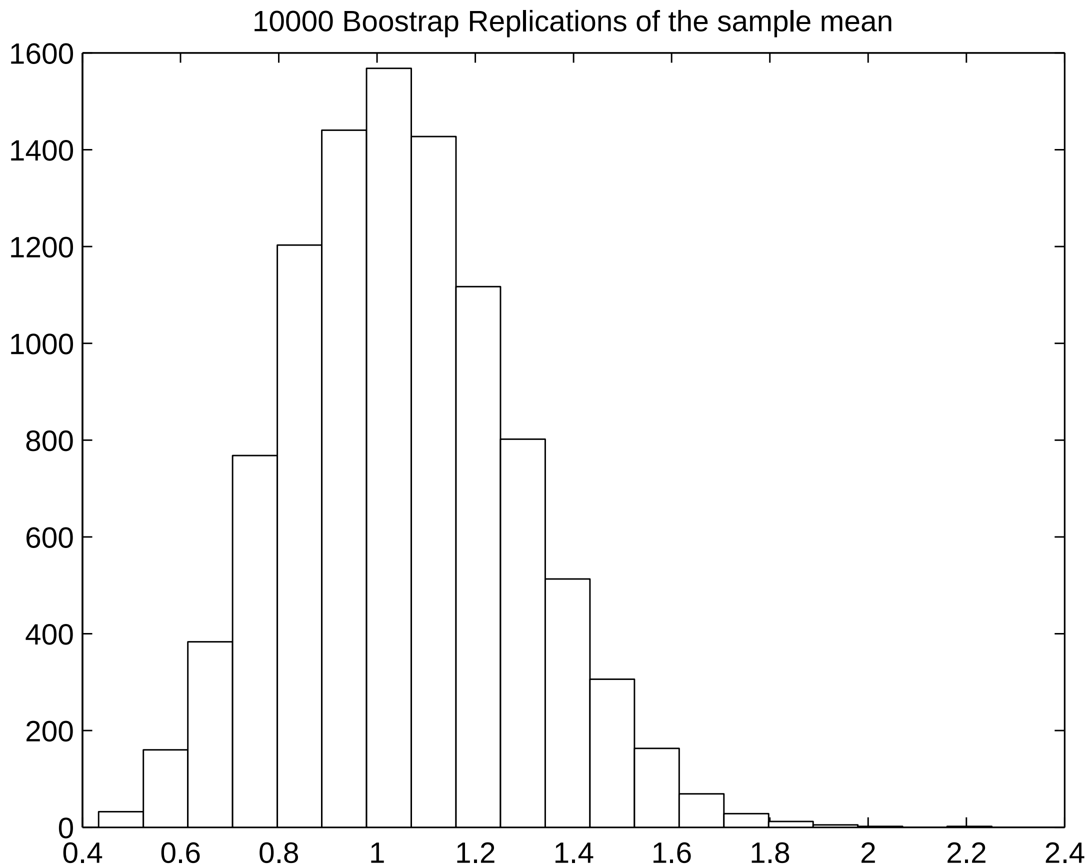
  where $\widehat{\theta}^*_L$ and $\widehat{\theta}^*_U$ are the lower $\alpha/2$ and upper $(1-\alpha/2)$ percentiles of the B bootstrap replications $\widehat{\theta}^*$, respectively. Practically, to find $\widehat{\theta}^*_L$ and $\widehat{\theta}^*_U$ you

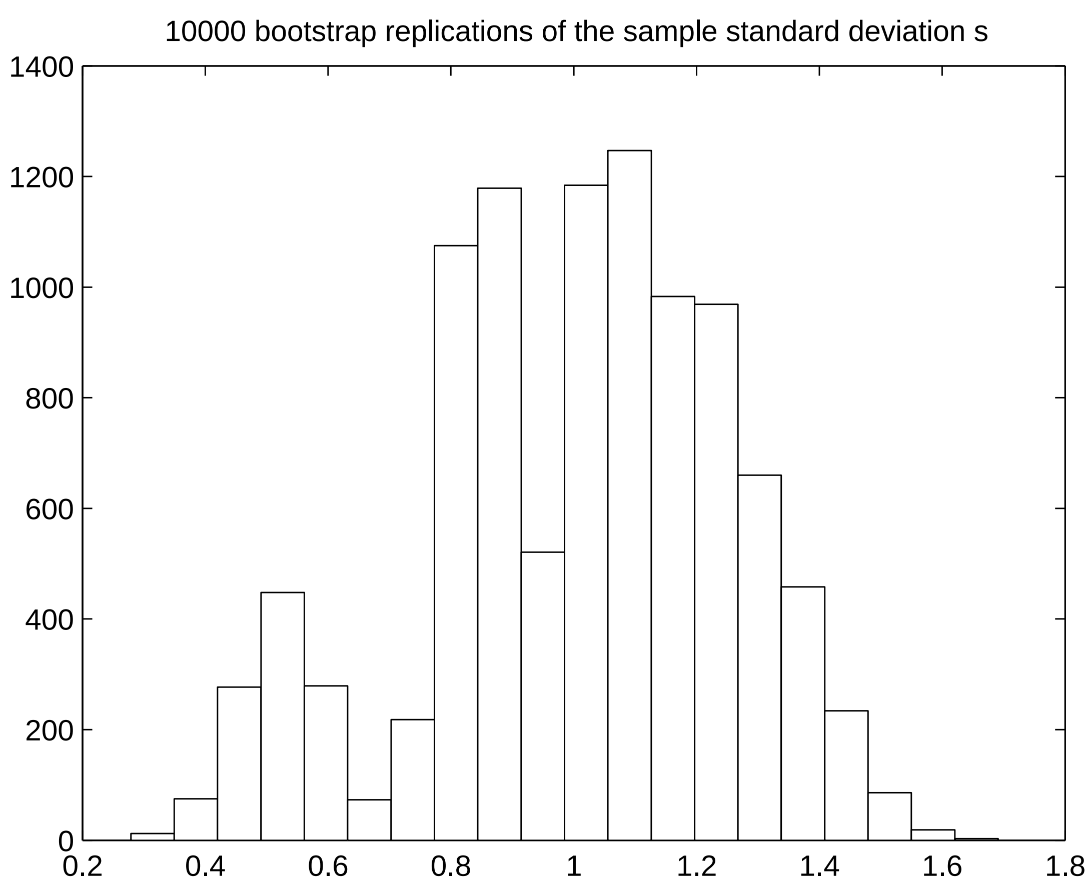
  1. Order the B bootstrap replications $\widehat{\theta}^*(1), \widehat{\theta}^*(2), \ldots, \widehat{\theta}^*(B)$ from smallest to largest.
  2. Calculate $L = B * \alpha/2$ and $U = B * (1-\alpha/2) + 1$.
  3. Find the L^{th} and U^{th} values in the ordered list of bootstrap replications.
  4. The L^{th} value is the lower confidence interval endpoint $\widehat{\theta}^*_L$ and the U^{th} value is the upper confidence interval endpoint $\widehat{\theta}^*_U$.

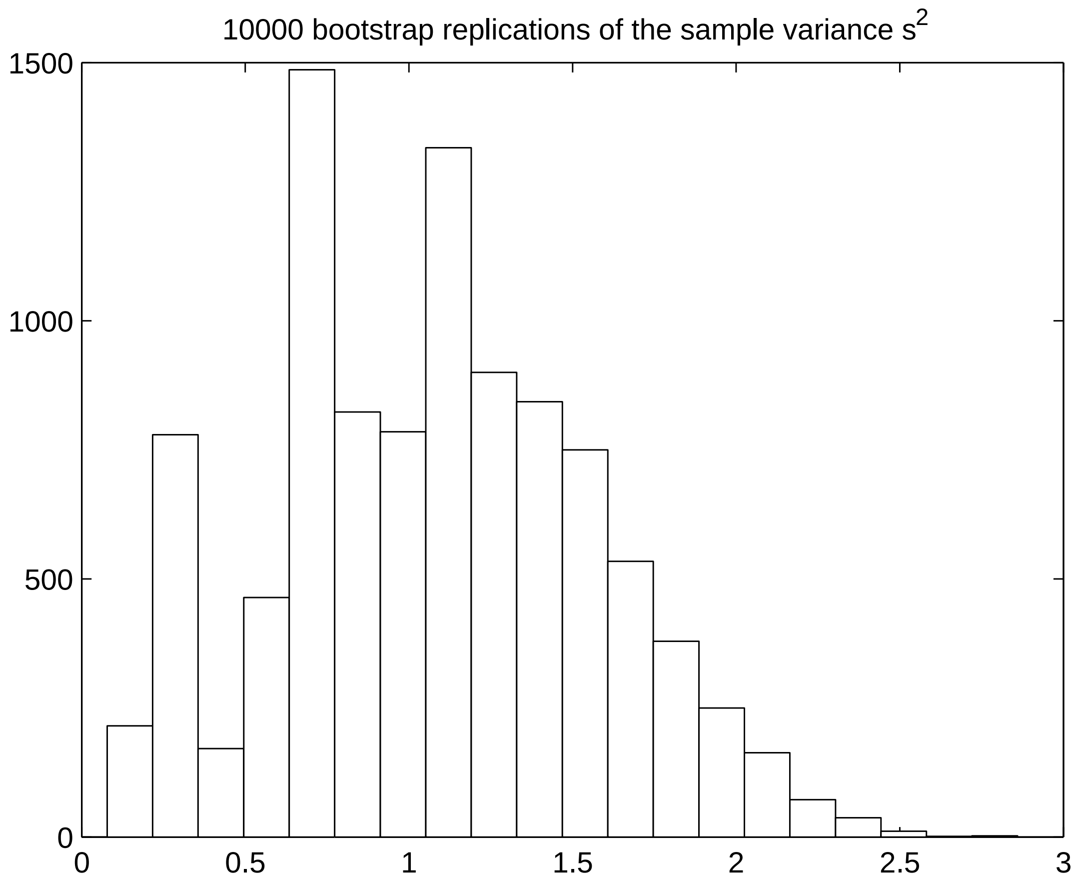
- There are improvements and corrections that can be applied to the percentile method when we do not believe the transformation f exists. We will consider two bias-corrected alternatives later in this section.

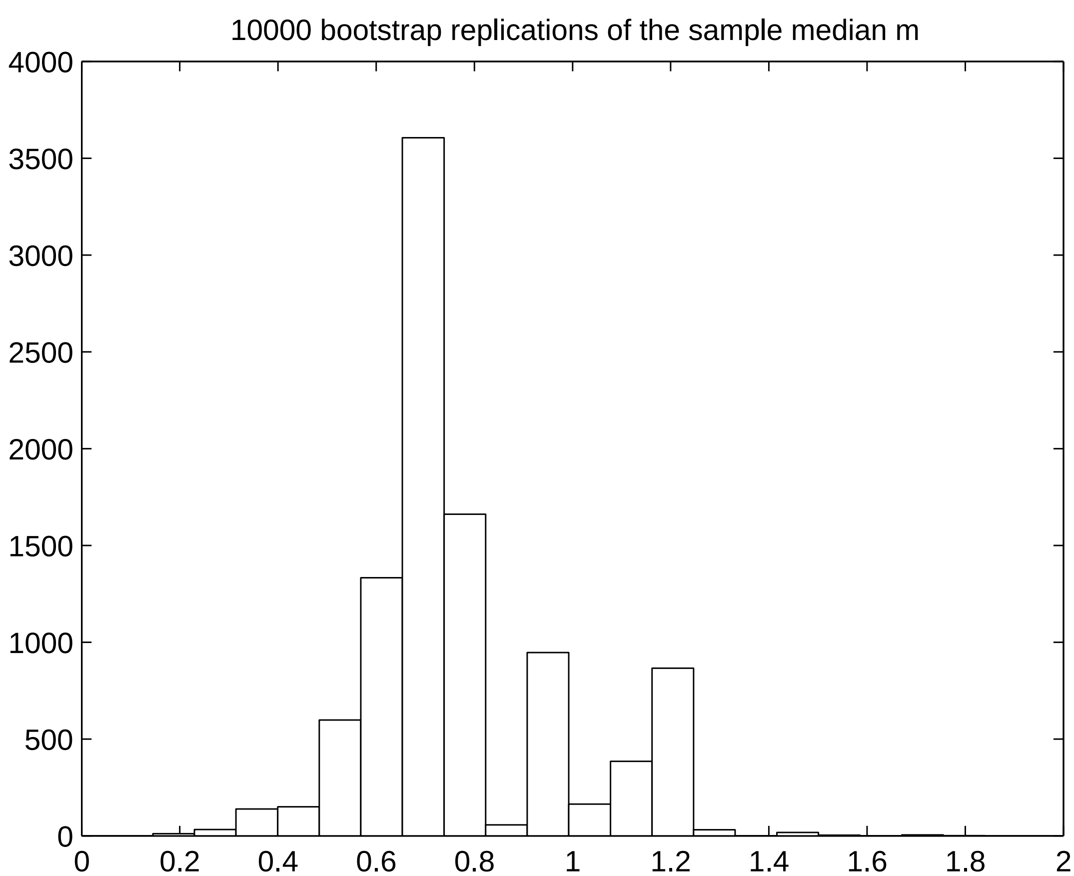
Bootstrapping Example: The Manly (2007) Data Set

```
The Data
  3.56    0.69    0.10   1.84   3.93   1.25   0.18   1.13    0.27   0.50
  0.67    0.01    0.61   0.82   1.70   0.39   0.11   1.20   1.21    0.72

theta(hat) values for xbar, s, s^2, s(with denominator n), median
      mean         s  variance    median
    1.0445    1.0597    1.1229    0.7050


The number of bootstrap samples B = 10000

Mean of the B bootstrap replications: theta(hat)^*_(.)
      mean         s   variance     median
    1.0499    1.0066    1.0759    0.7738



Bootstrap standard error:  s.e._B(theta(hat))
      mean         s   variance    median
    0.2323    0.2504    0.4845    0.2014



Bootstrap bias estimate:  bias(hat)_B(theta(hat))
      mean         s   variance    median
    0.0054   -0.0531   -0.0470    0.0688
```


10000 Boostrap Replications of the sample mean
1600
1400
1200
1000
800
600
400
200
0
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
2.2
2.4


10000 bootstrap replications of the sample standard deviation s
1400
1200
1000
800
600
400
200
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8


10000 bootstrap replications of the sample variance s2
1500
1000
500
0
0
0.5
1
1.5
2
2.5
3


10000 bootstrap replications of the sample median m
4000
3500
3000
2500
2000
1500
1000
500
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

```
-----------------------------------------------------------------
z-BASED CONFIDENCE INTERVALS
-----------------------------------------------------------------
95% confidence intervals -- z-based
      mean        s   variance    median
    0.5892    0.5690    0.1732    0.3103       <-- lower endpoint
    1.4998    1.5504    2.0726    1.0997       <-- upper endpoint


Bias-adjusted confidence intervals -- z-based
      mean        s   variance    median
    0.5838    0.6221    0.2202    0.2415       <-- lower endpoint
    1.4944    1.6035    2.1196    1.0308       <-- upper endpoint


-----------------------------------------------------------------
t-BASED CONFIDENCE INTERVALS
-----------------------------------------------------------------
95% confidence intervals -- t-based
      mean        s   variance    median
    0.5583    0.5357    0.1088    0.2835       <-- lower endpoint
    1.5307    1.5837    2.1371    1.1265       <-- upper endpoint


Bias-adjusted confidence intervals -- t-based
      mean        s   variance    median
    0.5529    0.5888    0.1558    0.2147       <-- lower endpoint
    1.5253    1.6368    2.1841    1.0576       <-- upper endpoint



-----------------------------------------------------------------
PERCENTILE CONFIDENCE INTERVALS
-----------------------------------------------------------------
95% percentile-based confidence intervals
      mean        s   variance    median
    0.6920    0.5135    0.2637    0.5000       <-- lower endpoint
    1.4535    1.3812    1.9077    1.2000       <-- upper endpoint



-----------------------------------------------------------------
BIAS CORRECTED CONFIDENCE INTERVALS (see Section 4.4.4)
-----------------------------------------------------------------

95% bias-corrected percentile-based confidence intervals
      mean        s   variance    median
    0.6450    0.4994    0.2494    0.4450       <-- lower endpoint
    1.5575    1.4606    2.1334    1.2100       <-- upper endpoint



---------------------------------------------------------------------
ACCELERATED BIAS CORRECTED CONFIDENCE INTERVALS (see Section 4.4.5)
---------------------------------------------------------------------

95% accelerated bias-corrected confidence intervals
      mean        s   variance    median
    0.6975    0.5308    0.2817    0.5000        <-- lower endpoint
    1.6560    1.5024    2.2573    1.2300        <-- upper endpoint
```

4.4.3 Bootstrap examples in R

- The ‘bootstrap’ package in R will generate bootstrap standard errors and estimates of bias.
- The ‘bootstrap’ R command will generate the bootstrap replications of a statistic and output the estimate, the standard error, and an estimate of bias.
- The ‘boot.ci’ R command will generate confidence intervals for the parameter of interest. I will consider three common bootstrap confidence intervals:
  1. The **percentile bootstrap confidence intervals**. These are generated by the procedure described in the notes.
  2. The **normal confidence intervals**. These intervals have the form

     $$\tilde{\theta} \pm z^* \, \text{s.e.}_{\text{boot}}(\widehat{\theta})$$

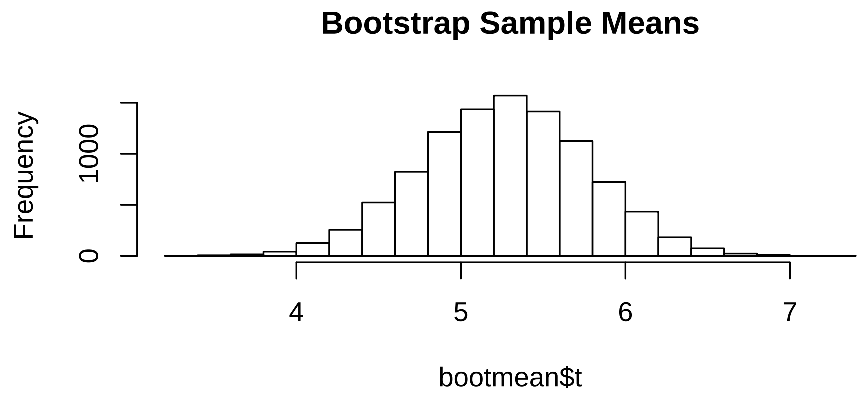
     which is the traditional z-based normal confidence interval except we add and subtract the margin of error about the bias-corrected estimate $\tilde{\theta}$.
  3. The **bias-corrected confidence interval**. These are percentile-based confidence intervals adjusted for the bias. That is, the endpoints of the intervals have bias adjustments.
- If you want a t-based confidence interval (which I recommend over a z-based interval), there are two possibilities:

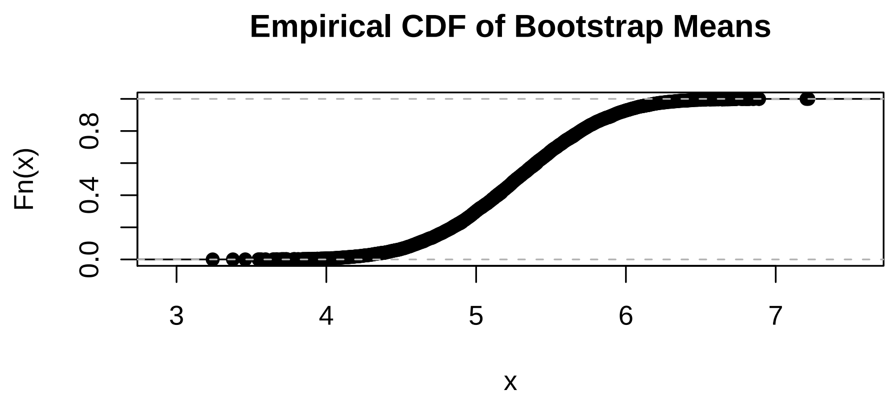
  $$\widehat{\theta} \pm t^* \, \text{s.e.}_{\text{boot}}(\widehat{\theta}) \qquad \text{and} \qquad \tilde{\theta} \pm t^* \, \text{s.e.}_{\text{boot}}(\widehat{\theta})$$

  If the estimate of bias is small relative to the standard error, use the interval centered at $\widehat{\theta}$. Otherwise, use the interval centered at the bias-corrected estimate $\tilde{\theta}$.

R code for Bootstrapping the Mean – symmetry present

```
library(boot)
y <- c(1,2.1,3.2,3.7,4.0,4.1,4.5,5.1,5.6,5.7,6.2,6.3,6.9,7.2,7.4,8.1,8.6)
y
n <- length(y)
n
thetahat = mean(y)
thetahat
Brep = 10000

# Bootstrap the sample mean
sampmean <- function(y,i) mean(y[i])

bootmean <- boot(data=y,statistic=sampmean,R=Brep)
bootmean
boot.ci(bootmean,conf=.95,type=c("norm"))
boot.ci(bootmean,conf=.95,type=c("perc"))
boot.ci(bootmean,conf=.95,type=c("bca"))

par(mfrow=c(2,1))

hist(bootmean$t,main="Bootstrap Sample Means")
plot(ecdf(bootmean$t),main="Empirical CDF of Bootstrap Means")
```





R output for Bootstrapping the Mean – symmetry present

```
 [1] 1.0 2.1 3.2 3.7 4.0 4.1 4.5 5.1 5.6 5.7 6.2 6.3 6.9 7.2 7.4 8.1 8.6
[1] 17
> thetahat
[1] 5.276471

> # Bootstrap the sample mean

ORDINARY NONPARAMETRIC BOOTSTRAP

Bootstrap Statistics :
    original       bias    std. error
t1* 5.276471 -0.008028824   0.5053402

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

Intervals :
Level      Normal
95%   ( 4.294,  6.275 )
Calculations and Intervals on Original Scale

Intervals :
Level     Percentile
95%   ( 4.259,  6.229 )
Calculations and Intervals on Original Scale
```

```
Intervals :
Level       BCa
95%   ( 4.245,  6.212 )
Calculations and Intervals on Original Scale
```

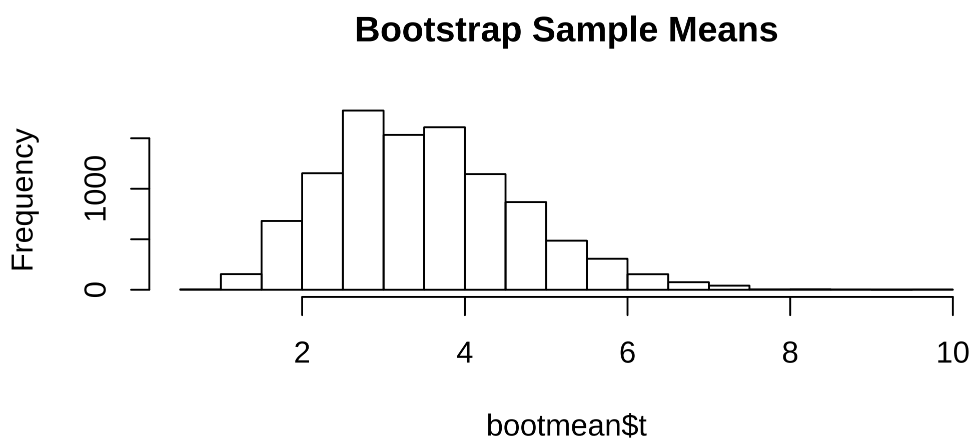
R code for Bootstrapping the Mean – skewness present

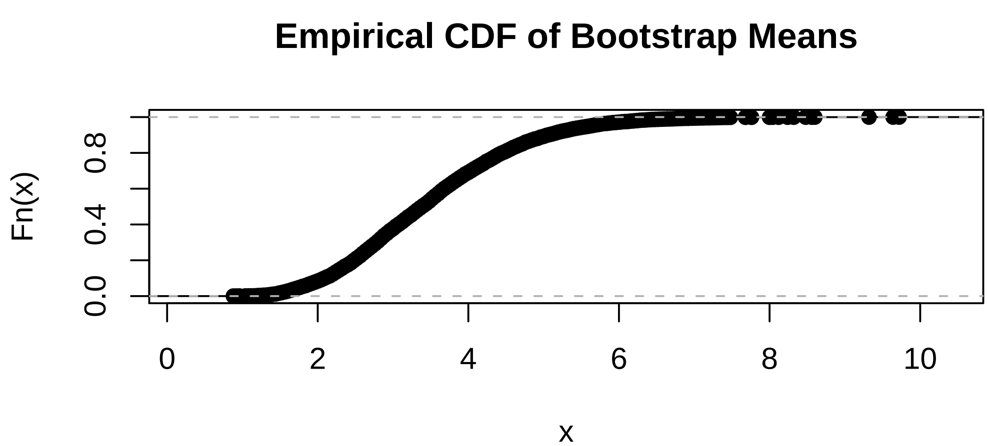
```
library(boot)
y <- c(2,2,1,4,1,0,5,3,1,6,0,0,3,1,3,0,3,0,2,20,0,2,3,1,25)
y
n <- length(y)
n
thetahat = mean(y)
thetahat
Brep = 10000

# Bootstrap the sample mean
sampmean <- function(y,i) mean(y[i])

bootmean <- boot(data=y,statistic=sampmean,R=Brep)
bootmean
boot.ci(bootmean,conf=.95,type=c("norm"))
boot.ci(bootmean,conf=.95,type=c("perc"))
boot.ci(bootmean,conf=.95,type=c("bca"))

par(mfrow=c(2,1))

hist(bootmean$t,main="Bootstrap Sample Means")
plot(ecdf(bootmean$t),main="Empirical CDF of Bootstrap Means")
```

R output for Bootstrapping the Mean – skewness present

```
[1] 2  2  1  4  1  0  5  3  1  6  0  0  3  1  3  0  3  0  2 20 0  2  3  1 25
[1] 25
> thetahat
[1] 3.52

> # Bootstrap the sample mean

ORDINARY NONPARAMETRIC BOOTSTRAP

Bootstrap Statistics :
    original    bias    std. error
t1*     3.52 -0.006596    1.188251


BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates
```

```
Intervals :
Level      Normal
95%   ( 1.198,  5.856 )
Calculations and Intervals on Original Scale


BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

Intervals :
Level     Percentile
95%   ( 1.56,  6.16 )
Calculations and Intervals on Original Scale


BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

Intervals :
Level       BCa
95%   ( 1.84,  7.18 )
Calculations and Intervals on Original Scale
```





4.4.4 Bias-Corrected Percentile Confidence Intervals

- One problem with using the bootstrap percentile method occurs when the assumption regarding the transformation to normality is not true.

- In this case, a confidence interval based on using the percentile method would not be appropriate. That is, the nominal (stated) confidence level is not close to the true confidence level.

- If the transformation f did exist, it is a monotonic increasing function f such that the transformed values $f(\widehat{\theta})$ are normally distributed with mean $f(\theta)$ and standard deviation 1. That is, $f(\widehat{\theta}) \sim N(\,f(\theta), 1)$. This implies that

$$\Pr\left\{f(\widehat{\theta}) > f(\theta)\right\} \;=\; \mathrm{P}(\widehat{\theta} > \theta) \;=\; 0.5.$$

  The probabilities are equal because the transformation is monotonic and increasing.

- Therefore, if such a transformation f exists we would expect that 50% of the bootstrap replications $(\widehat{\theta}^*(b), b = 1, 2, \ldots, B)$ would be greater that $\widehat{\theta}$. However, if the percentage is much higher or lower than 50%, we should consider removing bias.

- In other words, if B is large enough to adequately represent the distribution of the bootstrap replications, and the median of the bootstrap replications is not close to $\widehat{\theta}$, it may be necessary to modify the percentile bootstrap methods by adjusting for bias.

- To construct a bias-corrected confidence interval for θ, we relax the assumptions about the transformation f to be the following. We assume that a monotonic increasing function f exists for transforming $\widehat{\theta}$ such that the distribution of $f(\widehat{\theta})$ is normally distributed with mean $f(\theta) - z_0$ and standard deviation 1. That is, $f(\widehat{\theta}) \sim N(\,f(\theta) - z_0, 1)$, or, equivalently, $(f(\widehat{\theta}) - f(\theta) + z_0) \sim N(0, 1)$.

- This implies that $P(-z_{\alpha/2} \,<\, f(\widehat{\theta}) - f(\theta) + z_0 \,<\, z_{\alpha/2}) \;=\; 1 - \alpha$.

- Reordering of the terms yields the desired confidence interval for $f(\theta)$:

$$f(\widehat{\theta}) + z_0 - z_{\alpha/2} \,<\, f(\theta) \,<\, f(\widehat{\theta}) + z_0 + z_{\alpha/2}. \tag{11}$$

- By applying the inverse transformation f^{-1} to the confidence limits gives the confidence limits for θ. To apply this method we will need to estimate the constant z_0.

- Note that for any value t, we have

$$\begin{aligned}\Pr\{f(\widehat{\theta}) > t\} &= \Pr\{f(\widehat{\theta}) - f(\theta) + z_0 > t - f(\theta) + z_0\} \\ &= \Pr\{Z > t - f(\theta) + z_0\}\end{aligned}$$

  where $Z \sim N(0, 1)$. If we set $t = f(\theta)$, then

$$\Pr\{f(\widehat{\theta}) > f(\theta)\} \;=\; \Pr\{Z > z_0\}. \tag{12}$$

- Because f is monotonic and increasing, it follows from (12) that

$$\Pr\{\widehat{\theta} > \theta\} \;=\; \Pr\{Z > z_0\}.$$

Then, we assume that $\Pr\{\widehat{\theta} > \theta\}$ can be estimated by p, the proportion of bootstrap replications $\widehat{\theta}^*(b)$ that are greater than $\widehat{\theta}$. Thus, $z_0 \approx z_p$ where z_p is the value from from the $N(0,1)$ distribution having right-tail probability p.

- We now use z_p, the estimate of z_0, in (11) to find the value of p_U where

$$\begin{aligned} p_U &= \Pr\{f(\widehat{\theta}^*) < f(\widehat{\theta}) + z_0 + z_{\alpha/2}\} \\ &= \Pr\{f(\widehat{\theta}^*) - f(\widehat{\theta}) + z_0 < f(\widehat{\theta}) + z_0 + z_{\alpha/2} - f(\widehat{\theta}) + z_0\} \\ &= \Pr\{Z < 2z_0 + z_{\alpha/2}\} \end{aligned}$$

with $Z \sim N(0,1)$. This implies that the bootstrap upper confidence limit for $f(\theta)$ is the first bootstrap replication that is larger than p_U of the bootstrap replications $f(\widehat{\theta}^*)$. Recall: because the function f is unknown, we do not actually know the values of the transformed bootstrap replications $f(\widehat{\theta}^*)$. We only know that they exist.

- Then, once again we apply the assumption that f is monotonic and increasing, to find the bias-corrected upper confidence limit for θ. That is, **find that $\widehat{\theta}^*(b)$ value such that it is the first bootstrap replication that is larger than p_U of the B bootstrap replications $\widehat{\theta}^*$.**

- Similarly, we use z_p to find the value of p_L where

$$\begin{aligned} p_L &= \Pr\{f(\widehat{\theta}^*) < f(\widehat{\theta}) + z_0 - z_{\alpha/2}\} \\ &= \Pr\{f(\widehat{\theta}^*) - f(\widehat{\theta}) + z_0 < f(\widehat{\theta}) + z_0 - z_{\alpha/2} - f(\widehat{\theta}) + z_0\} \\ &= \Pr\{Z < 2z_0 - z_{\alpha/2}\} \end{aligned}$$

Thus, p_L is the proportion of bootstrap replications $\widehat{\theta}^*(b)$ that are less than $\widehat{\theta}$. To find the bias-corrected lower confidence limit for θ, **find that $\widehat{\theta}^*(b)$ such that it is the last bootstrap replication that is smaller than p_L of the B bootstrap replications $\widehat{\theta}^*$.**

- Therefore, the bias-corrected percentile confidence limit can be written as

$$\text{INVCDF}\{\Phi(2z_0 - z_{\alpha/2})\} \qquad \text{and} \qquad \text{INVCDF}\{\Phi(2z_0 + z_{\alpha/2})\}$$

where Φ is the standard normal CDF function and INVCDF is the inverse CDF of the empirical distribution of the B bootstrap replications $\widehat{\theta}^*(b), b = 1, 2, \ldots, B$.

4.4.5 Accelerated Bias-Corrected Percentile Confidence Intervals

- An alternative to a bias-corrected percentile confidence interval is the **accelerated bias-corrected percentile confidence interval**.

- The assumptions for the accelerated bias-corrected approach are less restrictive than the assumptions for the basic bias-corrected approach. We assume that a transformation $f(\widehat{\theta})$ of the estimator $\widehat{\theta}$ exists such that the distribution of $f(\widehat{\theta})$ is normal with mean $f(\theta) - z_0(1 + Af(\theta))$ and standard deviation $1 + Af(\theta)$. That is, $f(\widehat{\theta}) \sim N(\,f(\theta) - z_0(1 + Af(\theta)), 1 + Af(\theta))$ where z_0 and A are constants.

- Including the constant A allows for the standard deviation to vary linearly with $f(\theta)$. This additional flexibility will allow us to correct for this form of non-constant variance if it exists.

- Note: When $A = 0$, we are using the bias-corrected percentile approach.

- Standardizing $f(\widehat{\theta})$ by subtracting its mean and then dividing by its standard deviation implies

$$\Pr\left\{-z_{\alpha/2} \;<\; \frac{f(\widehat{\theta}) - f(\theta) + z_0(1 + Af(\theta))}{1 + Af(\theta)} \;<\; z_{\alpha/2}\right\} \;=\; \Pr(-z_{\alpha/2} < Z < z_{\alpha/2}) \;=\; 1 - \alpha$$

where $Z \sim N(0,1)$. This probability statement can be rewritten as

$$\Pr\left[\frac{f(\widehat{\theta}) + z_0 - z_{\alpha/2}}{1 - A(z_0 - z_{\alpha/2})} \;<\; f(\theta) \;<\; \frac{f(\widehat{\theta}) + z_0 + z_{\alpha/2}}{1 - A(z_0 + z_{\alpha/2})}\right] \;=\; 1 - \alpha \qquad (13)$$

or, more simply as

$$\Pr(L < f(\theta) < U) \;=\; 1 - \alpha$$

where L and U are the endpoints in (13).

- Let $f(\widehat{\theta}^*)$ denote a transformed bootstrap replicate. To approximate the lower limit L of $f(\theta)$ using bootstrapping, we assume that the bootstrap distribution of $f(\widehat{\theta}^*)$ approximates the distribution of $f(\widehat{\theta})$ which is $f(\widehat{\theta}^*) \sim N(\ f(\theta) - z_0(1 + Af(\theta)), 1 + Af(\theta))$.

- Therefore, we replace $f(\theta)$ in (13) with $f(\widehat{\theta}^*)$. The approximation is L^* where

$$\Pr\left[f(\widehat{\theta}^*) < L^*\right] \;=\; \Pr\left[f(\widehat{\theta}^*) < \frac{f(\widehat{\theta}) + z_0 - z_{\alpha/2}}{1 - A(z_0 - zat)}\right]$$

- After standardizing, we get

$$\begin{aligned}\Pr\left[f(\widehat{\theta}^*) < L^*\right] \;&=\; \Pr\left[\frac{f(\widehat{\theta}^*) - f(\widehat{\theta})}{1 + Af(\widehat{\theta})} + z_0 \;<\; \frac{z_0 - z_{\alpha/2}}{1 - A(z_0 - z_{\alpha/2})} + z_0\right] \qquad (14)\\ &=\; \Pr\left[Z \;<\; \frac{z_0 - z_{\alpha/2}}{1 - A(z_0 - z_{\alpha/2})} + z_0\right]\end{aligned}$$

where $Z \sim N(0,1)$.

- Equation (14) means that the probability of a transformed bootstrap replication $f(\widehat{\theta}^*)$ is less than the lower confidence limit for $f(\theta)$ equals the probability that a standard normal random variable is less than

$$z_L^* \;=\; \frac{z_0 - z_{\alpha/2}}{1 - A(z_0 - z_{\alpha/2})} + z_0$$

- Therefore, the lower confidence limit can be estimated by taking the value of the bootstrap distribution of $f(\widehat{\theta}^*)$ that is just greater than a fraction $\Phi(z_L^*)$. Although the form of transformation f is unknown, this is not a problem for finding the lower confidence limit of θ.

- Because of the assumption that f is monotonic and increasing, the lower confidence limit for θ is just the value of the bootstrap distribution of $\widehat{\theta}^*$ that is just greater than a fraction $\Phi(z_L^*)$.

- Using the same argument we can approximate the upper confidence limit for θ. That is, the upper confidence limit for θ is just the value of the bootstrap distribution of $\widehat{\theta}^*$ that is just greater than a fraction $\Phi(z_U^*)$ where

$$z_U^* \;=\; \frac{z_0 + z_{\alpha/2}}{1 - A(z_0 + z_{\alpha/2})} + z_0$$

- Therefore, the approximate $100(1-\alpha)\%$ accelerated bias-corrected bootstrap confidence interval for θ is

$$\text{INVCDF}\{\Phi(z_L)\} \;<\; \theta \;<\; \text{INVCDF}\{\Phi(z_U)\} \qquad (15)$$

  where INVCDF is the inverse of the empirical CDF of the bootstrap replications $\widehat{\theta}^*(b)$ for $b = 1, 2, \ldots, B$.

- The remaining problem is how to estimate the constants z_0 and A.

- z_0 can be estimated from the empirical CDF of the bootstrap replications $\widehat{\theta}^*$ by continuing to assume $f(\widehat{\theta}^*) \sim N(\, f(\theta) - z_0(1 + Af(\theta)), 1 + Af(\theta))$. Then

$$\begin{aligned}\Pr\left[f(\widehat{\theta}^*) > f(\widehat{\theta})\right] \;&=\; \Pr\left[\frac{f(\widehat{\theta}^*) - f(\widehat{\theta})}{1 + Af(\widehat{\theta})} + z_0 \;>\; z_0\right] \\ &=\; \Pr\left[Z > z_0\right]\end{aligned}$$

  where $Z \sim N(0,1)$. Because f is assumed to be monotonic and increasing it also holds that

$$\Pr\left[\widehat{\theta}^* > \widehat{\theta}\right] \;=\; \Pr\left[Z > z_0\right]$$

- Let p be the proportion of values in the bootstrap distribution of $\widehat{\theta}^*$ that are greater than $\widehat{\theta}$. Then z_0 can be estimated as $z_0 \;=\; z_p$ where z_p is the value such that

$$1 - \Phi(z_p) \;=\; p$$

  This is the same as the value derived for the bias-corrected percentile method.

- The final problem is estimation of the constant A. Unfortunately, A cannot be simply derived using probability statements like we did for z_0. Efron and Tibshirani (1993) recommend the following which uses jackknife replications.

- Let $\widehat{\theta}_{(i)}$ be the i^{th} jackknife replication of i^{th} and $\widehat{\theta}_{(\cdot)}$ be the mean of the n jackknife replications. Then a, the estimated value of A, is

$$a \;=\; \frac{\sum_{i=1}^{n}\left(\widehat{\theta}_{(\cdot)} - \widehat{\theta}_{(i)}\right)^3}{6\left[\sum_{i=1}^{n}\left(\widehat{\theta}_{(\cdot)} - \widehat{\theta}_{(i)}\right)^2\right]^{1.5}}$$

Table 3.2. Jackknife Calculations with the Estimation of the Population Standard Deviation from a Sample of Size 20

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	SD	PV	Log SD	PV′
Data	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.03 ←		0.032	
1		0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	0.879	3.959	−0.129	3.100
2	3.56		0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.056	0.586	0.055	−0.396
3	3.56	0.69		1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.036	0.971	0.035	−0.027
4	3.56	0.69	0.10		3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.043	0.840	0.042	−0.154
5	3.56	0.69	0.10	1.84		1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	0.813	5.202	−0.207	4.570
6	3.56	0.69	0.10	1.84	3.93		0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.059	0.544	0.057	−0.435
7	3.56	0.69	0.10	1.84	3.93	1.25		1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.040	0.898	0.039	−0.098
8	3.56	0.69	0.10	1.84	3.93	1.25	0.18		0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.059	0.527	0.058	−0.452
9	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13		0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.044	0.823	0.043	−0.170
10	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27		0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.052	0.671	0.051	−0.315
11	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50		0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.056	0.593	0.054	−0.389
12	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67		0.61	0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.031	1.062	0.031	0.060
13	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01		0.82	1.70	0.39	0.11	1.20	1.21	0.72	1.055	0.617	0.053	−0.366
14	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61		1.70	0.39	0.11	1.20	1.21	0.72	1.058	0.548	0.057	−0.431
15	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82		0.39	0.11	1.20	1.21	0.72	1.048	0.738	0.047	−0.251
16	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70		0.11	1.20	1.21	0.72	1.048	0.737	0.047	−0.252
17	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39		1.20	1.21	0.72	1.037	0.962	0.036	−0.037
18	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11		1.21	0.72	1.059	0.535	0.057	−0.444
19	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20		0.72	1.059	0.537	0.057	−0.442
20	3.56	0.69	0.10	1.84	3.93	1.25	0.18	1.13	0.27	0.50	0.67	0.01	0.61	0.82	1.70	0.39	0.11	1.20	1.21		1.057	0.575	0.055	−0.405
Jackknife estimate (mean)																						1.096 ←		0.133
Jackknife standard error																						0.273		0.290
Lower 95% confidence interval																						0.526		−0.473
Upper 95% confidence interval																						1.666		0.740

$\hat{\sigma} = 1.03285$ $\widehat{BIAS} = 19(1.02952 - 1.03285) = -.06327$

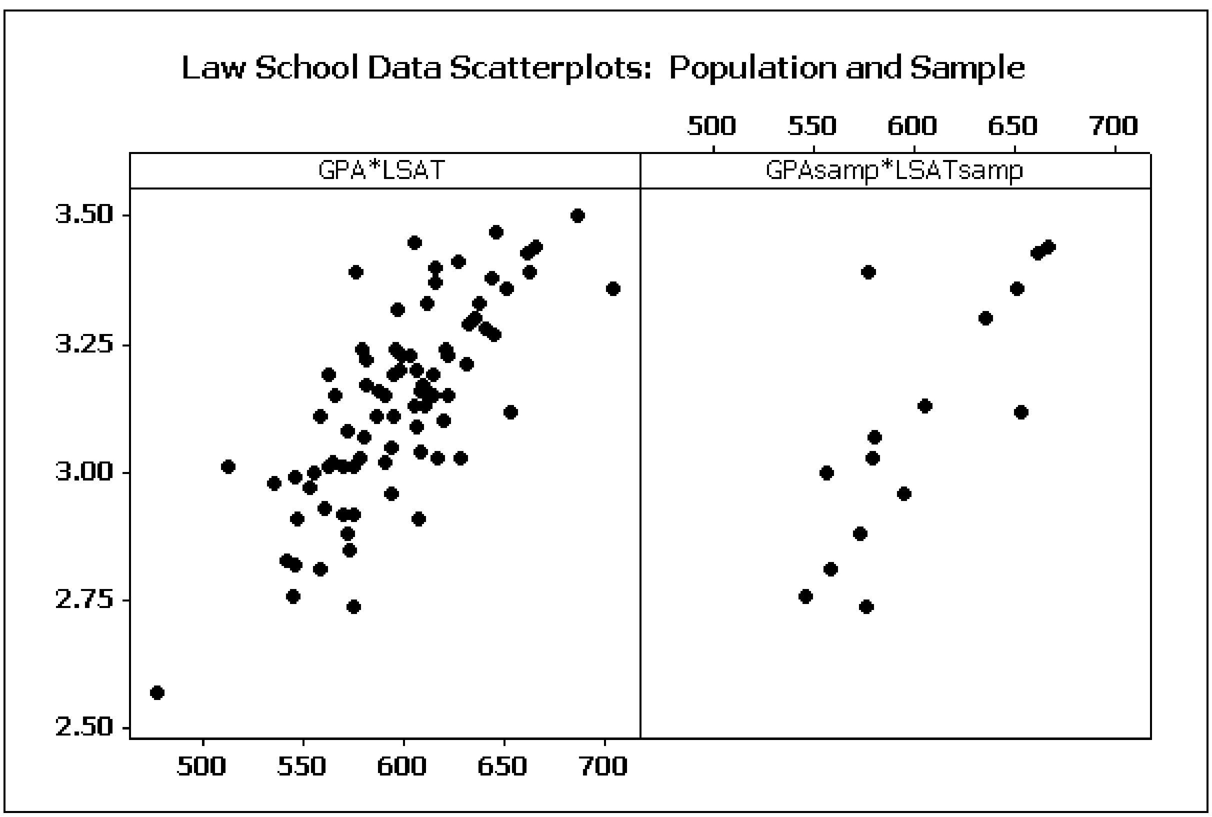
$\hat{\sigma}_{(\cdot)} = 1.02952$ $\Rightarrow \hat{\sigma}_{JACK} = \hat{\sigma} - \widehat{BIAS} = 1.03285 - (-.06327) = 1.09612$

Table 4.1: Law School data in Efron and Tibshirani (1993)

school	LSAT	GPA	school	LSAT	GPA	school	LSAT	GPA	school	LSAT	GPA
1	622	3.23	22	614	3.19	43	573	2.85	63	572	3.08
2	542	2.83	23	628	3.03	44	644	3.38	64	610	3.13
3	579	3.24	24	575	3.01	**(45)**	545	2.76	65	562	3.01
(4)	653	3.12	25	662	3.39	46	645	3.27	66	635	3.30
5	606	3.09	26	627	3.41	**(47)**	651	3.36	67	614	3.15
(6)	576	3.39	27	608	3.04	48	562	3.19	68	546	2.82
7	620	3.10	28	632	3.29	49	609	3.17	69	598	3.20
8	615	3.40	29	587	3.16	**(50)**	555	3.00	**(70)**	666	3.44
9	553	2.97	30	581	3.17	51	586	3.11	71	570	3.01
10	607	2.91	**(31)**	605	3.13	**(52)**	580	3.07	72	570	2.92
11	558	3.11	32	704	3.36	**(53)**	594	2.96	73	605	3.45
12	596	3.24	33	477	2.57	54	594	3.05	74	565	3.15
(13)	635	3.30	34	591	3.02	55	560	2.93	75	686	3.50
14	581	3.22	**(35)**	578	3.03	56	641	3.28	76	608	3.16
(15)	661	3.43	**(36)**	572	2.88	57	512	3.01	77	595	3.19
16	547	2.91	37	615	3.37	58	631	3.21	78	590	3.15
17	599	3.23	38	606	3.20	59	597	3.32	**(79)**	558	2.81
18	646	3.47	39	603	3.23	60	621	3.24	80	611	3.16
19	622	3.15	40	535	2.98	61	617	3.03	81	564	3.02
20	611	3.33	41	595	3.11	62	637	3.33	**(82)**	575	2.74
21	546	2.99	42	575	2.92						

Sampled schools have bold-faced school numbers.

Figure 4.1



Example: Bootstrapping a correlation coefficient ρ

- Reconsider the law school data in Table 4.1 and Figure 4.1 on the previous page.
- $\widehat{\rho} = r =$ Pearson correlation coefficient
- $\widehat{\xi} = \frac{1}{2} ln \left(\frac{1+r}{1-r} \right) =$ transformed Pearson correlation coefficient
- We will use R to generate bootstrap estimates of ρ and ξ for the law school data given on the previous page.

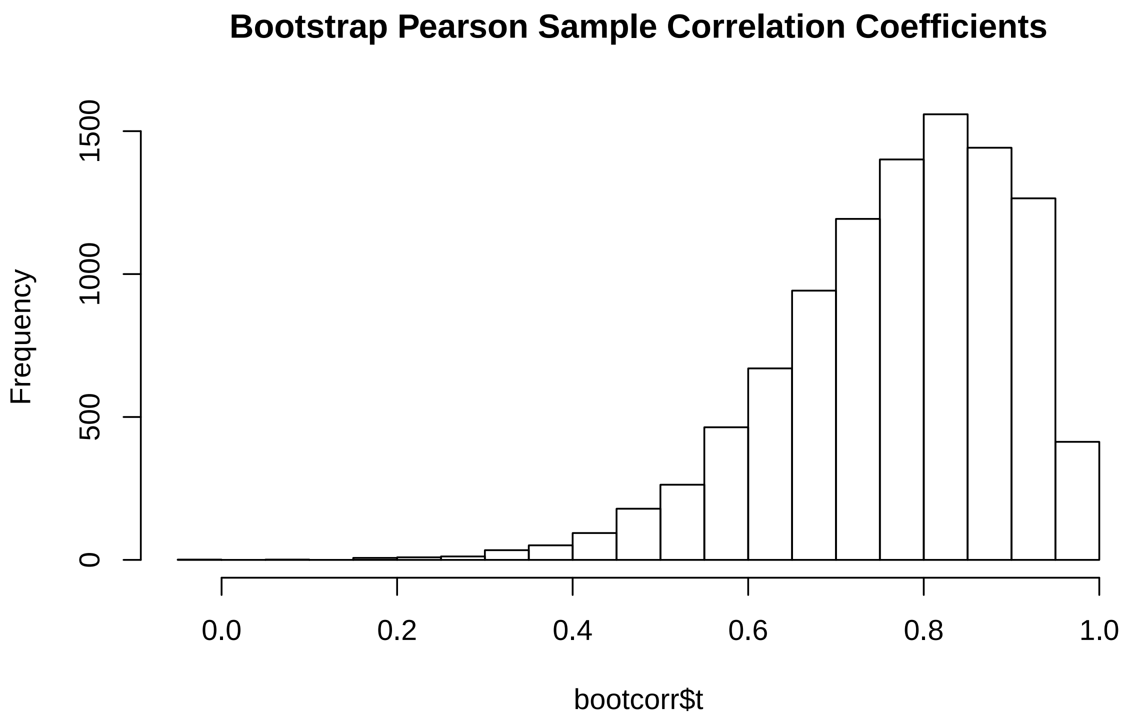
R output for Bootstrapping r and $\widehat{\xi}$

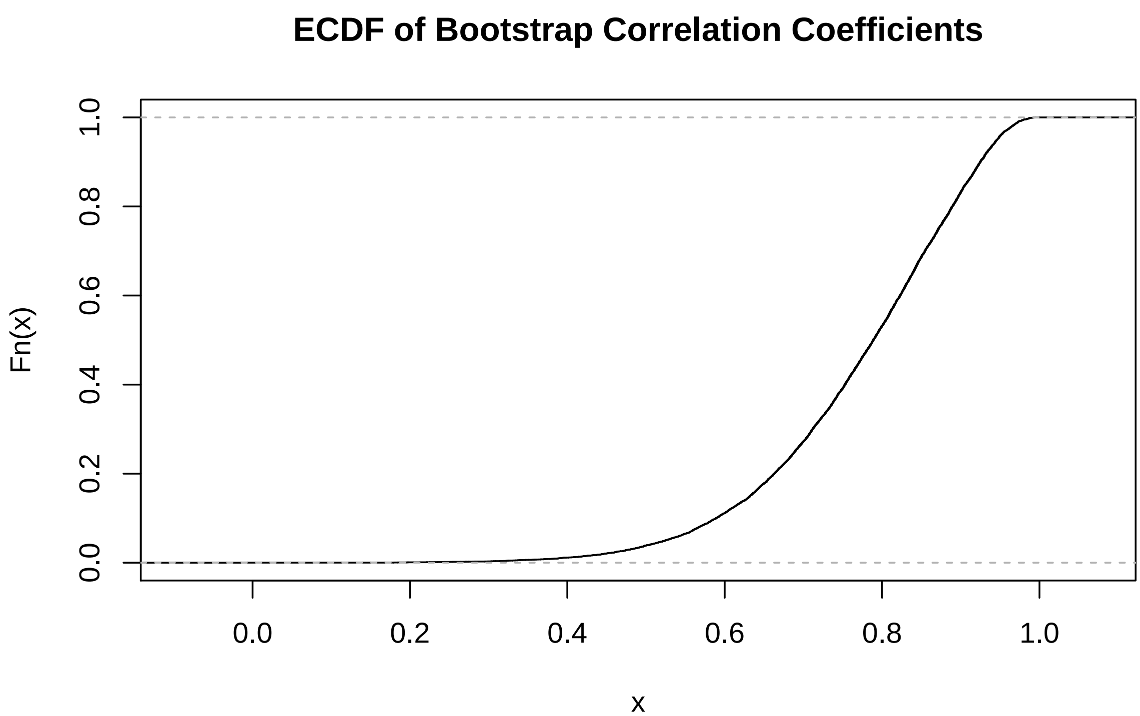
```
> # Bootstrap the Pearson correlation coefficient

ORDINARY NONPARAMETRIC BOOTSTRAP

Bootstrap Statistics :

     original       bias    std. error
t1* 0.7763745 -0.006956509   0.1324587

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

Intervals :
Level      Normal              Basic
95%   ( 0.5237,  1.0429 )   ( 0.5914,  1.0887 )

Level     Percentile            BCa
95%   ( 0.4641,  0.9613 )   ( 0.3369,  0.9403 )


> # Bootstrap the transformed Pearson correlation coefficient

ORDINARY NONPARAMETRIC BOOTSTRAP

Bootstrap Statistics :

    original     bias    std. error
t1* 1.036178 0.08097614   0.3794925

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates


Intervals :
Level      Normal              Basic
95%   ( 0.211,  1.699 )   ( 0.105,  1.565 )

Level     Percentile            BCa
95%   ( 0.507,  1.967 )   ( 0.356,  1.759 )
```

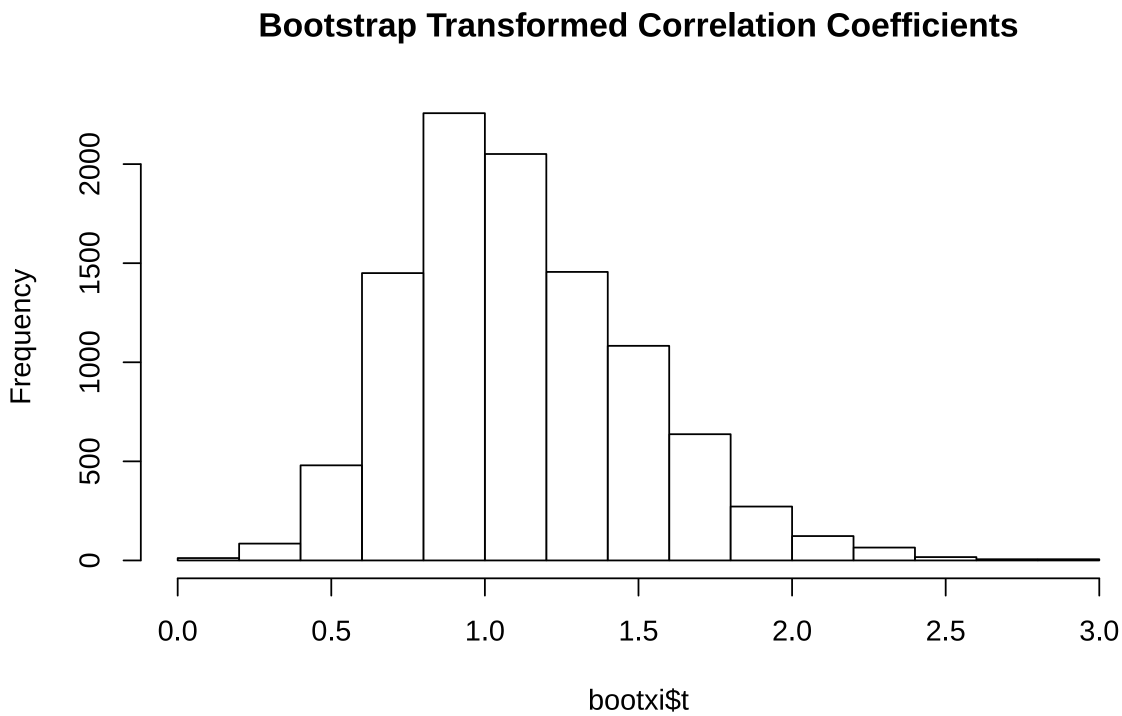
R code for Bootstrapping r and $\widehat{\xi}$

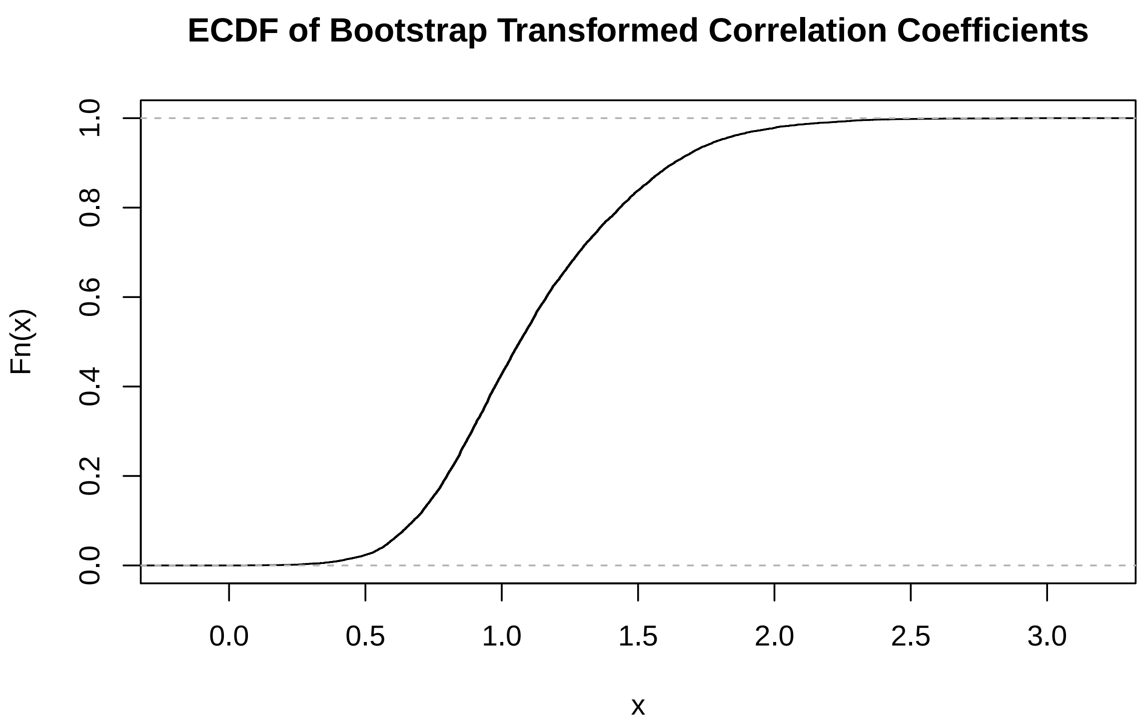
```
library(boot)

LSAT <- c(576,635,558,578,666,580,555,661,651,605,653,575,545,572,594)
GPA <- c(3.39,3.30,2.81,3.03,3.44,3.07,3.00,3.43,3.36,3.13,3.12,2.74,
2.76,2.88,2.96)

n = length(LSAT)
Brep = 10000

xy <- data.frame(cbind(LSAT,GPA))

# Bootstrap the Pearson correlation coefficient

pearson <- function(d,i=c(1:n)){
        d2 <- d[i,]
        return(cor(d2$LSAT,d2$GPA))
     }
bootcorr <- boot(data=xy,statistic=pearson,R=Brep)
bootcorr
boot.ci(bootcorr,conf=.95)

windows()
par(mfrow=c(2,1))
hist(bootcorr$t,main="Bootstrap Pearson Sample Correlation Coefficients")
plot(ecdf(bootcorr$t),main="ECDF of Bootstrap Correlation Coefficients")


# Bootstrap the transformed Pearson correlation coefficient

xihat <- function(dd,i=c(1:n)){
        dd2 <- dd[i,]
        return(.5*log((1+cor(dd2$LSAT,dd2$GPA))/(1-cor(dd2$LSAT,dd2$GPA))))
     }
bootxi <- boot(data=xy,statistic=xihat,R=Brep)
bootxi
boot.ci(bootxi,conf=.95)

windows()
par(mfrow=c(2,1))
hist(bootxi$t,main="Bootstrap Transformed Correlation Coefficients")
plot(ecdf(bootxi$t),main="ECDF of Bootstrap Transformed Correlation
Coefficients")
```

Here are histograms and ECDF plots of 10000 bootstrap replications of r and $\widehat{\xi}$.


Bootstrap Pearson Sample Correlation Coefficients
Frequency
0
500
1000
1500
0.0
0.2
0.4
0.6
0.8
1.0
bootcorr$t


ECDF of Bootstrap Correlation Coefficients
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
x


Bootstrap Transformed Correlation Coefficients
Frequency
0
500
1000
1500
2000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
bootxi$t


ECDF of Bootstrap Transformed Correlation Coefficients
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
x