Review: Sampling Distributions and Confidence Intervals

Bootstrap Methods

Useful when

- small sample
- or conditions not met
- or no procedure available

Idea: simulate the sampling distribution

Bootstrap Methods for μ

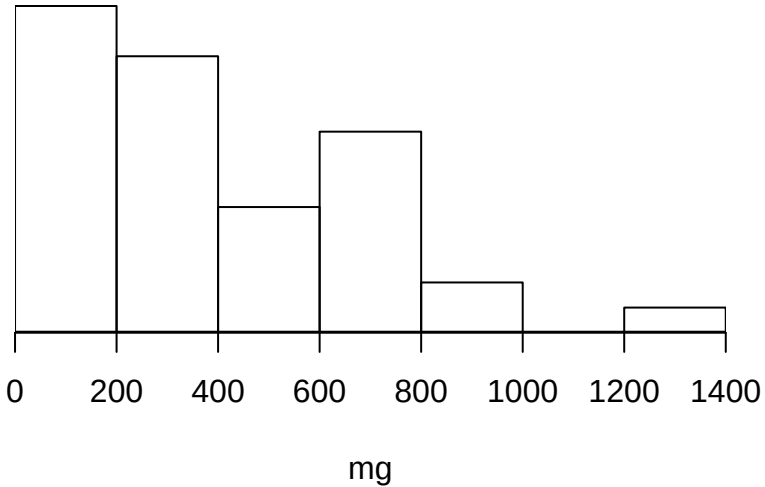
- random variable is not normally distributed
- n is small

Bootstrap Idea

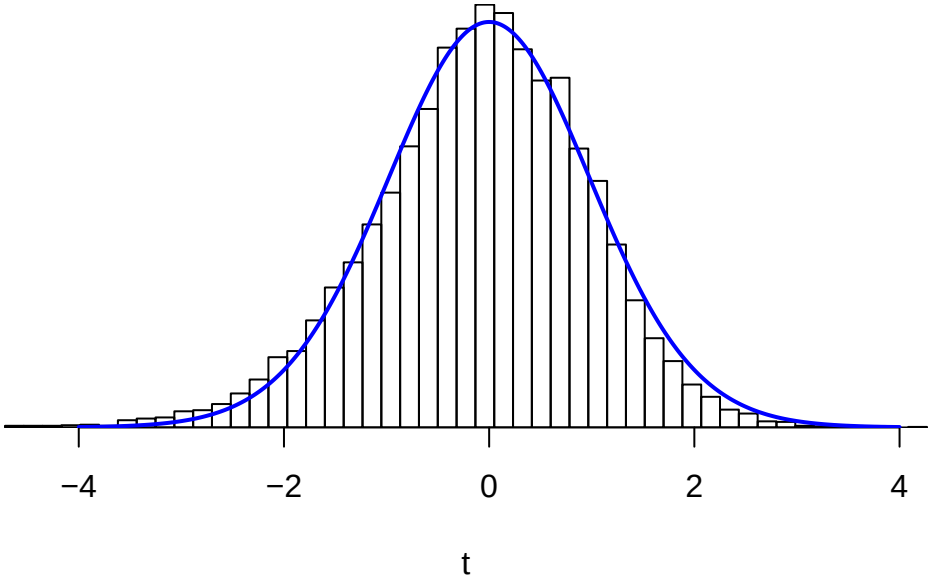
- original sample is the “pseudo” population
- resample with replacement
- $\Rightarrow$ approximate sampling distribution

Men's Cholesterol Levels

$n = 40$ and *skewed* distribution



Bootstrap CI simulated t distribution



The critical values

critical value	bootstrap	theory
$t_{.025}$	-2.3	-2.02
$t_{.975}$	1.84	2.02

Unfortunate Notation

$t_{.025}$

- the .025 or the .975 percentile?
- context
- if symmetry, then not important

Formula for the Confidence Interval

$$t_{.025} < t = \frac{\overline{x} - \mu}{s/\sqrt{n}} < t_{.975}$$

$$\overline{x} - t_{.975}\frac{s}{\sqrt{n}} < \mu < \overline{x} - t_{.025}\frac{s}{\sqrt{n}}$$

Formula in Ott is wrong, compare above to page 260

Results

Bootstrapped (good):

$$(395.23 - 84.95, 395.23 + 106.14) = (310.28, 501.37)$$

Theory (do not use):

$$(395.23 - 93.52, 395.23 + 93.52) = (301.71, 488.75)$$

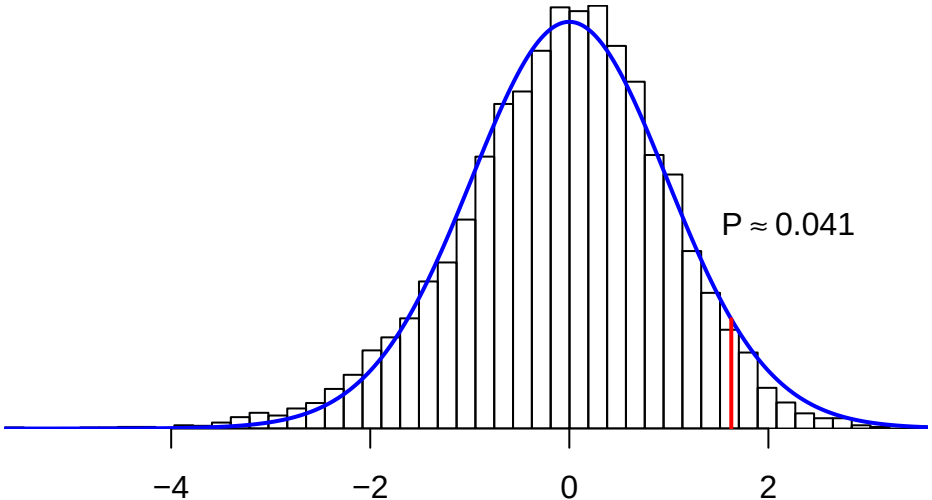
Results of the t-test for μ

```r
t.test( chol, mu = 320, alternative="greater")
```

```
##
##  One Sample t-test
##
## data:  chol
## t = 1.627, df = 39, p-value = 0.05589
## alternative hypothesis: true mean is greater than 320
## 95 percent confidence interval:
##  317.3259      Inf
## sample estimates:
## mean of x
##   395.225
```

Bootstrap t-test

Simulate test statistic with 10,000 resamples:



Result

- Cholesterol levels very skewed $\Rightarrow$ t-test inaccurate.
- Bootstrap t-test accounts for the skewness so reject H_0
- *Men's average cholesterol level is greater than 320 mg ($P = 0.041$)*