Central Limit Theorem

Grinnell College

March 4, 2024

Review

Bootstrapping involves the process of *resampling with replacement* from our original sample

When we compute a statistic on our bootstrapped sample (i.e., sample mean), we have a *bootstrapped sample statistic*

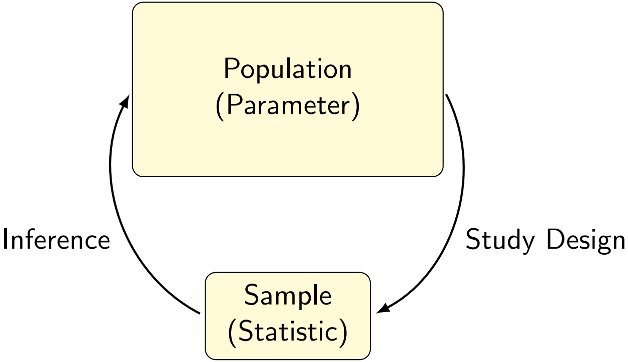
Repeating this process many many times gives us an estimate of the *sampling distribution*

Percentiles can be used on this bootstrapped sampling distribution without needing any further assumptions

Questions

1. Why are we constructing confidence intervals? What purpose do they serve?
2. Explain how the Point Estimate $\pm$ Margin of Error method works
3. What is the bootstrapping process? How is it used to compute confidence intervals?
4. Could I use Point Estimate $\pm$ Margin of Error on my bootstrapped sample means?
5. Explain how to create an 80% confidence interval using:
   - Point Estimate $\pm$ Margin of Error method
   - Bootstrapping
6. What determines the width of our confidence intervals?
7. What determines how frequently they contain the true parameter value, on average?

The Statistical Framework



Sequences and Limits

Consider a *sequence* of numbers:

$$1, \ \frac{1}{2}, \ \frac{1}{3}, \ \frac{1}{4}, \ \dots, \ \frac{1}{n}, \ \dots$$

Where are they going?

Sequences and Limits

In what sense are these numbers *going to zero?*

$$1, \ \frac{1}{2}, \ \frac{1}{3}, \ \frac{1}{4}, \ \ldots, \ \frac{1}{n} \to 0$$

- Will they ever reach zero?
- Is there a value of n in which it is "close enough?"
- Why would this be useful to know?

Sequences and Limits

What about these sequences?

$$\frac{10,000}{n}$$

$$\frac{1}{n^2}$$

$$\frac{n}{1+n^3}$$

Normal Distribution

If our variable X is normally distributed, with $X \sim N(\mu, \sigma^2)$, we have the curve:

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Normal Distribution

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

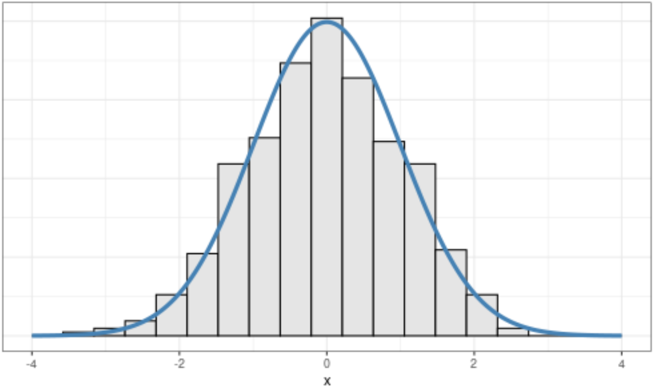
Note the *standardized variable*, $Z = \frac{x-\mu}{\sigma}$. Making this replacement gives a *standard normal distribution*,

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

which has a mean value of $\mu = 0$ and a standard deviation of $\sigma = 1$

Normal Distribution

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



The particulars of our sample may change

But we *assume* they are generated by normal, then we only need to know two things

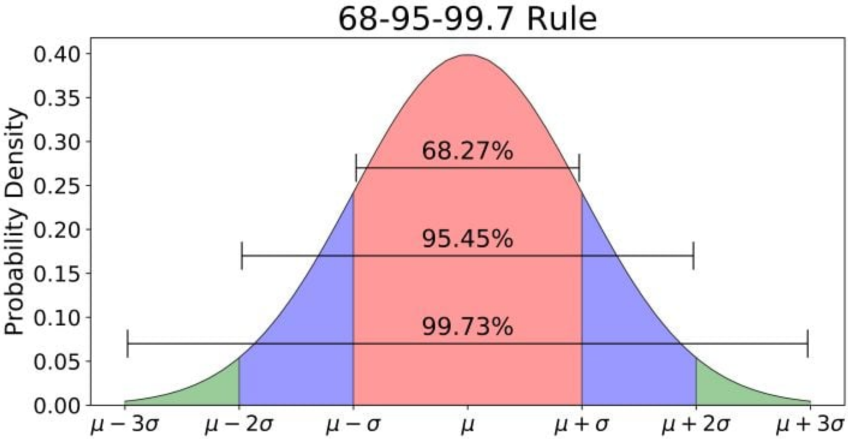
And just make statements about the general form rather than pertaining to specifics of sample

Central Limit Theorem

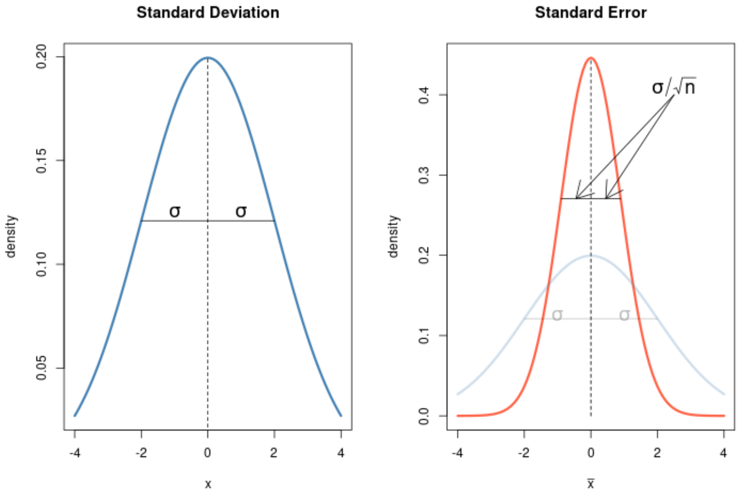
The **Central Limit Theorem (CLT)** is perhaps the most significant result in all of statistics:

1. A variable X has mean μ and standard deviation σ
2. The number of observations in our sample is n
3. The *sample mean* has a *sampling distribution* with mean μ and standard error $\sigma/\sqrt{n}$

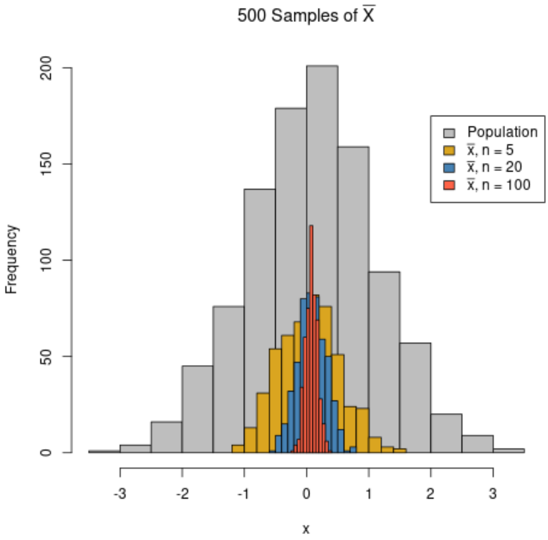
$$\overline{X} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right)$$


68-95-99.7 Rule
Probability Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
68.27%
95.45%
99.73%
μ − 3σ
μ − 2σ
μ − σ
μ
μ + σ
μ + 2σ
μ + 3σ

Standard Deviation and Standard Error



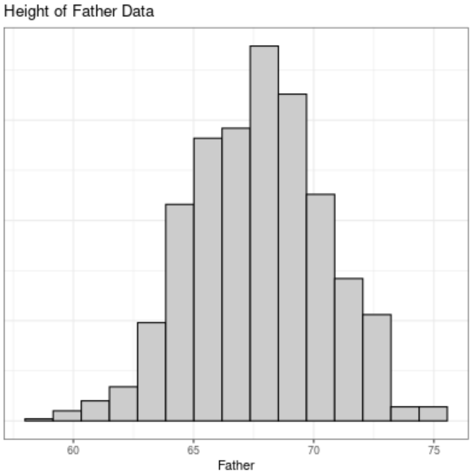
Sample Mean Distribution



Father Height Data

From our sample of $n = 1,078$, we find:

- $\overline{x} = 67.68$
- $\hat{\sigma} = 2.74$



Father Height Data

From our sample of $n = 1,078$, we find:

- $\overline{x} = 67.68$
- $\hat{\sigma} = 2.74$

To determine sampling distribution, we have

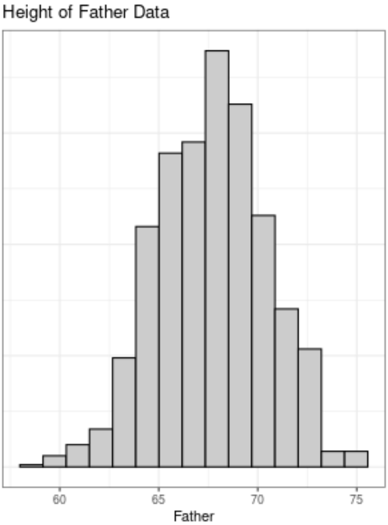
- Best prediction for mean is $\overline{x} = 67.68$
- Standard error of

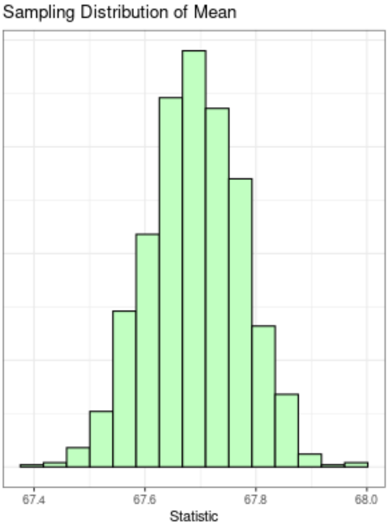
$$\text{Std. Error} = \frac{\hat{\sigma}}{\sqrt{n}} = 2.74/\sqrt{1,078} = 0.084$$

A 95% confidence interval would then look like:

$$67.69 \pm 2 \times 0.084 = (67.512, 67.848)$$

Father Height Data

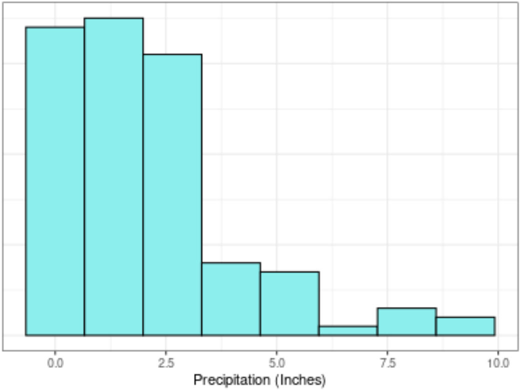




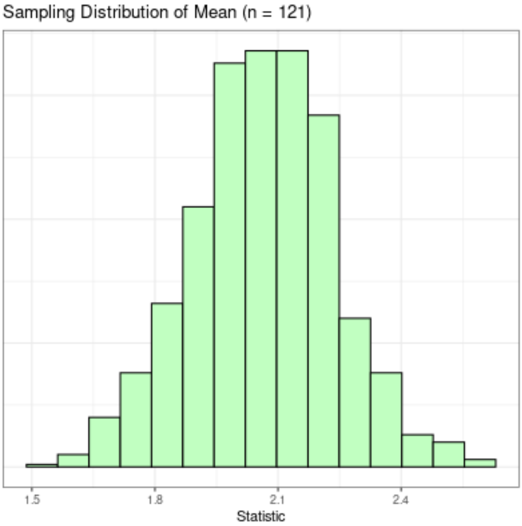
Non-normal Populations

What about our Grinnell rain data?

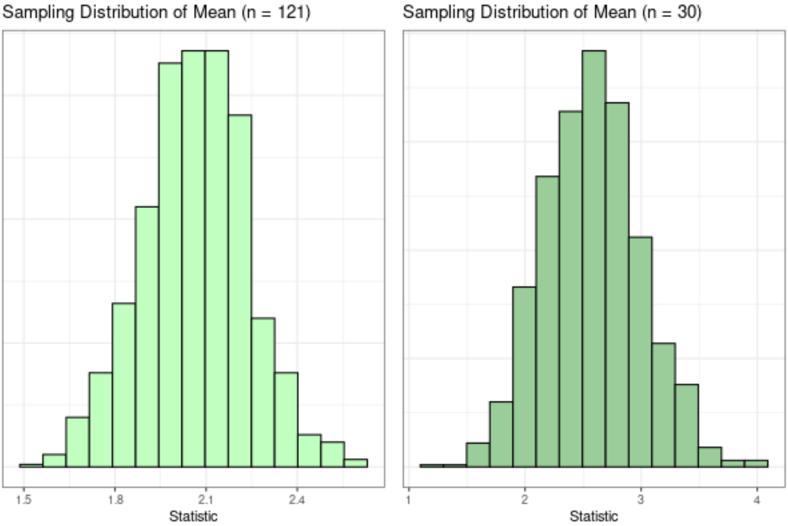
- $n = 121$
- $\overline{x} = 2.07$
- $\hat{\sigma} = 1.97$



Rain Mean Sampling Distribution



Changing sample size



Limits and Approximations

Review

The **Central Limit Theorem** *only* applies to sampling distribution of the mean

It *does not* require that the population follow a normal distribution