

# Confidence Intervals and Bootstrap

Stat 120

April 11 2023

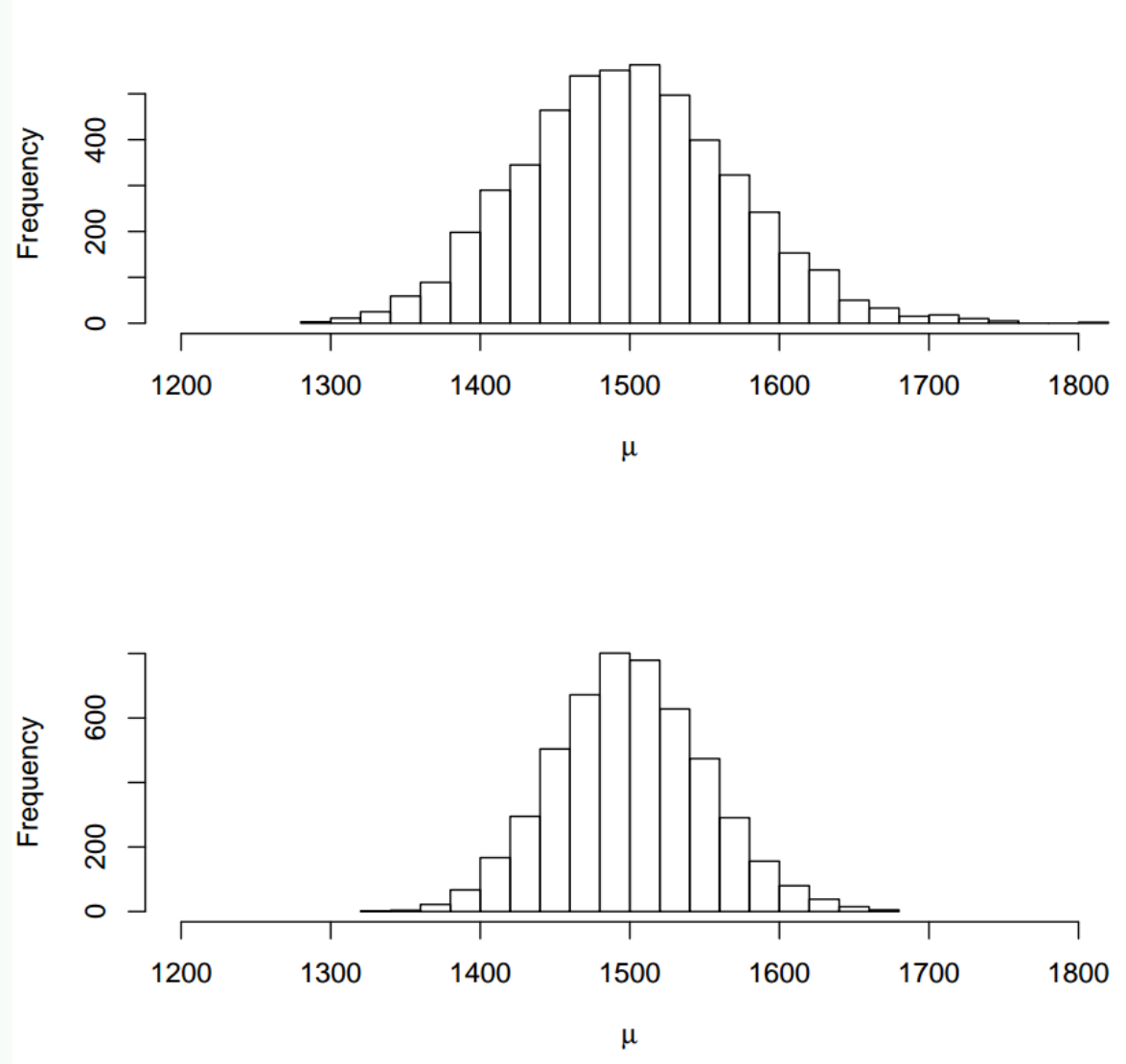
Let's start with a question!

*The higher the standard error of a statistic, the ..... the uncertainty surrounding the statistic.*

- 1. higher*
- 2. lower*

# Sampling distribution

Sampling distributions. Top  $n=50$ , Bottom  $n=100$



## Interval Estimation

- *Point estimates are almost always not accurate*
- *Uncertainty in point estimates measured by the Standard Error (SE)*
- *A plausible range of values for the population parameter is more reliable*
- *Interval Estimate: An interval estimate is an interval of numbers within which the parameter value is believed to fall*

# A Gallup Poll

VOTE FOR A MUSLIM FOR PRESIDENT

NO, WOULDN'T

YES, WOULD

38%

60%

SOURCE: GALLUP POLL; JUNE 2-7, 2015

***How accurate is an estimate of 60%?***

## Survey Methods

Results for this Gallup poll are based on telephone interviews conducted June 2-7, 2015, with a random sample of 1,527 adults, aged 18 and older, living in all 50 U.S. states and the District of Columbia. For results based on the total sample of national adults, the margin of sampling error is  $\pm 3$  percentage points at the 95% confidence level. All reported margins of sampling error include computed design effects for weighting.

# A Gallup Poll

VOTE FOR A MUSLIM FOR PRESIDENT

NO, WOULDN'T

YES, WOULD

38%

60%

SOURCE: GALLUP POLL; JUNE 2-7, 2015

**" ... the margin of sampling error is  $\pm 3$  percentage points at the 95% confidence level. "**

- ***Interval estimate:***  
 $60\% \pm 3\% = (57\%, 63\%)$
- ***The percentage of American adults who would vote for a Muslim for president is likely between 57% and 63%.***
- ***Would a majority of US adults vote for a qualified Muslim presidential candidate?***

## Margin of Error

- The **margin of error** measures how accurate a point estimate is likely to be in estimating a parameter.
- To determine the margin of error, we can use the statistic's **sampling distribution** and **standard error**

## Confidence Intervals

- A **confidence interval** is an interval containing the most believable values for a parameter
- A confidence interval is centered on the **point estimate** and extends a certain number of **standard errors** on either side of the estimate
- The **confidence level** tells us what percent of the intervals will contain the population parameter.

A 95% confidence interval will contain the true parameter for 95% of all samples.



## Gallup Poll Result Interpretation

***" ... the margin of sampling error is  $\pm 3$  percentage points at the 95% confidence level."***

***95% confidence means that 95% of all samples will yield a sample proportion that is within 3 percentage points of the population proportion***

## 95% Confidence Interval

*If the sampling distribution is relatively **symmetric** and **bell-shaped**, a 95% confidence interval can be estimated using*

$$\text{statistic} \pm 2 \times \text{SE}$$

*95% confidence means 95% of all samples will yield a statistic that is within 2 SE of the population parameter*

## Confidence Intervals are ...

- *always about the **population***
- *not **probability statements***
- *only about **population parameters**, not individual observations*
- *only reliable if the sample statistic they're based on is an **unbiased** estimator of the population parameter*

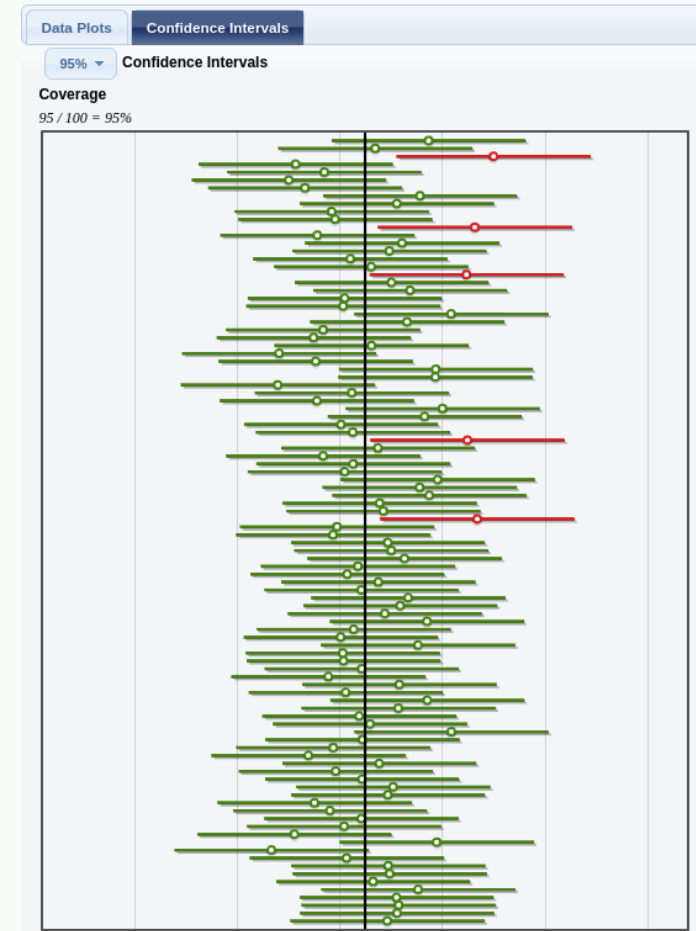
## A short demo

Let's all go to [Statkey](#) web app.



## Take Home Points

- The parameter is **fixed**
- The statistic is **random** (depends on the sample)
- The interval is also **random** (depends on the statistic)
- Confidence level is the **proportion** of intervals that capture the true parameter



Repeated Sampling of 100 95% Confidence Intervals,  
Truth = Vertical Line

What to do when we only have one sample - **BOOTSTRAP!**

- Repeated sampling is needed to compute the standard error of a sample statistic
- Can estimate the SE from a bootstrap distribution
- Use this SE to compute a confidence interval for an unknown parameter

# Bootstrap Distrubution using CarletonStats

```
library(CarletonStats)
X <- c(20,24,19,23,22,16)
boot(X)
```

```

** Bootstrap interval for statistic

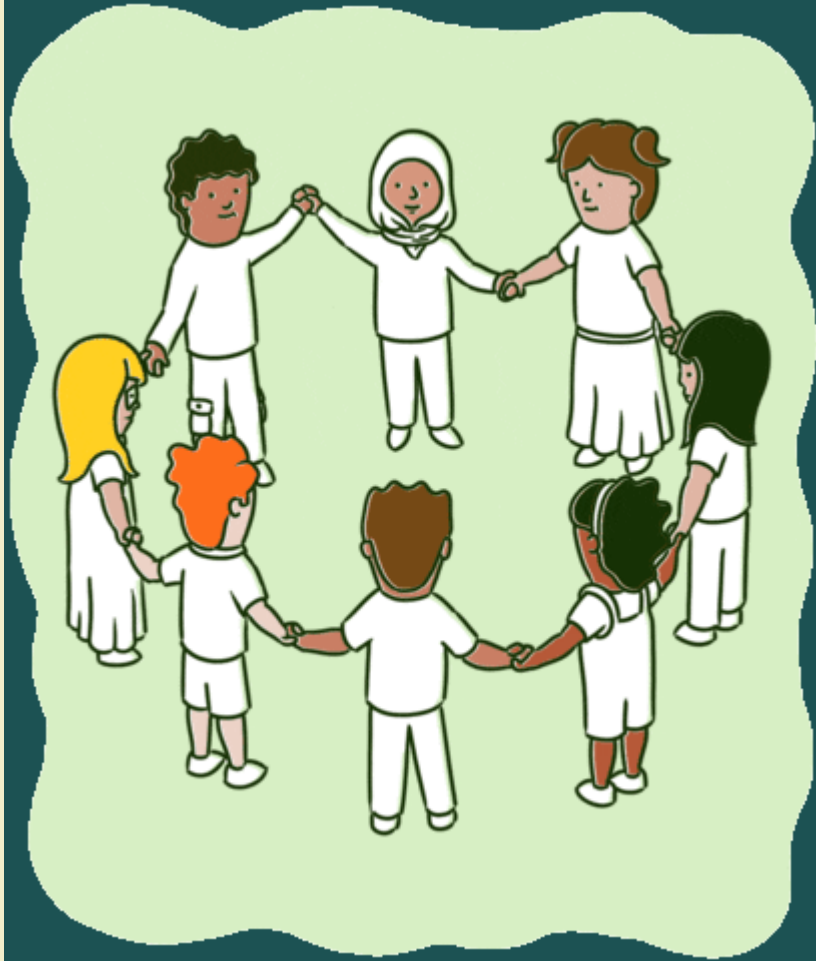
Observed  X : 20.66667
Mean of bootstrap distribution: 20.6599
Standard error of bootstrap distribution: 1.096

Bootstrap percentile interval
      2.5%      97.5%
18.33333 22.66667

      *-----*
```

# YOUR TURN 1

15:00



Go over the remaining portion of  
in class activity and let me  
know if you have any questions!