Null Hypothesis Testing and Bootstrapping

IS381 - Statistics & Probability for Data Analytics

Jason Bryer, Ph.D.

March 23, 2026

Hypothesis Testing

- We start with a null hypothesis (H_0) that represents the status quo.

- We also have an alternative hypothesis (H_A) that represents our research question, i.e. what we're testing for.

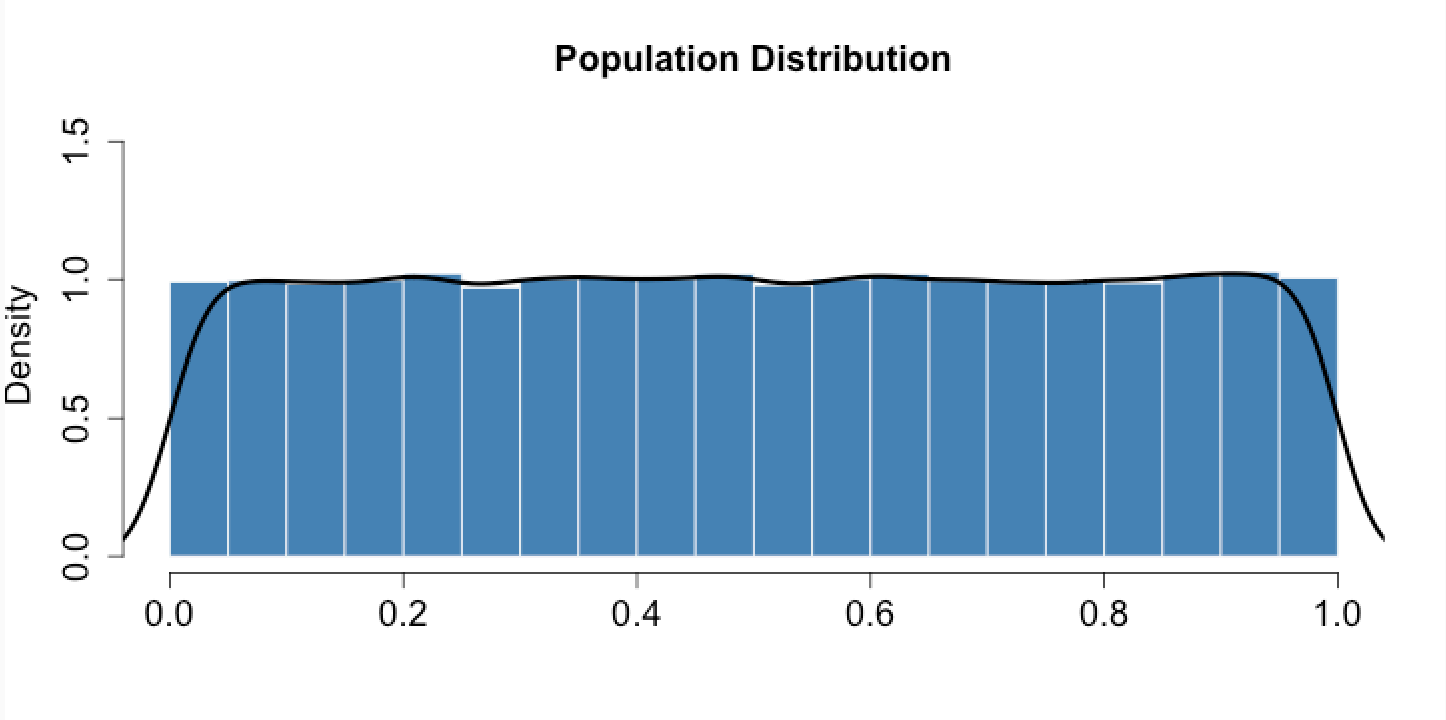
- We conduct a hypothesis test under the assumption that the null hypothesis is true, either via simulation or traditional methods based on the central limit theorem.

- If the test results suggest that the data do not provide convincing evidence for the alternative hypothesis, we stick with the null hypothesis. If they do, then we reject the null hypothesis in favor of the alternative.

Population Distribution (Uniform)

```
n <- 1e5
pop <- runif(n, 0, 1)
mean(pop)
```

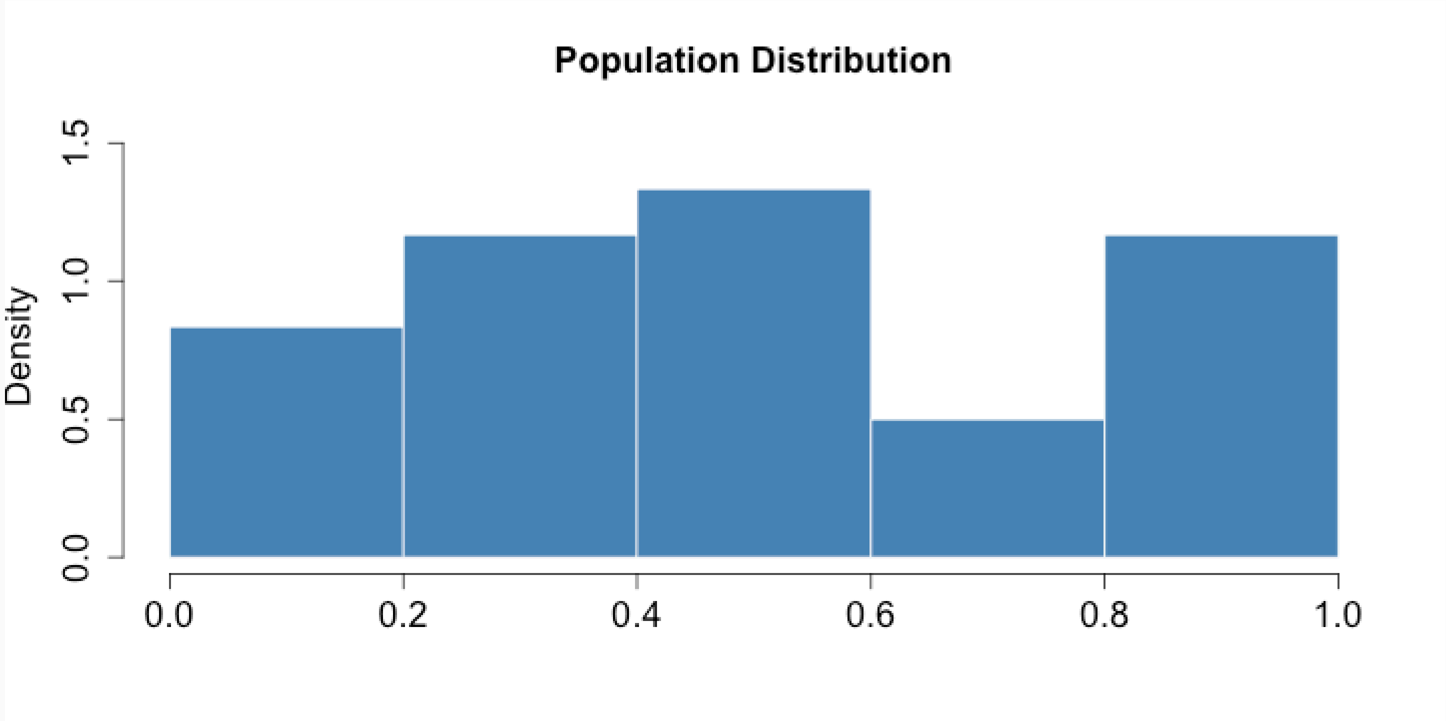
```
## [1] 0.5012701
```



Sample (n = 30)

```r
samp2 <- sample(pop, size=30)
mean(samp2)
```

```
## [1] 0.5129912
```



Hypothesis Testing (using CI)

H_0: The mean of `samp2` = 0.5
H_A: The mean of `samp2` $\neq$ 0.5

Using confidence intervals, if the *null* value is within the confidence interval, then we *fail* to reject the *null* hypothesis.

```
(samp2.ci <- c(mean(samp2) - 1.96 * sd(samp2) / sqrt(length(samp2)),
               mean(samp2) + 1.96 * sd(samp2) / sqrt(length(samp2))))
```

```
## [1] 0.4171924 0.6087900
```

Since 0.5 fall within 0.4171924, 0.60879, we *fail* to reject the null hypothesis.

Hypothesis Testing (using *p*-values)

$$\bar{x} \sim N\left(mean = 0.49, SE = \frac{0.27}{\sqrt{30} = 0.049}\right)$$

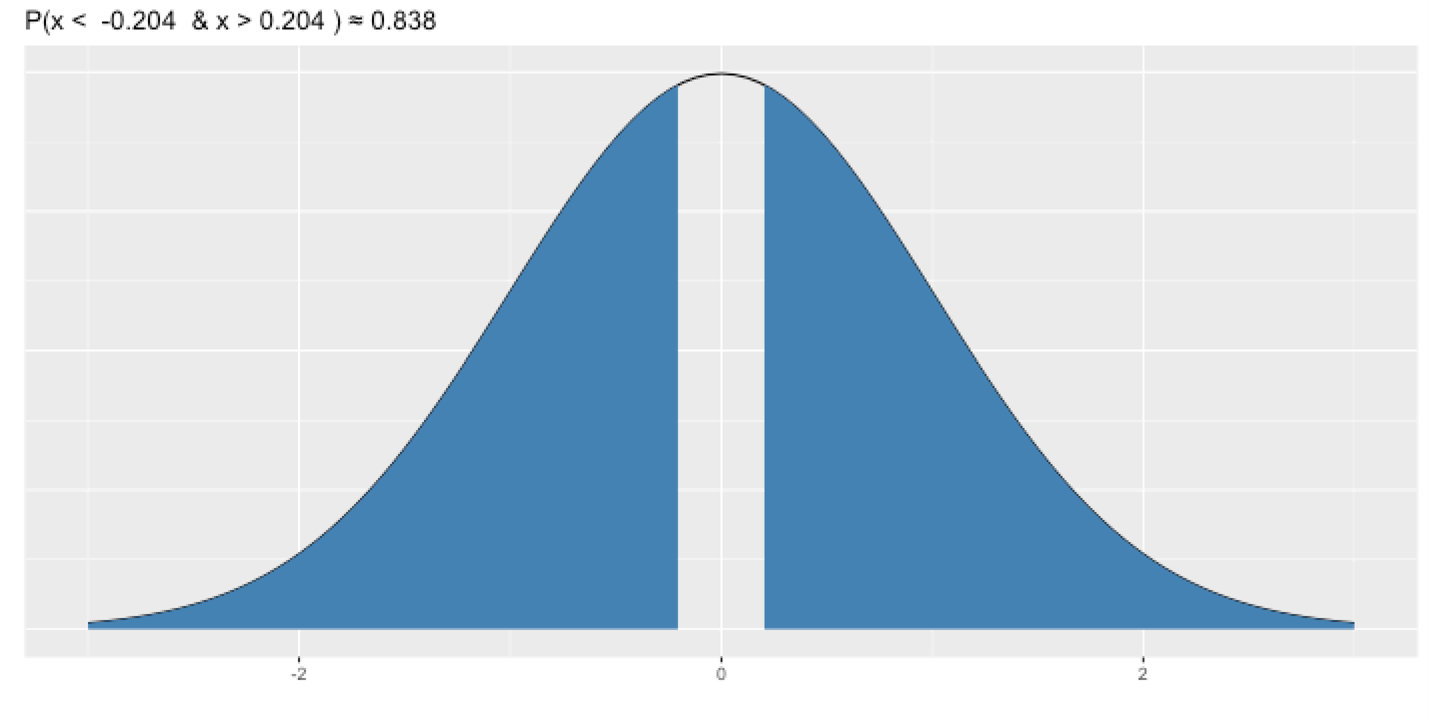
$$Z = \frac{\bar{x} - null}{SE} = \frac{0.49 - 0.50}{0.049} = -.204081633$$

```r
pnorm(-.204) * 2
```

```
## [1] 0.8383535
```

Hypothesis Testing (using *p*-values)

```
DATA606::normal_plot(cv = c(.204), tails = 'two.sided')
```



Type I and II Errors

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

	fail to reject H_0	**reject H_0**
H_0 true	✔	Type I Error
H_A true	Type II Error	✔

- Type I Error: **Rejecting** the null hypothesis when it is **true**.
- Type II Error: **Failing to reject** the null hypothesis when it is **false**.

Hypothesis Test

If we again think of a hypothesis test as a criminal trial then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

H_0 : Defendant is innocent
H_A : Defendant is guilty

Which type of error is being committed in the following circumstances?

- Declaring the defendant innocent when they are actually guilty

Type 2 error

- Declaring the defendant guilty when they are actually innocent

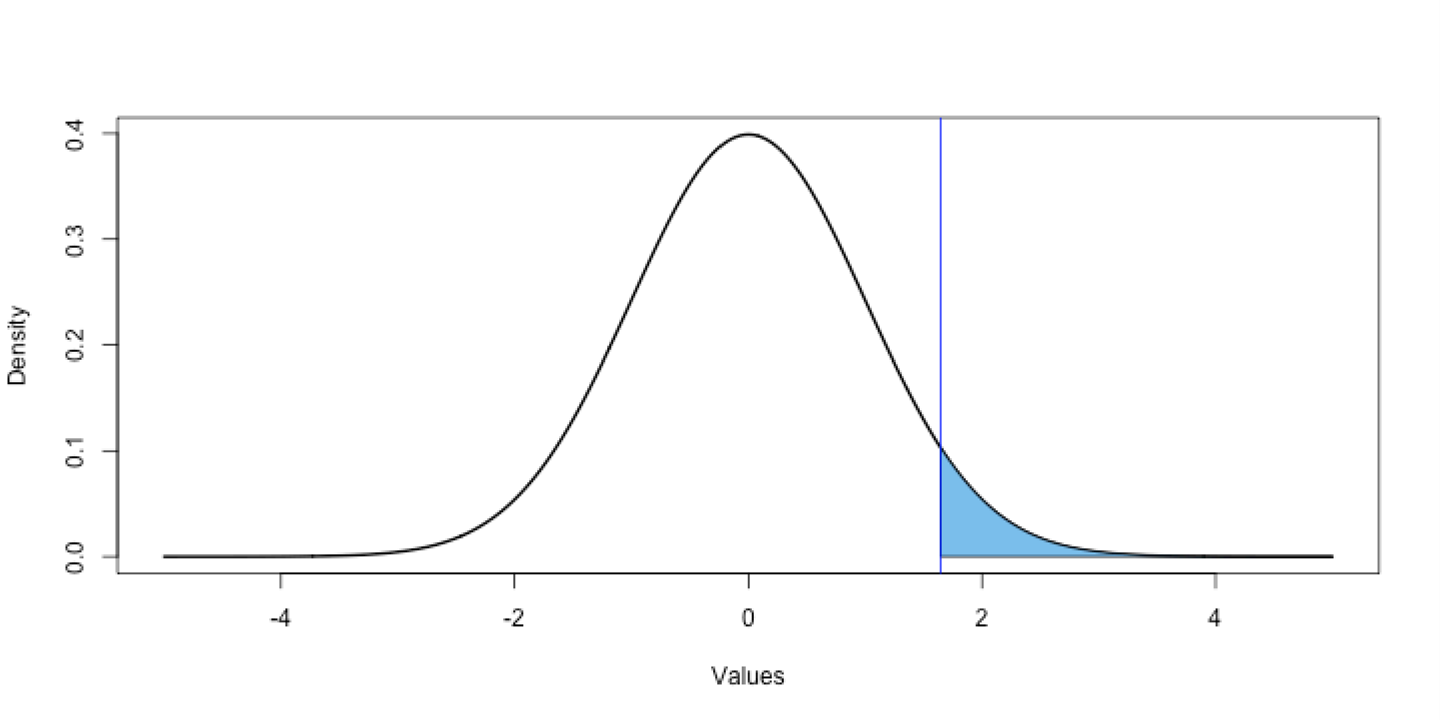
Type 1 error

Which error do you think is the worse error to make?

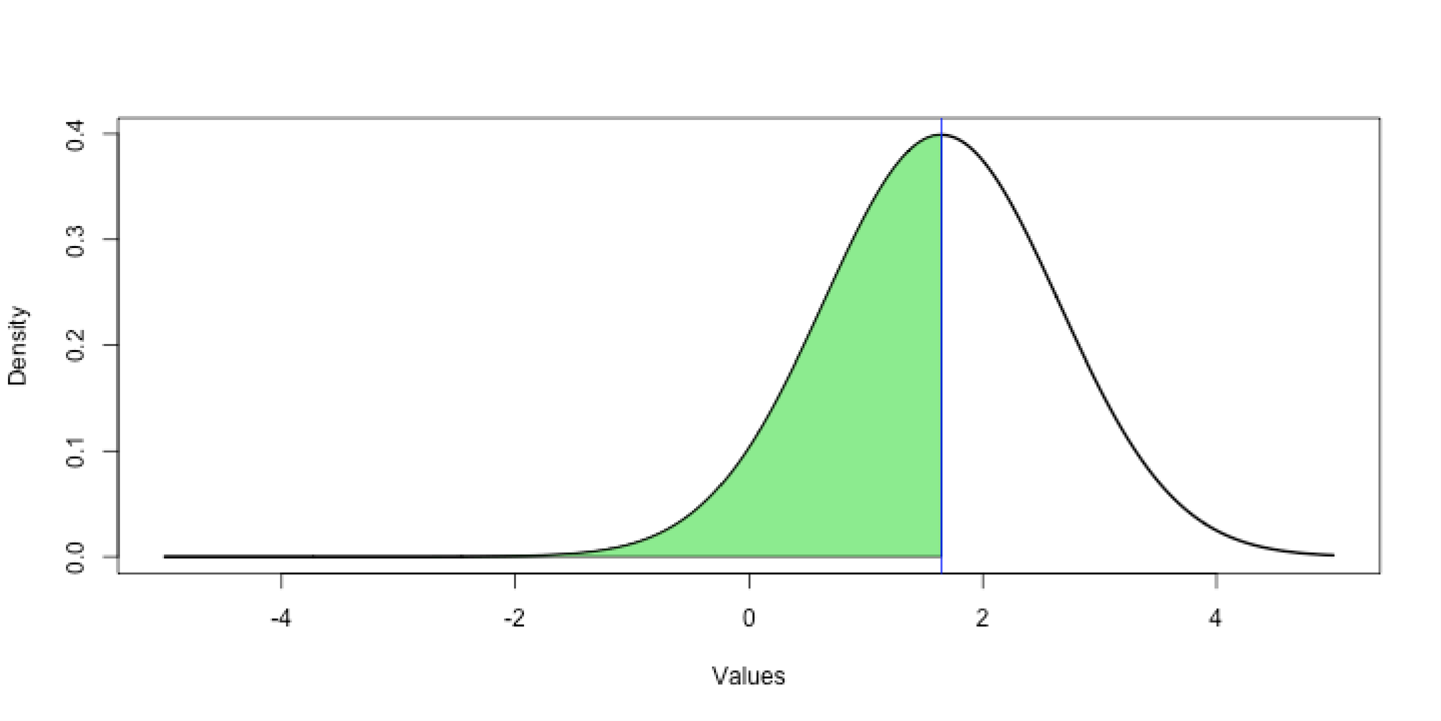
Null Distribution

```
(cv <- qnorm(0.05, mean=0, sd=1, lower.tail=FALSE))
```

```
## [1] 1.644854
```



Alternative Distribution



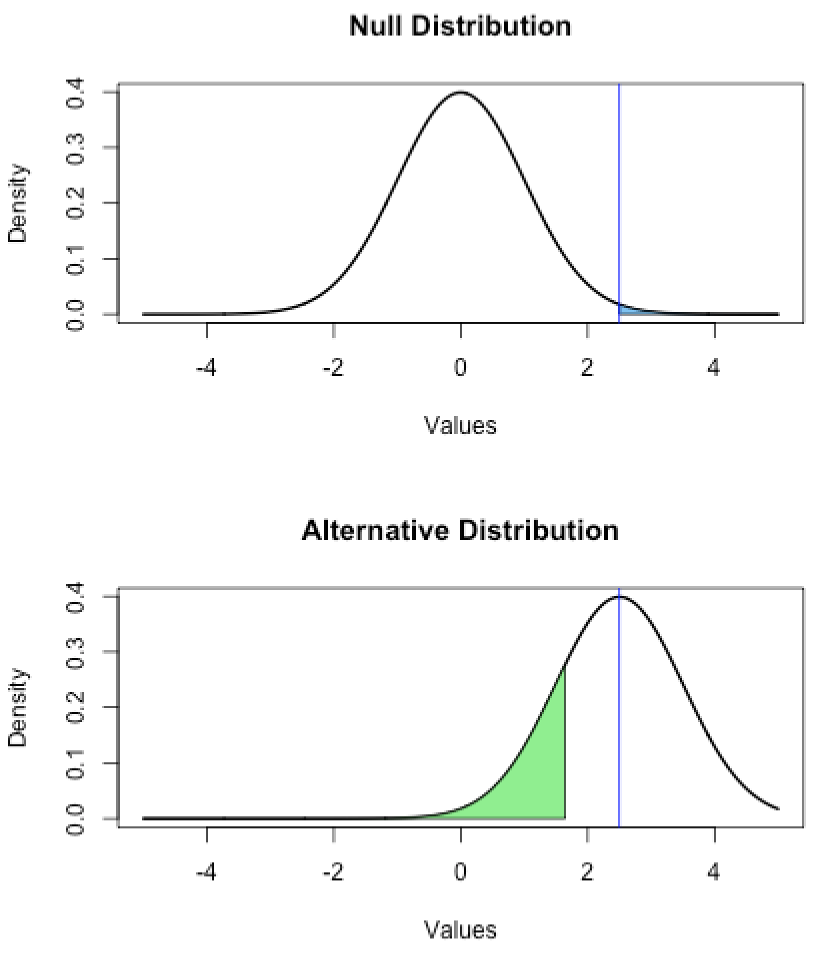
```
pnorm(cv, mean=cv, lower.tail = FALSE)
```

```
## [1] 0.5
```

Another Example (mu = 2.5)

```r
mu <- 2.5
(cv <- qnorm(0.05,
             mean=0,
             sd=1,
             lower.tail=FALSE))
```

```
## [1] 1.644854
```



Numeric Values

Type I Error

```r
pnorm(mu, mean=0, sd=1, lower.tail=FALSE)
```

```
## [1] 0.006209665
```

Type II Error

```r
pnorm(cv, mean=mu, lower.tail = TRUE)
```

```
## [1] 0.1962351
```

Shiny Application

Visualizing Type I and Type II errors: https://bcdudek.net/betaprob/

Why $p < 0.05$?

Check out this page: https://r.bryer.org/shiny/Why05/

See also:

Kelly M. *Emily Dickinson and monkeys on the stair Or: What is the significance of the 5% significance level?* Significance 10:5. 2013.

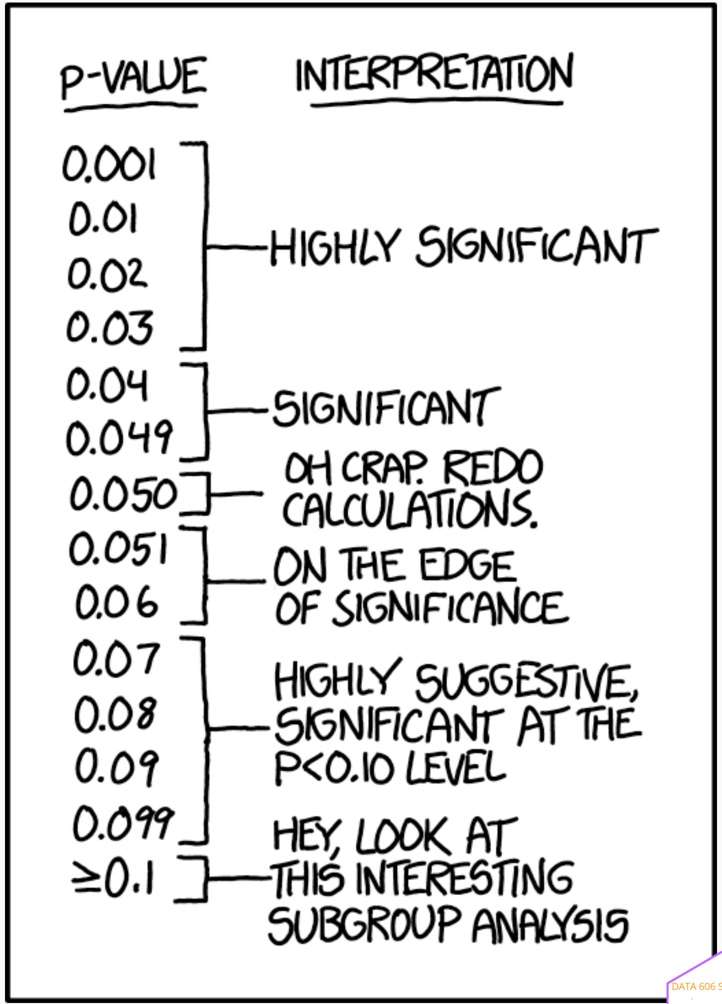
Statistical vs. Practical Significance

- Real differences between the point estimate and null value are easier to detect with larger samples.

- However, very large samples will result in statistical significance even for tiny differences between the sample mean and the null value (effect size), even when the difference is not practically significant.

- This is especially important to research: if we conduct a study, we want to focus on finding meaningful results (we want observed differences to be real, but also large enough to matter).

- The role of a statistician is not just in the analysis of data, but also in planning and design of a study.



P-VALUE	INTERPRETATION
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP. REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE P<0.10 LEVEL
0.08	
0.09	
0.099	
≥0.1	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS

Bootstrapping

Bootstrapping

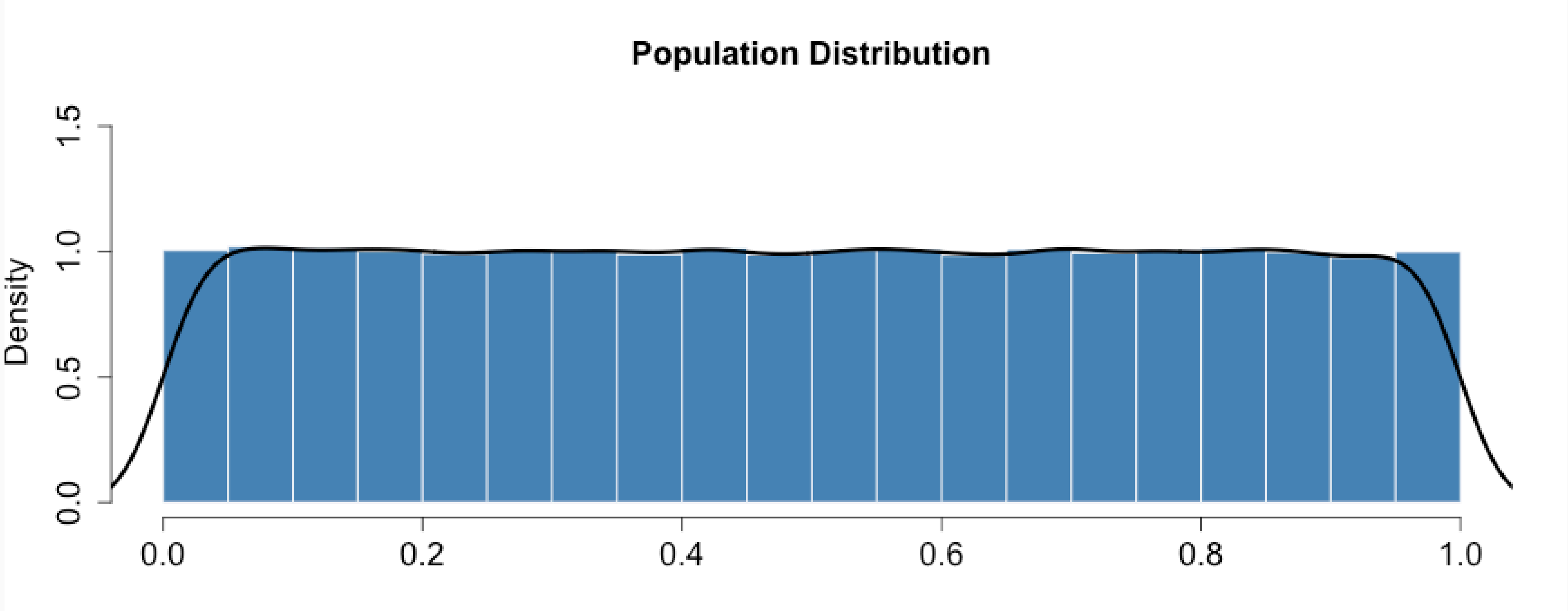
- First introduced by Efron (1979) in *Bootstrap Methods: Another Look at the Jackknife*.
- Estimates confidence of statistics by resampling *with* replacement.
- The *bootstrap sample* provides an estimate of the sampling distribution.
- The `boot` R package provides a framework for doing bootstrapping: https://www.statmethods.net/advstats/bootstrapping.html

Bootstrapping Example (Population)

Define our population with a uniform distribution.

```r
n <- 1e5
pop <- runif(n, 0, 1)
mean(pop)
```

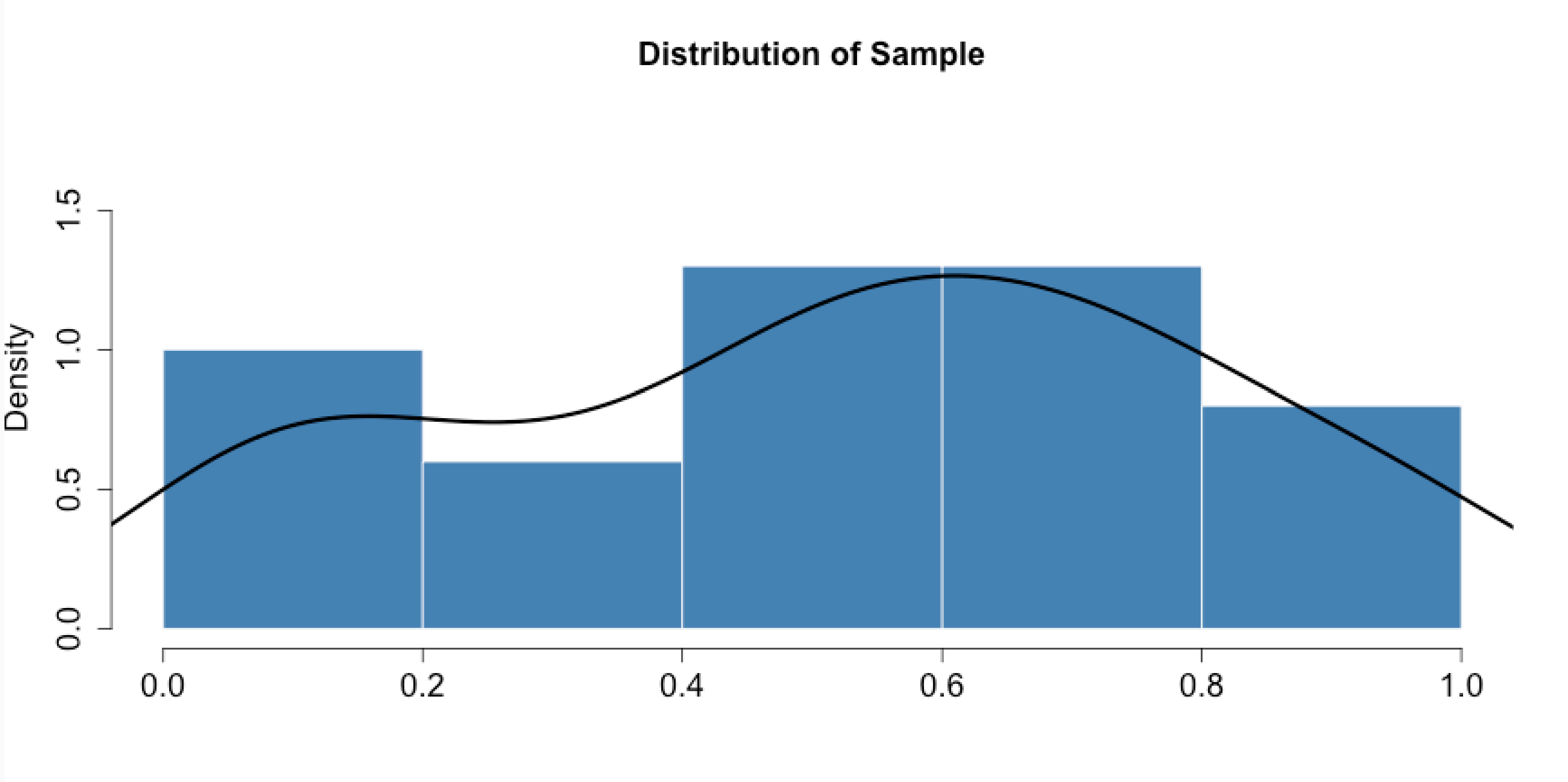
```
## [1] 0.4989141
```



Bootstrapping Example (Sample)

We observe one random sample from the population.

```
samp1 <- sample(pop, size = 50)
```



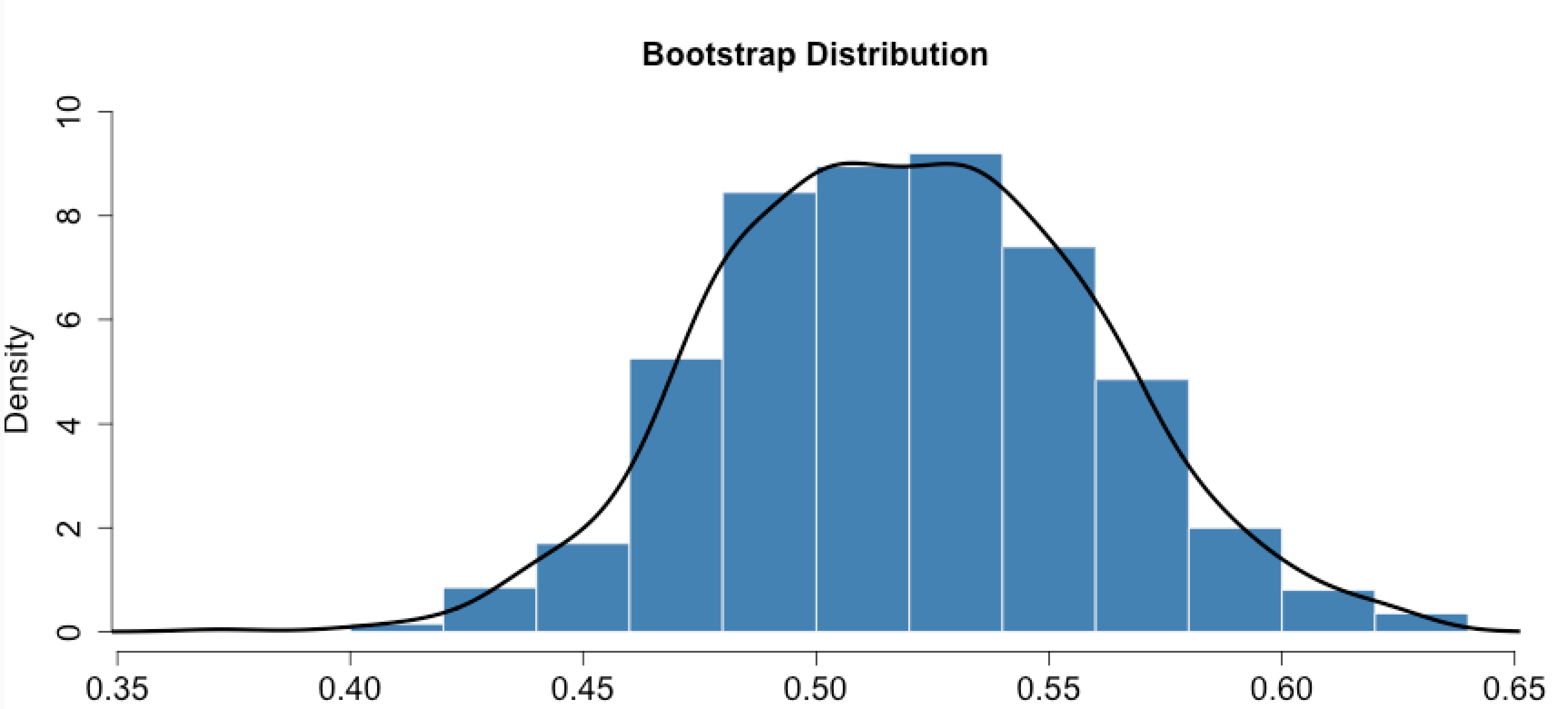
Bootsrapping Example (Estimate)

```r
boot.samples <- numeric(1000) # 1,000 bootstrap samples
for(i in seq_along(boot.samples)) {
    tmp <- sample(samp1, size = length(samp1), replace = TRUE)
    boot.samples[i] <- mean(tmp)
}
head(boot.samples)
```

```
## [1] 0.4837383 0.4680828 0.5178351 0.5398601 0.4843534 0.5365006
```

Bootsrapping Example (Distribution)

```
d <- density(boot.samples)
h <- hist(boot.samples, plot=FALSE)
hist(boot.samples, main='Bootstrap Distribution', xlab="", freq=FALSE,
     ylim=c(0, max(d$y, h$density)+.5), col='steelblue', border = "white",
     cex.main = 1.5, cex.axis = 1.5, cex.lab = 1.5)
lines(d, lwd=3)
```



95% confidence interval

```
c(mean(boot.samples) - 1.96 * sd(boot.samples),
  mean(boot.samples) + 1.96 * sd(boot.samples))
```

```
## [1] 0.4421955 0.5972504
```

Bootstrapping is not just for means!

```
boot.samples.median <- numeric(1000) # 1,000 bootstrap samples
for(i in seq_along(boot.samples.median)) {
    tmp <- sample(samp1, size = length(samp1), replace = TRUE)
    boot.samples.median[i] <- median(tmp) # NOTICE WE ARE NOW USING THE median FUNCTION!
}
head(boot.samples.median)
```

```
## [1] 0.6509364 0.5525239 0.4820703 0.5737473 0.5525239 0.5387680
```

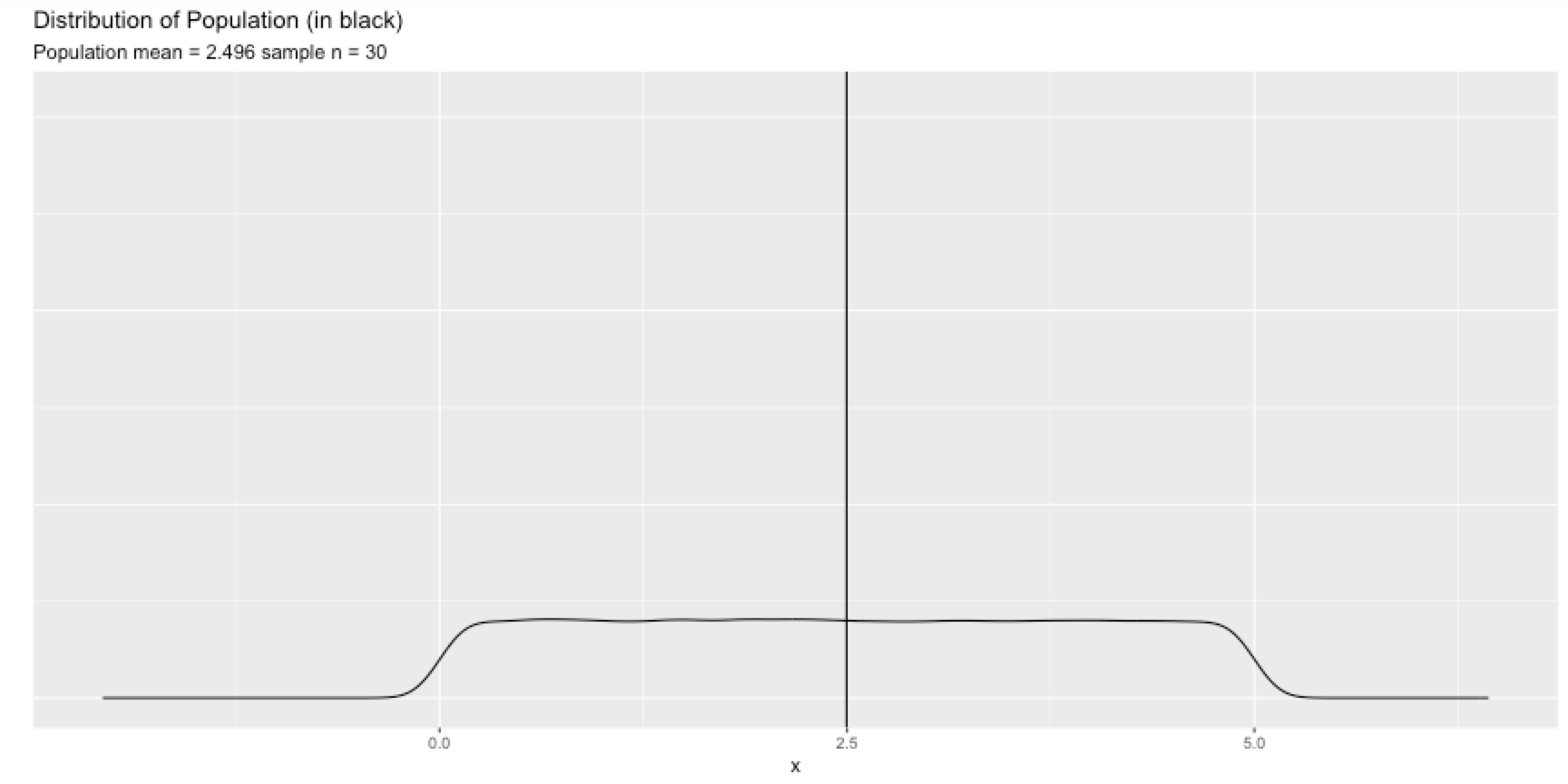
95% confidence interval for the median

```
c(mean(boot.samples.median) - 1.96 * sd(boot.samples.median),
  mean(boot.samples.median) + 1.96 * sd(boot.samples.median))
```

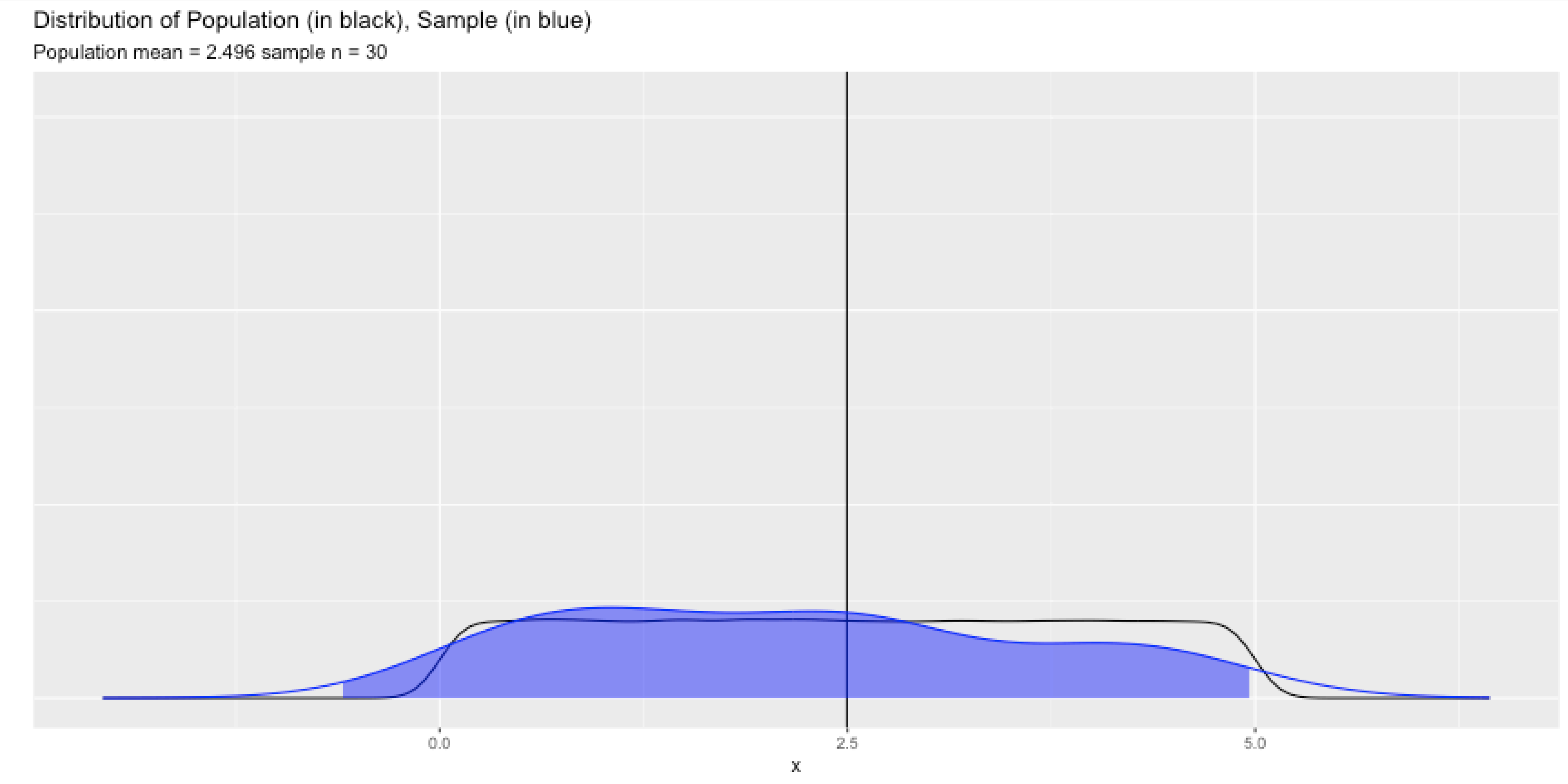
```
## [1] 0.4536408 0.6484383
```

Review

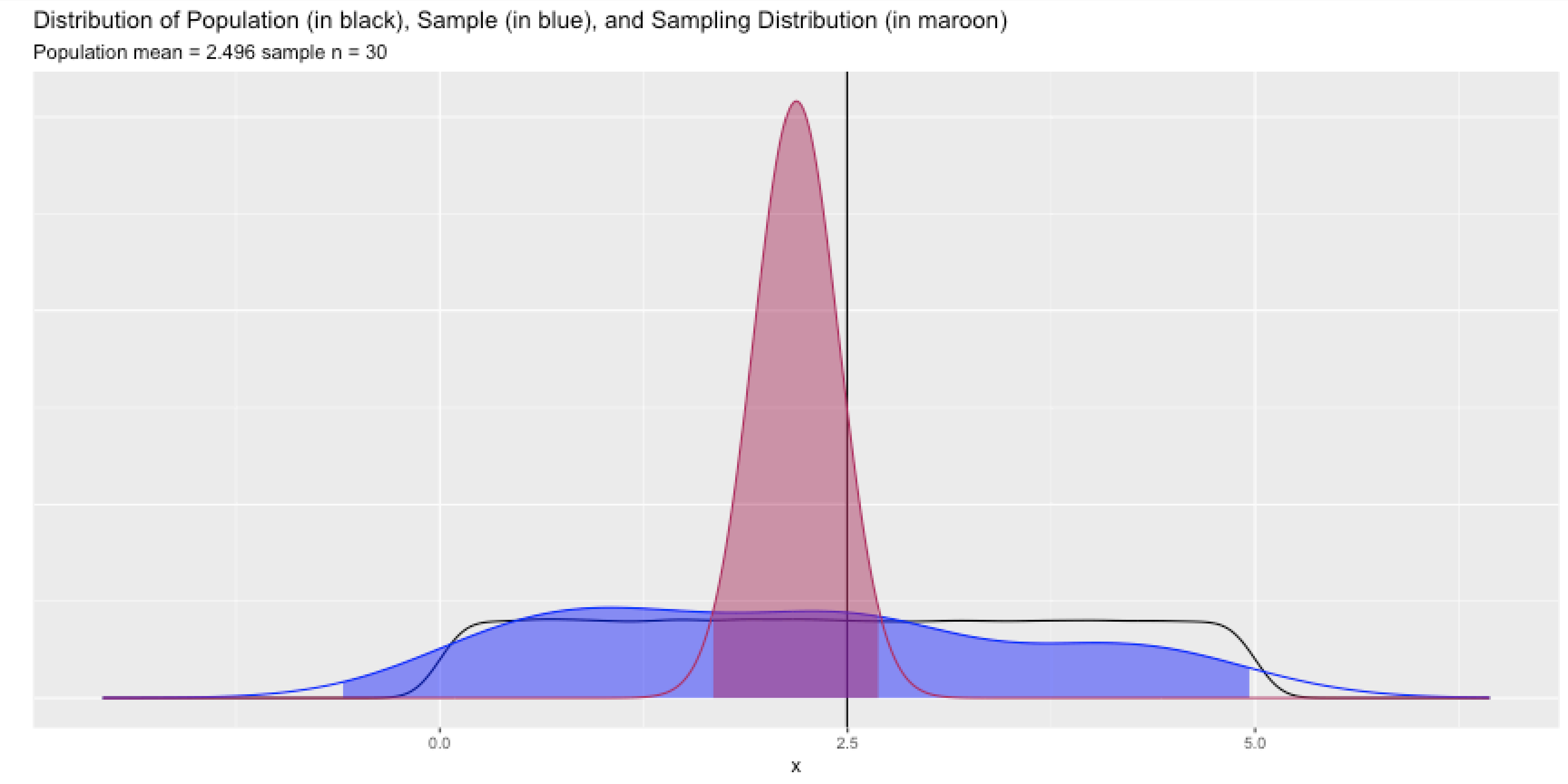
Review: Sampling Distribution



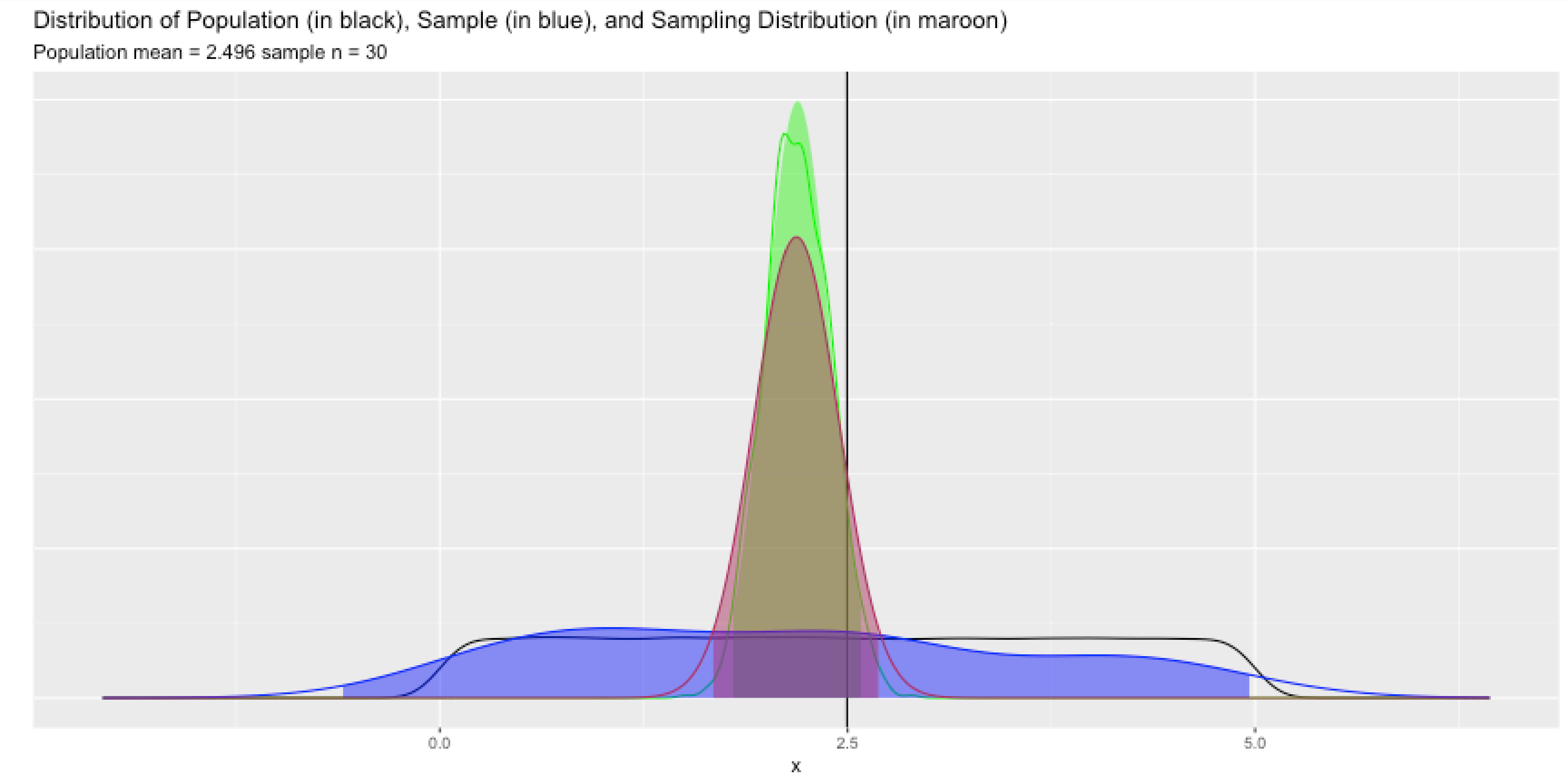
Review: Sampling Distribution



Review: Sampling Distribution



Review: Add Bootstrap Distribution



One Minute Paper

1. What was the most important thing you learned during this class?
2. What important question remains unanswered for you?

https://forms.gle/bz8GvYWfKdMggRv38