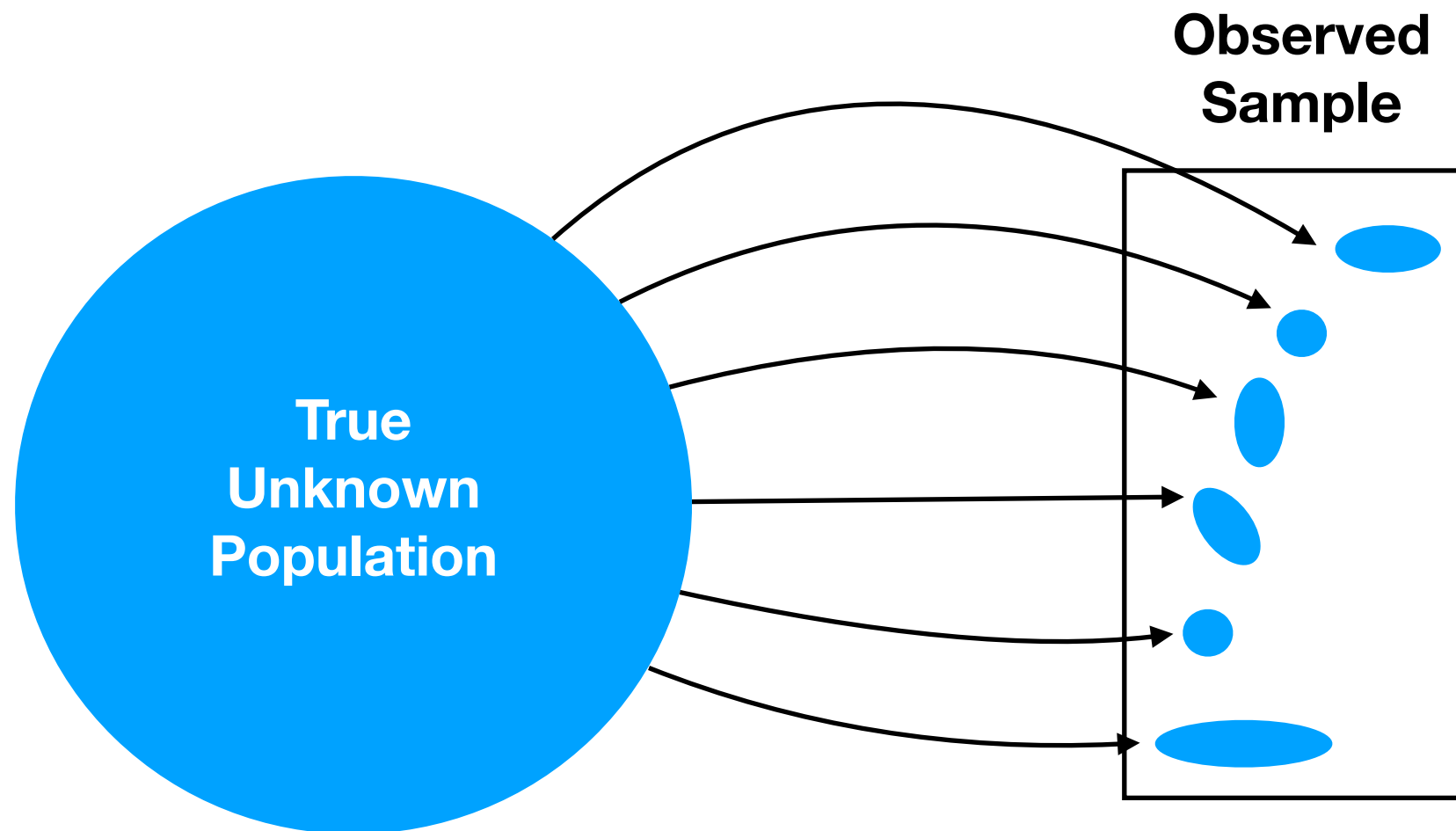


Assessing Confidence

Using Bootstrapping

Samples v Populations



Trying to learn about this but have observed this.

Bootstrapping

Let's say we have a sample of values...

```
[2.148, 3.141, 2.668, 3.647, 6.799, 5.728,  
 3.346, 3.318, 2.305, 2.051]
```

...with mean...

3.5151

How confident are we that this value is a good representation of the mean value for the entire population from which we've sampled?

Bootstrapping

If we know the distribution from which the data were sampled, we sometimes have a formula available to tell us about the 95% confidence interval on our estimate.

[ASIDE: What is a 95% confidence interval?]

But what if we don't know what kind of distribution our data were drawn from?

Bootstrapping

In that situation, we need to “pull ourselves up by our bootstraps”, without a formula.

Brad Efron, a Stanford statistician, gave us a procedure to do just that.

If we generate a series of pseudo-replicate datasets, by **sampling with replacement** from the original dataset, we can get an estimate of our confidence in the empirical estimate.

Bootstrapping

[2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]

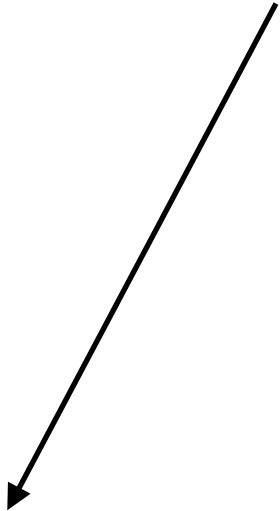


[2.148,

Example Pseudoreplicate

Bootstrapping

[2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]

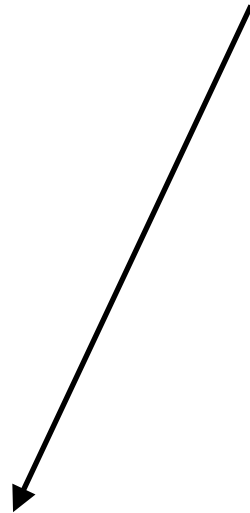
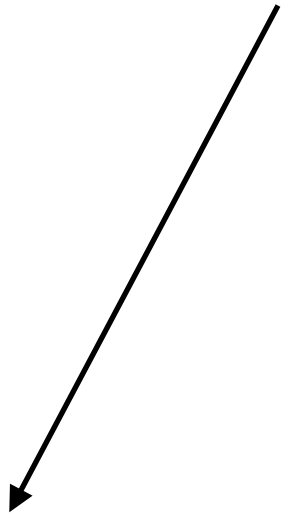


[2.148, 2.668,

Example Pseudoreplicate

Bootstrapping

[2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]

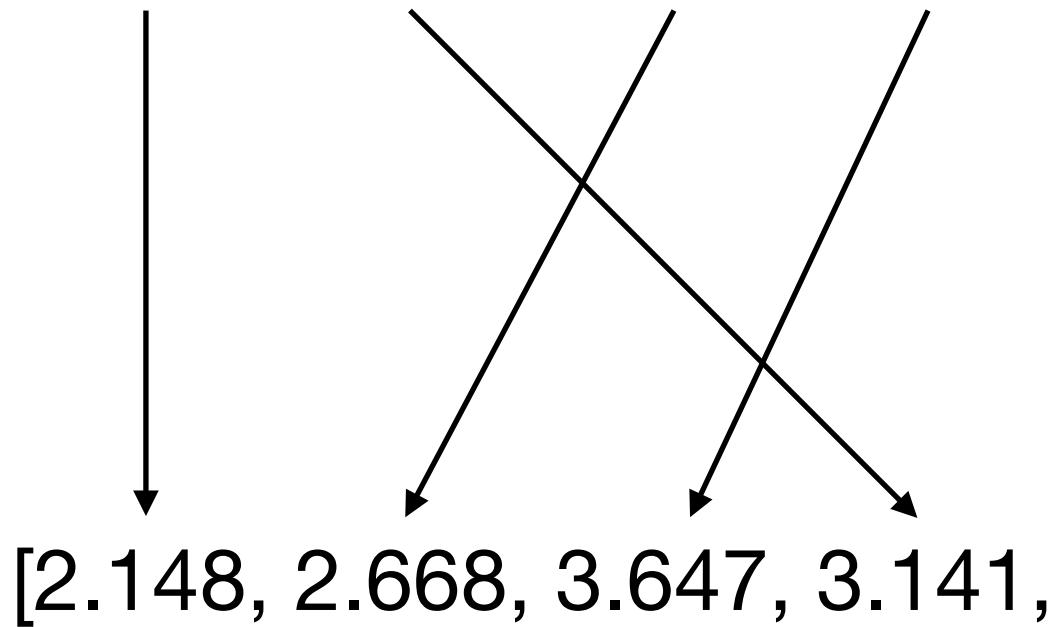


[2.148, 2.668, 3.647,

Example Pseudoreplicate

Bootstrapping

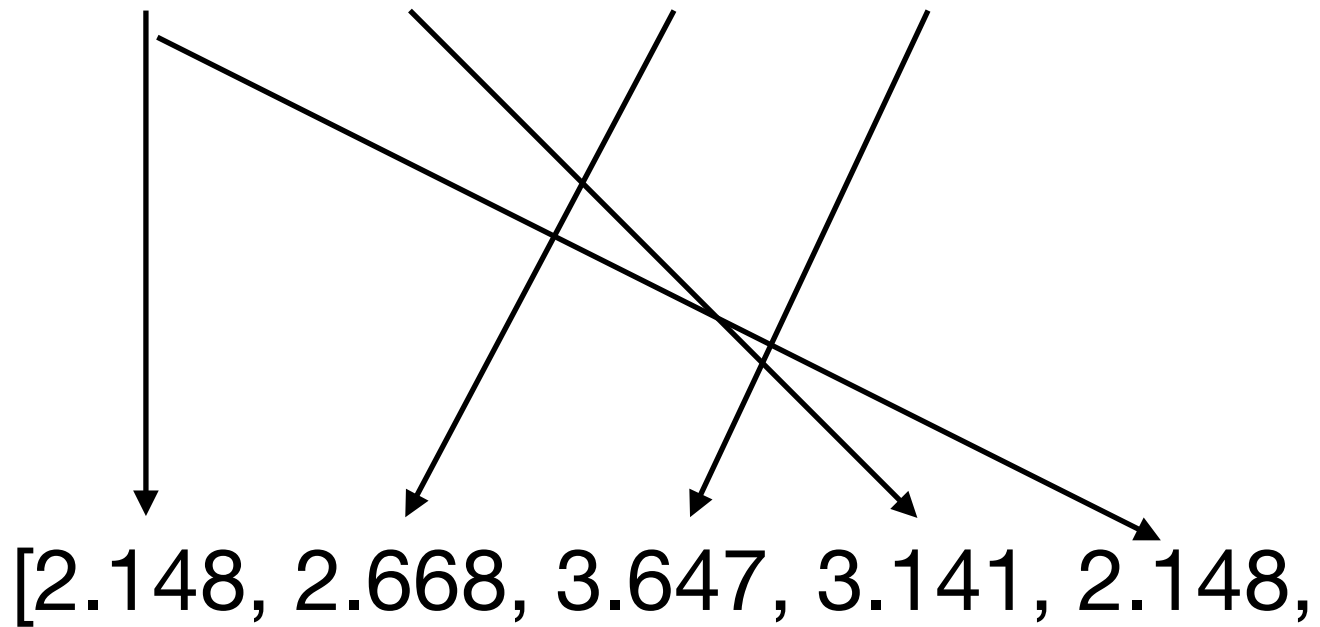
[2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]



Example Pseudoreplicate

Bootstrapping

[2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]



Example Pseudoreplicate

Bootstrapping

[2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]

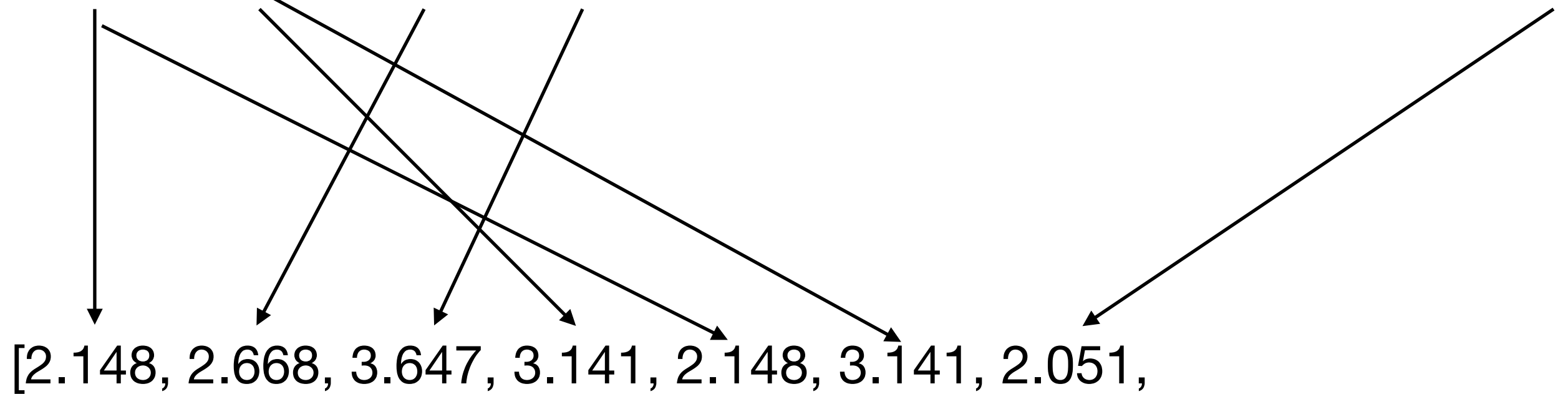
The diagram illustrates the selection of a pseudoreplicate from a bootstrap sample. It shows two rows of numbers. The top row is the original data: [2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]. The bottom row is a bootstrap sample: [2.148, 2.668, 3.647, 3.141, 2.148, 3.141, ...]. Arrows indicate the selection process: a vertical arrow points from the first element (2.148) of the original data to the first element (2.148) of the bootstrap sample. Four diagonal arrows point from the second, third, fourth, and fifth elements of the original data (3.141, 2.668, 3.647, 3.141) to the second, third, fourth, and fifth elements of the bootstrap sample (2.668, 3.647, 3.141, 2.148), respectively. The sixth and seventh elements of the bootstrap sample (2.148, 3.141) are not connected by arrows, indicating they were not selected in this particular pseudoreplicate.

[2.148, 2.668, 3.647, 3.141, 2.148, 3.141,

Example Pseudoreplicate

Bootstrapping

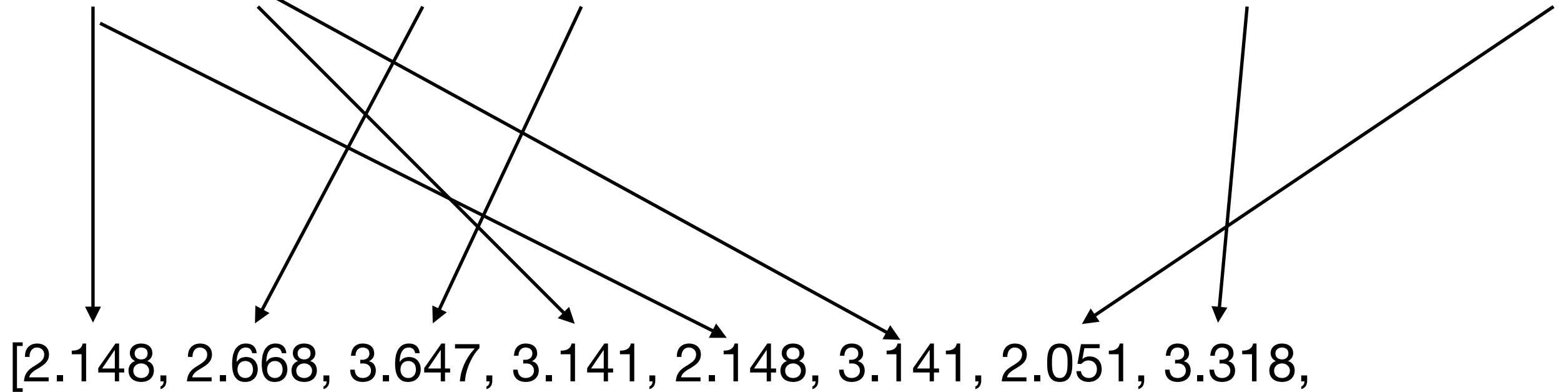
[2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]



Example Pseudoreplicate

Bootstrapping

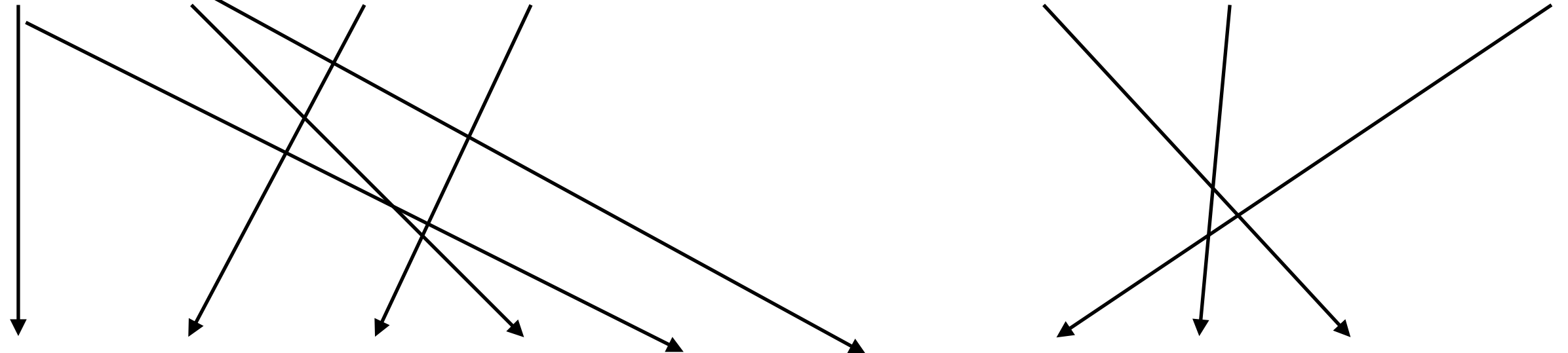
[2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]



Example Pseudoreplicate

Bootstrapping

[2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]

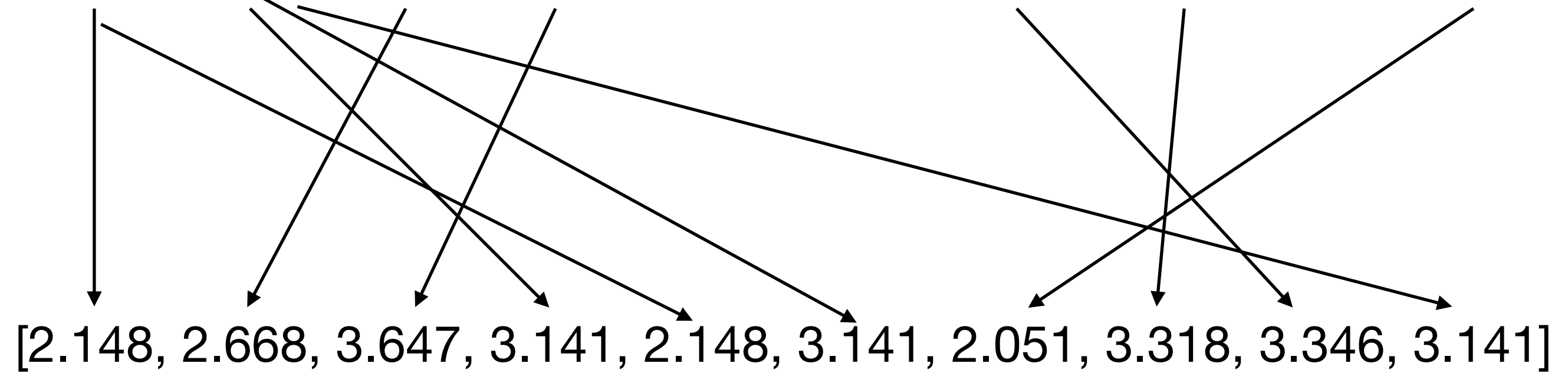


[2.148, 2.668, 3.647, 3.141, 2.148, 3.141, 2.051, 3.318, 3.346,

Example Pseudoreplicate

Bootstrapping

[2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]



Example Pseudoreplicate

Bootstrapping

[2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]

[3.318, 2.668, 2.668, 2.668, 2.305, 3.346, 2.148, 2.051, 2.668, 6.799]

[2.148, 2.148, 2.148, 3.141, 2.148, 3.318, 2.051, 5.728, 3.647, 3.647]

[6.799, 3.318, 6.799, 3.141, 6.799, 5.728, 3.346, 3.141, 2.668, 5.728]

[2.051, 6.799, 2.305, 5.728, 3.318, 3.647, 6.799, 3.141, 2.668, 3.346]

[2.668, 3.346, 3.346, 2.305, 2.148, 3.346, 3.647, 3.318, 3.346, 2.305]

[2.305, 3.318, 2.148, 3.647, 2.148, 2.668, 6.799, 6.799, 3.141, 3.141]

[3.318, 2.668, 3.318, 3.141, 5.728, 3.318, 2.305, 2.668, 5.728, 3.141]

[6.799, 5.728, 6.799, 2.051, 2.668, 2.051, 2.305, 5.728, 3.647, 3.141]

[2.051, 2.668, 5.728, 2.148, 3.318, 2.148, 2.148, 2.148, 2.305, 3.141]

[3.141, 2.305, 2.668, 3.141, 3.346, 2.148, 3.318, 2.051, 3.141, 5.728]

Bootstrapping

[2.148, 3.141, 2.668, 3.647, 6.799, 5.728, 3.346, 3.318, 2.305, 2.051]

Observed Mean: 3.5151

Bootstrap
Means

[3.318, 2.668, 2.668, 2.668, 2.305, 3.346, 2.148, 2.051, 2.668, 6.799]	3.0639
[2.148, 2.148, 2.148, 3.141, 2.148, 3.318, 2.051, 5.728, 3.647, 3.647]	3.0124
[6.799, 3.318, 6.799, 3.141, 6.799, 5.728, 3.346, 3.141, 2.668, 5.728]	4.7467
[2.051, 6.799, 2.305, 5.728, 3.318, 3.647, 6.799, 3.141, 2.668, 3.346]	3.9802
[2.668, 3.346, 3.346, 2.305, 2.148, 3.346, 3.647, 3.318, 3.346, 2.305]	2.9775
[2.305, 3.318, 2.148, 3.647, 2.148, 2.668, 6.799, 6.799, 3.141, 3.141]	3.6114
[3.318, 2.668, 3.318, 3.141, 5.728, 3.318, 2.305, 2.668, 5.728, 3.141]	3.5333
[6.799, 5.728, 6.799, 2.051, 2.668, 2.051, 2.305, 5.728, 3.647, 3.141]	4.0917
[2.051, 2.668, 5.728, 2.148, 3.318, 2.148, 2.148, 2.148, 2.305, 3.141]	2.7803
[3.141, 2.305, 2.668, 3.141, 3.346, 2.148, 3.318, 2.051, 3.141, 5.728]	3.0987

Bootstrapping

Observed Mean: 3.5151

Bootstrap
Means

3.0639

3.0124

4.7467

3.9802

2.9775

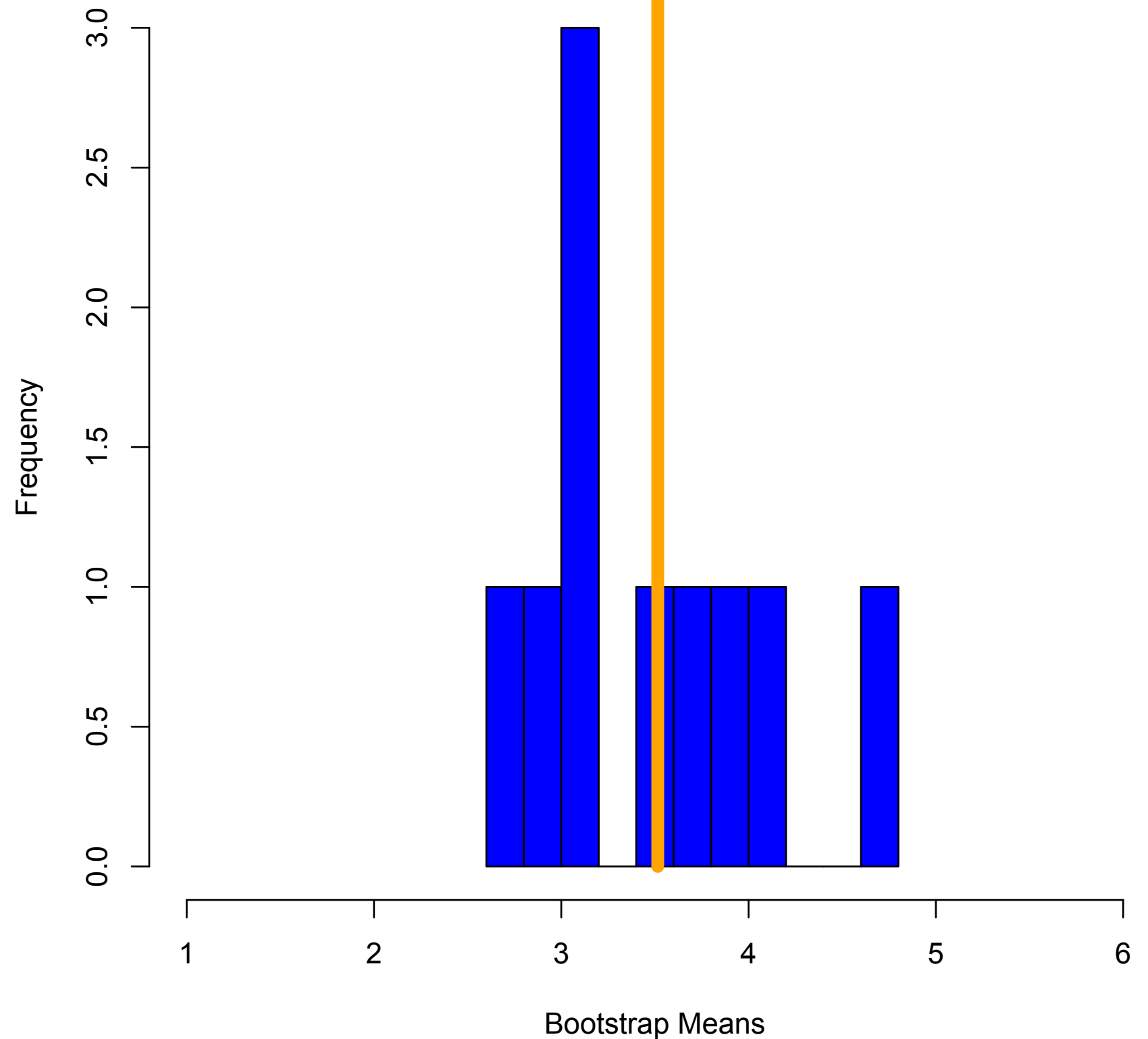
3.6114

3.5333

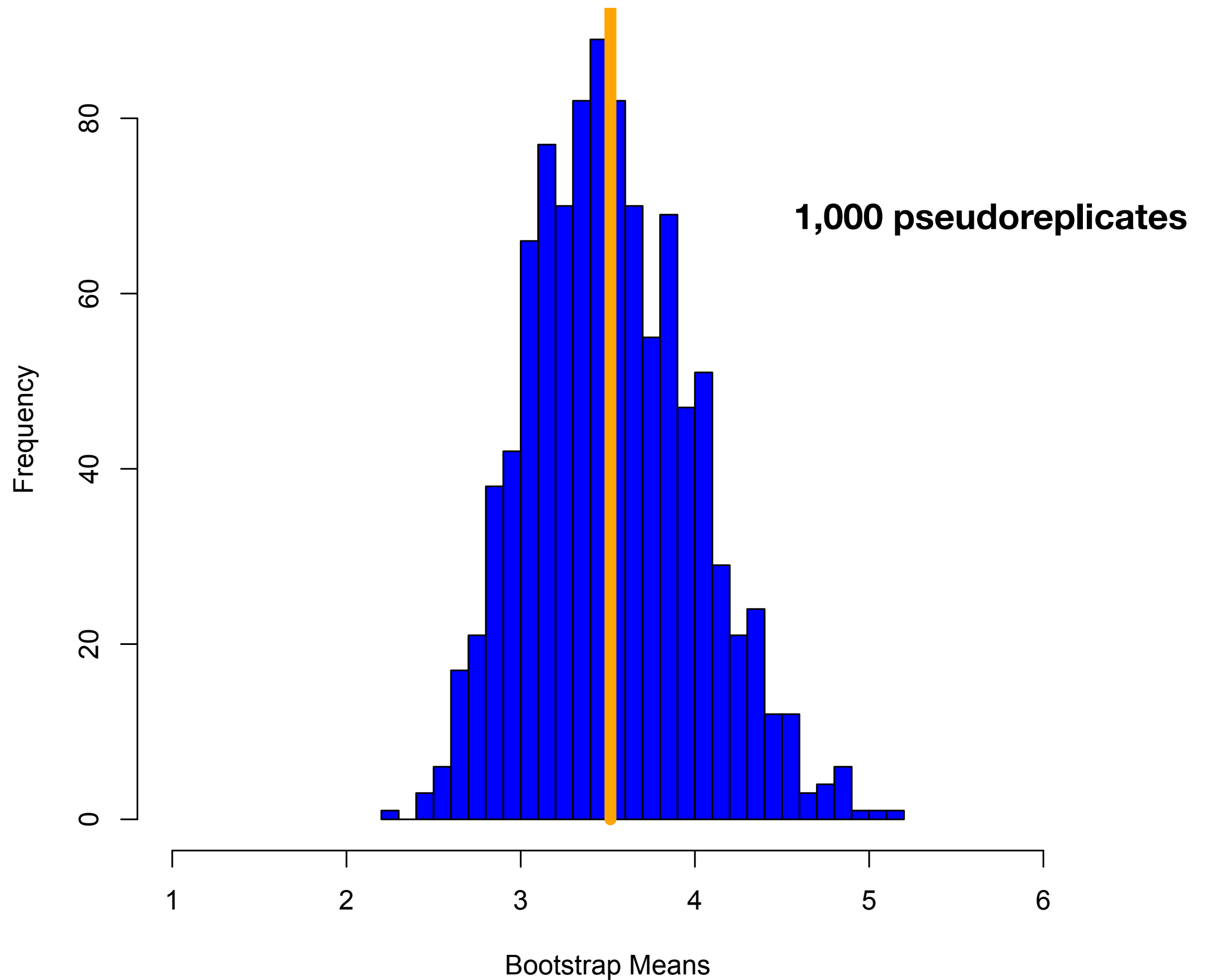
4.0917

2.7803

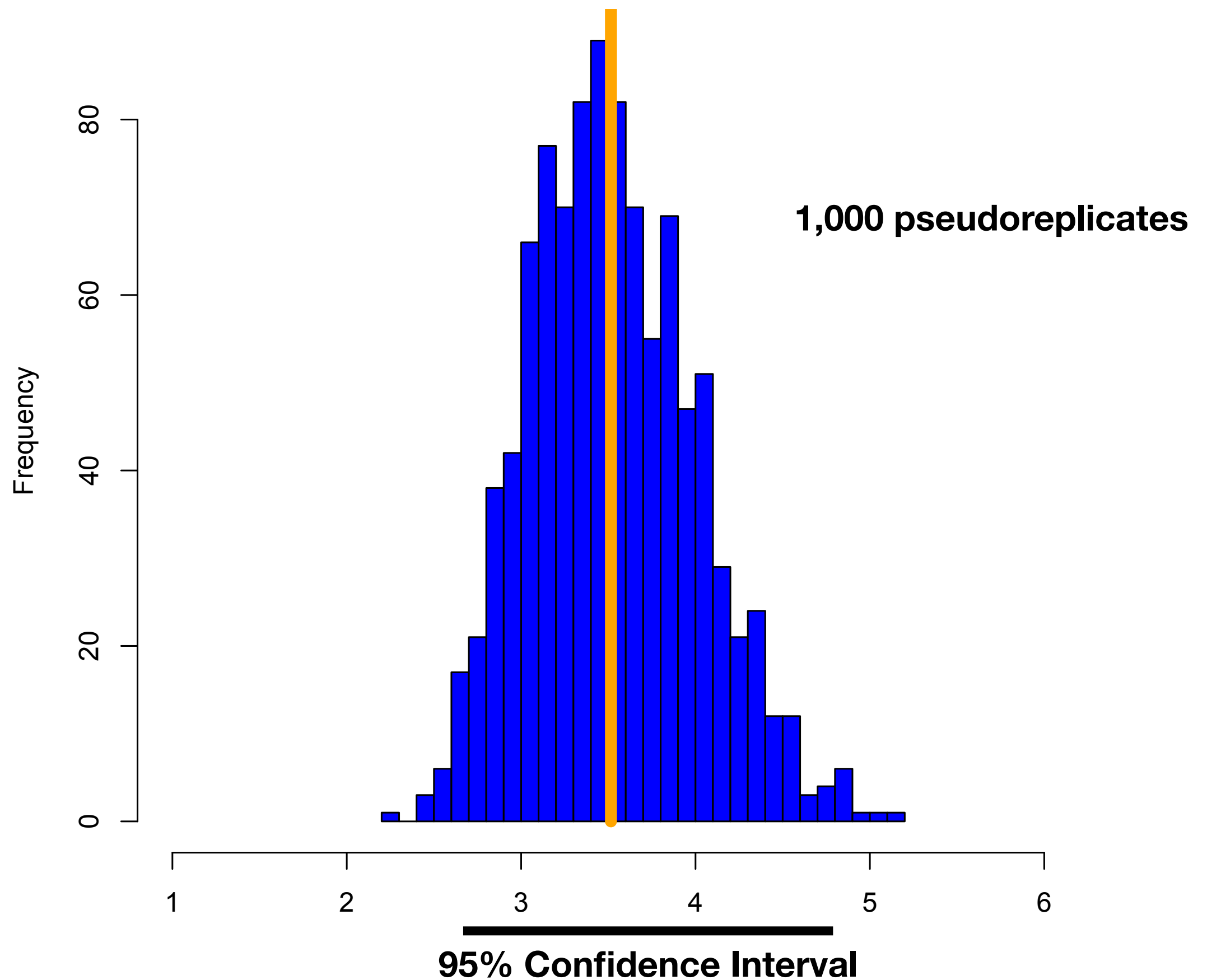
3.0987



Bootstrapping



Bootstrapping



Bootstrap Animations

<https://seeing-theory.brown.edu/frequentist-inference/index.html#section3>

<https://www.stat.auckland.ac.nz/~wild/BootAnim/>