

Bootstrap

Grinnell College

April 1, 2026

Review Questions

- ▶ Why can I not directly compare two t-statistics with different sample sizes?
- ▶ What are three equivalent ways to test a hypothesis?
- ▶ Why can I directly compare two p-values from different sample sizes?
- ▶ Draw a picture illustrating why $|t| < C$ implies $p > \alpha$
- ▶ TRUE/FALSE: Changing H_0 will also change my p-value

Repeated Samples

The confidence intervals we constructed of the form:

$$\text{Point Estimate} \pm \text{Margin of Error}$$

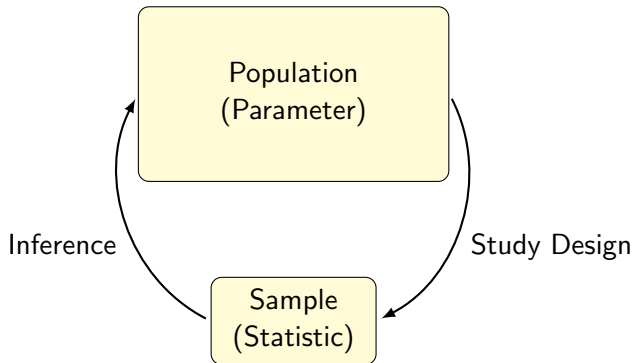
- ▶ Relied on assumptions about our *sampling process*
- ▶ Examined what might happen if we could repeat sampling ad infinitum

There are, naturally, some limitations:

- ▶ We are limited to collecting a single sample
- ▶ Our assumptions may be tenuous. What if our statistic doesn't not follow the CLT?

It would be helpful to have a more general method of constructing intervals with similar properties we had before

The Statistical Framework



Bootstrapping

Somewhat amazingly, we can get around this problem with a resampling technique known as **bootstrapping**

Bootstrapping refers to an algorithmic process whereby, for a sample of size n , we resample *with replacement* and compute a new statistic on the bootstrapped sample.

Instead of drawing more samples from our *original population*, we treat our sample as an *estimate* of the population and instead draw bootstrapped samples from our original sample

Bootstrapping

“Pick yourself up from bootstraps”



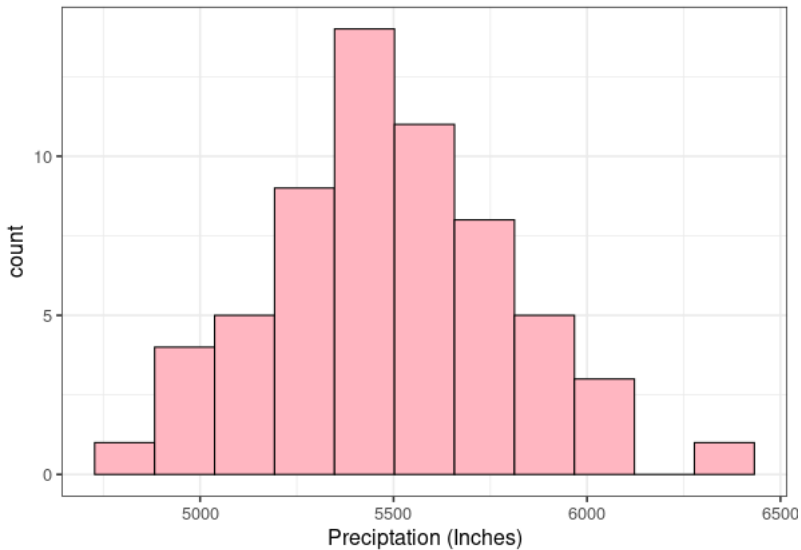
original sample

1 2 3 4 5 6 7 8 9 bootstrap sample 1

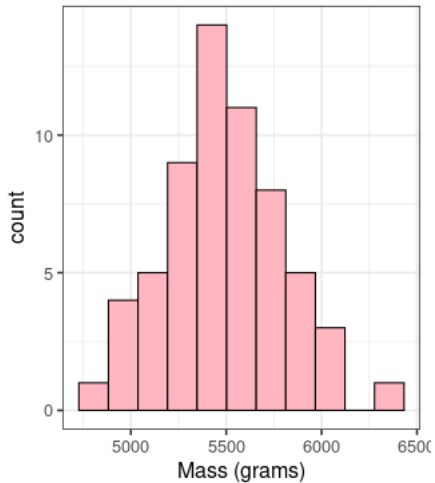
1 2 3 4 5 6 7 8 9 bootstrap sample 2

1 2 3 4 5 6 7 8 9 bootstrap sample 3

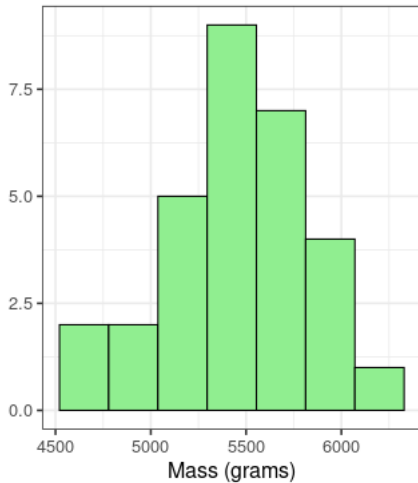
Body Mass Male Gentoo Penguins (N = 61)



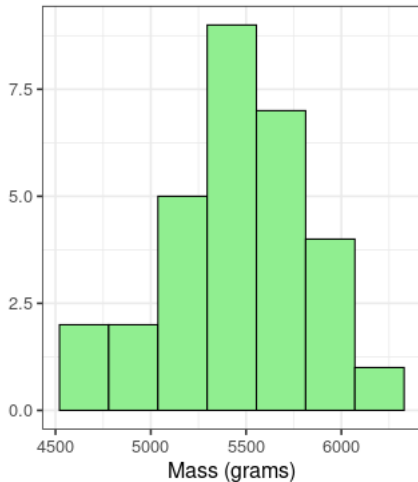
Male Gentoo Penguins (N = 61)



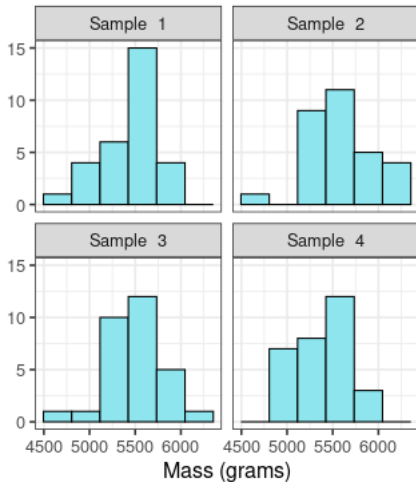
Penguin Sample (n = 30)



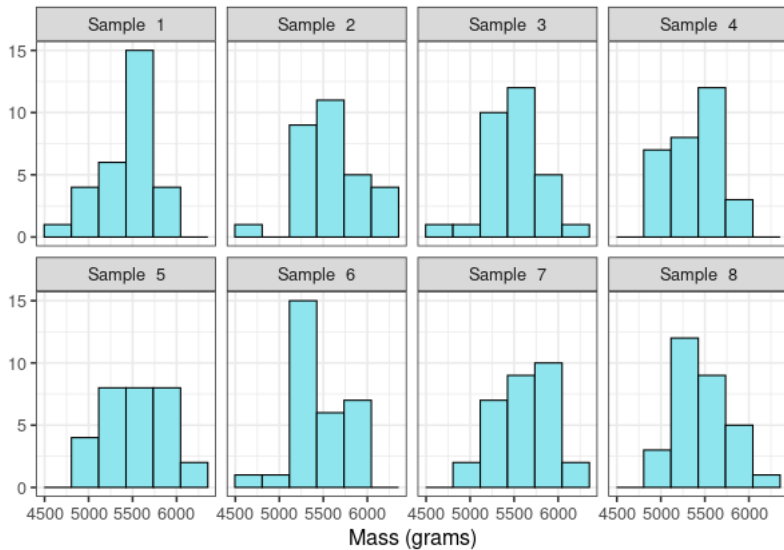
Penguin Sample (n = 30)



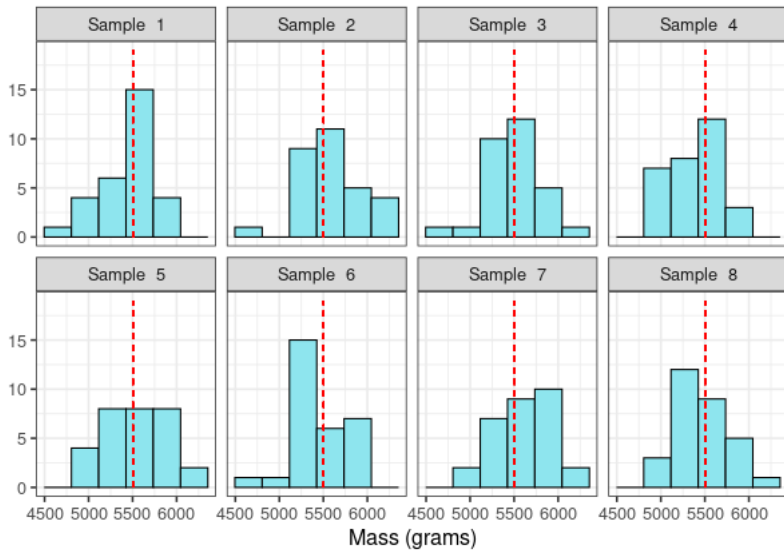
Bootstrapped Samples (n = 30)



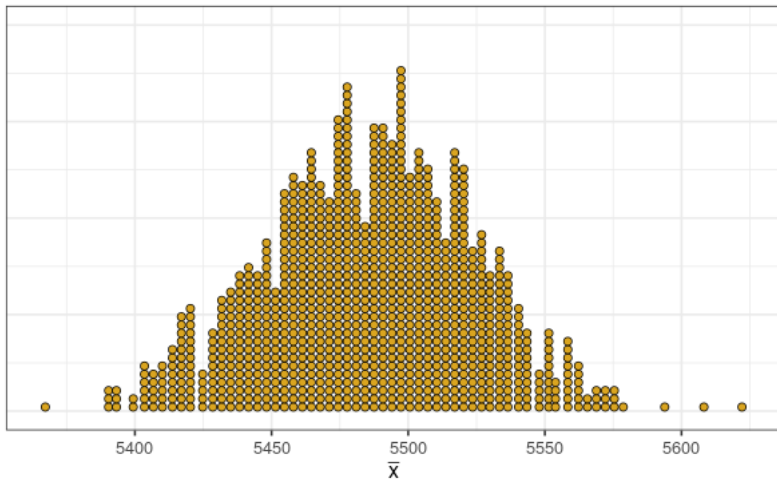
Bootstrapped Samples (n = 30)



Bootstrapped Samples (n = 30)



Means from 1000 Bootstraps



Bootstrapped Sampling Distribution

The collection of bootstrapped statistics gives us an estimate of the *sampling distribution*

- ▶ What values did we see?
- ▶ How frequently did they appear?

In this case, because we were bootstrapping the sample mean we find that the sampling distribution looks approximately normal

In general, we should expect that our bootstrapped sampling distribution to be theoretically identical to the *true* sampling distribution, whatever that may be

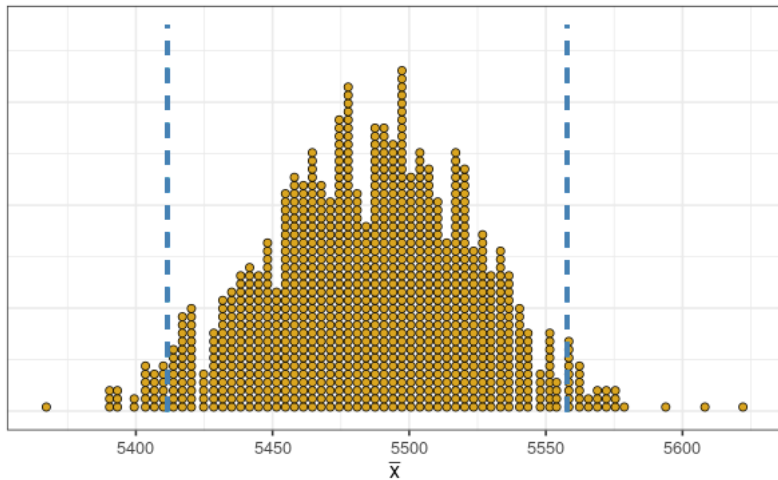
Quantiles

Recall that Monday we introduced the `qnorm()` and `qt()` functions to find us the quantiles of our sampling distribution when the distribution was known

If we don't know the distribution, we can use a vector of values, along with the `quantiles()` function to perform a similar task

```
1 > ## Find mean/se
2 > mean(mass_sample); se(mass_sample)
3 [1] 5484.8
4 [1] 40.096
5 >
6 > ## Find critical value from t
7 > quants95 <- c(0.025, 0.975)
8 > qt(quants95, df = 29)
9 [1] -2.0452  2.0452
10 >
11 > ## Point +/- MOE
12 > 5484.8 + c(-2.0452, 2.0452)*40.096
13 [1] 5402.8 5566.8
14 >
15 > ## Using quantile() on bootstrap
16 > quantile(boot_sample, probs = quants95)
17 2.5% 97.5%
18 5408.6 5565.2
```

Bootstrapped Sample Means

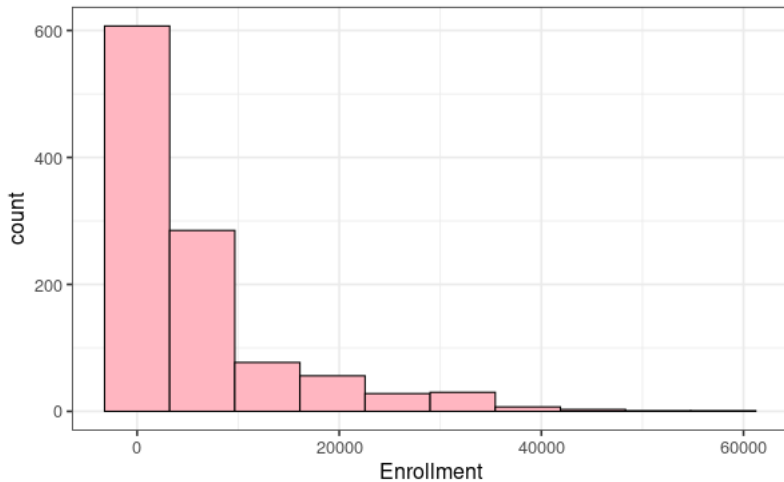


When the sampling distribution is approximately normal, the quantiles of the bootstrap should match closely with those computed using the margin of error method

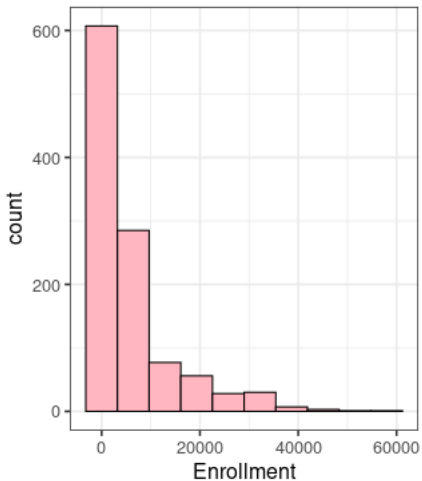
Bootstrapping is especially appropriate when our sampling distribution is not normal:

- ▶ The population variable is highly skewed
- ▶ The number of observations in our sample is not large enough for CLT approximation
- ▶ We want the sampling distribution of a statistic that is *not* normally distributed

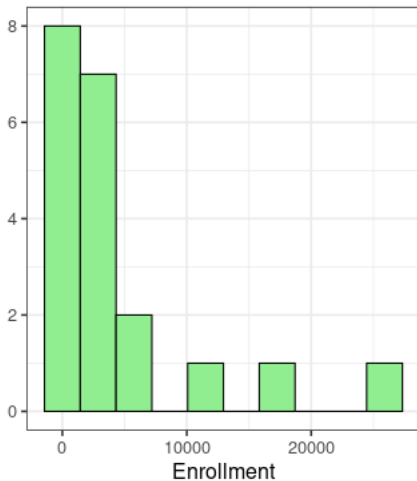
College Enrollment (N = 1095)



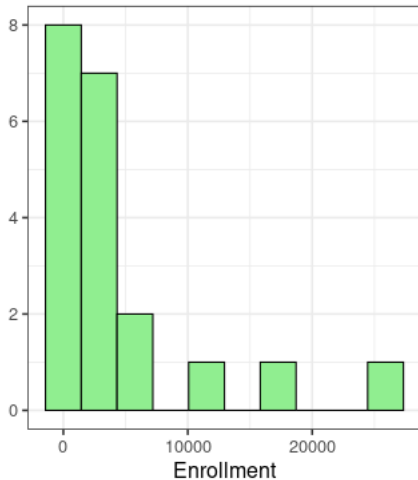
Enrollment (N = 1095)



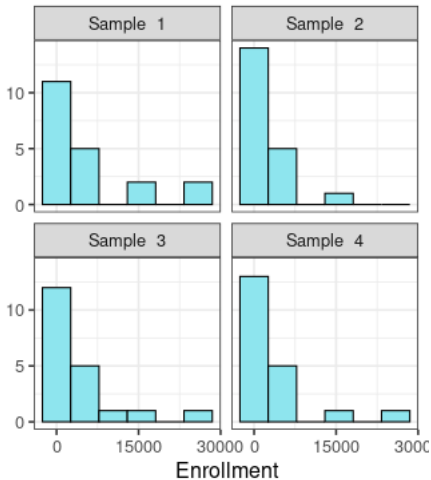
Enrollment Sample (n = 20)



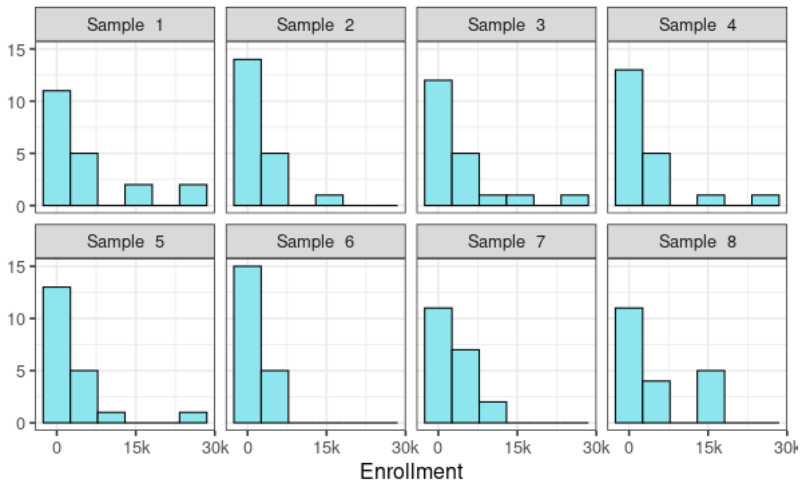
Enrollment Sample (n = 20)



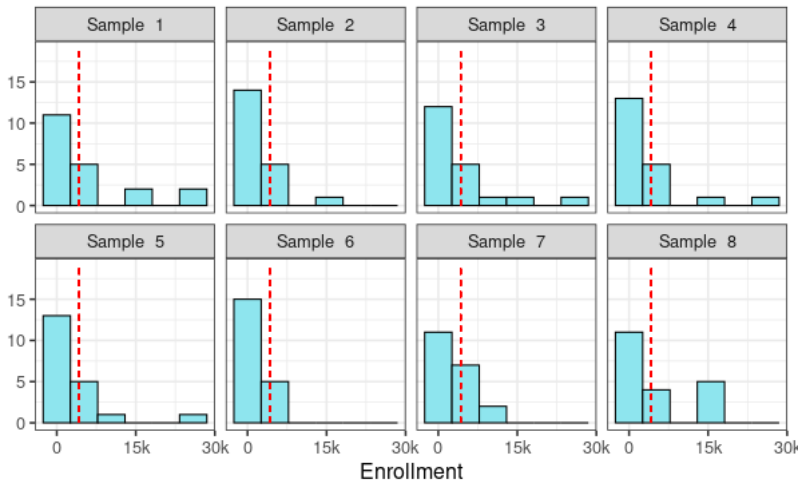
Bootstrapped Samples (n = 20)



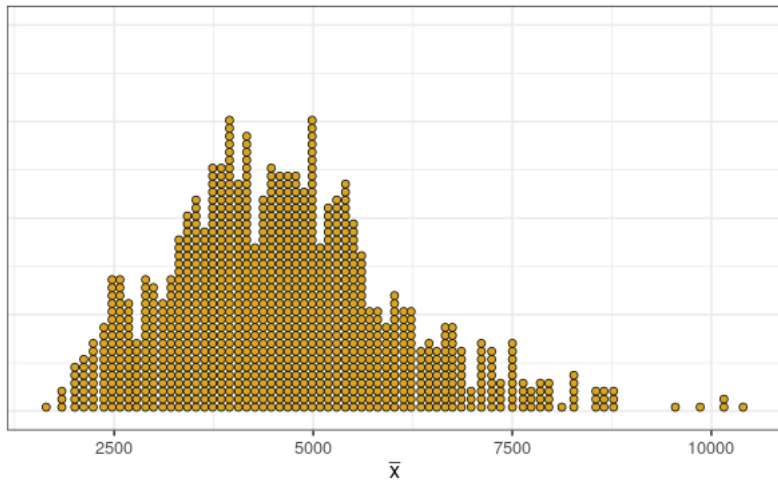
Bootstrapped Enrollment Samples (n = 20)



Bootstrapped Enrollment Samples (n = 20)



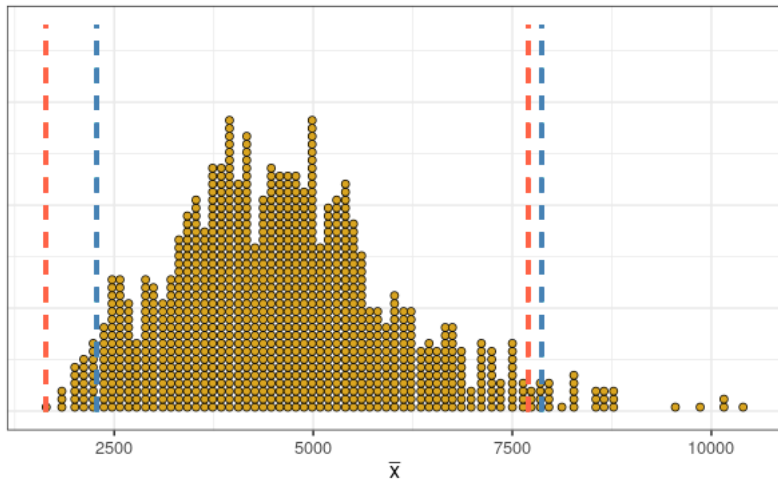
Means from 1000 Bootstraps



Creating confidence intervals for college enrollment

```
1 > quants <- c(0.025, 0.975)
2 > qt(quants, df = 20-1)
3 [1] -2.093  2.093
4 >
5 > ## Point estimate +/- MOE
6 > 4669.4 + c(-2.093, 2.093) * 1447.3
7 [1] 1640.2 7698.6
8 >
9 > ## Find quantiles directly from sample distribution
10 > quantile(enrollment_bootstrap, c(0.025, 0.975))
11   2.5%   97.5%
12 2279.8 7867.2
```

Bootstrapped Sample Means



Frequently Unasked Questions

Concern: What if our sample sucks?

Concern: If we can use bootstrapping, why do we bother with CLT?

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*

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