Estimation, Part 3

The Bootstrap

Estimation

Recall: We can construct a confidence interval for a parameter if we understand what the sampling distribution of that parameter looks like.

Eg. Using *t*-distributions for the sampling distribution of the mean

Problems:

- We have to make assumptions about the population of interest to use particular sampling distributions.
- It's not always easy to find the sampling distribution for certain parameters.

Bootstrap Estimation

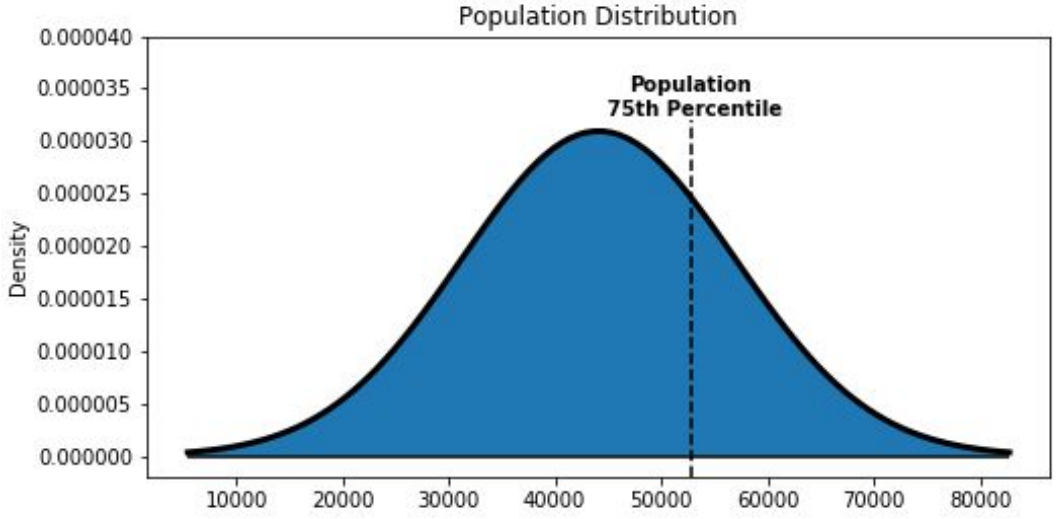
Big Idea:

- Rather than analytically trying to determine the sampling distribution, approximate it using our sample.
- Draw samples from the approximation (by resampling with replacement from the original sample) and look at how the relevant sample statistics are distributed.
- Once we have a good idea about the sampling distribution, use this information to construct a confidence interval.

Bootstrap Estimation

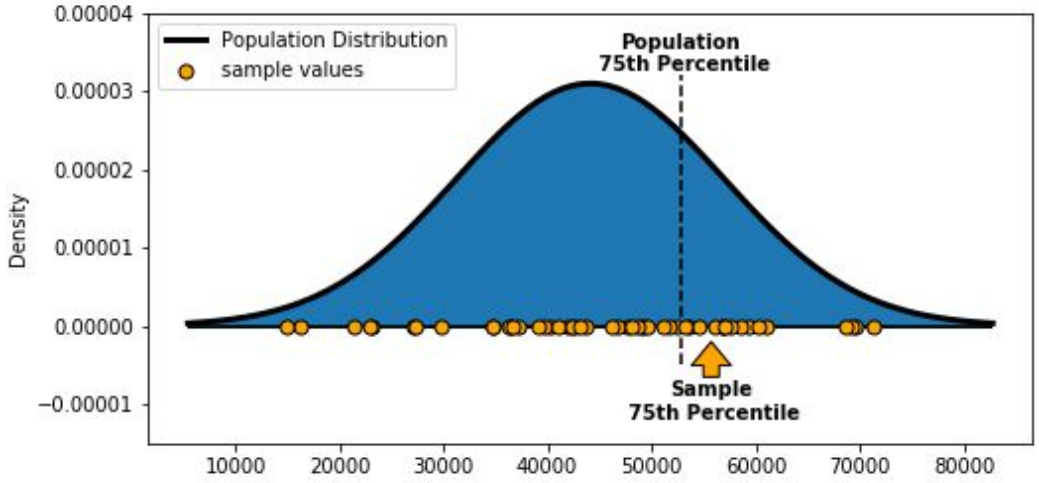
Example: Start by looking at our (unknown) population distribution.

Goal: Estimate the 75th percentile of the population values.



Bootstrap Estimation

Start by taking a sample of size 50 from the population and find the sample 75th percentile:



The sample gives us a *point estimate* of the 75th percentile, which in this case is too high.

Bootstrap Estimation

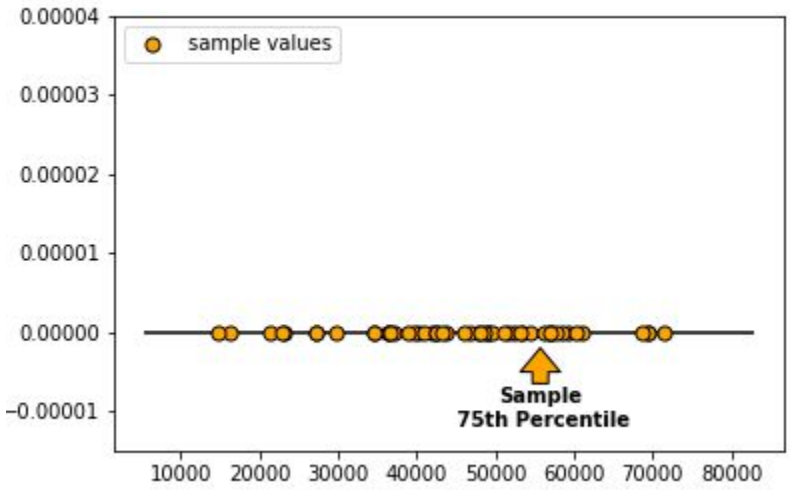
How good can we expect the point estimate to be?

To answer that, we need to know what the distribution of all possible point estimates for the 75th percentile look like.

But, we only have our single sample...

Bootstrap Estimation

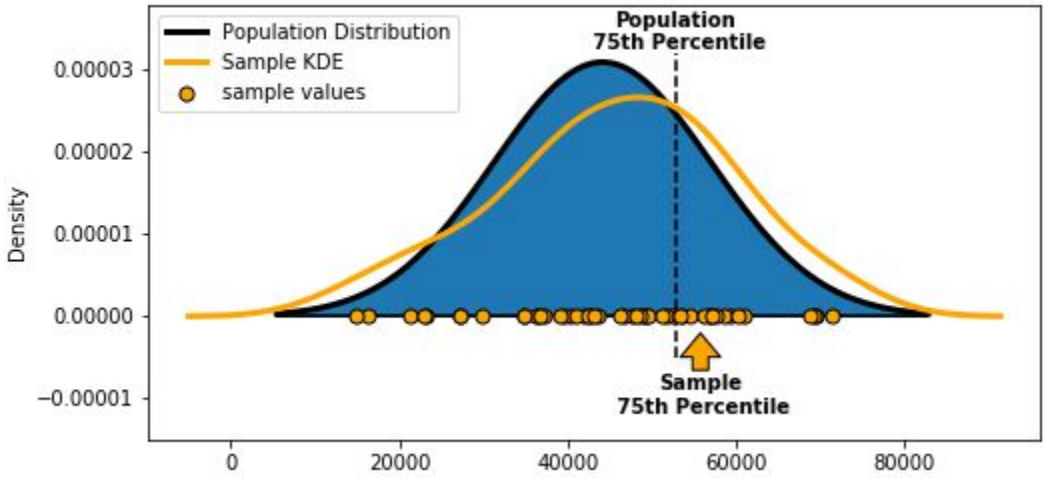
Based on our sample, how can we approximate the population distribution?



NASHVILLE SOFTWARE SCHOOL

Bootstrap Estimation

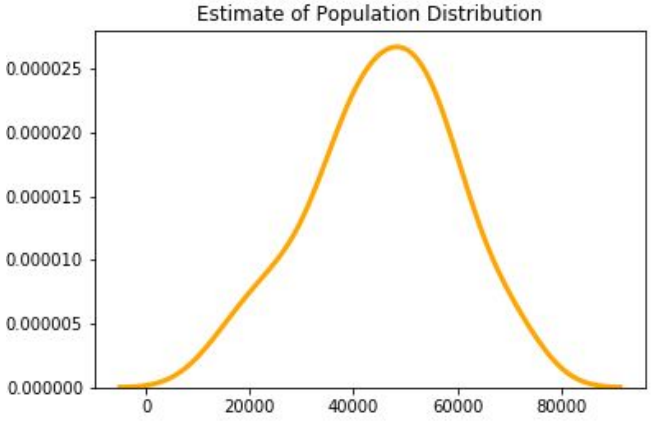
One way would be to use Kernel Density estimation:



NASHVILLE SOFTWARE SCHOOL

Bootstrap Estimation

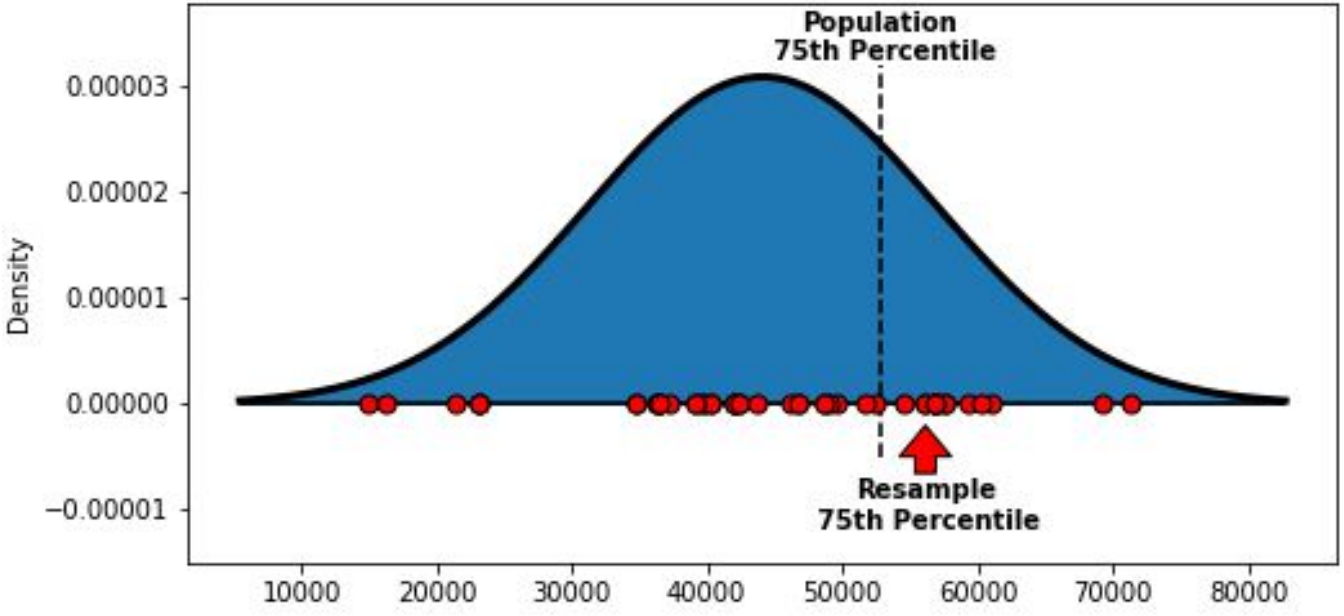
If we want to see what possible sample look like, we can use the distribution determined by our sample:



This is usually done by repeatedly *resampling* from our sample - drawing samples of the same size from the original sample **with replacement.**

Bootstrap Estimation

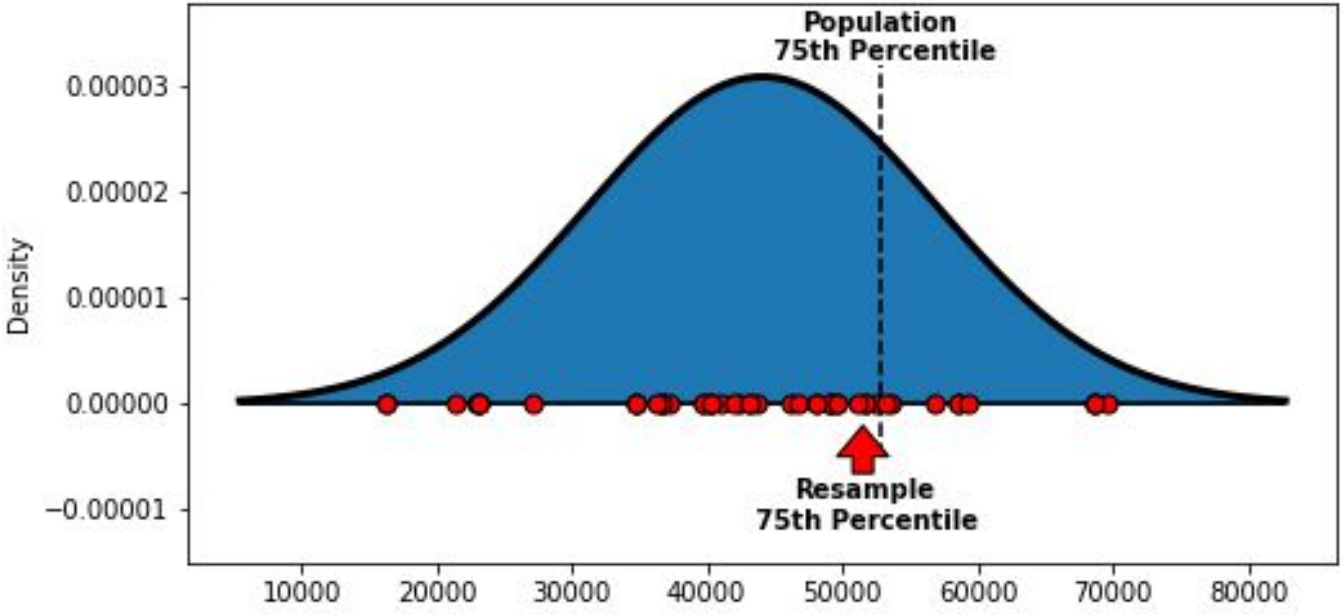
One possible resample:



NASHVILLE SOFTWARE SCHOOL

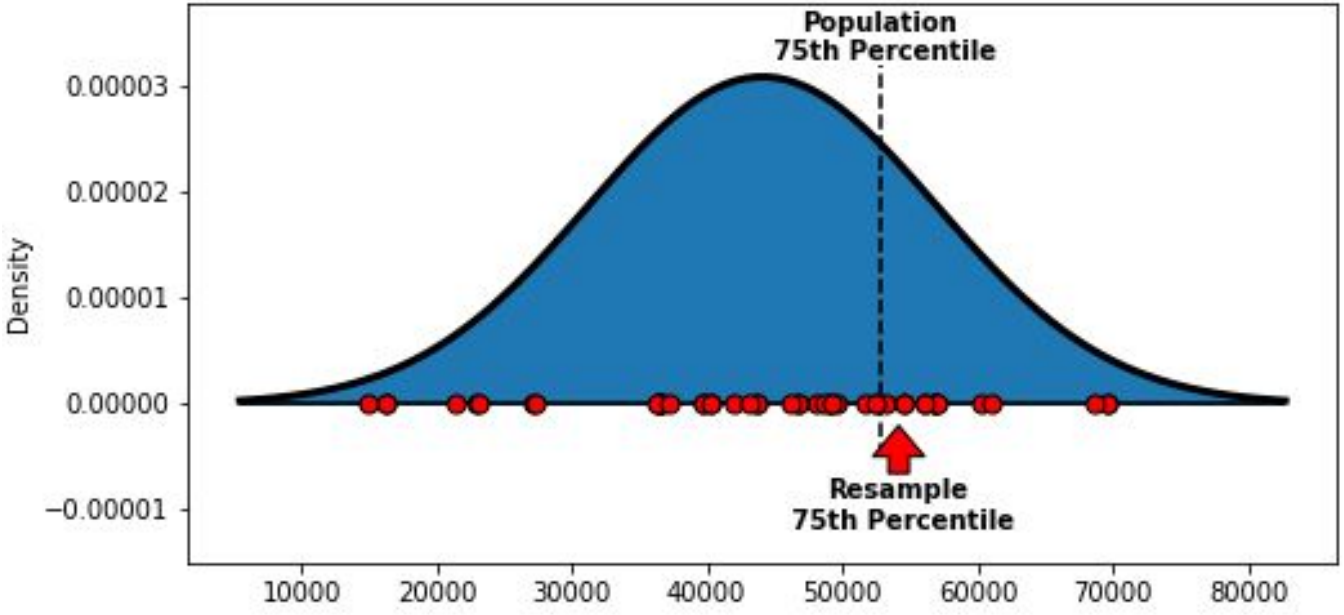
Bootstrap Estimation

Another possible resample:



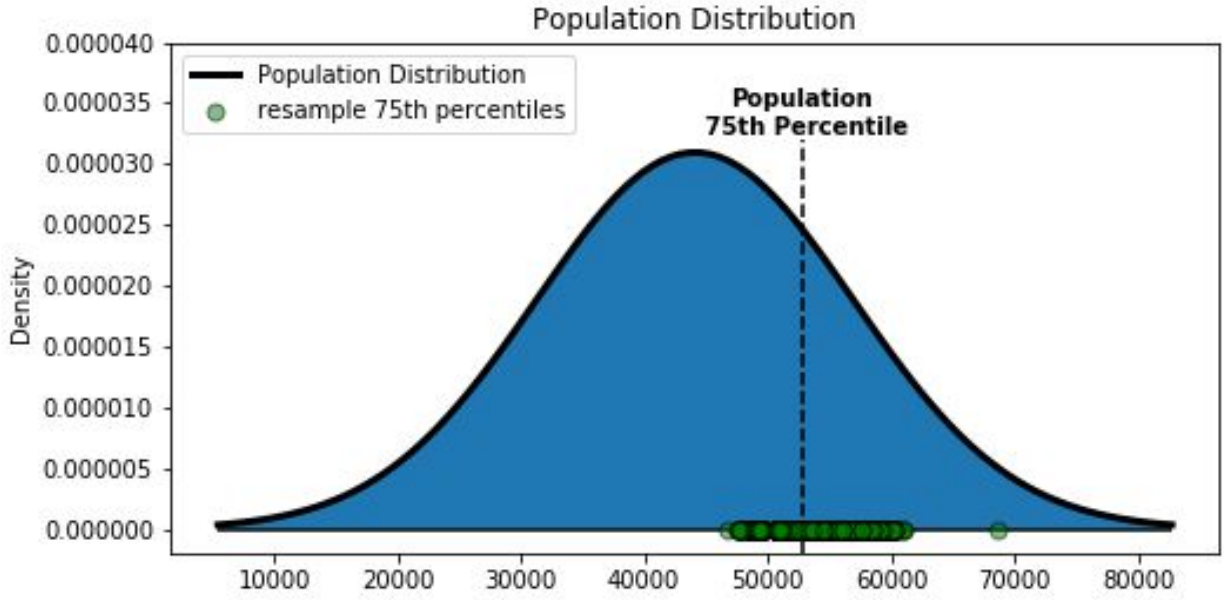
Bootstrap Estimation

Yet another possible resample:



Bootstrap Estimation

By take a large number (10,000 or so) resamples, we get an idea of the distribution of sample 75th percentiles:



NASHVILLE SOFTWARE SCHOOL

Bootstrap Estimation

Recipe for bootstrap 95% confidence interval:

1. Draw a sample of size *n*.
2. Find the **point estimate**, *m*, for the statistic of interest (eg. mean)
3. Draw a large number (say, 10,000) of samples with replacement of size *n* from the original sample.
4. For each new sample, calculate the statistic of interest.
5. Find the 2.5th and 97.5th percentile, *a* and *b* of the calculated statistics.
6. The 95% confidence interval is:

$$[m - (b - m), m + (m - a)]$$