Announcements
Percentiles
The 80th percentile is the value in a set that is at least as large as 80% of the elements in the set.

For $s = [1, 7, 3, 9, 5]$, $\text{percentile}(80, s)$ is 7.

The 80th percentile is ordered element 4: $(80/100) \times 5$

For a percentile that does not exactly correspond to an element, take the next greater element instead.
The percentile Function

- The $p$th percentile is the value in a set that is at least as large as $p\%$ of the elements in the set.
- Function in the datascience module:
  \[
  \text{percentile}(p, \text{values})
  \]
- $p$ is between 0 and 100
- Returns the $p$th percentile of the array
Discussion Question

Which are True, when \( s = [1, 7, 3, 9, 5] \)?

\[
\begin{align*}
\text{percentile}(10, s) &= 0 \\
\text{percentile}(39, s) &= \text{percentile}(40, s) \\
\text{percentile}(40, s) &= \text{percentile}(41, s) \\
\text{percentile}(50, s) &= 5
\end{align*}
\] (Demo)
Estimation
Inference: Estimation

- How big is an unknown parameter?

- If you have a census (that is, the whole population):
  - Just calculate the parameter and you’re done

- If you don’t have a census:
  - Take a random sample from the population
  - Use a statistic as an estimate of the parameter

(Demo)
Variability of the Estimate

- One sample ➔ One estimate
- But the random sample could have come out differently
- And so the estimate could have been different
- Main question:
  - How different could the estimate have been?
- The variability of the estimate tells us something about how accurate the estimate is:
  \[ \text{estimate} = \text{parameter} + \text{error} \]
Where to Get Another Sample?

- One sample ➔ One estimate
- To get many values of the estimate, we needed many random samples
- Can’t go back and sample again from the population:
  - No time, no money
- Stuck?
The Bootstrap
The Bootstrap

- A technique for simulating repeated random sampling

- All that we have is the original sample
  - ... which is large and random
  - Therefore, it probably resembles the population

- So we sample at random from the original sample!
Why the Bootstrap Works

population

sample

resamples

All of these look pretty similar, most likely.
Key to Resampling

● From the original sample,
  ○ draw at random
  ○ with replacement
  ○ as many values as the original sample contained

● The size of the new sample has to be the same as the original one, so that the two estimates are comparable

(Demo)