# Predictions from Samples

Referenced from lesson [Randomness and Sample Size](https://www.bootstrapworld.org/materials/spring2021/en-us/courses/data-science/lessons/random-samples/index.shtml) (Spring, 2021)

1. In the Definitions Area of the [Expanded Animals Starter File](https://code.pyret.org/editor#share=1d3HuG_LjdX9HpfQPCmVmDG9mYml4nOX8&v=ebd213d), define the following samples:

tiny-sample = random-rows(big-animals-table, 10)  
small-sample = random-rows(big-animals-table, 20)  
medium-sample = random-rows(big-animals-table, 40)  
large-sample = random-rows(big-animals-table, 80)

1. Click run and make a pie-chart of the species in the tiny-sample.

* What animals are in the sample?
* Click run for a new random sample and make another pie-chart of species in the tiny-sample. What animals are in the sample?
* Click run for a new random sample and make another pie-chart of species in the tiny-sample. Based on these samples, how many species of animals do you think are at the shelter?
* Which species do you think are the most at the shelter?

1. What did you learn from taking multiple samples that you wouldn’t have known if you’d only taken a single sample?
2. Now use small-sample to make a pie-chart of the species.

* What animals are in the sample?
* Click run for a new random sample and make another pie-chart of species in the small-sample. What animals are in the sample?

1. Now that you’ve seen small-sample, how has your sense of the distribution of the species changed?
2. Now use medium-sample to make a pie-chart of the species. If there are about 400 animals at the shelter, how many of each species would you predict there to be.
3. Now use large-sample to make a pie-chart of the species. If there’s anything you’d like to change about your prediction now that you’ve seen large-sample, record it here.
4. Let’s see how accurate your prediction is…​ *feel free to click run and build a few more pie charts from your samples if you want to collect more information first!* When you’re ready, make a pie-chart of animals-table-2.

* Which predictions were closest?
* Which predictions were off?
* Were there any surprises?

1. In the real world, we usually don’t have access to a whole dataset to check predictions against! How could we test…​

-*Every giraffe on the planet?* -*Everyone who has ever come in contact with a covid-positive person?* -*Every person who identifies as queer?*

1. **What strategies can we use to make sure that predictions from samples are as close to accurate as possible?**

*These materials were developed partly through support of the National Science Foundation, (awards 1042210, 1535276, 1648684, and 1738598).  Bootstrap:Data Science by the* [*Bootstrap Community*](https://bootstrapworld.org/community/index.shtml) *is licensed under a* [*Creative Commons 4.0 Unported License*](https://creativecommons.org/licenses/by-nc-nd/4.0/)*. This license does not grant permission to run training or professional development. Offering training or professional development with materials substantially derived from Bootstrap must be approved in writing by a Bootstrap Director. Permissions beyond the scope of this license, such as to run training, may be available by contacting* [*contact@BootstrapWorld.org*](https://www.bootstrapworld.org/materials/spring2021/en-us/courses/data-science/lessons/random-samples/pages/mailto:contact@BootstrapWorld.org)*.*

Last updated 2021-10-12 13:35:38 -0400