Sampling: Takeaways 🖻

by Dataquest Labs, Inc. - All rights reserved © 2019

Syntax

• Sampling randomly a **Series** object:

```
### Sampling 10 sample points ###
sample_10 = Series.sample(10)

### Sampling 500 sample points ###
sample_500 = Series.sample(500)
```

• Making the generation of random numbers predictable using the random_state parameter:

```
### Sampling 10 sample points in a reproducible way ###
sample_10 = Series.sample(10, random_state = 1)

### Using a different value for `random_state` ###
sample_10_different = Series.sample(10, random_state = 2)
```

Concepts

- The set of *all* individuals relevant to a particular statistical question is called a **population**. A smaller group selected from a population is called a **sample**. When we select a smaller group from a population, we do **sampling**.
- A **parameter** is a metric specific to a population, and a **statistic** is a metric specific to a sample. The difference between a statistic and its corresponding parameter is called **sampling error**. If the sampling error is low, then the sample is **representative**.

- To make our samples representative we can try different sampling methods:
 - Simple random sampling.
 - Stratified sampling.
 - Cluster sampling.
- When we describe a sample or a population, we do **descriptive statistics**. When we try to use a sample to draw conclusions about a population, we do **inferential statistics** (we *infer* information from the sample about the population).

Resources

- The Wikipedia entry on sampling.
- <u>The Wikipedia entry</u> on samples.
- The Wikipedia entry on populations.



Takeaways by Dataquest Labs, Inc. - All rights reserved © 2019