Sampling: Takeaways ⤢

Syntax

- Sampling randomly a series object:

```
sample_10 = Series.sample(10)
```

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

```
sample_10 = Series.sample(10, random_state=1)
```

Concepts

- We call the set of *all* individuals relevant to a particular statistical question a **population**. We call a smaller group selected from a population 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**
  - **Proportional stratified sampling**
  - **Cluster sampling**

- Simple random sampling requires us to choose the individuals in the populations randomly — all individuals must have equal chances of being selected.

- Stratified sampling requires us to organize our data into different groups (strata) and then sample randomly from each group. Unlike simple random sampling, stratified sampling ensures we get a sample that has observations for all the categories of interest.

- Proportional stratified sampling requires us to consider the proportions in the population when we divide the data into strata.

- Cluster sampling requires us to list all the data sources (all the clusters) we can find and then randomly pick a few from which to collect data. Once we've made our choice, we can perform simple random sampling on each cluster.

- 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.
- [The Wikipedia entry](#) on samples.
- [The Wikipedia entry](#) on populations.