

How does systematic sampling work in Pandas, and can you provide examples?

Authored by
stats writer

April 21, 2024

RECOMMENDED CITATION

stats writer (2024). *How does systematic sampling work in Pandas, and can you provide examples?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=137811>

Systematic sampling is a method used in Pandas, a Python library for data analysis, to select a subset of data from a larger dataset in a systematic and unbiased manner. This method involves selecting every n th data point from the larger dataset, where n is the sampling interval. This ensures that the selected subset is representative of the entire dataset and reduces the risk of bias.

To implement systematic sampling in Pandas, the dataset needs to be sorted in a random order. Then, the sampling interval needs to be specified and the desired number of samples can be selected using the "iloc" function. For example, if we have a dataset of 1000 rows and we want to select every 10th row, we can use the following code in Pandas:

```
df.sample(frac=1).iloc
```

This will randomly shuffle the dataset and select every 10th row, resulting in a subset of 100 rows.

Overall, systematic sampling in Pandas is a useful tool for obtaining a representative sample from a larger dataset and can be easily implemented using the "iloc" function.

Systematic Sampling in Pandas (With Examples)

Researchers often take samples from a population and use the data from the sample to draw conclusions about the population as a whole.

One commonly used sampling method is systematic sampling, which is implemented with a simple two step process:

1. Place each member of a population in some order.
2. Choose a random starting point and select every n th member to be in the sample.

This tutorial explains how to perform systematic sampling on a pandas DataFrame in Python.

Example: Systematic Sampling in Pandas

Suppose a teacher wants to obtain a sample of 100 students from a school that has 500 total students. She chooses to use systematic sampling in which she places each student in alphabetical order according to their last name, randomly chooses a starting point, and picks every 5th student to be in the sample.

The following code shows how to create a fake data frame to work with in Python:

```
import pandas as pd
import numpy as np
import string
import random

#make this example reproducible
np.random.seed(0)

#create simple function to generate random last names
def randomNames(size=6,
chars=string.ascii_uppercase):
```

```
return ''.join(random.choice(chars) for _ in range(size))
```

```
#create DataFrame
```

```
df = pd.DataFrame({'last_name': ,  
'GPA': np.random.normal(loc=85, scale=3, size=500)})
```

```
#view first six rows of DataFrame
```

```
df.head()
```

```
last_name GPA
```

```
0 PXGPIV 86.667888
```

```
1 JKRRQI 87.677422
```

```
2 TRIZTC 83.733056
```

```
3 YHUGIN 85.314142
```

```
4 ZVUNVK 85.684160
```

And the following code shows how to obtain a sample of 100 students through systematic sampling:

```
#obtain systematic sample by selecting every 5th row
```

```
sys_sample_df = df.iloc
```

```
#view first six rows of DataFrame
```

```
sys_sample_df.head()
```

```
last_name gpa
```

3 ORJFW 88.78065
8 RWPSB 81.96988
13 RACZU 79.21433
18 ZOHKA 80.47246
23 QJETK 87.09991
28 JTHWB 83.87300

#view dimensions of data frame
sys_sample_df.shape

(100, 2)

Notice that the first member included in the sample was in the first row of the original data frame. Each subsequent member in the sample is located 5 rows after the previous member.

And from using shape() we can see that the systematic sample we obtained is a data frame with 100 rows and 2 columns.

Types of Sampling Methods

Cluster Sampling in Pandas

Stratified Sampling in Pandas