

Veri Analizi



sampling

5.01.2024

Muhammet Gazel 180401123

Örnekleme

-Örnekleme Nedir?

Örnekleme, büyük bir popülasyondan rastgele seçilmiş bir alt küme oluşturma sürecidir. Bu süreç, popülasyonun genel özelliklerini daha küçük bir örneklem üzerinden anlamamızı sağlar. Örneklem seçimi, istatistiksel analizlerde güvenilir sonuçlar elde etmek ve genellemeler yapmak için kritik öneme sahiptir.

-Örnekleme Türleri:

1. Rastgele Örnekleme (Simple Random Sampling) Her bir elemanın seçilme olasılığı eşittir ve rastgele seçilir.
2. Stratifiye Örnekleme (Stratified Sampling): Popülasyon önceden tanımlanmış alt gruplara (stratejilere) ayrılır ve her bir strata rastgele örnekler alınır.
3. Kümelenmiş Örnekleme (Cluster Sampling) Popülasyon kümeler halinde gruplara ayrılır ve rastgele kümeler seçilir.

-Örnekleme ve Veri Analizi Arasındaki Bağlantı:

Birçok veri analizi çalışması, örneklemeler üzerinde yapılır. Örnekleme, popülasyon içinden rastgele seçilmiş bir alt kümeden veri toplamak anlamına gelir. Bu, bir hipotez testi veya bir regresyon analizi yapmadan önce veriyi anlamak için kullanılır. Örneğin, bir pazarlamacı, hedef kitlesi yerine rastgele bir örnek üzerinde testler yaparak pazarlama stratejilerini değerlendirebilir.

-İki Örneklem Karşılaştırma ve Anlamı:

İki örneklem karşılaştırma, istatistiksel anlamlılığı belirlemek için kullanılır. Bu, iki farklı grup arasında istatistiksel bir fark olup olmadığını anlamak için yapılır. Örneğin, bir ilaç şirketi, bir tedavi yöntemi ve plasebo arasındaki etkileri karşılaştırarak ilacın gerçek etkilerini belirlemeye çalışabilir.

-Python ile Örnekleme ve Karşılaştırma:

Python, veri analizi için güçlü bir dil olarak öne çıkıyor. İstatistiksel örnekleme ve karşılaştırmalar yapmak için `numpy`, `random` ve `matplotlib` gibi kütüphaneleri kullanabiliriz. Bu kütüphaneler, örnekleme yapmak, istatistiksel analizler yapmak ve sonuçları görselleştirmek için kullanılabilir.

Verdiğiniz örnek kodlara dayanarak, bu örnek ile ilgili bazı temel istatistikleri hesaplayabiliriz. Örneğin, bir örneklem veri setinin ortalama değeri, medyanı ve standart sapması gibi istatistikleri hesaplayabiliriz.

```
```python
import numpy as np

Örneklem veri seti
sample = [23, 45, 67, 32, 12, 56, 78, 43, 21, 34]
```

- `np.mean()` ile örneklem veri setinin ortalaması hesaplanıyor.
- `np.median()` ile örneklem veri setinin medyanı bulunuyor.
- `np.std()` ile örneklem veri setinin standart sapması hesaplanıyor.

Bu istatistikler, veri setinin merkezi eğilimi (ortalama, medyan) ve yayılımı (standart sapma) hakkında bilgi sağlar ve veri setinin özetlenmesine yardımcı olur. Bu, veri setinizin genel özellikleri hakkında bir fikir edinmek için kullanışlı bir başlangıç noktası olabilir.

bir popülasyon oluşturuyoruz:

```
```python
numbers=[]
size=10000
for i in range(size):
    number=random.gauss(50,15)
    numbers.append(number)
```
```

Bu popülasyon üzerinden örneklem almak için bir fonksiyon kullanıyoruz:

```
```python
def get_sample_from_a_population(size=100,pop_1=[]):
    if len(pop_1)==0:
        numbers=generate_population_with_size_mean_and_sd(10000,50,15)
    else:
        numbers=pop_1
    size_of_sample=size
    sample=[]
    for i in range(size_of_sample):
        number=random.choice(numbers)
        sample.append(number)
    return sample
```
```

Daha sonra, bu örneklem verileri üzerinde istatistiksel analizler yapabilir ve iki örneklem arasındaki farkları inceleyebiliriz:

```
```python
pop_1=generate_population_with_size_mean_and_sd(20000,30,20)
sample_1=get_sample_from_a_population(90,pop_1)
np.mean(sample_1),np.std(sample_1),len(sample_1)
```
```

İki örneklem arasındaki farkı hesaplamak içinse şu kodu kullanıyoruz:

```
```python
def get_two_sample_difference():
    # İki farklı örneklem üzerinden istatistiksel karşılaştırma yapılıyor
    # ... (Kodun devamı)
    return mean_1-mean_2
```
```

### Değiştirmeli ve Değiştirmesiz Örneklemeyi:

- Bağlantılı veri analitiği ve istatistiksel analizlerde sıkça kullanılan örneklem yöntemleri, popülasyonun belirli bir bölümünü veya alt kümesini temsil etmek için önemlidir. Değiştirmeli ve değiştirmesiz örneklemeler, bu tekniklerin iki ana türüdür.

#### Değiştirmeli Örneklem

Değiştirmeli örneklem, her bir gözlemin seçildikten sonra popülasyona geri konulduğu bir örneklem yöntemi. Bu, aynı gözlemin birden fazla kez seçilebileceği anlamına gelir. Bu yöntem, rastgele örneklem ve bootstrap yöntemleri gibi tekniklerde kullanılır.

Python'da değiştirmeli örneklem yapmak için `numpy` kütüphanesini kullanabiliriz. Örnek bir kod örneği aşağıdaki gibidir:

```
```python
import numpy as np
```

Değiştirmesiz Örneklem

Değiştirmesiz örneklem, her bir gözlemin seçildikten sonra popülasyondan çıkarıldığı bir örneklem yöntemi. Bu, aynı gözlemin bir kez seçilebileceği ve bir daha seçilemeyeceği anlamına gelir. Bu yöntem, rastgele örneklem alma ve örneklemleme yaparken öğelerin tekrar seçilmemesi gereken durumlarda kullanılır.

Python'da değiştirmesiz örneklem yapmak için `random` modülünü kullanabiliriz. Örnek bir kod örneği şu şekildedir:

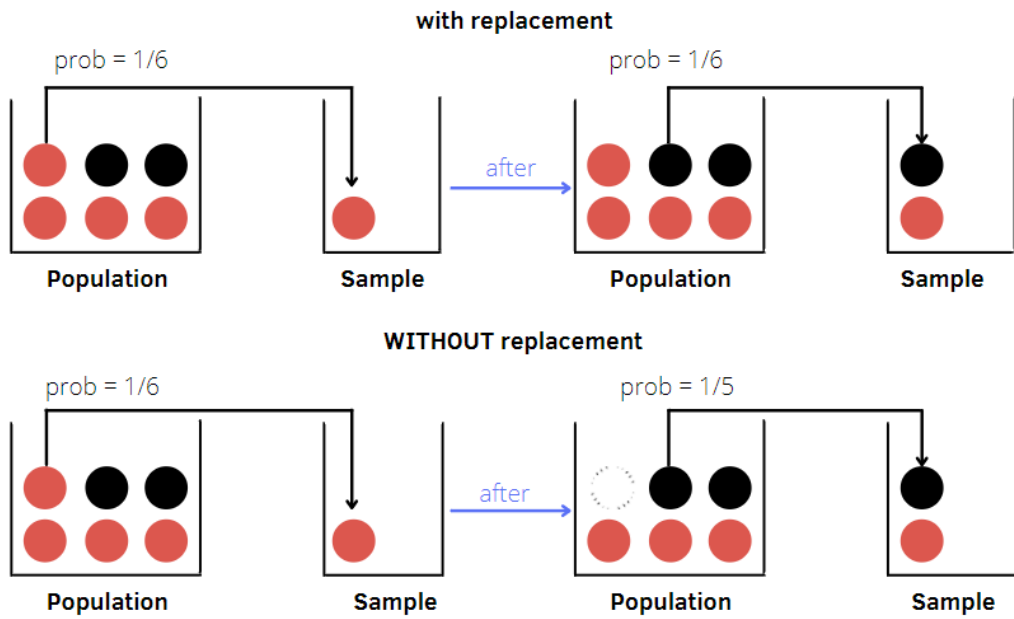
```
```python
import random
```

eklediğimiz örnekte:

`get_sample_from_a_population` fonksiyonu içinde, örneklem alma işlemi `random.choice()` kullanılarak gerçekleştiriliyor. Bu işlem, verilen bir popülasyondan (`pop_1`) belirli bir boyutta (`size`) örneklem almayı amaçlıyor. Ancak, bu fonksiyon herhangi bir örneklem türü belirtmiyor, yani bu örneklem işlemi değiştirmeli ya da değiştirmesiz olabilir.

`generate_population_with_size_mean_and_sd` fonksiyonu içinde de örneklem verisi oluşturuluyor, ancak burada da örneklem yöntemi açıkça belirtilmiş değil.

Örnekleme işlemlerini belirli bir örneklem türüyle uyumlu hale getirmek istiyorsanız, örneklem fonksiyonları içinde yer alan `random.choice()` fonksiyonunun `replace` parametresini kullanarak bu yöntemi değiştirebilirsiniz. Örneğin, `random.choice(numbers, replace=True)` şeklinde kullanılarak değiştirmeli örneklem yapabilirsiniz.



### -örnekleme hatası

Örnekleme sürecinde bazı hatalar ortaya çıkabilir ve bu hatalar, örneklem verilerinin popülasyonu temsil etme kabiliyetini etkileyebilir. Bu duruma "örnekleme hatası" denir.

Örnekleme hatası, örneklem alınırken yapılan hatalardan kaynaklanabilir. Örneğin, yanlış bir örneklem büyüklüğü seçmek, rastgele seçim yaparken önyargılı olmak veya popülasyonu yeterince temsil etmeyen bir örneklem seçmek örnekleme hatalarına örnek olarak verilebilir.

Örneğin, yazdığınız kodlarda örneklem alırken rastgele seçme işlemini `random.choice()` fonksiyonu ile yapıyorsunuz. Bu durumda, bazı durumlarda örneklem alırken bazı değerlerin daha fazla seçilme olasılığı olabilir ve bu da örnekleme hatalarına yol açabilir.

eklediğimiz örnekte::

```
def get_sample_from_a_population(size=100, pop_1=[]):
 if len(pop_1) == 0:
 numbers = generate_population_with_size_mean_and_sd(10000, 50, 15)
 else:
 numbers = pop_1
 size_of_sample = size
 sample = []
 for i in range(size_of_sample):
 number = random.choice(numbers)
 sample.append(number)
 return sample
```

Bu fonksiyon, belirtilen büyüklükte bir örneklem seçmek için kullanılıyor. Ancak, örneklem büyüklüğünü seçerken popülasyonun gerçek özelliklerini temsil edebilecek kadar büyük bir örneklem alınmamış olabilir. Örnekleme hatası, bu tür durumlarda örneklem verilerinin popülasyonu yeterince temsil etmeme riskini taşır.

Örnekleme hatasını azaltmak için daha büyük örneklem büyüklükleri seçmek, rastgele seçimlerde önyargısız olmak ve popülasyonun farklı özelliklerini yansıtan bir örneklem oluşturmak önemlidir. Bu, daha güvenilir istatistiksel sonuçlar elde etmenize yardımcı olabilir.