

Temel Kavramlar

Kitle: Belirli bir karakteristik özelliğe sahip birimler topluluğuna kitle (**population**) denir.

Örneklem: Kitleyi temsil edebilecek nitelikte birimlerin oluşturduğu herhangi bir alt küme *örneklem* (**sample**) olarak tanımlanır.

Örnekleme: İlgilendiğimiz kitleden örneklem çekme işlemine **örnekleme (sampling)** adı verilir. Örnekleme yapmak için kullanılan ve literatürde yaygın olarak kullanılan bazı örnekleme yöntemleri (**sampling techniques**) aşağıdaki gibidir:

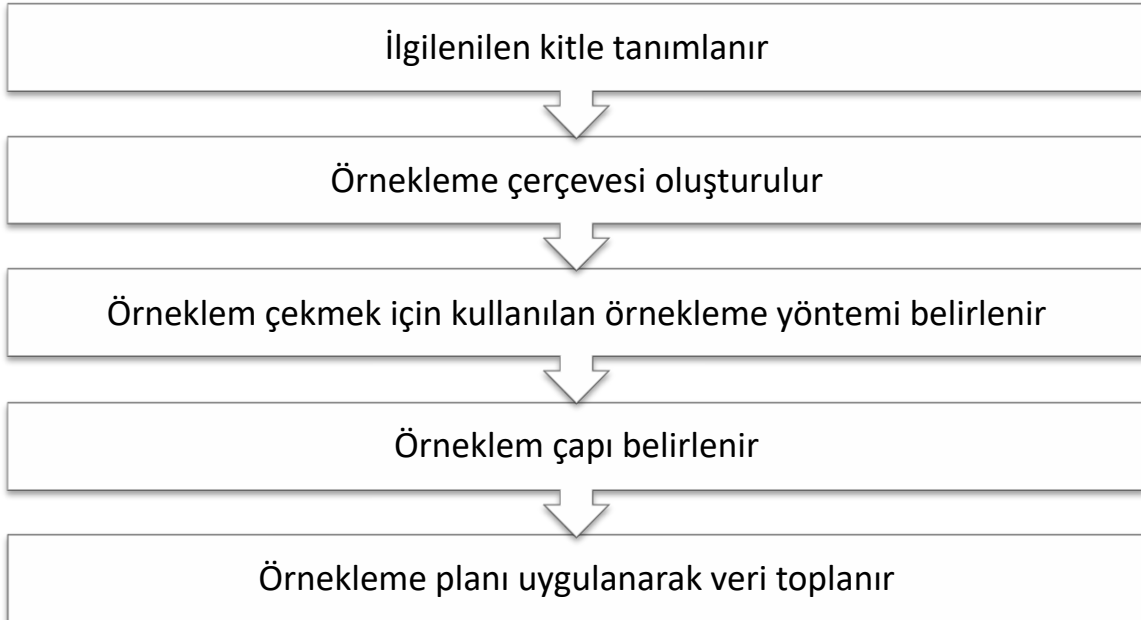
- Basit rastgele örnekleme (**simple random sampling**)
- Tabakalı rastgele örnekleme (**stratified sampling**)
- Sistemik örnekleme (**systematic sampling**) ve
- Küme örnekleme (**cluster sampling**)

Örnekleme, kitle parametrelerini

- Toplam (Y)
- Ortalama (\bar{Y})
- Oran (P)

tahmin etmek için bazı kısıtlar altında örneklem çekme işlemidir.

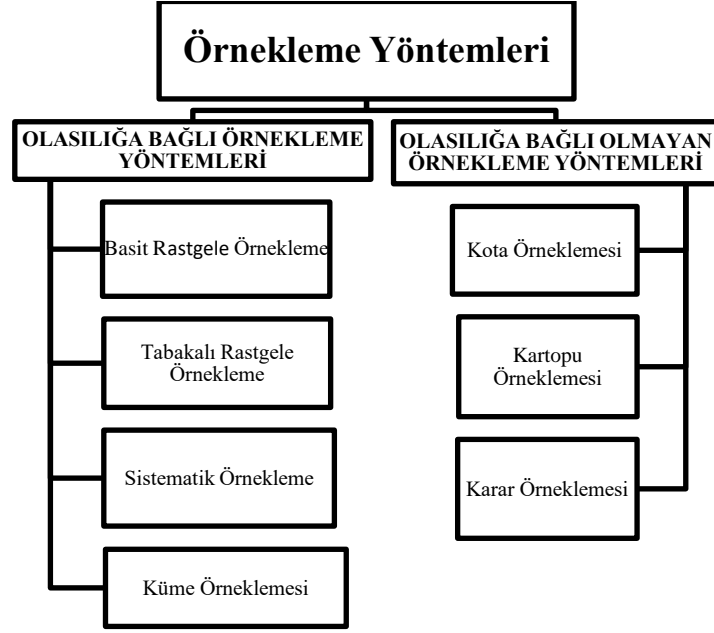
Örnekleme sürecinin aşamaları aşağıda gösterildiği gibidir:



Örnekleme Teorisinin Amacı: Kitleyi en iyi şekilde temsil edebilecek örnekleme seçmek ve bu örnekleme yardımıyla kitle parametrelerini tahmin etmektir. Bu amaca ulaşmak için kitlenin yapısına en uygun örnekleme yönteminin belirlenmesi oldukça önemlidir.

Örnekleme Birimlerinin Seçilme Olasılığına Göre Örnekleme Yöntemleri

Örnekleme yöntemleri olasılığa bağlı ve olasılığa bağlı olmayan olmak üzere iki gruba ayrılırlar.



Olasılığa Bağlı Örnekleme Yöntemleri:

- Kitledeki her birimin örnekleme çekilme olasılığı sıfırdan farklı bilinen bir değere eşittir.
- Kitledeki örnekleme birimlerinden oluşan bir liste (çerçeve) gereklidir.

Olasılığa Bağlı Olmayan Örnekleme Yöntemleri:

- Uygulama maliyeti az, uygulama süresi kısadır.
- Örnekleme birimlerinden oluşan bir liste (çerçeve) gerekli değildir.
- Örnekleme seçimi araştırmacının kişisel değerlendirmeleri ışığında yapılır.

Örnekleme Dışı Hatalar

- Kayıt hataları
- Ölçüm hataları
- Kitlede yer almasına rağmen çerçevede yer almayan birimlerden kaynaklanan hatalar

Örneklem Seçimi

- Bilgisayardan rastgele sayılar üretmek yada
- Rastgele sayılar tablosu kullanarak

yapılabilir.

Daha basit bir diğer yöntem ise kitleyi oluşturan birimlerin 1'den N'e kadar numaralandırılması ve bu birimlerin içinden n tanesinin rastgele olarak seçilmesidir.

Rastgele Sayılar Tablosu Kullanmanın Amacı

Örneklem seçimindeki yanlılığı önlemektir. Çünkü, rastgele sayılar tablosundaki rakamlar birbirinden bağımsızdır ve seçilme şansları eşittir.

Beklenen Değer

$y_1, y_2, \dots, y_n, P(Y = y)$ kesikli dağılımından rastgele bir örneklem olsun.

Y rastgele değişkeninin beklenen değeri

$$E(Y) = \sum_{i=1}^n y_i P(Y_i = y_i)$$

olarak tanımlanır.

Burada,

$$\sum_{i=1}^n P(Y_i = y_i) = 1$$

olarak varsayılır.

Beklenen Değerin Bazı Özellikleri

Y rastgele değişken ve c bir sabit olmak üzere

- $E(Y + c) = E(Y) + c$
- $E(cY) = cE(Y)$

olarak tanımlanır.

Y ve Z rastgele değişkenler ve d bir sabit olmak üzere

- $E(Y + Z) = E(Y) + E(Z)$
- $E(cY + dZ) = cE(Y) + dE(Z)$

olarak tanımlanır.

Y ve Z rastgele değişkenler olsun

- $E(YZ) = E(Y)E(Z)$ (Y ve Z Bağımsız ise)
- $E(YZ) = E(Y)E(Z) + Cov(Y, Z)$ (Y ve Z Bağımsız değil ise)

olarak tanımlanır.

Varyans

$y_1, y_2, \dots, y_n, P(Y = y)$ kesikli dağılımından rastgele bir örneklem olsun.

Y rastgele değişkeninin varyansı

$$\sigma^2 = E(Y - E(Y))^2 = E(Y - \bar{Y})^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2$$

olarak tanımlanır.

Burada,

$$E(Y) = \bar{Y}$$

dır.

Varyansın Bazı Özellikleri

Y rastgele değişken ve c bir sabit olmak üzere

- $Var(Y + c) = Var(Y)$
- $Var(cY) = c^2 Var(Y)$

olarak tanımlanır.

Y ve Z rastgele değişkenler olsun

- $Var(Y + Z) = Var(Y) + Var(Z)$ (Y ve Z Bağımsız ise)
- $Var(Y - Z) = Var(Y) + Var(Z)$ (Y ve Z Bağımsız ise)
- $Var(Y + Z) = Var(Y) + Var(Z) + 2Cov(Y, Z)$ (Y ve Z Bağımsız değil ise)
- $Var(Y - Z) = Var(Y) + Var(Z) - 2Cov(Y, Z)$ (Y ve Z Bağımsız değil ise)

Yansızlık

Kitle parametresinin tahmin edicisi $\hat{\theta}$, kitle parametresi θ için yansız bir tahmin edici ise

$$E(\hat{\theta}) = \theta$$

olarak tanımlanır. Burada,

θ : Kitle parametresini

$\hat{\theta}$: Kitle parametresinin tahmin edicisini

gösterir.

Eğer

$$E(\hat{\theta}) \neq \theta$$

ise kitle parametresinin tahmin edicisi $\hat{\theta}$, kitle parametresi θ için yanlı bir tahmin edicidir ve yan miktarı (Bias)

$$\text{Bias} = E(\hat{\theta}) - \theta$$

şeklinde gösterilir.