

# Bilişimde İstatistiksel Analiz

## Ders 6

Ali Mertcan KOSE Msc.

`amertcankose@ticaret.edu.tr`

İstanbul Ticaret Üniversitesi



İSTANBUL TİCARET  
ÜNİVERSİTESİ

# Neden Örneklem?

- Örneklemde çalışmak kitlede çalışmaktan daha kolaydır.
- Kitle üzerinde çalışmak çok daha masraflı olabilir.
- Çoğu durumda tüm kitleye ulaşmak mümkün değildir.
- Örneklem sonuçları daha doğru olabilir. Çünkü daha az sayıda kişi ile (örnek ile) çalışılacağından, araştırma daha özenli yürütülebilir
- Eğer örneklem olasılıksal yöntemlerle seçiliyorsa, yapılan örnekleme hatasının kestirimini de bulmak mümkündür.

# Neden Örneklem?

- Örneklemede temel amaç seçilen örneklemin kitleyi temsil edebilecek özellikte olmasıdır.
- Örneklemin kitleyi tümüyle temsil etmesi beklenir. Ancak gerekli değildir.
- Örneklemin, kitleyi ulaşmak istediğimiz bilgide farklılık yaratabilecek etkenler yönünden temsil edebilecek özellikte olması yeterlidir.

# Neden Örneklem?

- Örneklemekten Yaralanma Koşulları
  - Uygun Örneklem Büyüklüğü
  - Uygun Örneklem Yöntemi

**Uygun örneklem büyüklüğü kullanılacak istatistiksel yöntem temel alınarak hesaplanmalıdır.** Temel ilke olarak örneklem büyüklüğü araştırmanın birincil hipotezi(ve dolayısıyla amacı) için hesaplanır. İdeal olarak ise, her önemli değişken için ayrı bir örneklem büyüklüğü hesaplanır ve en büyük tahmini yapacak olan seçilir.

Uygulanabilecek her bir istatistiksel yöntem için farklı örneklem genişliği formülleri bulunmaktadır. Burada ise en basit **kitle ortalamasının** ve **oranının** kestirilmesinde kullanılacak örneklem büyüklüğü formülleri ele alınacaktır.

# Uygun Örneklem Büyüklüğü

- 1 Kitle ortalamasının kestiriminde kullanılacak örneklem büyüklüğü formülü:

- Kitle büyüklüğü (N) bilinmiyorsa;

$$n = \frac{Z_{\alpha/2}^2 \times \sigma^2}{d^2}$$

- Kitle büyüklüğü (N) biliniyorsa;

$$n = \frac{N \times Z_{\alpha/2}^2 \times \sigma^2}{d^2 \times (N-1) + Z_{\alpha/2}^2 \times \sigma^2}$$

$Z_{\alpha/2}$  = iki yönlü hipoteze göre z tablo değeri

$\sigma^2$  = Kitle varyansı

d = Ortalama göre yapılmak istenen  $\pm$  sapma

N = Kitle büyüklüğü

# Uygun Örneklem Büyüklüğü

- ② Kitle oranının kestiriminde kullanılacak örneklem büyüklüğü formülü:

- Kitle büyüklüğü (N) bilinmiyorsa;

$$n = \frac{Z_{\alpha/2}^2 \times P \times (1-P)}{d^2}$$

- Kitle büyüklüğü (N) biliniyorsa;

$$n = \frac{N \times Z_{\alpha/2}^2 \times P \times (1-P)}{d^2 \times (N-1) + Z_{\alpha/2}^2 \times P \times (1-P)}$$

$Z_{\alpha/2}$  = iki yönlü hipoteze göre z tablo değeri

$P$  = İncelenecek k olayın görülüş sıklığı

$d$  = İncelenecek k olayın görüş sıklığına göre yapılmak istenen  $\pm$  sapma

$N$  = Kitle büyüklüğü

# Uygun Örneklem Büyüklüğü

Bu formüller basit rasgele örnekleme yöntemi için geçerli formüllerdir. Diğer yöntemlerde farklı formüller kullanılır. Formüllerdeki çoğu bilinmeyen araştırmacı tarafından öngörülen ve/veya daha önceki çalışmalar (pilot çalışmalar) yardımıyla elde edilen değerler olabilmektedir.



# Örnek 1

Bir araştırmacı yeni doğan bebeklerin ortalama doğum ağırlığını tahmin etmeye çalışıyor. Eğer %95 güvenirlilikle bebek ağırlıklarını gerçek ortalamadan (kitle ortalaması)  $\pm 250$  gr sapma ile tahmin etmeye çalışıyorsa, örnekleme kaç yeni doğan alınmalıdır. (Kitle Standart Sapması 700 gr olarak kabul ediliyor)

$$Z_{\alpha/2} = Z_{0.025} = 1.96$$

$$\sigma^2 = 700^2$$

$$d = 250 \text{ gr}$$

# Örnek 1

$$n = \frac{1.96^2 \times 700^2}{250^2} = 30$$

Eğer  $N = 60$  ise;

$$n = \frac{60 \times 1.96^2 \times 700^2}{250^2 \times (59) + 1.96^2 \times 700^2} = 20$$

$d = 400$  gr, alınırsa gerekli örneklem genişliği  $9.97 \cong 10$

## Örnek 2

Bir bölgede 50 yaş ve üstü yetişkenlerde 0.40 oranında olduğu tahmin edilen malnütrisyon sıklığını 0.04 hata ve 0.95 olasılıkla kestirebilmek için kullanılacak uygun örneklem büyüklüğü nedir?

$$Z_{\alpha/2} = Z_{0.025} = 1.96$$

$$P = 0.40$$

$$d = 0.04$$

$$n = \frac{1.96^2 \times 0.40 \times (0.60)}{0.04^2} = 577$$

Eğer bu bölgede 5000 yetişkin olduğu öngörülürse;

$$n = \frac{5000 \times 1.96^2 \times 0.40 \times (0.60)}{0.04^2 \times (4999) + 1.96^2 \times 0.40 \times (0.60)} = 517$$

Sadece örneklem sayısının yeterli olması elde edilecek sonuçların geçerli ve güvenilir olması için yeterli değildir, bunun kadar önemli bir diğer konu da örneklemin yapısının ilgilenilen konu çerçevesinde toplumun yapısına uygun olmasıdır.

Örneklem süreci yöntem açısından iki temel gruba ayrılır

- Olasılıklı örnekleme yöntemleri
  - Basit Tesadüfi Örneklem Yöntemi
  - Sistematiik Örneklem Yöntemi
  - Tabakalı Örneklem Yöntemi
  - Küme Örneklem Yöntemi

- Olasılıksız örnekleme yöntemleri
  - Keyfi Örnekleme Yöntemi
  - Kota Örnekleme Yöntemi
  - Dilim Örnekleme Yöntemi
  - Kartopu Örnekleme Yöntemi

# Olasılıklı Örnekleme Yöntemleri

- Olasılıklı örnekleme yöntemlerinde örnekleme seçilecek örnek birimlerine eşit şans verilir.
- Örnek birimlerine eşit şans verilerek kitledeki değişkenliğin örneklemede korunması sağlanır. Böylece örneklemin kitleyi temsil yeteneği arttırılmış olur.
- Kitledeki her örnek birimine örnekleme seçilme yönünden eşit şans verebilmek için kitledeki birimler arasından *rasgele* seçim yapılır.
- Rasgeleliği sağlayabilmek için rasgele sayılar tablosu ya da rasgele sayı üreten bilgisayar yazalımlarından yararlanılır.

Örneklemede rasgelelik, kitledeki her deneğe örnekleme seçilme yönünden *eşit şans* verilmesidir.

Bu şansın eşitlenememesi durumunda; örneklemeden elde edilecek sonuçlardaki hatalar rasgele olmayacağı için sonuçlar *yanlı* olur.

Örneklemede yansız sonuçlar elde edebilmek için rasgelelik koşullarına uyulmalıdır.

# Basit Tesadüfi Örneklem Yöntemi

- Kitle geniş bir coğrafyaya yayılmamışsa homojenlik durumu söz konusu ise
- $N$  tane birim içerisinde  $n$  birim rasgele olarak seçilir.
- Burada her bir örneklem biriminin seçilme olasılığı eşittir.
- Bu yöntemin uygulanabilmesi için kitleye ilişkin çerçevenin bulunması gerekir.



# Basit Tesadüfi Örneklem Yöntemi

- iadeli(yerine koyarak)  $\rightarrow N^n \rightarrow \frac{1}{N^n} \rightarrow$  herhangi bir örnekleme seçme olasılığı
- iadesiz(yerine konmadan)  $\rightarrow \binom{N}{n} \rightarrow \frac{1}{\binom{N}{n}} \rightarrow$  herhangi bir örnekleme seçme olasılığı

**Örneklem Hatası**  $|\hat{\theta} - \theta|$ : istatistikle parametre arasındaki farka verilen isimdir.

$$|\hat{\theta} - \theta| \rightarrow |\hat{x} - \mu|$$

parametre varsa örneklem hatasına gerek yoktur.

# Basit Tesadüfi Örneklem Yöntemi

N büyüklüğündeki bir kitleden n büyüklüğünde bir örnekleme basit tesadüfi örnekleme yöntemiyle seçmek için;

- 1 Kitlede 1'den N'ye tüm denekler sıralanarak numaralandırılır.

1,2,3,4,5,6,7,8,9,10,.....,N-1,N

- 2 Gerekli örneklem genişliği(n) hesaplanır.
- 3 Rasgele sayılar tablosu, Excel ya da herhangi bir yazılım ile n sayı rasgele olarak seçilir.
- 4 Seçilen rasgele sayılara karşılık gelen denekler çalışmaya alınır.
- 5 Kitle parametreleri tahmin edilir.

# Basit Tesadüfi Örneklem Yöntemi

- 1 Kitle ortalamasının kestiriminde kullanılacak örneklem büyüklüğü formülü:

- Kitle büyüklüğü (N) bilinmiyorsa;

$$n = \frac{Z_{\alpha/2}^2 \times \sigma^2}{d^2}$$

- Kitle büyüklüğü (N) biliniyorsa;

$$n = \frac{N \times Z_{\alpha/2}^2 \times \sigma^2}{d^2 \times (N-1) + Z_{\alpha/2}^2 \times \sigma^2}$$

$Z_{\alpha/2}$  = iki yönlü hipoteze göre z tablo değeri

$\sigma^2$  = Kitle varyansı

d = Ortalama göre yapılmak istene  $\pm$  sapma

N = Kitle büyüklüğü

# Basit Tesadüfi Örneklem Yöntemi

- ② Kitle oranının kestiriminde kullanılacak örneklem büyüklüğü formülü:

- Kitle büyüklüğü (N) bilinmiyorsa;

$$n = \frac{Z_{\alpha/2}^2 \times P \times (1-P)}{d^2}$$

- Kitle büyüklüğü (N) biliniyorsa;

$$n = \frac{N \times Z_{\alpha/2}^2 \times P \times (1-P)}{d^2 \times (N-1) + Z_{\alpha/2}^2 \times P \times (1-P)}$$

$Z_{\alpha/2}$  = iki yönlü hipoteze göre z tablo değeri

$P$  = İncelenecek k olayın görülüş sıklığı

$d$  = İncelenecek k olayın görüş sıklığına göre yapılmak istenen  $\pm$  sapma

$N$  = Kitle büyüklüğü

Eğer kitle kendi içinde birbirinden farklı gruplara ayrılmışsa, her bir gruptan yeterli sayıda birimin seçilmesinde garanti altına almak için sistematiik örneklemeye kullanılır. Yani burada grupların ölçülen nicelik bakımından farklılık gösterdiği varsayılmaktadır.

# Sistematik Örnekleme Yöntemi

- Kitledeki denek sayısı  $N$ , örnekleme seçilecek denek sayısına  $n$ 'ye bölünür.

$$k = \frac{N}{n}$$

- $1-k$  arasında rasgele bir sayı seçilir( $R$ ).
- $R$ 'ye  $k$  eklenerek seçilecek diğer sıra numaraları belirlenir.

1 2 3 4 ...  $i$  ...  $k$  ...  $i+k$  ...  $i+2k$  ...  $i+3k$  ...  $N$

# Tabakalı Örneklem Yöntemi

Bu örneklem, toplanmak istenen bilginin doğruluğunu etkileyecek faktörler olduğunda, kitleyi bu faktör gruplarına göre tabakalara ayırarak her tabakadan ayrı ayrı örneklem seçerek yapılır.

Her tabakan ayrı örneklem seçerek, tabakların kitledeki değişkenliği örnekleme de korunarak örneklemin kitleyi temsil yeteneği artırılmış olur.

# Tabakalı Örneklem Yöntemi

Tabaka	Tabaka Büyüklüğü	Örneklem Büyüklüğü
1	$N_1$	$n_1$
2	$N_2$	$n_2$
...	...	...
k	$N_k$	$n_k$
Toplam	$N$	$n$



## örneklemeyi Paylaştırma Yöntemleri

- Eşit Paylaştırma

$$n_i = \frac{n}{L}$$

- Oranlı Paylaştırma

$$n_i = \frac{N_i}{N} \times n$$

- Neyman Paylaştırma

$$n_i = \frac{N_i \times \sigma_i}{\sum_j^L N_j \times \sigma_j} \times n$$

- En uygun Paylaştırma

$$n_i = \frac{N_i \times \sigma_i / \sqrt{C_i}}{\sum_j^L N_j \times \sigma_j / \sqrt{C_j}} \times n$$

- **Yararları**

- Eğer tabakalama iyi yapılmış ise daha doğru bilgi elde etme olanağı sağlar.
- Her tabakadan alınan örneklemin kendi tabakasını temsil yeteneği olduğundan her tabaka için ayrı sonuç elde etme olanağı sağlar

- **Eksikleri**

- Örneklem hatasını hesaplamak zordur.
- Eğer tabaka örneklem büyüklükleri küçük ise bilginin doğruluğu azalır.

# Tabakalı Örneklem Yöntemi

Tabakalı örneklemeden iyi sonuç alabilmek için,

- Tabakalar, kendi içinde homojen,
- Tabakalar, kendi aralarında heterojen olmalıdır.

- Kitledeki deneklerin listelenemediği, bu nedenle tek tek deneklere ulaşmanın olanaksız olduğu durumlarda kullanılan örnekleme yöntemidir.
- Bu yöntemde kitle, birbirine benzer deneklerden oluşan kümelere (denek grupları) ayrılır.
- Bu yöntemde, denek seçme yerine küme seçilerek örneklem oluşturulur.

# Küme Örnekleme Yöntemi

- Kümeler kendi aralarında benzer oldukları gibi her biri ayrı ayrı kitleyi temsil edebilir özellikle olmalı ya da kümeler bu özelliği taşıyacak biçimde oluşturulmalıdır.
- Küme örnekleme özellikle saha araştırmalarında deneklere(kişilere) ulaşmanın zor olduğu durumlarda kullanılır. Bu durumda sınıflar, köyler, sokaklar gibi deneklerin bir arada bulunduğu birimler küme olarak bellirlenir.
- Bu yöntemde, denek seçme yerine küme seçilerek örneklem oluşturulur.

Özellikle sosyal bilimlerde yapılan araştırmalarda, araştırmada yer alacak bazı birimlerle ilgili daha ayrıntılı bilgiye ihtiyaç duyulabilir. Bu durumda, olasılıklı olmayan örneklem yöntemleri kullanılarak, birimler örneğe daha kolay ve ucuz biçimde seçilebilir. Olasılıklı olmayan örneklem yöntemleri ile seçim işlemi bazı sübjektik kriterlere göre yapılır. Diğer ifadeyle, birimlerin örneğe seçimi, araştırmacının kişisel görüşüne bağlıdır. Bu nedenle olasılıklı örneklem yöntemlerinin maliyetleri daha düşüktür.

Olasılıklı olmayan örneklem yöntemlerinde elde edilen tahminlere ilişkin örneklem hataları ve yapılan tahminin hassaslığı ancak sübjektif yargılara göre yorumlanabilir. Bu nedenle özellikle tanımlayıcı istatistiklere ihtiyaç duyulduğunda ve elde edilen sonuçların yığına belli bir güvenirlikle genellemesine ihtiyaç duyulmadığı durumlarda, olasılıklı olmayan örneklem yöntemlerinin kullanılması tercih edilmelidir. Başlıca olasılıklı olmayan örneklem yöntemleri, keyfi, kota, dilim ve kartopu olmak üzere 4 tanedir.

**Keyfi Örneklem Yöntemi:** Aynı zamanda kolay ulaşılabilir örneklem olarak da adlandırılır. Özellikle bütçe ve zamanın çok kısıtlı olduğu araştırmalarda olasılıklı örneklem yöntemlerinin kullanılmasının mümkün olmadığı durumlarda kullanılır. Keyfi örneklem ile araştırmacı en kolay ulaşabileceği birimleri örneğe dahil eder.

**Kota Örneklem Yöntemi:** Evrenin belirli bir grubundan ya da alt grubundan belirli sayıda bireyin seçilmesidir.



**Dilim Örneklem Yöntemi:** Evren çok geniş, örneklem birimlerine ulaşmak çok yüksek maliyetliyse tercih edilebilir. Evren dilimlere ayrılır ve bir dilim seçilerek araştırmaya dahil edilecek birimler elde edilir. Evren homojen ise örneklemin evreni temsili söz konusu olabilir.

**Kartopu Örneklem Yöntemi:** Özellikle ilgilenilen birimlerin zor ulaşabilir veya gizlenmiş olduğu durumlarda kullanılması tercih edilen bir örneklem yöntemi.

İlk aşamada evrende yer alan bir birey rasgele olarak seçilir. Bu birey birinci birimdir. Bu bireye aynı evren tanımında yer alan bir tanıdığı sorulur. Var ise bu bireye ulaşılır. Bu süreç istenilen örneklem hacmine ulaşıncıya kadar sürdürülür.