

# İstatistik 2

## Bölüm 6

### Genel Tekrar 6

# Veri Toplama Yöntemleri

## ○ Sayımlar (Census)

- > Populasyonu oluşturan her bir üye ile ilgili bilgiler tek tek toplanır.
- > Oldukça maliyetli ve çok uzun zaman gerektirmektedir. Bu nedenle pratik değildir.

## ○ Saha Araştırmaları (Sample Survey)

- > populasyonu temsil ettiği düşünülen bir kesit üzerinde çalışılarak populasyon parametrelerinin tahmin edilmesi ve bunlara ilişkin çıkarımlar yapılır.

# Veri Toplama Yöntemleri

## ○ Deneysel Araştırmalar (Experiment)

- > Araştırmacıların neden-sonuç ilişkilerini anlamak için yaptıkları kontrollü çalışmalardır.
- > Araştırmacı hangi gruba kaç deneğin düşeceğini ve hangi grubun hangi tedaviyi alacağını kendisi belirlediğinden kontrollü denemeler adını alır.
- > Analiz aşamasında gruplarda "bağımlı değişken" karşılaştırılır. Araştırmadan elde edilen sonuca göre girişimin (bağımsız değişken) bağımlı değişkeni etkileyip etkilemediği belirlenir.

# Veri Toplama Yöntemleri

## ● Gözlemsel çalışma (Observational study)

- > Deneysel çalışmalara benzer olarak gözlemsel çalışmalarda da neden-sonuç ilişkisi incelenir.
- > Fakat arařtırmacı hangi deneklerin hangi gruba atanacağını ve hangi grubun hangi tedaviyi alacağını belirleyemez.
- > Örneğin saha arařtırmalarında deneklere bir girişimde bulunulmaz bu nedenle saha arařtırmaları gözlemsel çalışmalara örnektir.

# Veri Toplama Yöntemlerinin Avantaj ve Dezavantajları

## ● Kaynaklar

- > Büyük örneklerde örneklem seçilmesi yoluyla yapılan çalışmalar sayımlara göre avantajlıdır. İyi seçilmiş bir örneklem yoluyla çok iyi parametre tahminleri yapılabilir, bunun yanında çok az bir güç kaybıyla araştırma daha çabuk ve ucuza mal edilir.

## ⦿ Genellemeneme

- > Örneklemeden elde edilen sonuçların populusyona genellenmesidir.
- > Genellemeneme yapılabilmesi için örnekleme ait bireylerin temsil ettiđi populusyondan rasgele seçilmiř olması gerekir.

## ⦿ Çıkarım

- ⦿ Neden-sonuç ilişkisi de denekler gruplara rasgele atıldığı durumda incelenebilir.
  - > “Deneysel çalışmalar” araştırmacılara deneklerin gruplara atanması imkanını verdiğiinden neden-sonuç ilişkisi incelemede uygun dizaynlardır.

# Temel Örnekleme Yöntemleri ve Örneklem Büyüklüğü Seçiminin Önemi

Ankara Üniversitesi, SBF İstatistik 2 Ders Notları Prof. Dr. Onur Özsoy



# populasyon

## ■ Tanım

Belirli bir özelliği taşıyan birimlerin tümünün oluşturduğu topluluk olarak tanımlanır.

## ■ Örnek

Türkiye 'de yapılacak olan seçim sonuçları hakkında fikir sahibi olmak için yapılacak çalışmada populasyon, oy verme hakkına sahip tüm yetişkinlerden oluşmaktadır.

# ÖRNEKLEM

## ■ Tanım

populasyonda ilgilenilen herhangi bir olayı incelemek için populasyondan seçilen ve populasyonu temsil ettiği varsayılan az sayıda birim içeren topluluğa ÖRNEKLEM denir.

## ■ Örnek

Seçmenlerin genel eğiliminin belirlenmesi için populasyondan seçilen ve görüşleri alınan bir milyon Türk, çalışma için seçilen örnektir.

# ÖRNEKLEM ÇERÇEVESİ

## ■ Tanım

Örnekleme seçilecek birimlerin tanımlanmasına ÖRNEKLEM ÇERÇEVESİ denir.

## ■ Örnek

Oy verme yetkisine sahip bütün bireylerin listesi örneklem çerçevesini oluşturur.

# ÖRNEKLEME BİRİMİ

## ■ Tanım

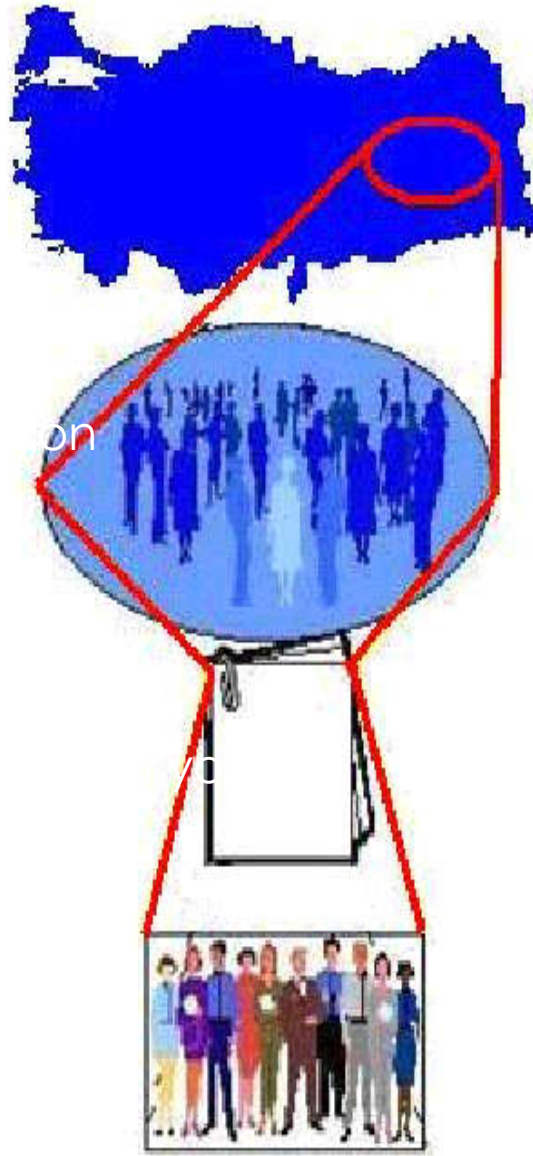
Örneklem çerçevesinde listelenen ve örnekleme seçilecek birey, aile, materyal ve benzer birimlere ÖRNEKLEME BİRİMİ denir.

## ■ Örnek

Türkiye’de oy verme hakkına sahip birey örnekleme birimidir.

# ÖRNEKLEME

Bireylerin populasýondan seçilme işlemlerine ÖRNEKLEME denir.



# ÖRNEKLEMENİN AMACI

- Örnekleme, bütünün sayılabilen ve/veya sayılamayan özellikleri hakkında istatistiksel bilgiye, az miktarda seçilen kısımdan ulaşmak için kullanılır.
- Örneklemeyle yapılan bir araştırmanın başarısı popülasyondaki gerçek durumu yansıtma derecesiyle ölçülür.

## ***Arařtırmanın başarılı olabilmesi;***

- Seçilen örneklemin populasyonu en iyi biçimde temsil etmesine,
  - Tüm deneklere ulaşılabilmesine,
  - Yanıt almada başarılı olunmasına,
  - Veri toplarken tarafsız olunmasına,
  - Hatasız tam doğru ve yararlı veri derlenmesine,
  - Deęerlendirmenin ve sonuçların yorumlanmasının doğru biçimde yapılmasına
- baęlıdır.



# ÖRNEKLEMENİN AVANTAJLARI

- Bilgiyi örnekten elde etmek, bütün yığından elde etmekten daha ucuzdur.
- Bilgi örneklemeyle daha çabuk toplanır.
- Örneklemeyle daha ayrıntılı veriler elde edilebilir.
- Başka bir şekilde elde edilemeyecek veriler elde edilebilir.

# ÖRNEKLEM BÜYÜKLÜĞÜ

Örnekleme ile yapılan arařtırmalarda temel hedef, eęer bilinmeyen populasyon deęerlerini tahmin etmek istiyorsak bu deęerleri gerçeęe en yakın biçimde tahmin edebilmek, eęer grupları karşılařtırıyorsak güvenilir ve geçerli biçimde karara varabilmektir. Bu hedefe ulařabilmek için yeterli örneklem büyüklüęü önemli bir öğedir.

## **Örnekleme ve araştırmanın diğer aşamalarıyla ilgili önemli noktalar:**

- Uygun örnekleme yönteminin kullanılması
- Deneklerin yansız ve eşit şansa seçilmesi
- Örnekleme alınan tüm deneklere ulaşılması
- Doğru araştırma deseni seçilmesi
- Doğru ölçüm
- Doğru veri derleme
- Doğru analiz
- Doğru yorum

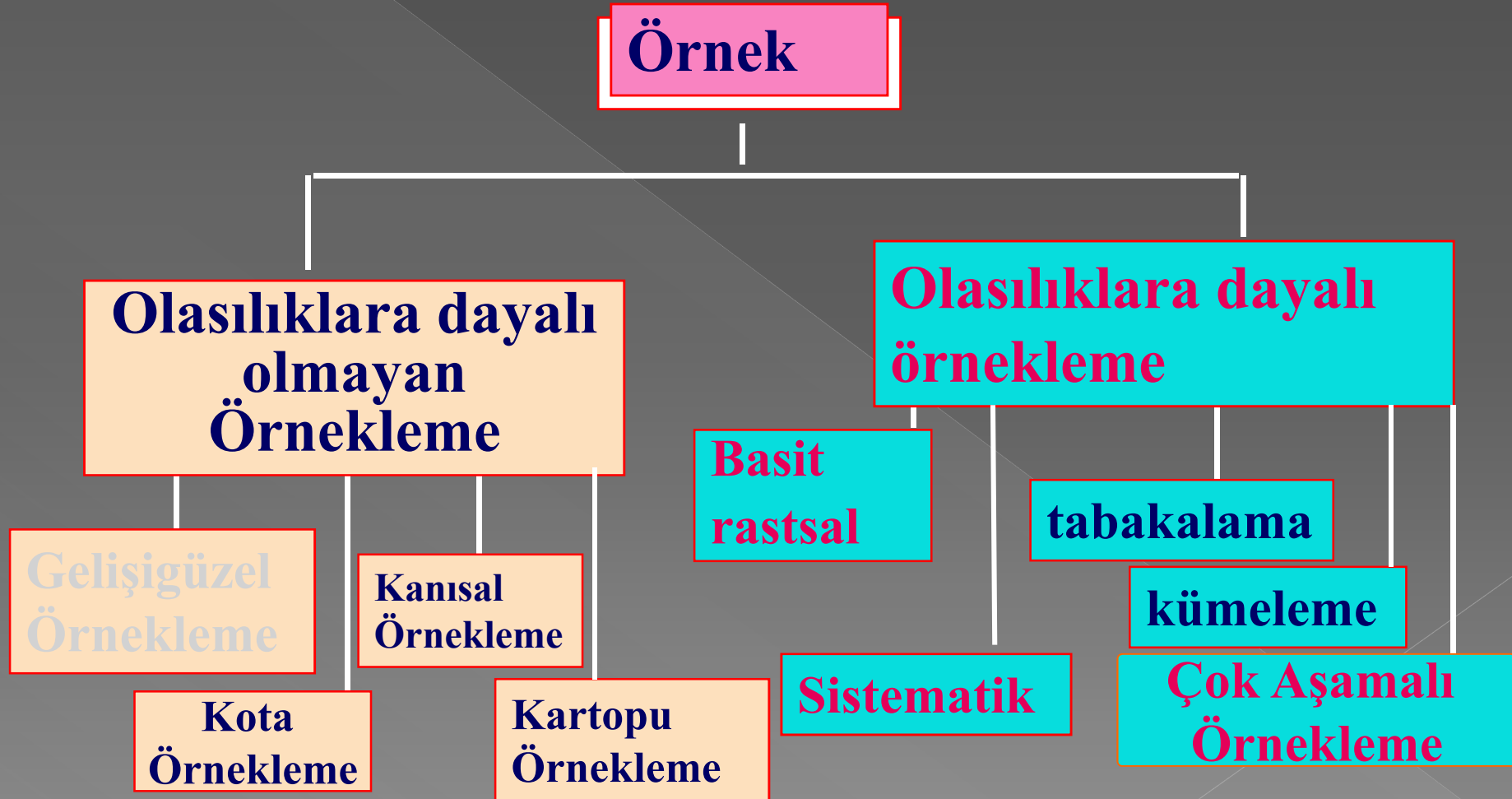
# Yeterli örneklem büyüklüğü neden önemlidir?

- **Bilimsel nedenler:** tahmin edilen populasyon parametresinin güven aralığı ne kadar dar olursa gerçek populasyon değerinin o denli isabetli olarak tahmin edildiği kabul edilir.
- **Ekonomik nedenler:** Yetersiz büyüklükte yada gereğinden daha büyük bir örneklem kaynakların ziyan olmasına neden olur.
- **Etik nedenler:** Yetersiz büyüklükte bir örneklem yanlış sonuçlar elde edilmesine buda toplumsal maliyete sebep olabilir.

# Örneklem büyüklüğünün saptanması için gerekli bilgiler

- Araştırmanın amacı
- populasyon parametresini tahmin etmek
- Hipotez testi yapmak
- Yanıt değişkeni
- Araştırma deseni
- Araştırma denekleri ile ilgili bilgiler
- Önemlilik düzeyi
- Testin gücü
- Kullanılacak istatistiksel yöntemler
- Olayın görüş sıklığı yada standart sapması
- Yapılacak sapma miktarı
- Kullanılacak örnekleme yöntemi
- Örnekleme birimi
- Kontrol altında tutulması gereken değişkenler
- Maliyet, zaman, uygulanabilirlik

# Örnekleme Yöntemleri



# Olasılıklı Örnekleme

Olasılıklı örnekleme yöntemleri popülasyondaki bireylerin bilinen ve eşit olasılıkla örnekleme seçilmelerini sağlayan yöntemlerdir.

Her araştırmada öncelik olasılıklı örnekleme yöntemlerine verilmelidir.

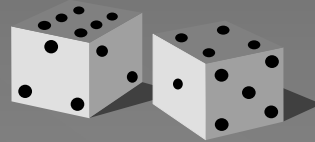
# BASİT rastsal ÖRNEKLEME

- Basit rastsal örnekleme, populusyona ait her bir üyenin oluşturulacak kesit örnek içinde bulunma şansının eşit olmasıdır
- 1. Örnekte yer alan her bir üye aynı populusyondan gelmeli,
- 2. Örnekte yer alan her bir üye aynı populusyondan bağımsız bir şekilde seçilmelidir (seçimler rastsal olarak ve yerine koymadan yapılmalıdır).



# Basit rastsal Örnekleme

- Populasyon üyelerinin her birinin örnek içinde yer alma şanslarının eşit olması durumudur.
- Seçim yerine koyma veya koymadan yapılabilir.
- rastsal sayılar tablosu kullanılabilir.



# Basit rastsal Örnekleme

Avantajları:

- Uygulanması kolaydır,
- Seçim olasılıklı olduğundan her bireye eşit olasılıkla seçilme şansı verir,
- populasyon çok büyük değilse seçim işlemleri kolaydır,
- İstatistiksel değerlendirme ağırlıksız yapıldığından işlemler ve sonuçların yorumu kolaydır

# Basit rastasal Örnekleme

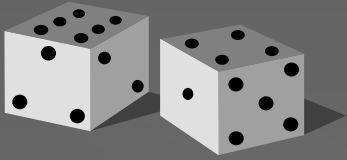
Dezavantajları:

- populyasyondaki bireyleri listelemek gerekir.
- populyasyondaki birey sayısı çok olduğunda ve bireylerin tümüne ulaşılması mümkün olmadığında kullanılması uygun değildir.
- populyasyondaki bireyleri belirli bazı özelliklere göre sınıflayamaz.
- populyasyon heterojen yapı gösteriyorsa varyans büyük olacaktır.

# Sistemantik Örneklenme

- Örnekte yer alacak gözlem sayısı belirlenir:  $n$
- Populasyon üye sayısı  $k$  sayıda gruba ayrılır.

$$k = N/n$$



- 1. Gruptan rastsal olarak bir seçim yapılır (diyelim birinci gruptan 3. Üye seçildi)

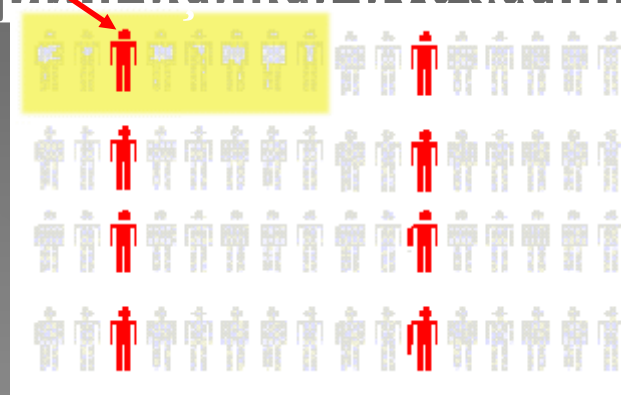
- daha sonra her bir gruptan üçüncü üye secilir

$$N = 64$$

$$n = 8$$

$$k = 8$$

1. grup



# Sistematik Örnekleme

## Avantajları:

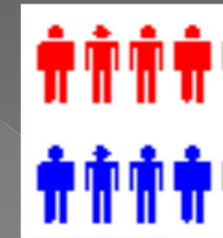
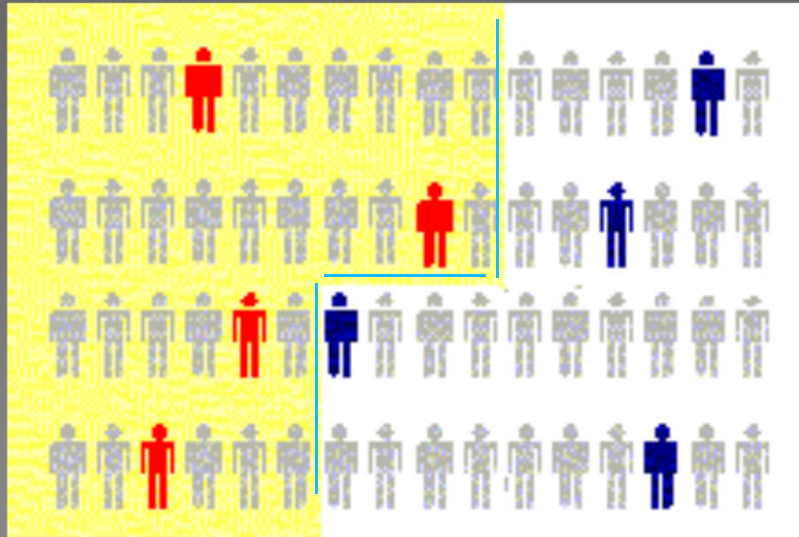
- Seçim olasılıklı olduğu için her bireye eşit olasılıkla seçilme şansı verir.
- popülasyon çok büyük olsa da seçim işlemleri kolaydır.
- İstatistiksel değerlendirme ağırlıksız yapıldığından işlemler ve sonuçların yorumu kolaydır.

## Dezavantajları:

- Örnekleme seçimini yapabilmek için popülasyondaki bireylerin listesi ve ulaşılabilir olmaları gereklidir.
- popülasyondaki bireyleri belirli bazı özelliklere göre sınıflayamaz.

# Tabakalama Yöntemi ile Örnekleme

- Populasyon, üyelerin ortak özelliklerine bağlı olarak 2 veya daha fazla sayıda gruba ayrılır. Bu gruplar bir biri ile çakışmamalı
- Her bir gruptan rastsal örnekleme ile örnekler oluşturulur
- İki veya daha fazla örnek birleştirilir.



# Tabakalama Yöntemi ile Örnekleme

Avantajları:

populasyonda tabakalara bağlı varyasyon söz konusu ise popülasyonu iyi derecede temsil eder ve popülasyon değerlerini gerçeğe yakın olarak tahmin eder.

Dezavantajları:

Tabakalardaki birey sayıları bilinmediğinde ve bireylere ulaşmak mümkün olmadığında seçim işlemleri güçleşir.

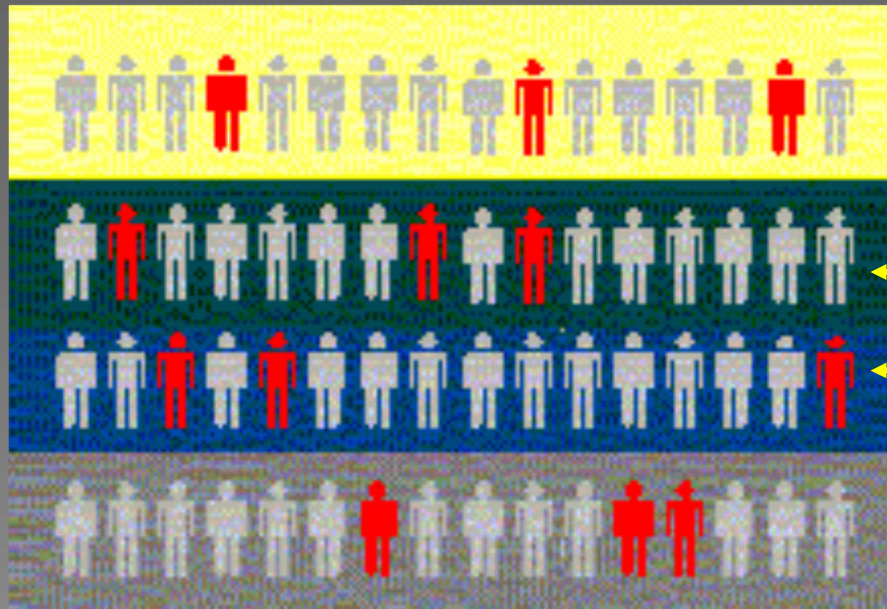
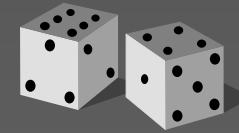
# Kümeleme Yöntemi ile Örnekleme

Basit Rastgele, Sistemantik, Tabakalama yöntemlerinin kullanılabilmesi için araştırma konusuna giren bireylere tek tek ulaşmak gerekir. Kümeleme yönteminde ise örnekleme birimi birey yada aile değildir. Örnekleme birimi bir grup, bir demet yada bir kümedir. Küme birimi olarak sokak, semt, mahalle yada coğrafi bölgeler kullanılabilir.



# Kümeleme Yöntemi ile Örnekleme

- Populasyon çok sayıda kümeye bölünür,  
her bir küme populasyonu temsil eder
- basit örnekleme yapılır
- örnekler birleştirilir.



Populasyon 4  
kümeye  
bölünmüştür

# Kümeleme Yöntemi ile Örneklem

## Avantajları:

- Diğer yöntemlere kıyasla daha az kaynağa ihtiyaç duyulur.
- Örneklem çerçevesi eksik olduğunda da uygulanabilir.

## Dezavantajları:

- Kümeler, popülasyonu iyi şekilde temsil etmeyebilir.
- Uygulanan analiz yöntemleri Basit Rasgele Örneklemeden daha karışıktır.

# ÇOK AŞAMALI ÖRNEKLEME

Bazı durumlarda bir araştırmada birden çok sayıda örnekleme yöntemi kullanılabilir yada örnekleme seçimi birden çok aşamada yapılabilir.

**ÖRNEK:** Çocukların aşılama oranını incelemek için yapılacak bir çalışmada önce bölge, sonra il örneklemesi, sonra ilçe, köy yada semt örneklemesi ve en sonunda da aile örneklemesi yapılabilir.

# ÇOK AŞAMALI ÖRNEKLEME

- Çok aşamalı örnekleme yönteminde her aşamada aynı örnekleme yöntemi kullanılabilceği gibi, farklı örnekleme yöntemleri de kullanılabilir. Her aşamada farklı bir örnekleme çerçevesi oluşturulur.

# OLASILIKLI OLMAYAN ÖRNEKLEME YÖNTEMLERİ

Bireylerin populasasyondan bilinen belirli bir olasılıkla seçilmediği yöntemlere “Olasılıklı Olmayan Örneklem Yöntemleri” adı verilir.

# GELİŞİGÜZEL ÖRNEKLEME

Herhangi bir konuda popülasyondaki bireylerin fikrini öğrenmek yada bir uygulama konusunda görüşlerini almak üzere araştıracının o anda önüne çıkan bireylerle görüşmesidir.

ÖRNEK: Cebeci'de caddede bekleyerek rastgele seçilecek insanlara herhangi bir ürüne ilişkin görüşlerini sorarak bilgi toplamak.

# KANISAL ÖRNEKLEME

Hangi grubun yada örnekleme birimlerinin popülasyonu daha iyi temsil edeceğine ve örnekleme alınması gerekeceğine araştırmacı kişisel kanısını kullanarak karar verir.

Bu nedenle araştırma sonuçları yanlı olabilir.

# KOTA ÖRNEKLEME

populasyonun belirli bir grubundan yada alt grubundan belirli sayıda bireyin seçilmesidir. Bu yöntemle seçilen örneklem de populasyonu iyi şekilde temsil edemeyebilir.



# KARTOPU ÖRNEKLEME

Çalışmaya alınma kriterlerini sağlayan bir kişi belirlendikten sonra, kriterleri sağlayan başka bireylere ulaşabilmek için o kişiden bilgi alınır.

Kartopu örnekleme yöntemi, popülasyondaki bireylere ulaşmak çok zor yada imkansız olduğu durumda kullanışlıdır.

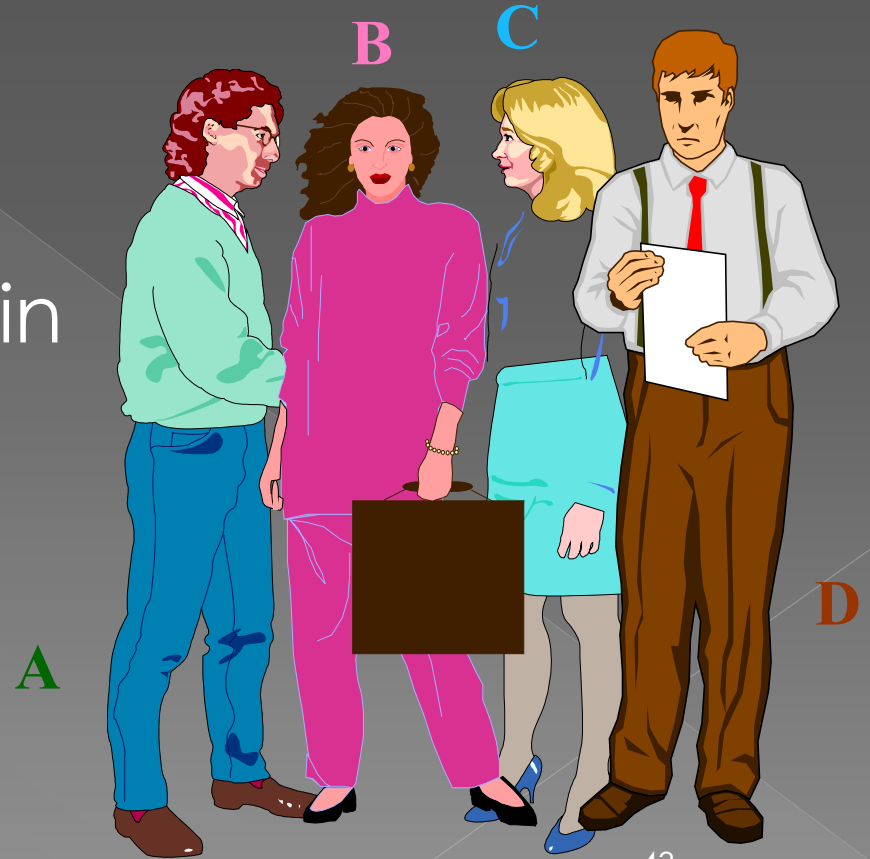
ÖRNEK:Evsizlerle ilgili bir çalışmada, bölgedeki tüm evsizleri listelemek mümkün değildir. Bir yada birkaç evsizi ulaşılarak, çalışmaya alınabilecek diğer evsizler belirlenebilir.

# Örnekleme dağılımı

- Örnek istatistikleri populasyon parametrelerinin tahmini için kullanılır.
- Problem: Farklı örnekler farklı tahmine neden olur
  - > Örnek büyüdükçe yapılan tahmin daha etkin olur Ancak büyük örneğin maliyeti yüksektir.

# Örnekleme Dağılımının Oluşturulması

- ◉ Elimizde bir populasyon olsun
- ◉  $N=4$
- ◉ rastsal değişken,  $X$ , populasyon üyelerinin yaşları olsun
- ◉  $X: 18, 20, 22, 24$

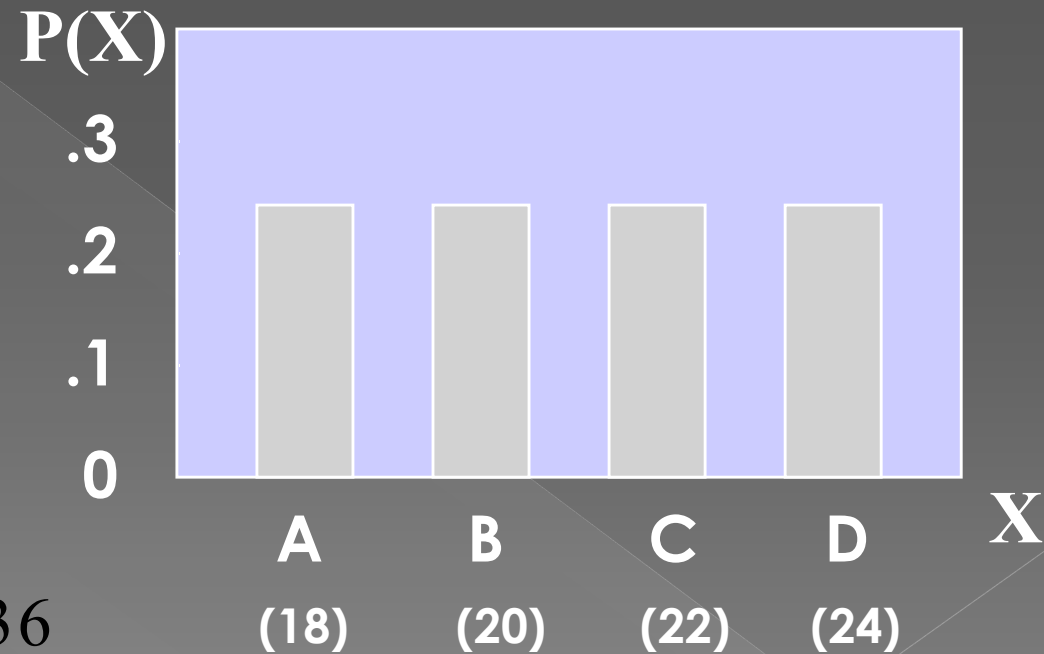


# Örnekleme Dağılımının Oluşturulması

Populasyona ait üyeler için tanımsal istatistikler

$$\begin{aligned}\mu &= \frac{\sum_{i=1}^N X_i}{N} \\ &= \frac{18 + 20 + 22 + 24}{4} = 21\end{aligned}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}} = 2.236$$



Düzensün dağılım

# Örnekleme Dağılımının Oluşturulması

$n=2$  üyeden oluşan olası bütün

örnekler

1. Göz.	2. Gözlem			
Göz.	18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 örnek ortalaması



1. Göz.	2. Gözlem			
Göz.	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Yerine koyarak yapılan örnekleme

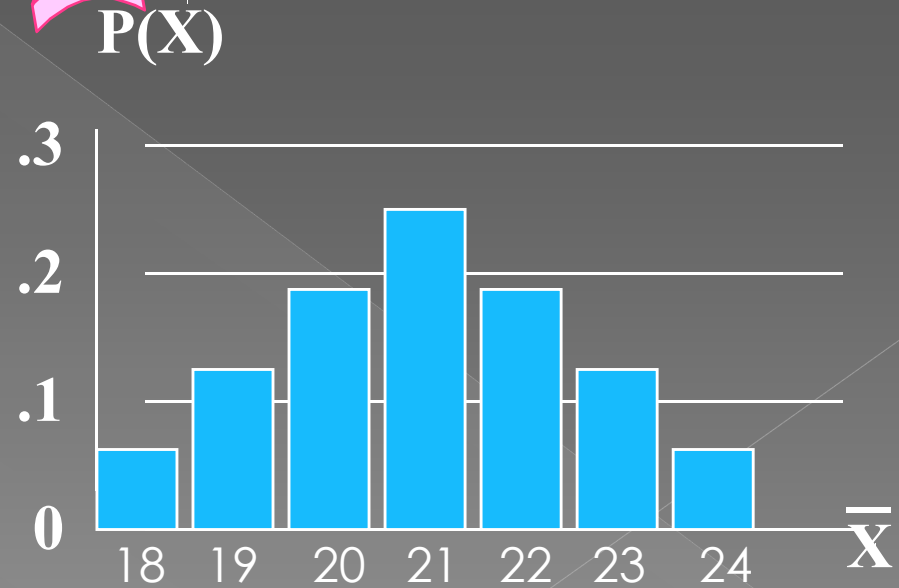
# Örnekleme Dağılımının Oluşturulması

## Örneklerin örnekleme dağılımları

16 örnek ortalaması

Örnek ortalamalarının dağılımı

1. Göz.	2. Gözlem			
Göz.	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24



# Örnekleme Dağılımının Oluşturulması

Örnekleme dağılımı için tanımsal istatistikler

$$\mu_{\bar{X}} = \frac{\sum_{i=1}^N \bar{X}_i}{N} = \frac{18+19+19+L+24}{16} = 21$$

$$\sigma_{\bar{X}} = \sqrt{\frac{\sum_{i=1}^N (\bar{X}_i - \mu_{\bar{X}})^2}{N}}$$
$$= \sqrt{\frac{(18-21)^2 + (19-21)^2 + L + (24-21)^2}{16}} = 1.58$$

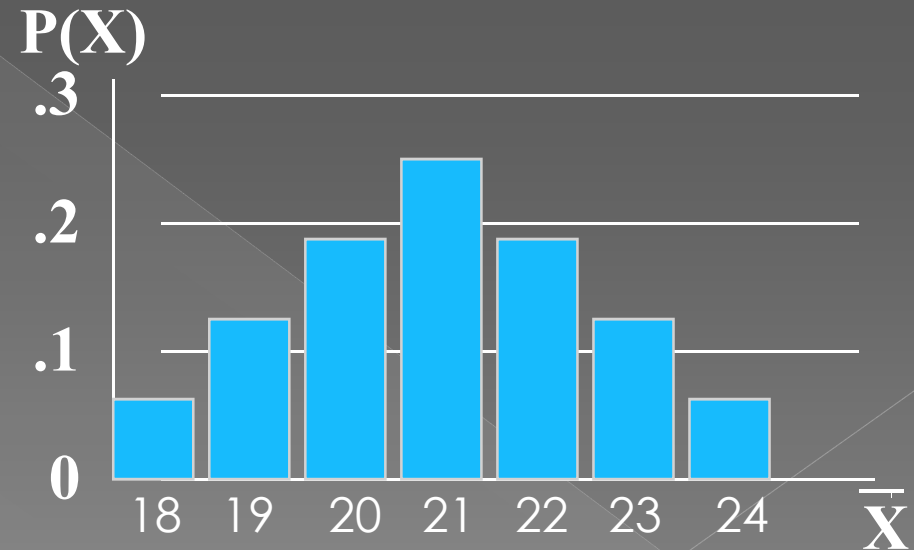
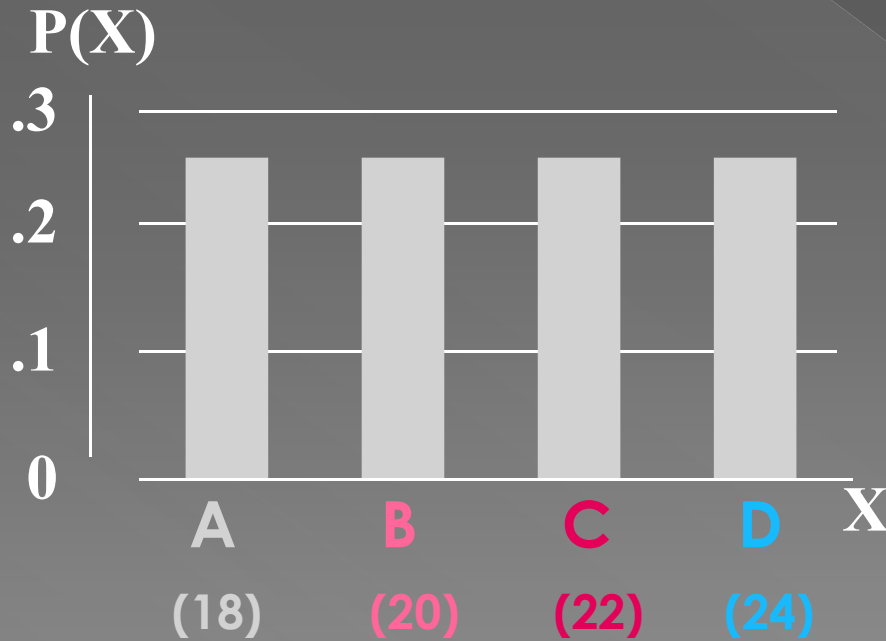
# Örnekleme Dağılımının Oluşturulması

Populasyon  
 $N = 4$

$$\mu = 21 \quad \sigma = 2.236$$

Örnekleme dağılımı  
 $n = 2$

$$\mu_{\bar{X}} = 21 \quad \sigma_{\bar{X}} = 1.58$$





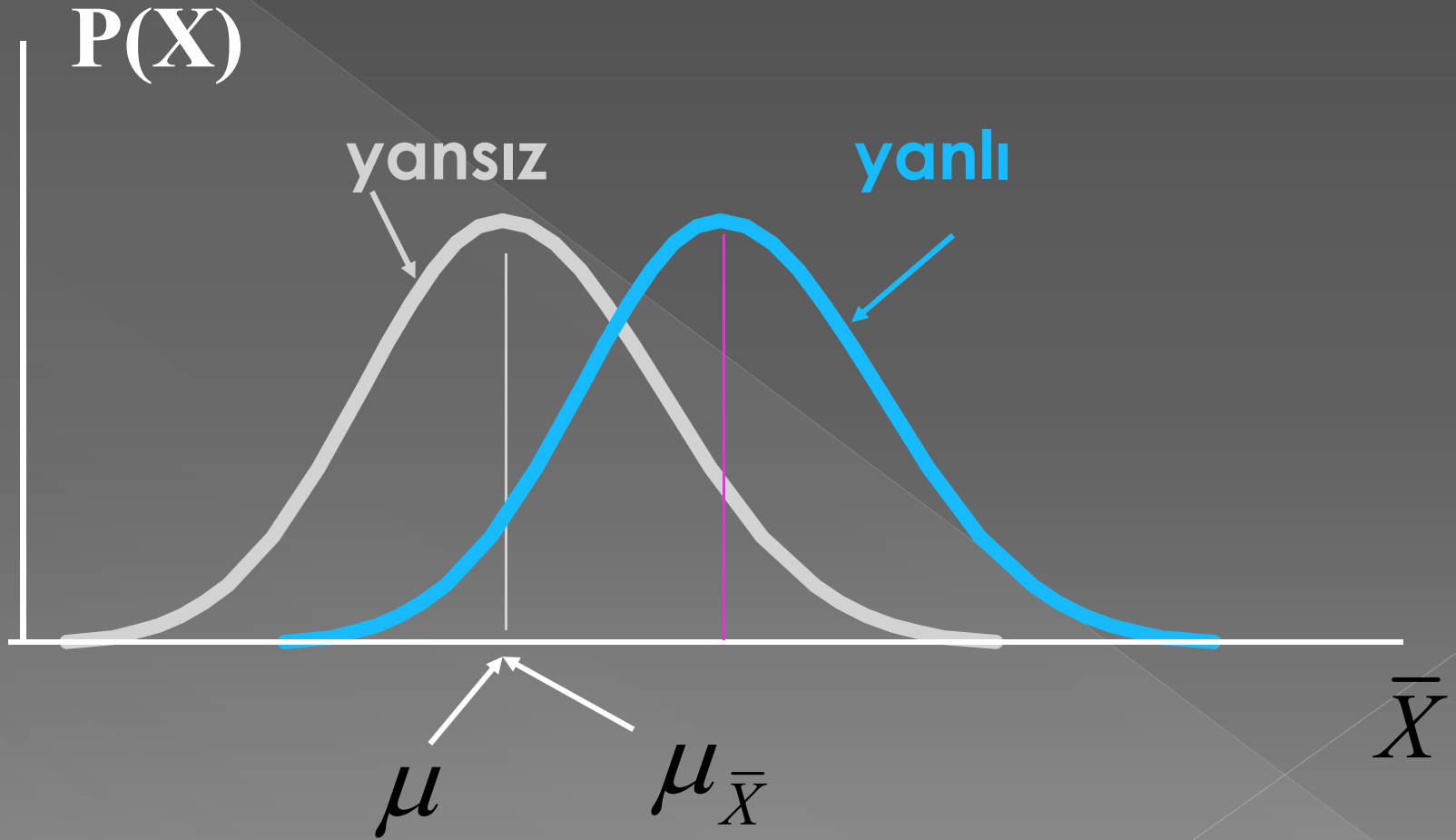
# Sayısal Özetlerin Özellikleri

- ◉  $\mu_{\bar{X}} = \mu$ 
  - > bu  $\bar{X}$  'nin tarafsız tahmin edici olduğunu gösterir

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

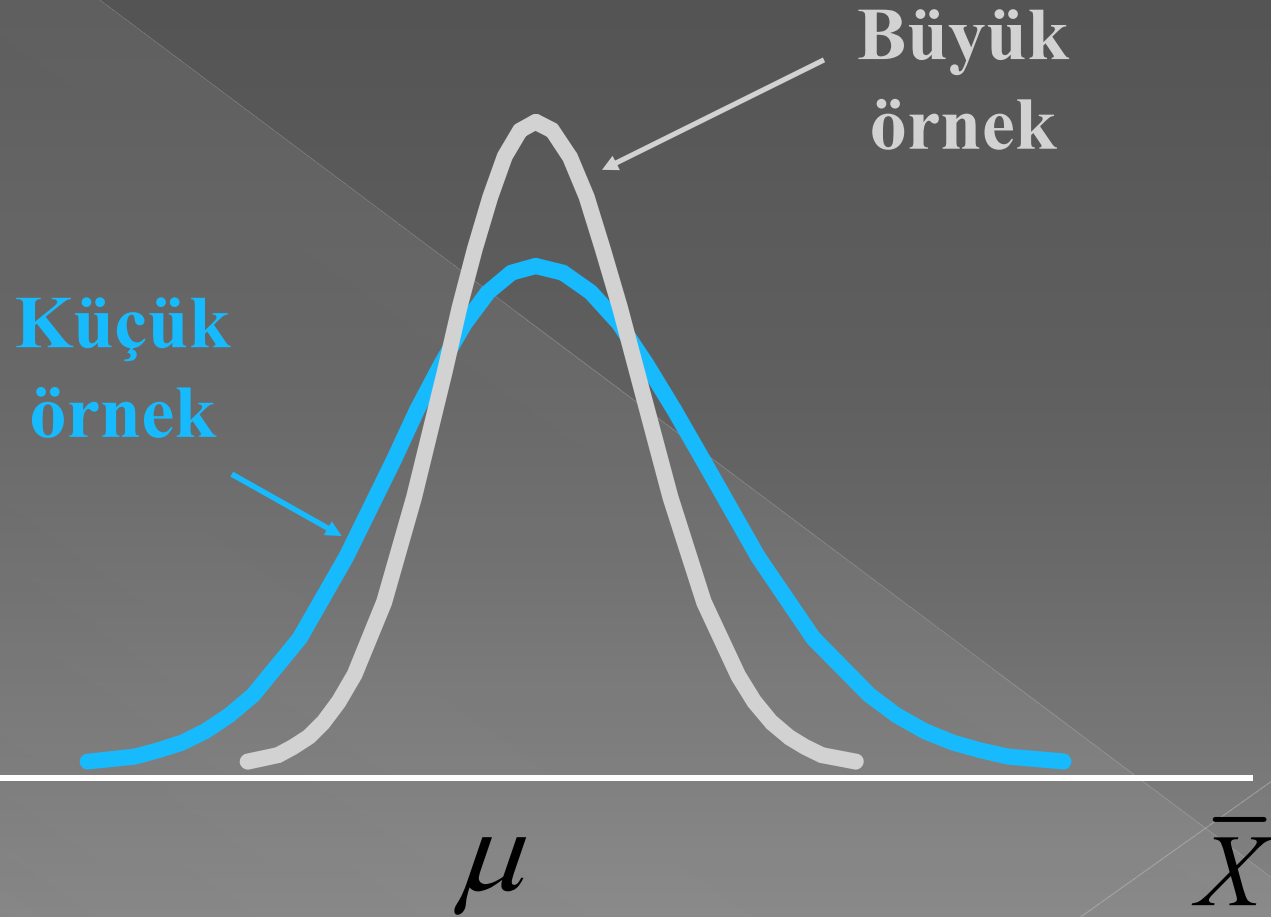
- ◉ Yerine koyarak yapılan örnekleme:
  - >  $n$  arttıkça,  $\sigma_{\bar{X}}$  azalır

# Yansızlık



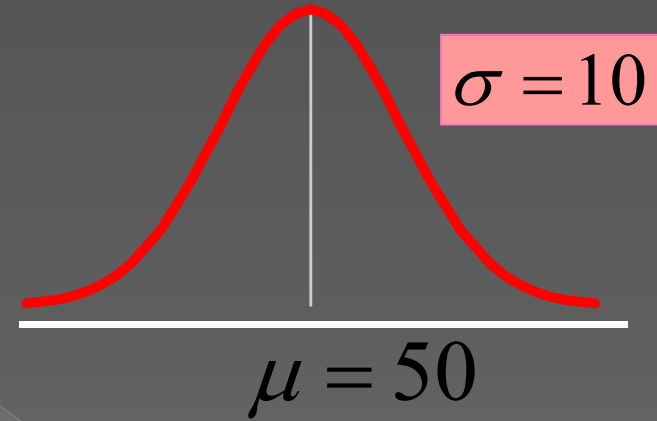
# Büyük ve Küçük Örnek

$P(X)$



# Populasyonun Normal Dağılmış Olması

Normal Populasyon



Merkezi Eğilim

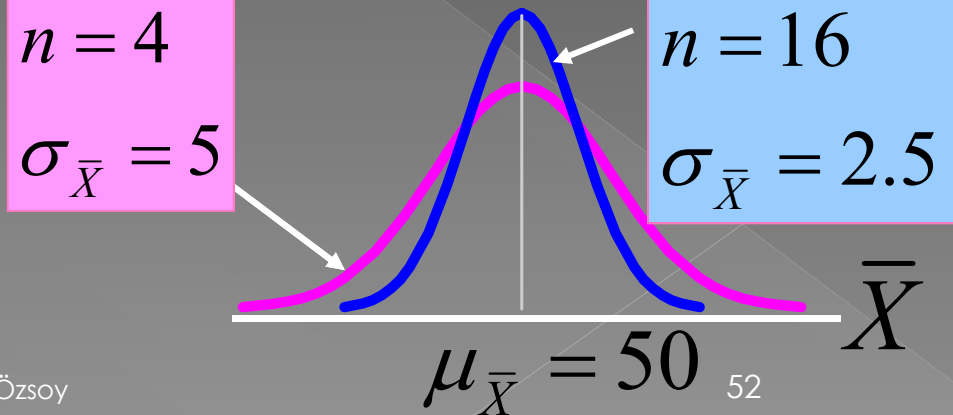
$$\mu_{\bar{X}} = \mu$$

yayılım

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Yerine koyarak  
örnekleme

Örnek dağılımı



# Populasyonun Normal Dağılmamış Olması

Merkezi Eğilim

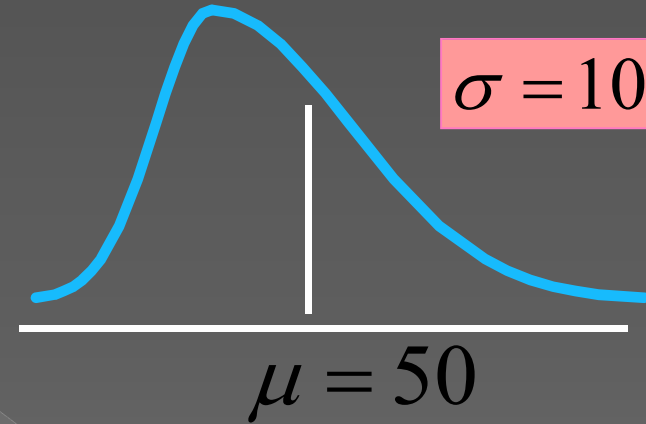
$$\mu_{\bar{X}} = \mu$$

yayılım

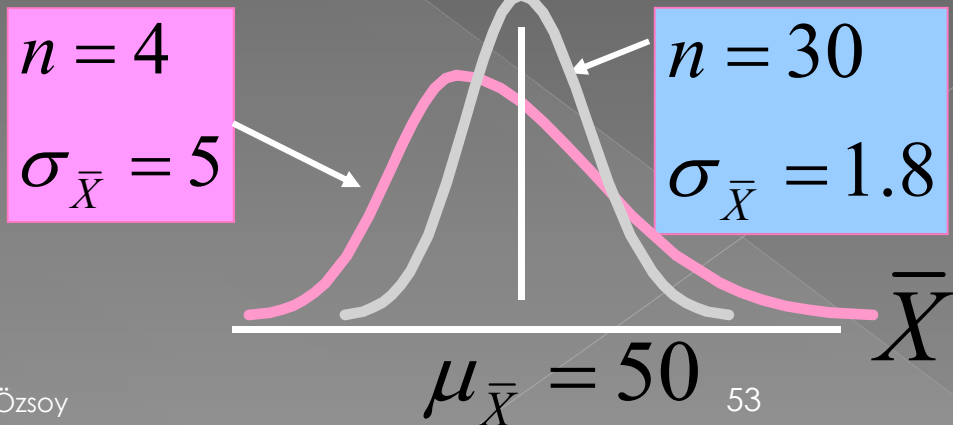
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Yerine koyarak  
örnekleme

Normal Populasyon

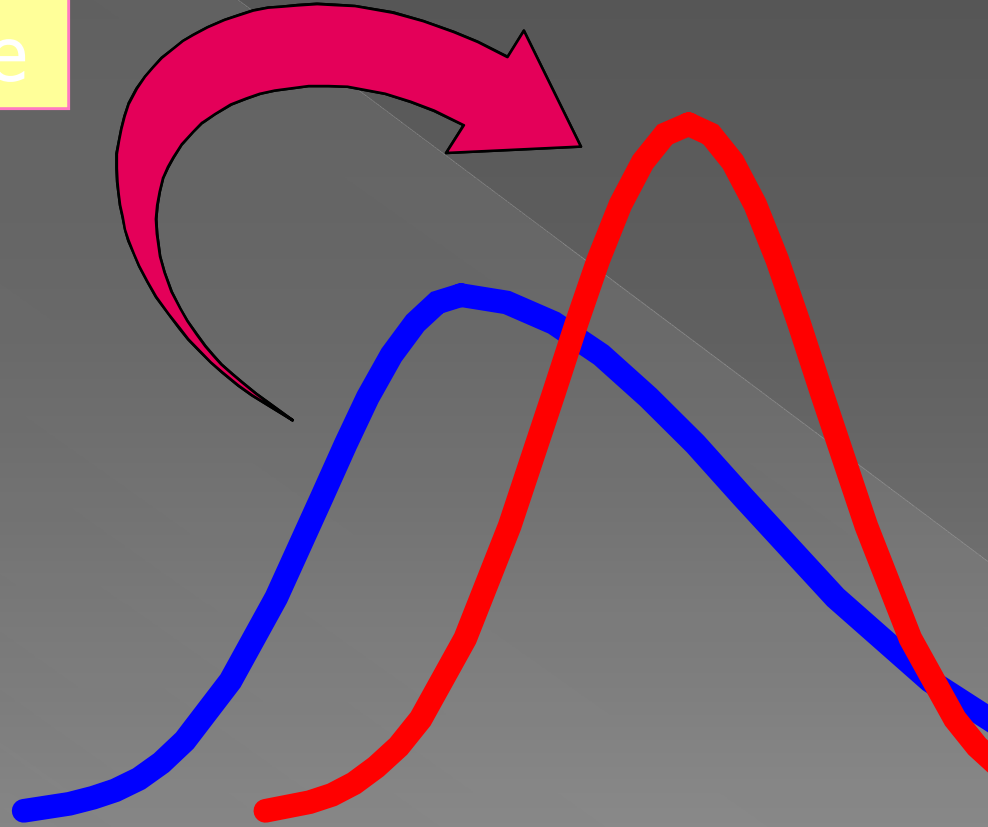


Örnek dağılımı



# Merkezi Limit Teoremi (MLT)

Örnek  
büyüdükçe



Ana  
populasyon  
un şekline  
bağlı  
kalmaksızın  
örnekleme  
dağılımının  
şekli  
normal  
dağılıma  
yaklaşır

$\bar{X}$

# Örnek Ne kadar Büyük Olmalı?

- Bir çok dağılım için,  $n > 30$
- Çok iyi simetrik dağılımlar için,  $n > 15$
- Normal popülasyona ait örnekleme dağılımı normal dağılıma sahiptir.

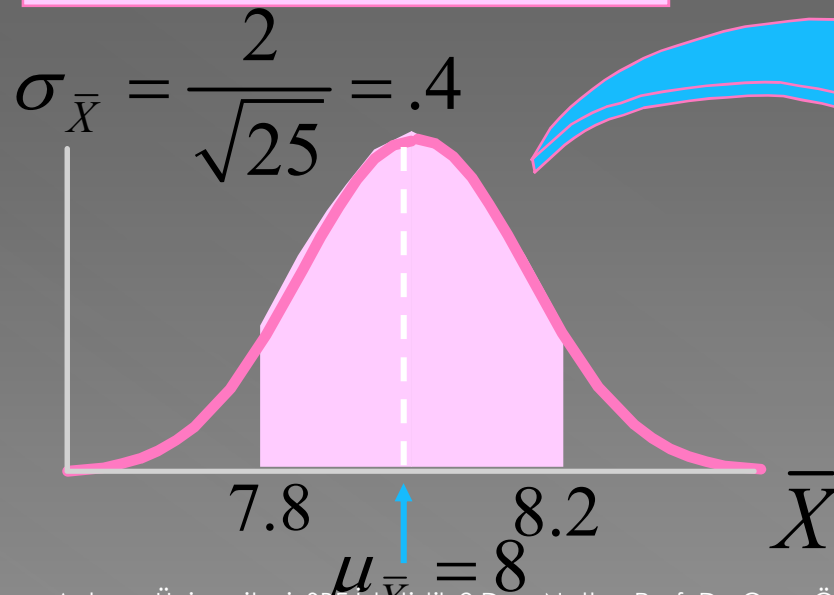
Örnek:

$$\mu = 8 \quad \sigma = 2 \quad n = 25$$

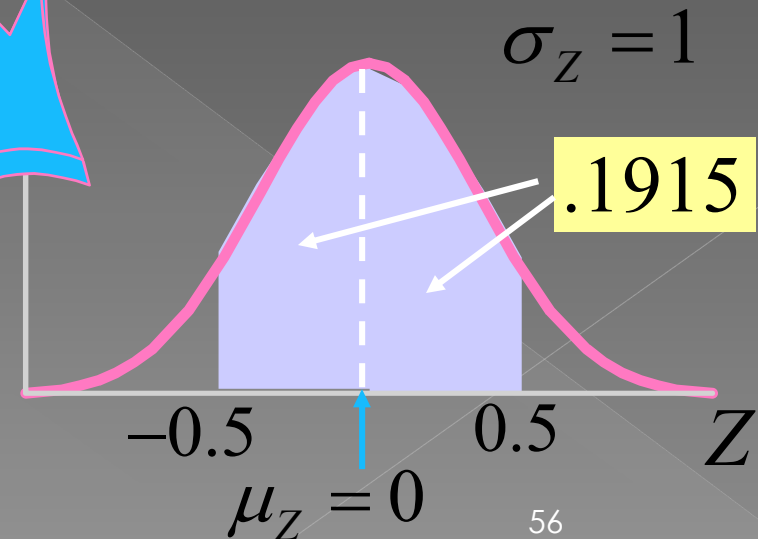
$$P(7.8 < \bar{X} < 8.2) = ?$$

$$P(7.8 < \bar{X} < 8.2) = P\left(\frac{7.8 - 8}{2/\sqrt{25}} < \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < \frac{8.2 - 8}{2/\sqrt{25}}\right)$$
$$= P(-.5 < Z < .5) = .3830$$

Örnekleme dağılımı



Standart normal dağılım





# Örnek Oranı İçin Örnekleme Dağılımı

- Kategorik değişken
  - > Örnek: cinsiyet, eğitim durumu gibi
- Populasyon oranı ( $p$ ) ile gösterilir
- Örnek oranı tahmin için kullanılır

$$\bar{p} = \frac{X}{n} = \frac{\text{başarılı sonuç sayısı}}{\text{örnek büyüklüğü}}$$

- Olası iki sonuç olması durumunda  $X$  binom olasılık dağılımına sahiptir.

# Örnekleme ve Örnekleme Dağılımı

- Normal dağılıma sahiptir

$$np \geq 5$$

$$n(1-p) \geq 5$$

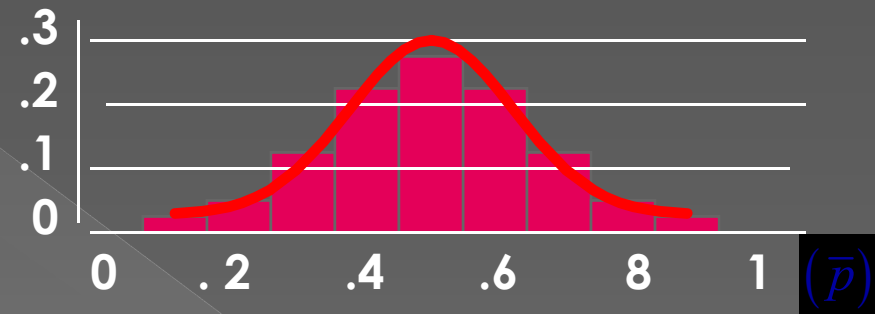
Ortalama:

$$\mu_{\bar{p}} = p$$

> standart sapma:

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Örnekleme dağılımı

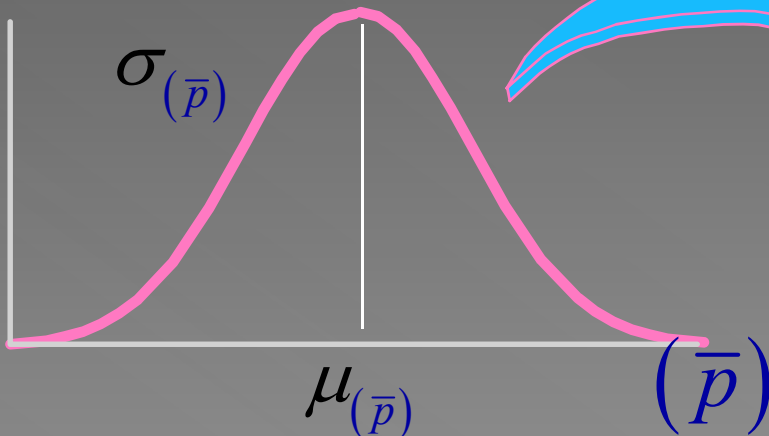


$p = \text{populasyon oranıdır}$

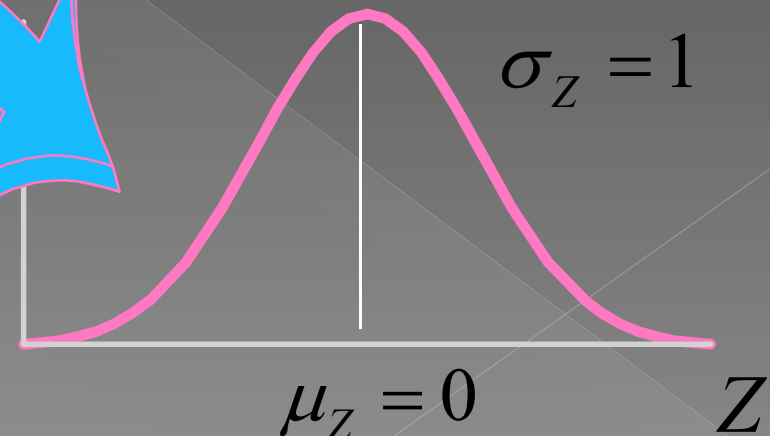
# Örnek oranının Standart Normal Dağılıma Dönüştürülmesi

$$Z \cong \frac{p_{(\bar{p})} - \mu_{(\bar{p})}}{\sigma_{(\bar{p})}} = \frac{p_{(\bar{p})} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Örnek dağılımı



Standart normal dağılım

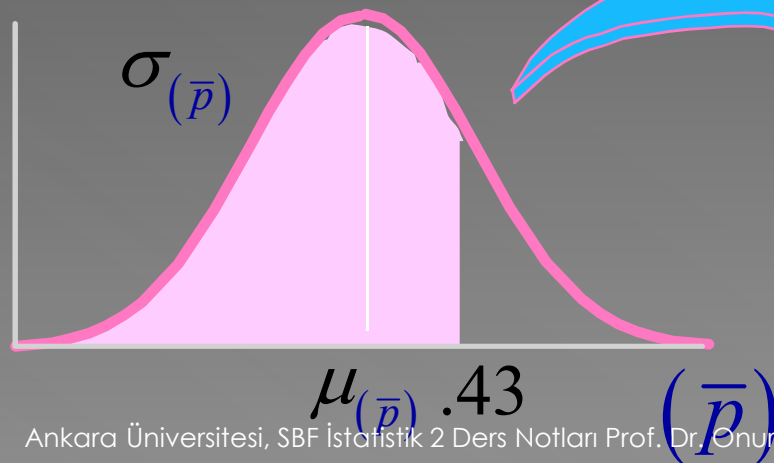


# Örnek:

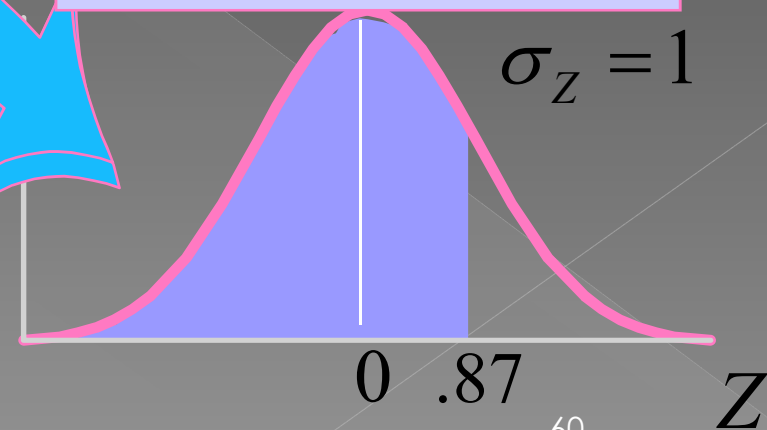
$$n = 200 \quad p = .4 \quad P(\bar{p} < .43) = ?$$

$$P(\bar{p} < .43) = P\left(\frac{(\bar{p}) - \mu_{(\bar{p})}}{\sigma_{(\bar{p})}} < \frac{.43 - .4}{\sqrt{\frac{.4(1-.4)}{200}}}\right) = P(Z < .87) = .8078$$

Örnek dağılımı



Standart normal dağılım



# Sonlu Populasyondan Örnek

$$n > .05 N \quad \text{veya} \quad n / N > .05$$

İse sonlu populasyon düzeltme faktörü (SPDF) kullanılır

> SPDF ile standart sapma

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}}$$