



Örnekleme Mantığı

Yaşar Tonta

H.Ü. Bilgi ve Belge Yönetimi Bölümü

tonta@hacettepe.edu.tr

<http://yunus.hacettepe.edu.tr/~tonta/tonta.html>



Not

- ➔ Bu slaytlarda yer alan bilgiler BBY 207 Sosyal Bilimlerde Arařtırma Yöntemleri dersi için hazırlanmıřtır. Slaytlarda atıf yapılan kaynakların okuma listesinde tam bibliyografik bilgileri verilmektedir. Bazı kaynaklardan (özellikle Babbie, Karasar, Schutt, Kaptan) daha yoğun olarak yararlanılmıřtır. Tüm alıntılar için spesifik olarak her zaman kaynak gösterilmemiřtir.



Plan

- ➔ SPSS Ödevi I
- ➔ Ara sınav
- ➔ Örneklem mantığı



AMERİKAN BAŞKANLIK SEÇİMİ TAHMİNLERİ - 1996

Tarih	Şirket	Oy yüzdesi (%)			
		Clinton	Dole	Perot	Diğer
28-31/10	Hotline/Battleground	49	40	9	2
30/10-2/11	CBS/NY Times	54	35	9	2
31/10-3/11	Pew Research Center	52	38	8	2
1 - 3 / 11	Reuter/Zeugby	49	41	8	2
1 - 3 / 11	Harris	51	39	9	1
2 - 3 / 11	ABC	52	39	7	2
2 - 3 / 11	NBC/WSJ	51	38	9	2
3 - 4 / 11	Gallup/CNN/USA Today	51	38	9	2
	Seçim Sonuçları	49	41	9	2

Örnekleme

- ➡ 90 milyon Amerikan seçmenin oy verme davranışını doğruya yakın tahmin etmek bu şirketler kaç kişiyle görüşmüş olabilir?
- ➡ Hiç kimse her şeyi gözleyemez
- ➡ Örnekleme: Neyin gözlenip neyin gözlenmeyeceğine karar verme süreci
- ➡ Örneklemden evrene genelleme
- ➡ Olasılık örnekleme
- ➡ Rastgele seçim



Tarihçe

- ➡ ABD Başkanlık seçimleri
- ➡ Literature Digest dergisi telefon rehberi ve otomobil kayıtlarından yararlanarak örneklem seçer
- ➡ 6 eyaletteki seçmenlere kartpostal göndererek hangi partiye oy vereceklerini sorar
- ➡ 1924, 1928 ve 1932 seçim sonuçlarını doğru tahmin eder



Tarihçe ...

- ➔ Literary Digest 1936 seçimlerinde 10 milyon kişiye kartpostal gönderir
- ➔ Cumhuriyetçi aday Alf London'ın %57'lik oy oranıyla kazanacağını tahmin eder
- ➔ Böyle bir ABD Başkanı tanıyor musunuz?
- ➔ Peki rakibini tanıyor musunuz?
- ➔ Franklin D. Roosevelt %61 oyla seçildi

Niye Böyle Oldu?

- ➔ Kartpostalların geri dönüş oranı %22
- ➔ Bunlar Cumhuriyetçiler miydi?
- ➔ Cumhuriyetçiler postaneye yakın mı oturuyorlardı?
- ➔ Demokratlar kamu oyu yoklamalarına güvenmiyor muydu?



Yanıt...

- ➔ Seçilen örneklem çerçevesi
- ➔ Telefon aboneleri ve otomobil sahipleri
- ➔ 1929 ekonomik bunalımı
- ➔ Herkes telefon abonesi ve otomobil sahipleri gibi düşünmüyor



Gallup

- ➡ George Gallup 1936 seçim sonuçlarını doğru tahmin eder
- ➡ “Kota örnekleme” tekniğini kullanır
- ➡ Kota örnekleme “evren”in özelliklerini bilmeye dayanır
- ➡ Gallup bu yöntemi kullanarak 1936, 1940 ve 1944 seçimlerini doğru tahmin eder
- ➡ Gallup 1948’de yanılır..



Neden?

- ➔ Tahminciler işi erken bırakır
- ➔ Çok fazla kararsız vardır
- ➔ Daha da önemlisi örneklem evreni temsil etmez
- ➔ Evren hakkındaki bilgiler 1940 nüfus sayımına dayanır
- ➔ Ama 2. Dünya Savaşı kentlere göçü artırır
- ➔ 1948'e gelindiğinde nüfus dağılımı değişir
- ➔ Kentlerde yaşayanlar daha demokrat eğilimlidir
- ➔ Harry Truman başkan seçilir

Örnekleme Teknikleri

Olasılığa dayanmayan
Örnekleme teknikleri

Kolaycı

Yargısal

Kartopu

Kota

Olasılığa dayanan
Örnekleme teknikleri

Basit rastgele

Sistematik

Tabakalı

Kümeleme



Mevcut “denek”lere dayanma (kolayda örnekleme)

- ➔ Sokaktaki vatandaşlara sorulur
- ➔ Riskli
- ➔ Genelleme yapılırken dikkatli olunmalı

Amaçlı ya da yargısal örnekleme

- ➔ Evreni bilmeye dayalı örneklem seçimi
- ➔ Özellikle anketlerin tasarlanmasında kullanılır
- ➔ Anketteki yetersizlikleri ortaya çıkarır
- ➔ Bir ön testtir

Kartopu örnekleme

- ➔ Kazara örnekleme!
- ➔ Daha çok ön arařtırmalarda (niteliksel) kullanılır
- ➔ Özel bir evren
- ➔ Bulunması zor insanlarla ilgili olarak yapılır (evsizler, tinerciler, kaçak işçiler, vs.)
- ➔ Birkaç kişi belirlenir, onlardan örnekleme seçilerek diğer kişilerin bilgilerine ulaşılır
- ➔ “Kartopu” terimi denek sayısının giderek artması nedeniyle kullanılır

Kota örnekleme

- ➔ Temsil sorunu
- ➔ Bir matris ya da tabloyla işe başlanır
- ➔ Her gözdeki deęişken için veriler toplanır (erkek/kadın, çeşitli yaş gruplarına göre dağılım, eğitim düzeyleri vs.)
- ➔ Her gözdeki veriler evrene oranlanır
- ➔ Kota çerçevesi doğru olmalı (güncel bilgi gerekli)
- ➔ Örneklem ögelerinin seçiminde önyargı olmamalı

Olasılıęa Dayalı Örneklem Seçimi

- ➔ Evrenin tüm bireyleri aynı özellikleri taşıyorsa o zaman tek bir örnek almak yeterli olurdu
- ➔ Ama değil

Örnekleme seçiminde bilinçli / bilinçsiz önyargılar I

- ➔ HÜ öğrencileriyle ilgili bir araştırma
- ➔ İlk rastlanan kişilerle görüşmek
- ➔ Kişisel eğilimler (“Cool”/çalışkan)
- ➔ Dengeli bir örneklem seçmek isteseniz bile bu dengenin oranlarını bilmeyebilirsiniz
- ➔ Her 10. öğrenciyle görüşmek (gene de temsil etmeyebilir)

Örnekleme seçiminde bilinçli / bilinçsiz önyargılar II

- ➔ Önyargı = örnekleme seçilen deneklerin “tipik” olmaması, yani evreni temsil etmemesi
- ➔ Radyo / TV / web sayfalarında yapılan kamu oyu yoklamaları bu türden (çoğu kimsenin haberi olmuyor)
- ➔ SMS gönderilerek yapılan yarışmalar bu türden
- ➔ Genelleme yapmak yanlış

Olasılıęa dayalı örnekleme seçimi

- ➔ Örnekleme seçilen deneklerin özellikleri evreni oluşturan deneklerin özelliklerine yakınsa evreni temsil edebilir
- ➔ **Temel ilke: Evreni oluşturan her ögenin örnekleme seçilme şansı eşit olmalı**



Olasılıđa dayalı örneklem seçiminin avantajları

- ➔ Evreni temsil edebilme özelliđi diđer yöntemlerle seçilen örneklemlerden daha fazla
- ➔ Örneklemin evreni temsil düzeyi doğru olarak hesaplanabilir

Örnekleme Teknikleri

Olasılığa dayanmayan
Örnekleme teknikleri

Kolaycı

Yargısal

Kartopu

Kota

Olasılığa dayanan
Örnekleme teknikleri

Basit rastgele

Sistematik

Tabakalı

Kümeleme

Örnekleme Tasarım Türleri

- ➔ Basit rastgele örnekleme
- ➔ Sistematiik örnekleme
- ➔ Tabakalı örnekleme
- ➔ Kümeleme örnekleme

Basit rastgele örneklem

- ➔ Rastgele sayılar tablosundan seçilerek yapılır
- ➔ Zahmetli



Sistematiik örnekleme

- ➔ Bir listeden her k 'inci öge seçilir
- ➔ Başlangıç değeri rastgele alınır
- ➔ Örneklem arası:
 - evren büyüklüğü / örneklem büyüklüğü
- ➔ Örneklem oranı:
 - örneklem büyüklüğü / evren büyüklüğü
- ➔ Listedeki ögeler devirsel olmamalı

Tabakalı örnekleme

- ➔ Evreni temsil yeteneđi, daha yüksek
- ➔ Büyük örnekleme küçük standart hata üretir
- ➔ Benzeşik evren küçük standart hata üretir
- ➔ Tabakalı örnekleme ikinci faktöre dayanır
- ➔ Tabakalama homojenleştirir
- ➔ Sıralanmış listeden sistematik örnekleme seçimi de tabakalı örnekleme sonucunu verir
- ➔ Aynı listeden basit rastgele örnekleme seçersek tabakalama kaybolur.



Küme örnekleme

- ➔ Listeleme
- ➔ Örnekleme



- ➔ Büyüklükle orantılı olasılıksal örnekleme
- ➔ Orantısız örnekleme ve ağırlıklandırma



Örneklem Büyüklüğünü Etkileyen Faktörler

- ➔ Kararın önem derecesi
- ➔ Arařtırmanın türü
- ➔ Değişken sayısı
- ➔ Analiz türü
- ➔ İncelenen olayın evrendeki yoğunluğu
- ➔ Yanıt oranları
- ➔ Mali kaynak kısıtlılığı
- ➔ Alt grup sayısı
- ➔ Evrenin varyansı



Terminoloji



Öge

- ➔ Hakkında bilgi toplanan ve analizin temelini oluřturan birim (örneğin, kişiler ya da belirli özellikleri taşıyan kişiler)
- ➔ Öge örneklem seçiminde kullanılır
- ➔ Analiz birimi veri analizi için kullanılır

Evren

- ➔ Arařtırılacak ögelerin toplamı
- ➔ “HÜ öğrencileri”
 - Hangi öğrenciler?
 - 2005-2006 öğretim yılında öğrenim gören öğrenciler?
 - 2005-2006 öğretim yılında okula kayıtlı ve en az 1 kredilik ders yüklenmiş lisans öğrencileri?



Arařtırma Evreni

- ➔ Örneklemin gerçekte içinden seçileceđi ögelerin toplamı
- ➔ Evrenin kuramsal tanımında yer alan tüm ögelerin örnekleme seçilmesi garanti deđil
- ➔ Bazı öğrenciler öğrenci listesinde yer almayabilir (özel öğrenciler, kayıt donduranlar, geçici olarak uzaklařtırılanlar, vs.)
- ➔ Telefon numaraları gizli olan aboneler

Örneklem Birimi

- Örneklemin belli aşamalarında seçim için düşünülen öge ya da ögeler seti
- Tek aşamalı örnekleme örnekleme birimi öge ile hatta analiz birimiyle aynıdır
- Daha karmaşık örneklemlerde farklı düzeylerdeki örnekleme birimleri kullanılabilir
- Mahalleler – haneler- yetişkinler



Birincil
Örneklem
birimi



İkincil
Örneklem
birimi



Üçüncül
Örneklem
birimi

Örnekleme Çerçevesi

- ➔ Örneklemin ya da örneklemin belirli bir aşamasının seçileceđi örneklem birimlerinin geçerli listesi
- ➔ Tek aşamalı örneklem çerçevesi araştırma evrenidir
- ➔ Öğrenci listesi
- ➔ Şehirdeki mahallelerin listesi



Gözlem Birimi

- ➔ Veri toplama birimi, hakkında bilgi toplanacak ögeler seti
- ➔ Genellikle analiz birimi ile gözlem birimi aynıdır



Deęiřken

- ➔ Birbirinden ayrı özellikler seti:
cinsiyet, yař, medeni durum, vs.



Parametre

- ➔ Verilen bir deęiřkenin bir evrendeki özet tanımı
- ➔ Ortalama yař, ortalama eęitim süresi, vs.



İstatistik

- ➔ Bir deęişkenin örneklemdaki deęerini özetleyen tanım
- ➔ Örneklemden çıkan yaş ortalaması istatistiktir
- ➔ Örneklem istatistikleri evren parametrelerini tahmin etmek için kullanılır

Örnekleme Hatası

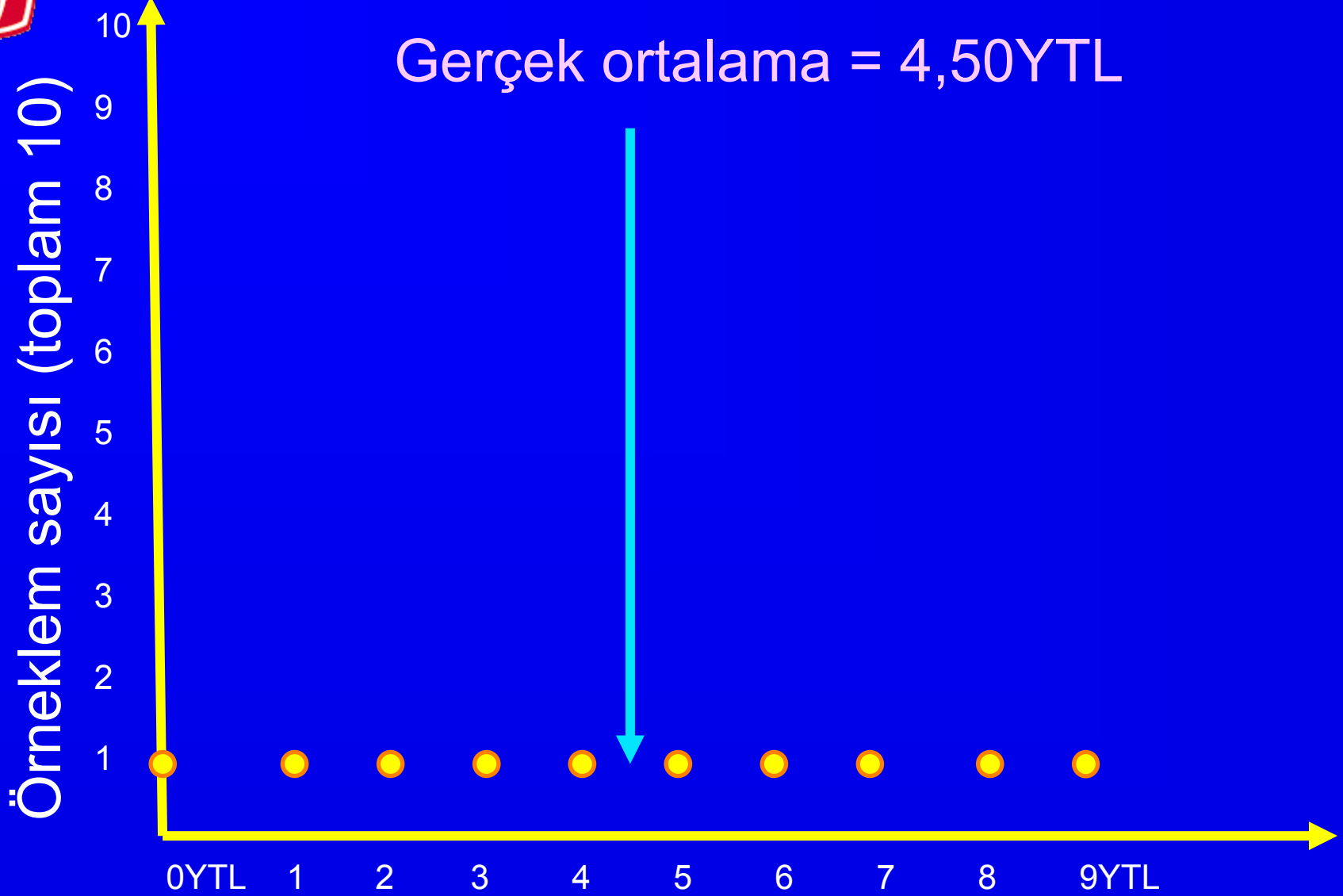
- ➔ Örnekleme istatistiđi nadiren evren parametresine eşit çıkar
- ➔ Örnekleme hatası hesaplanabilir

Güven Düzeyleri ve Güven Aralıkları

- ➔ Örneklem istatistiklerinin doğruluđu evren parametresine uzaklığına göre açıklanır
- ➔ %95 güvenle örneklem istatistiđinin evren parametresinden ± 5 veya ± 1 puan aralığında olması

Olasılık Örneklem Kuramı ve Örneklem Dağılımı

- ➔ Cebindeki para miktarı 0 ile 9 YTL arasında deęişen 10 kiři var
- ➔ Ortalama 4,5 YTL
- ➔ 10 kiřinin tümüne sormadan ortalamayı nasıl tahmin edebiliriz?
- ➔ Örneklem seçerek



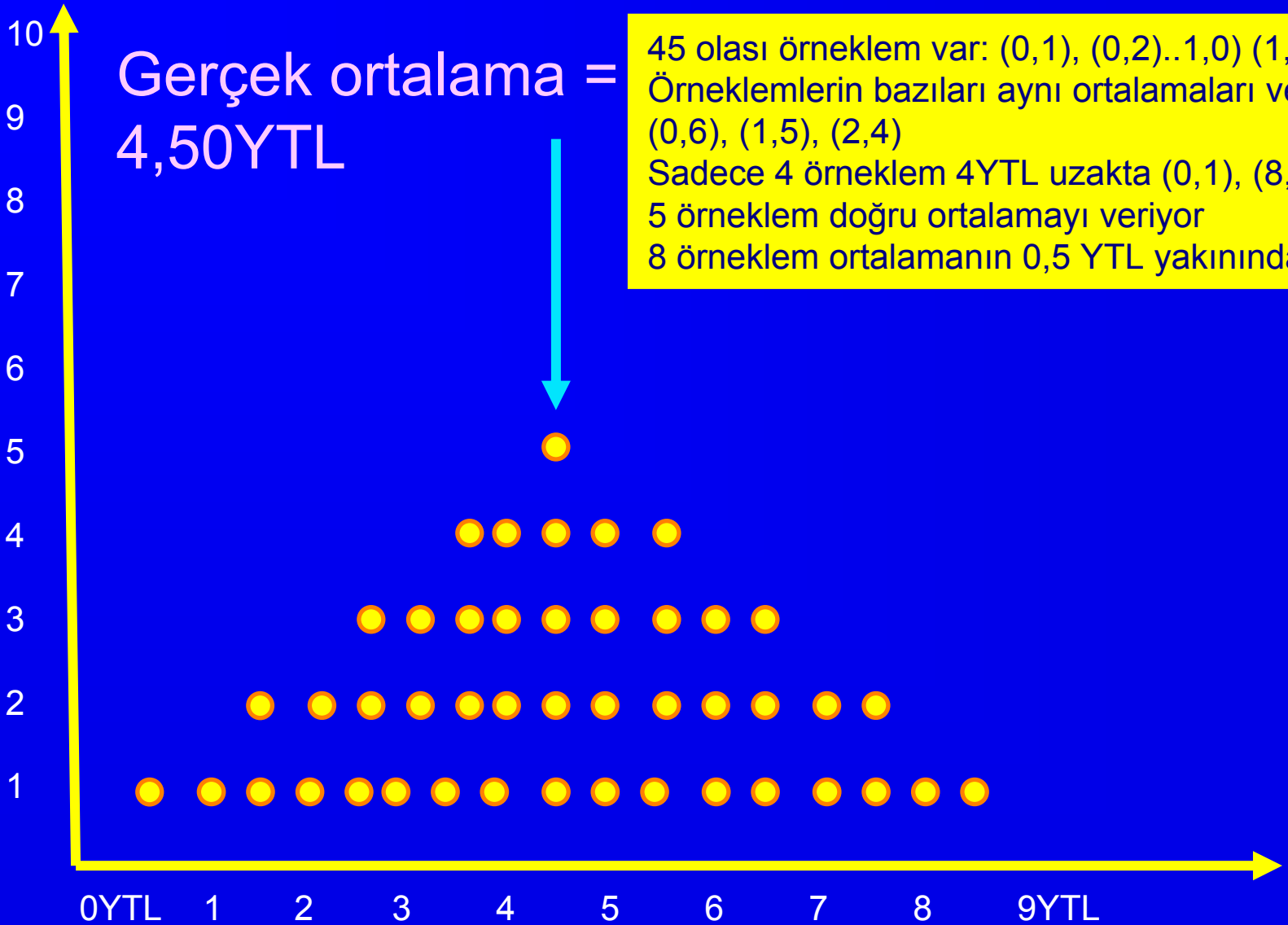
Tahmini ortalama (örneklem büyüklüğü = 1)



Örneklem sayısı (toplam 45)

Gerçek ortalama =
4,50YTL

45 olası örneklem var: (0,1), (0,2)..1,0) (1,9) ...
Örneklemlerin bazıları aynı ortalamaları veriyor
(0,6), (1,5), (2,4)
Sadece 4 örneklem 4YTL uzakta (0,1), (8,9)
5 örneklem doğru ortalamayı veriyor
8 örneklem ortalamanın 0,5 YTL yakınında

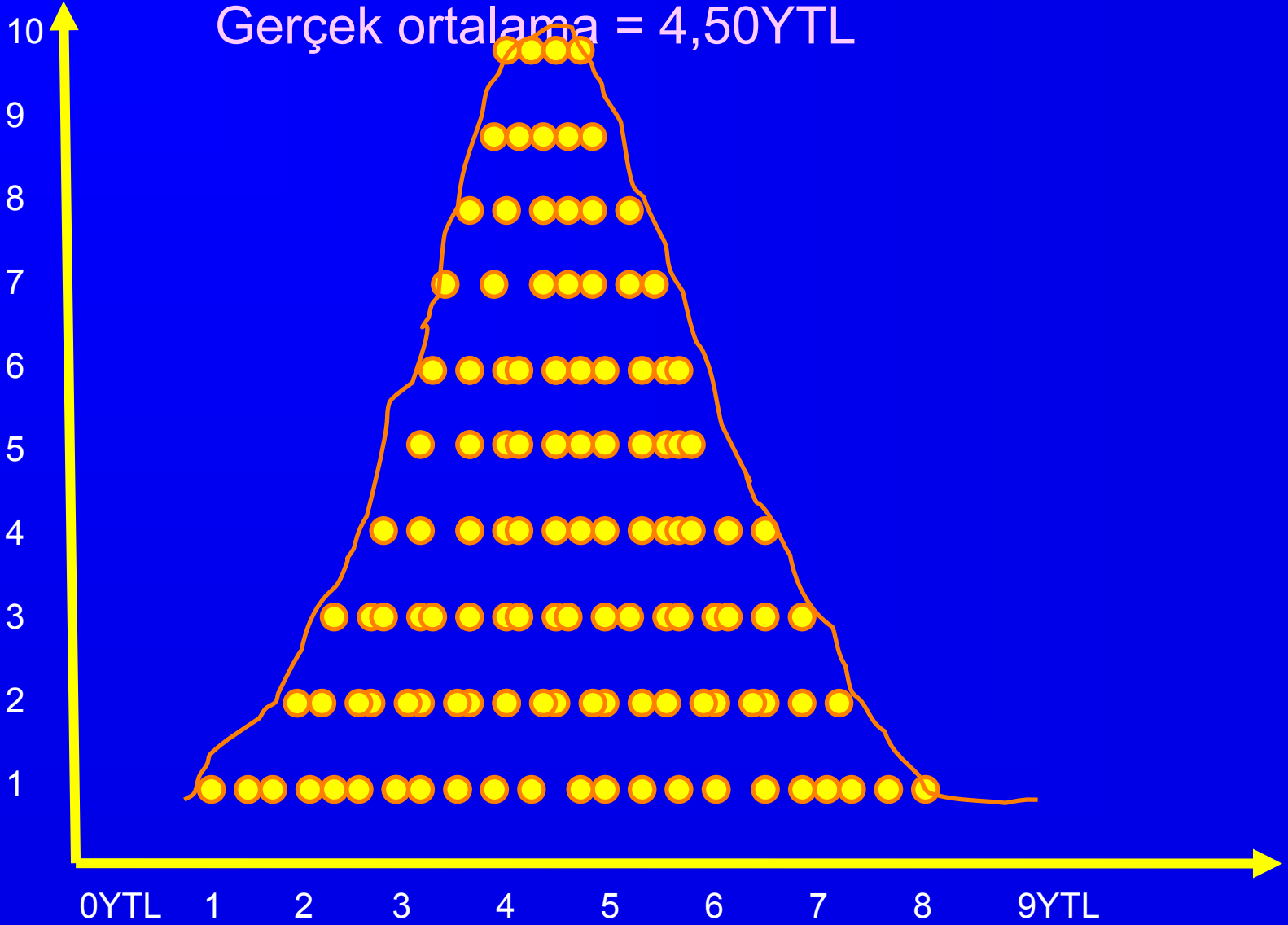


Tahmini ortalama (örneklem büyüklüğü = 2)



Gerçek ortalama = 4,50YTL

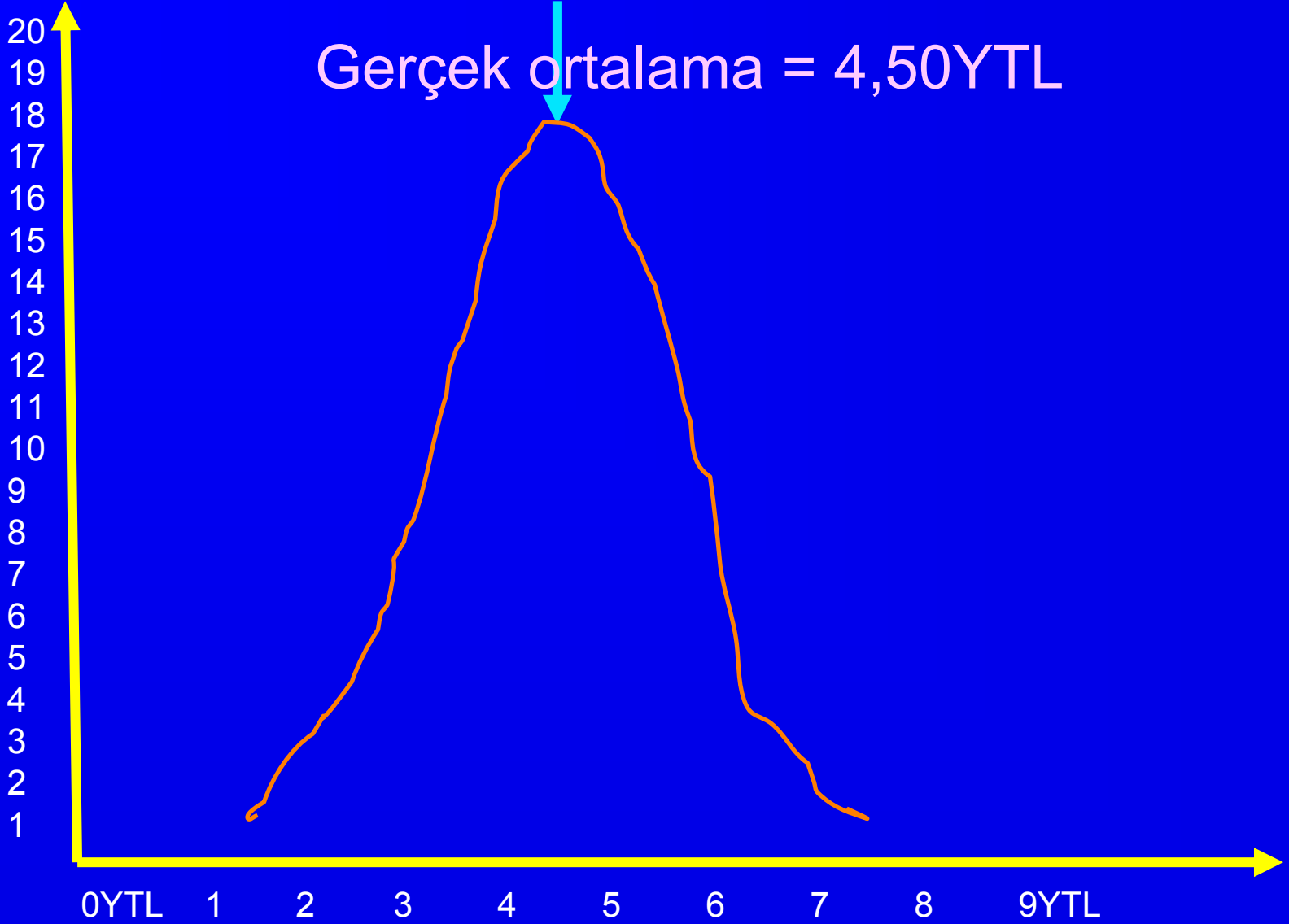
Örneklem sayısı (toplam 120)



Tahmini ortalama (örneklem büyüklüğü = 3)



Örneklem sayısı (toplam 210)



Tahmini ortalama (örneklem büyüklüğü = 4)



Örneklem sayısı (toplam 252)

20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1

0YTL

1

2

3

4

5

6

7

8

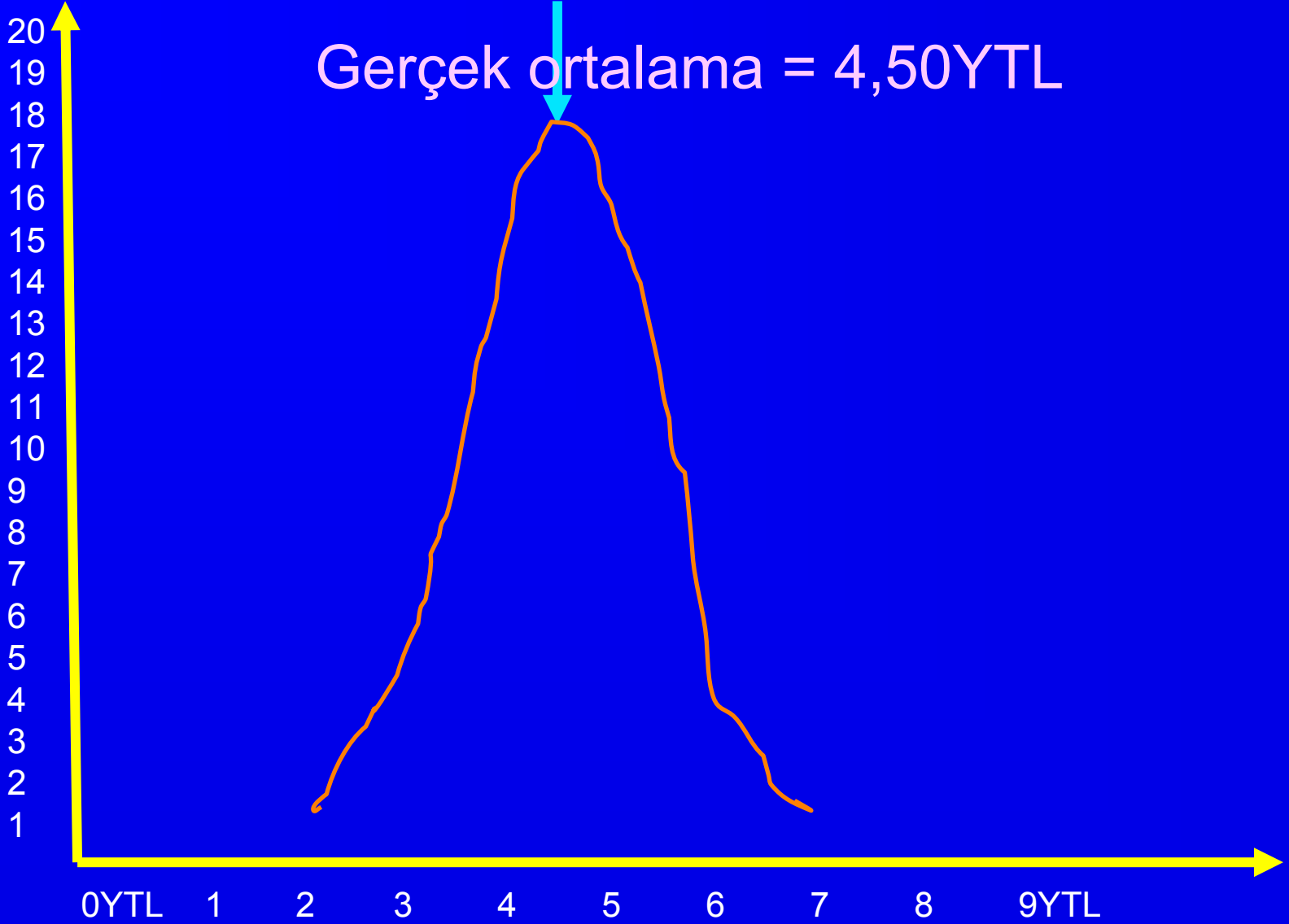
9YTL

Gerçek ortalama = 4,50YTL

Tahmini ortalama (örneklem büyüklüğü = 5)



Örneklem sayısı (toplam 252)



Tahmini ortalama (örneklem büyüklüğü = 5)



Örneklem sayısı (toplam 210)

20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1

0YTL 1 2 3 4 5 6 7 8 9YTL

Gerçek ortalama = 4,50YTL

Tahmini ortalama (örneklem büyüklüğü = 6)

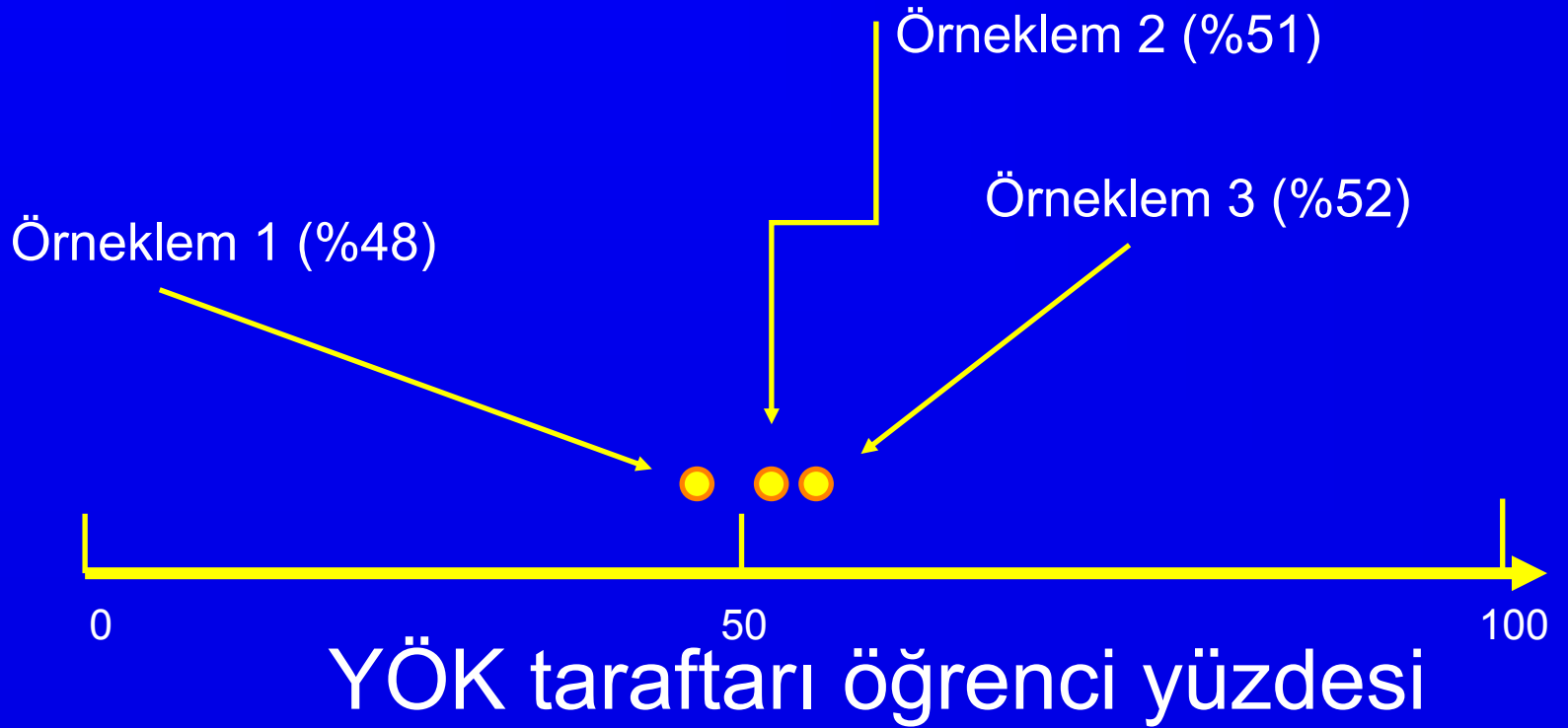


Öğrencilerin YÖK'e karşı tutumu

- ➔ Daha gerçekçi bir örnek verelim
- ➔ Diyelim ki HÜ öğrencilerinin YÖK'e karşı tutumunu araştırıyoruz
- ➔ YÖK'e (__) karşıyım (__) taraftarım

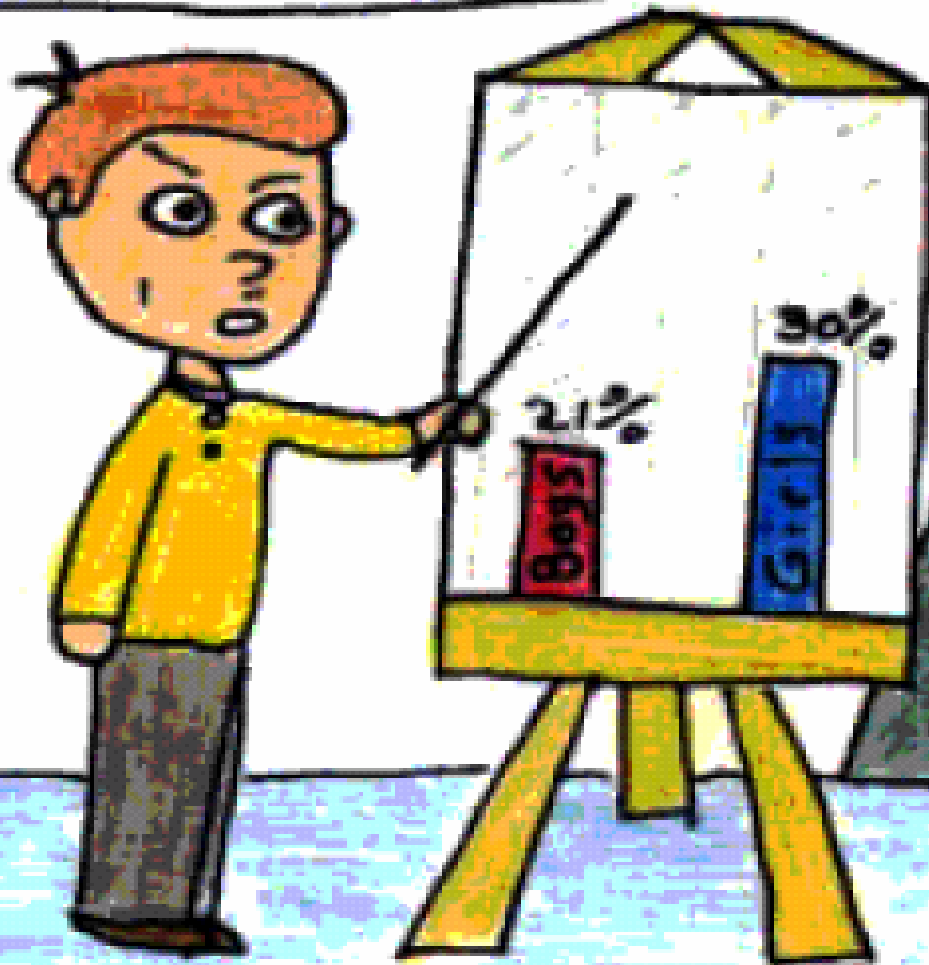
Öğrencilerin YÖK'e karşı tutumu...

- Arařtırma evreni: 20.000 HÜ öğrencisi
- Örneklem çerçevesi: Öğrenci listesi
- Ögeler: Birey olarak öğrenciler
- Değişken: Öğrencinin YÖK'e karşı tutumu (ikili değişken)
- Her öğrenciye bir numara verelim
- Rastgele 100 öğrenci seçelim





21% of the boys and 30% of the girls support me; therefore I'll get 51% of the vote.



VOTE FOR ME
AS PRESIDENT
OF THE
MATH CLUB

Ben Shabat



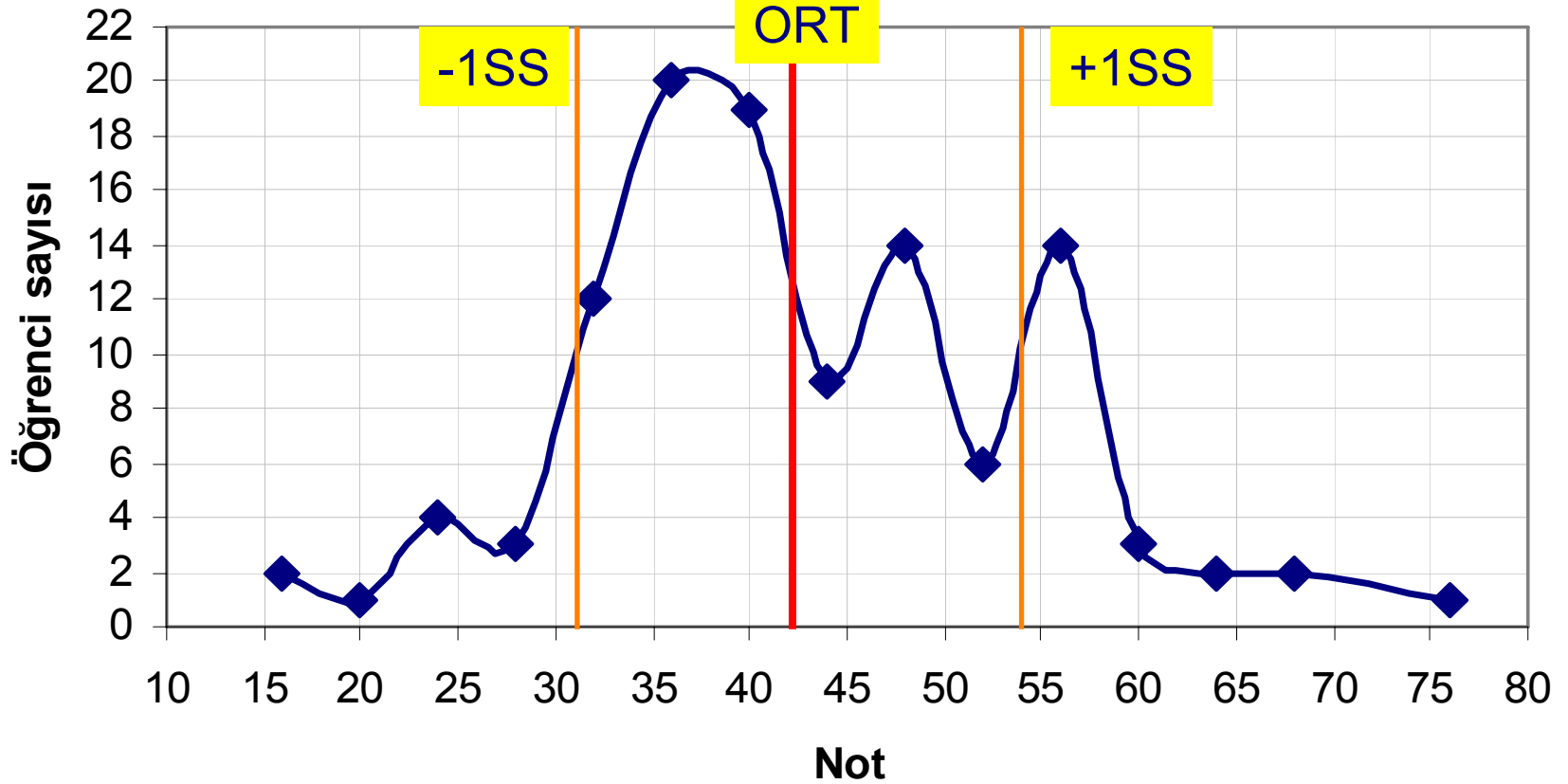
Örnekleme Dağılımı

- ➔ Rastgele seçilmiş 10 kişinin not ortalamasını alsanız bu, sınıf ortalamasını tam olarak yansıtmayabilir (eksik ya da fazla olabilir). Ama normal dağılım söz konusuysa çıkan değerin ortalamaya yakın olması lazım. Örnekleme artırırsanız daha isabetli örneklem ortalaması tutturabilirsiniz.

- ➔ Örnekleme dağılımı ile ilgili hareketli örnek:
http://www.ruf.rice.edu/%7Elane/stat_sim/sampling_dist/index.html



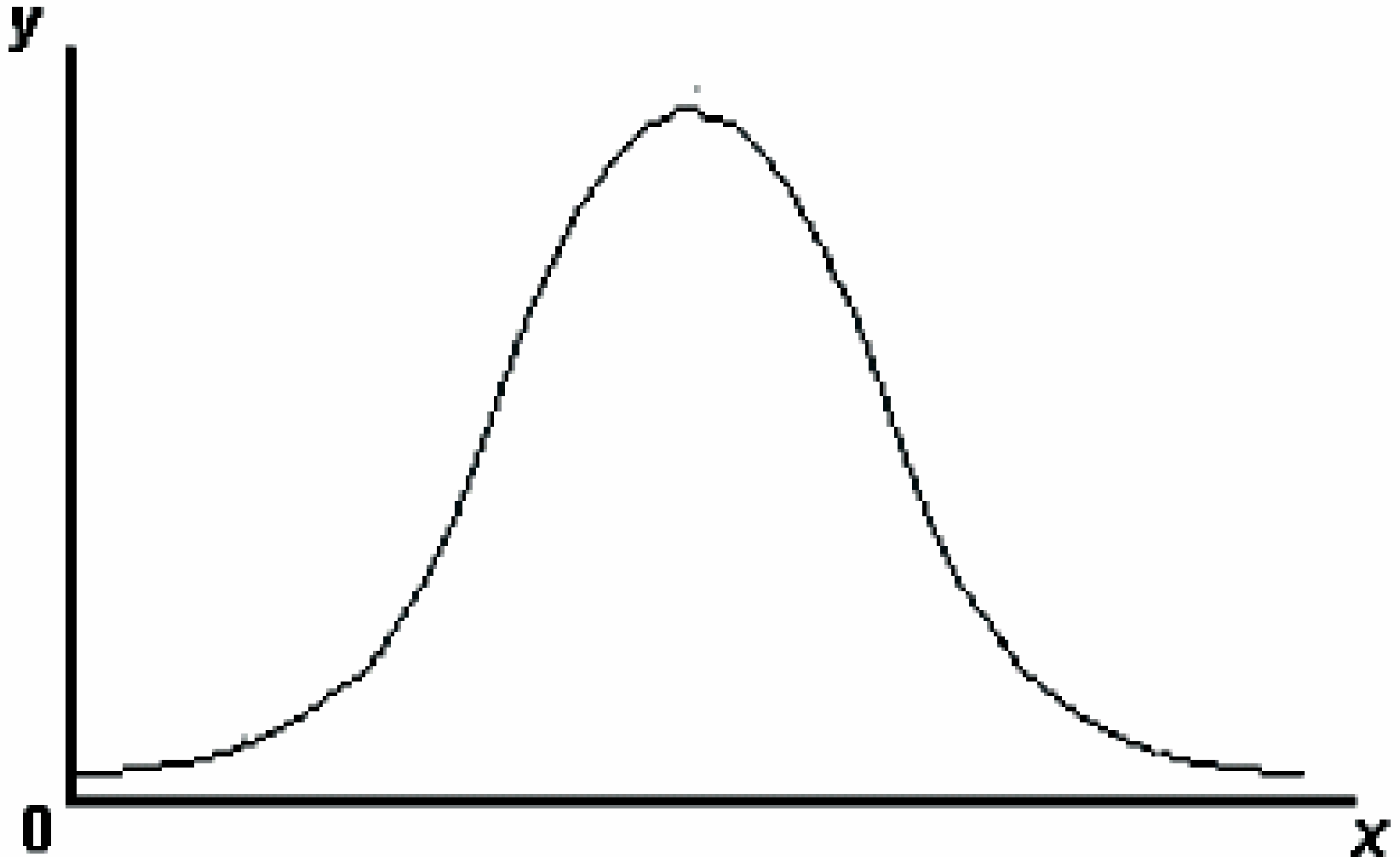
BBY 207 Sınav Sonuçları



$X=42,7$; $SS=11,2$

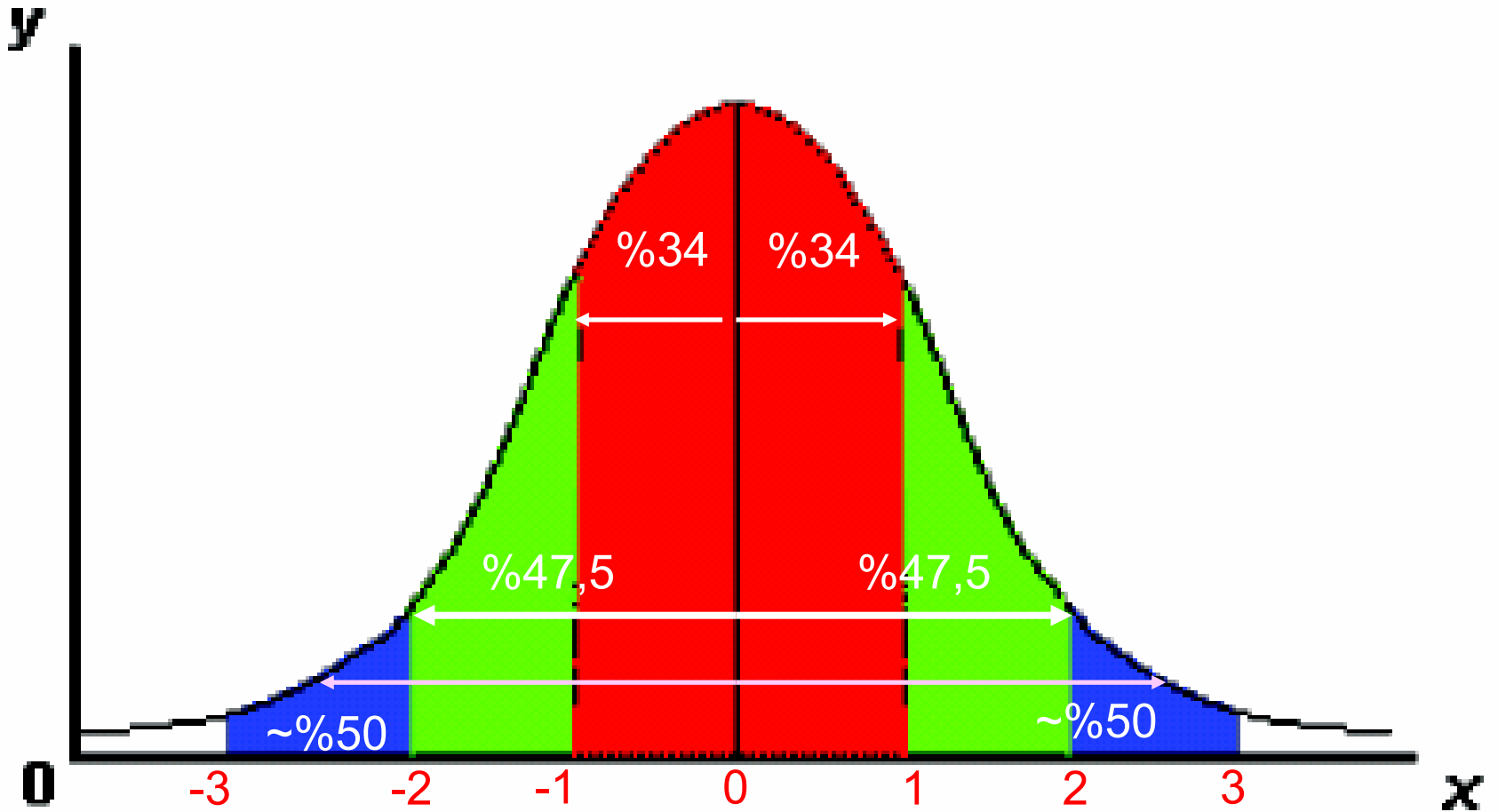


Normal Dağılım





Standart Sapma



©Copyright, Robert Niles, <http://www.robertniles.com/stats/stddev.shtml>

Olasılık Kuramı

➡ Örneklem istatistiklerinin evren parametresine ne kadar yakın olduğunu verir

➡ $s = \sqrt{p * q / n}$

➡ s: standart hata

➡ n = örneklem büyüklüğü

➡ p = birşeyin olma olasılığı

➡ q = birşeyin olmama olasılığı

YÖK örneđi

- ➔ Varsayalım örneklemdaki 100 öğrencinin yarısı YÖK taraftarı, yarısı deđil
- ➔ Formülü uygulayarak standart hatanın 0.05 olduğunu hesaplarız (yani %5)
- ➔ 100 örneklemden 68'i parametrenin ± 1 standart hata (yani %5) altında ya da üstündedir
- ➔ 100 örneklemden 95'i parametrenin ± 2 standart hata (yani %10) altında ya da üstündedir
- ➔ 100 örneklemden 99.9'u parametrenin ± 3 standart hata (yani %15) altında ya da üstündedir
- ➔ Yani 1000 örneklemden sadece biri %65'in üzerinde ya da %35'in altında bir örneklem istatistiđi verir (evren parametresinin %50 olduğunu hatırlayın)

Standart Hata

- ➔ Evren parametresiyle örneklem büyüklüğünün bir ölçüsüdür
- ➔ Örneklem büyüklüğü arttıkça standart hata azalır (4 kat artarsa SH yarıya düşer, yani örneklem dağılımlarının ortalamaları evren parametresine daha yakınlaşır)



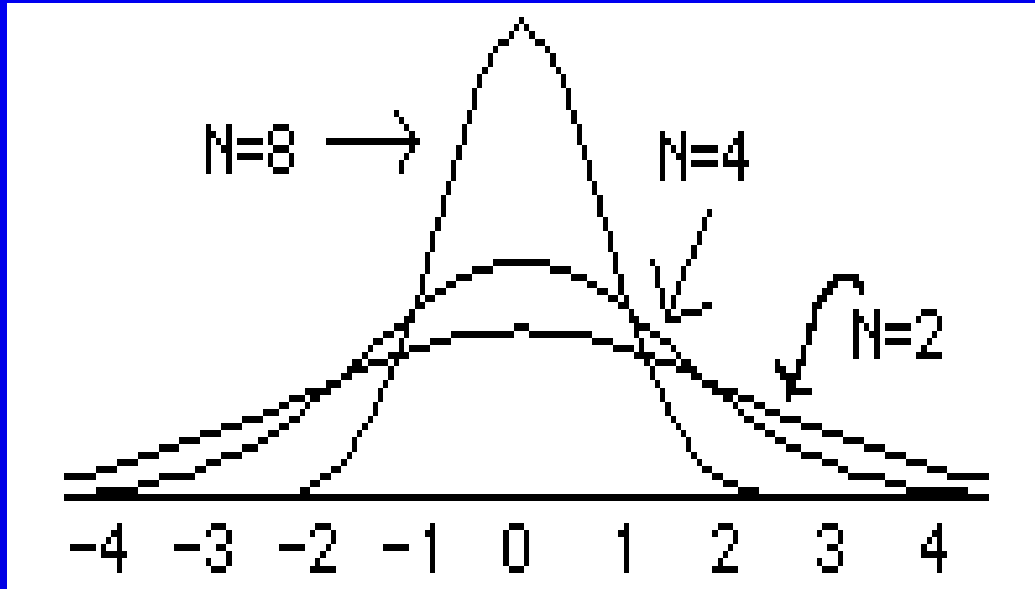
Örneklem Dağılımı II

Örneklem büyüklüğü arttıkça standart hata azalır.

Ortalaması μ , SS'si σ olan bir evrenden bir örneklem seçerseniz,

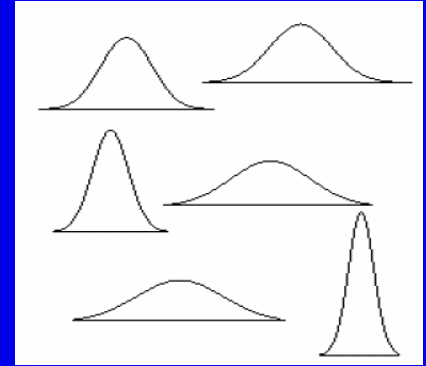
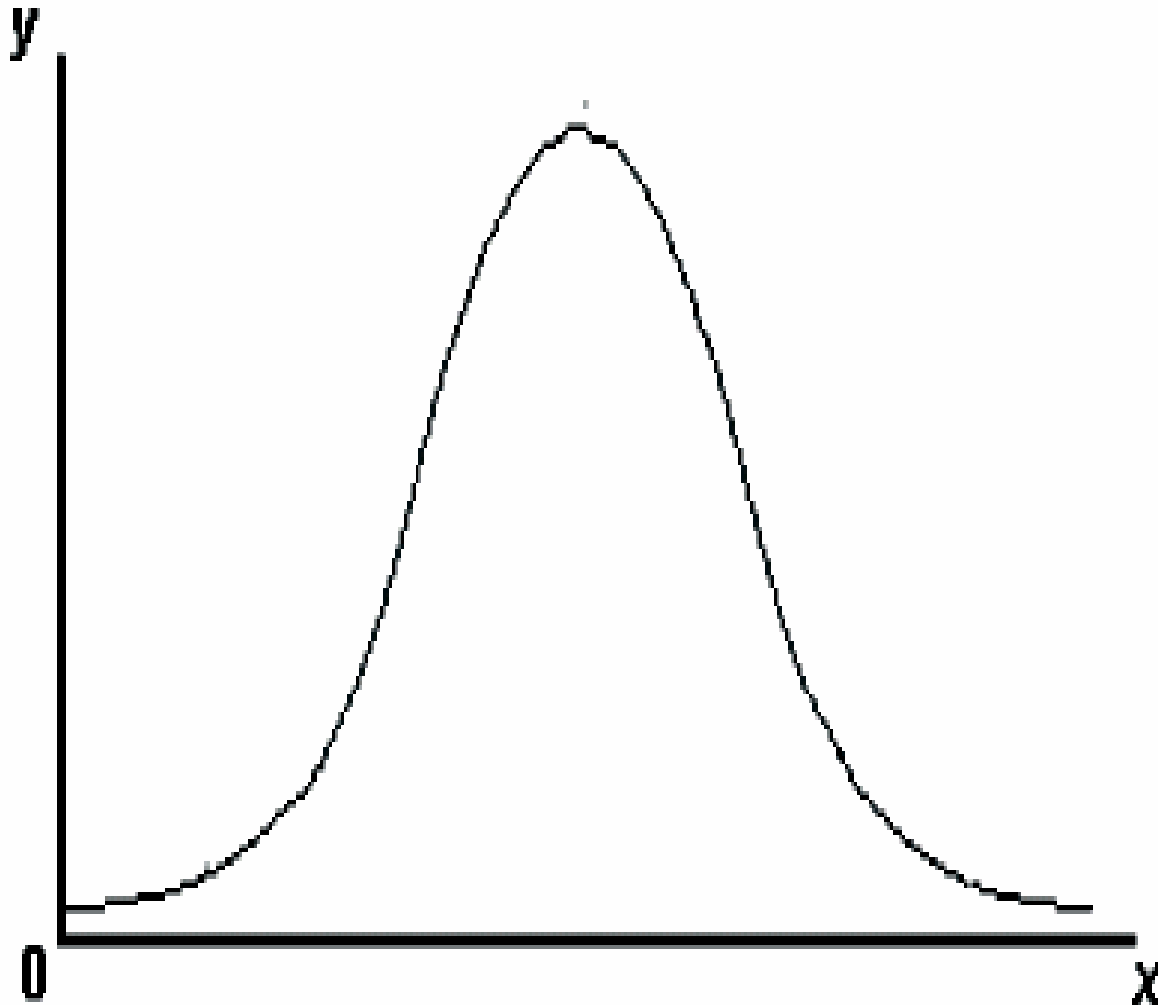
Örneklemin ortalaması μ , SS'si $\frac{\sigma}{\sqrt{N}}$ olur (N = örneklem büyüklüğü)

Örneklemin standart sapması ortalamasının standart hatası olarak bilinir.





Normal Dağılım



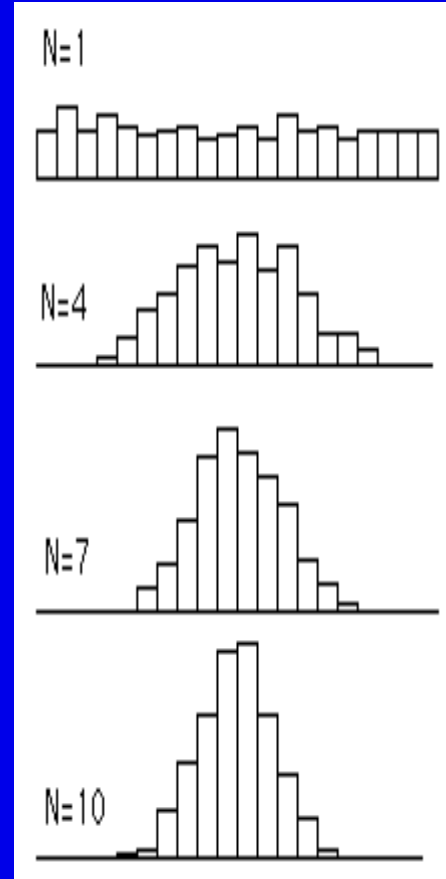
$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

μ = ortalama
 σ = standart sapma
 π = 3.14159
 e = 2.718282.



Merkezi Limit Teoremi

- ➔ Bilgisayar normal dağılım gösteren bir evrenden N sayı seçiyor ve ortalamaları hesaplıyor. Örneklem büyüklüğü (N) 1, 4, 7 ve 10 için bilgisayar bu işlemi 500 defa tekrarlıyor.
- ➔ N arttıkça dağılım normalleşiyor
- ➔ N arttıkça dağılım daha tekbiçim oluyor
- ➔ Eğer evrendeki herkes aynı görüşteyse her örneklem aynı sonucu verir



$$s = \sqrt{p * q / n}$$



Merkezi Limit Teoremi II

- ➡ Evren parametresini bilmiyoruz (bilsek niye örneklem alıp evren parametresini tahmin etmeye çalışalım!)
- ➡ Örneklemelerin büyüklüğü sınırlı (birkaç yüz ya da en fazla 1000-2000 denek)
- ➡ Sadece bir örneklem seçiyoruz



???

- ➔ Tek bir rastgele örneklem seçerek elde edilen istatistik evren parametresinin $\pm 1SH$ 'lık sınırlar içinde olma olasılığı %68.
- ➔ Buna güven düzeyi deniyor
- ➔ Yani %68 güvenle örneklem tahmini (istatistik) evren parametresinin $\pm 1SH$ içindedir
- ➔ Ya da %95 güvenle örneklem tahmini (istatistik) evren parametresinin $\pm 2H$ içindedir
- ➔ Ama evren parametresini bilmiyoruz?
- ➔ O zaman örneklem istatistiğini evren parametresi olarak kabul ediyoruz.



YÖK örneđi

- ➔ %95 güven düzeyinde öğrencilerin %40-%60 arasında (± 2 SH) YÖK taraftarı olduđu söylenebilir
- ➔ %40-%60 güven aralıđıdır
- ➔ %68 güven düzeyinde güven aralıđı %45-%55 olur
- ➔ Örneklem istatistiđine dayalı tahmin verilirken hem güven düzeyi hem de güven aralıđı belirtilmelidir
- ➔ Standart hata oranına karar verdikten sonra örneklem büyüklüđu saptanabilir