



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



BBY 207: Sosyal Bilimlerde Araştırma Yöntemleri I

OBAMA-MCCAIN TAHMİNLERİ - 2008

Poll Source

		OBAMA	MCCAIN
Rasmussen Daily Tracking	October 25-27, 2008	51%	46%
Reuters/CSPAN/Zogby	October 25-27, 2008	49%	45%
Rasmussen Daily Tracking	October 26-28, 2008	50%	47%
Reuters/CSPAN/Zogby	October 27-29, 2008	50%	43%
Rasmussen Daily Tracking	October 29 - October 31, 2008	51%	47%
Gallup Daily Tracking	October 29-31, 2008	52%	41%
Pew Center for Research	October 29-November 1, 2008	49%	42%
Investor's Business Daily/TIPP	October 29-November 1, 2008	47%	45%
ABC News/Washington Post	October 30-November 1, 2008	54%	43%
CNN	October 30-November 1, 2008	51%	44%
Rasmussen Daily Tracking	October 30-November 1, 2008	51%	46%
Zogby Daily Tracking	October 31, 2008	49%	44%
Gallup	October 31-November 2, 2008	53%	40%
NBC News/Wall Street Journal	October 31-November 2, 2008	51%	43%
Rasmussen Daily Tracking	October 31-November 2, 2008	52%	46%

SONUÇ
OBAMA %52,9
MCCAIN %45,7

Kaynak:
http://en.wikipedia.org/wiki/Nationwide_opinion_polling_for_the_United_States_presidential_election,_2008



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

Örneklem Teknikleri

Olasılıęa dayanmayan
Örneklem teknikleri

Kolaycı

Yargısal

Kartopu

Kota

Olasılıęa dayanan
Örneklem 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 değişken 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 isteseydik 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**

Olasılıęa dayalı örneklem seçiminin avantajları

- ➔ Evreni temsil edebilme özellięi dięer yöntemlerle seçilen örneklemelerden 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



Örneklem Tasarım Türleri

- ➔ Basit rastgele örneklem
- ➔ Sistemantik örneklem
- ➔ Tabakalı örneklem
- ➔ Kümeleme örneklem



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 örneklem

- ➔ Evreni temsil yeteneği, daha yüksek
- ➔ Büyük örneklem küçük standart hata üretir
- ➔ Benzeşik evren küçük standart hata üretir
- ➔ Tabakalı örneklem ikinci faktöre dayanır
- ➔ Tabakalama homojenleştirir
- ➔ Sıralanmış listeden sistematik örneklem seçimi de tabakalı örneklem sonucunu verir
- ➔ Aynı listeden basit rastgele örneklem 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

- ➔ Araştırılacak 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?



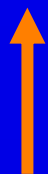
Arařtırma Evreni

- ➔ 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

- ➡ Örneklemde belli aşamalarında seçim için düşünülen öge ya da ögeler seti
- ➡ Tek aşamalı örneklemede örneklem birimi öge ile hatta analiz birimiyle aynıdır
- ➡ Daha karmaşık örneklemlerde farklı düzeylerdeki örneklem 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



Deęişken

- ➔ Birbirinden ayrı özellikler seti:
cinsiyet, yaş, 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 istatistiği nadiren evren parametresine eşit çıkar
- ➔ Örnekleme hatası hesaplanabilir

Güven Düzeyleri ve Güven Aralıkları

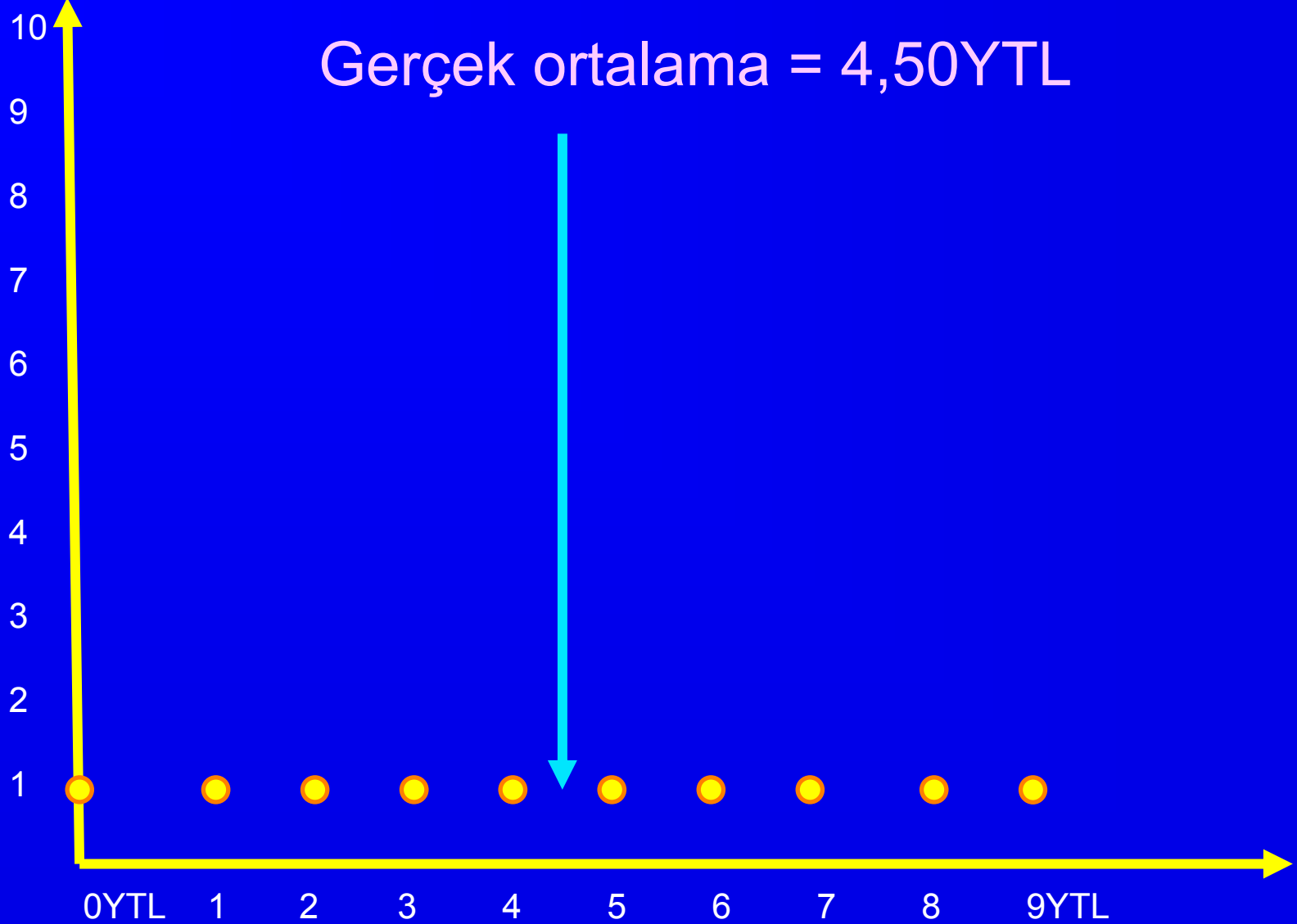
- ➔ Örneklem istatistiklerinin doğruluęu evren parametresine uzaklıęına göre açıklanır
- ➔ %95 güvenle örneklem istatistięinin evren parametresinden ± 5 veya ± 1 puan aralıęında olması

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



Örneklem sayısı (toplam 10)



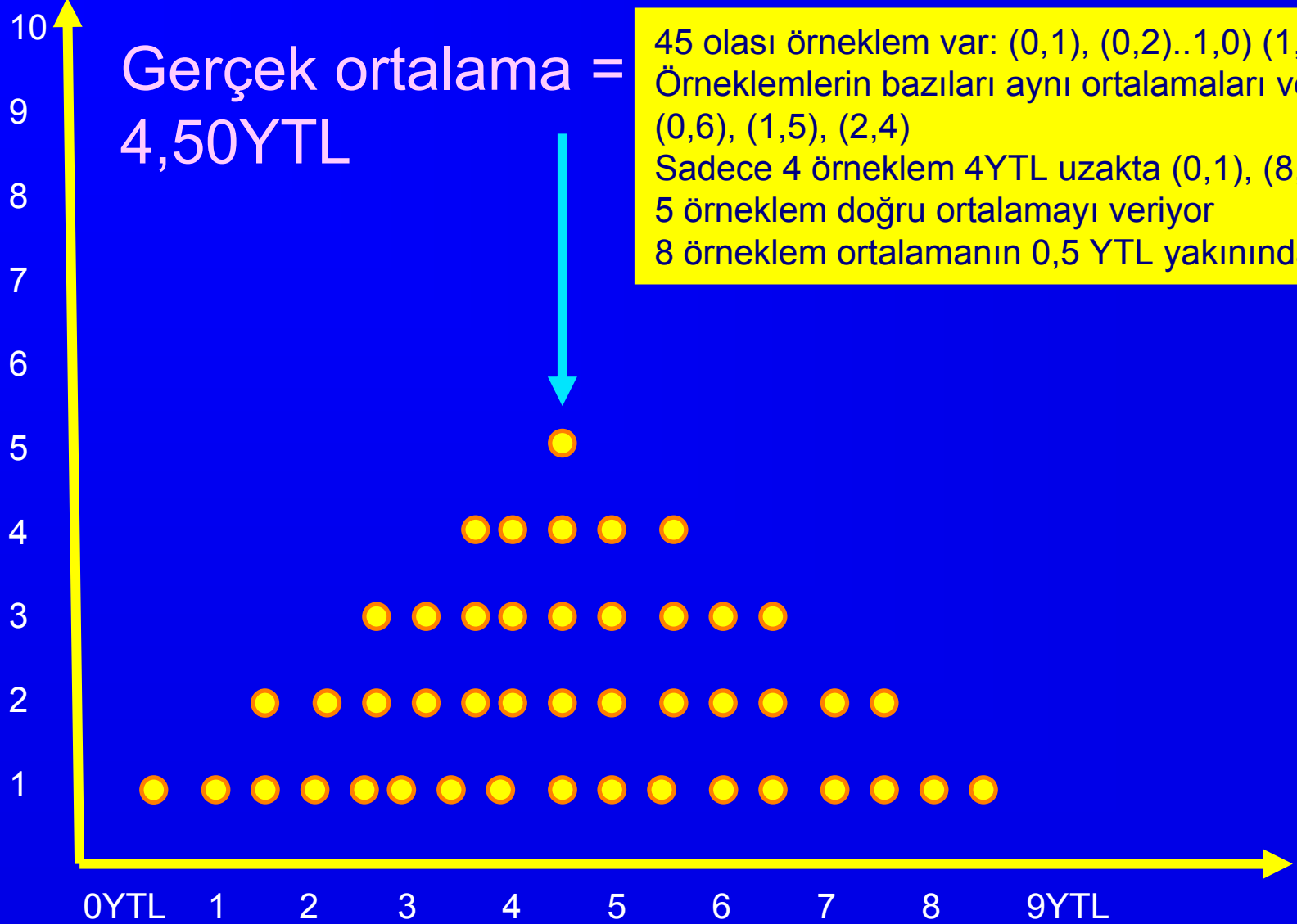
Tahmini ortalama (örneklem büyüklüğü = 1)



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

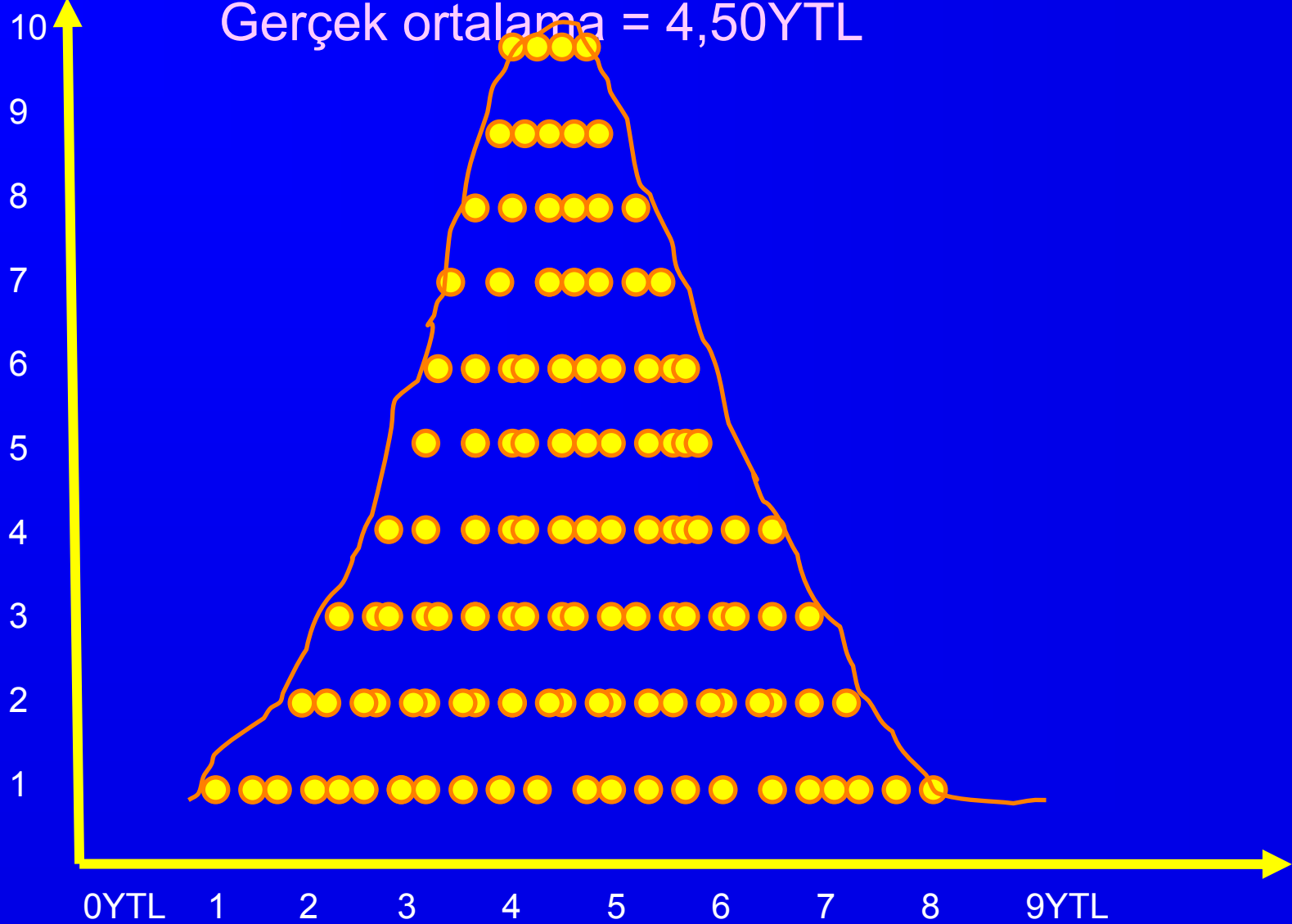


Tahmini ortalama (örneklem büyüklüğü = 2)



Gerçek ortalama = 4,50YTL

Örneklem sayısı (toplam 120)



Tahmini ortalama (örneklem büyüklüğü = 3)



Örneklem sayısı (toplam 210)

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 = 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)

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 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)

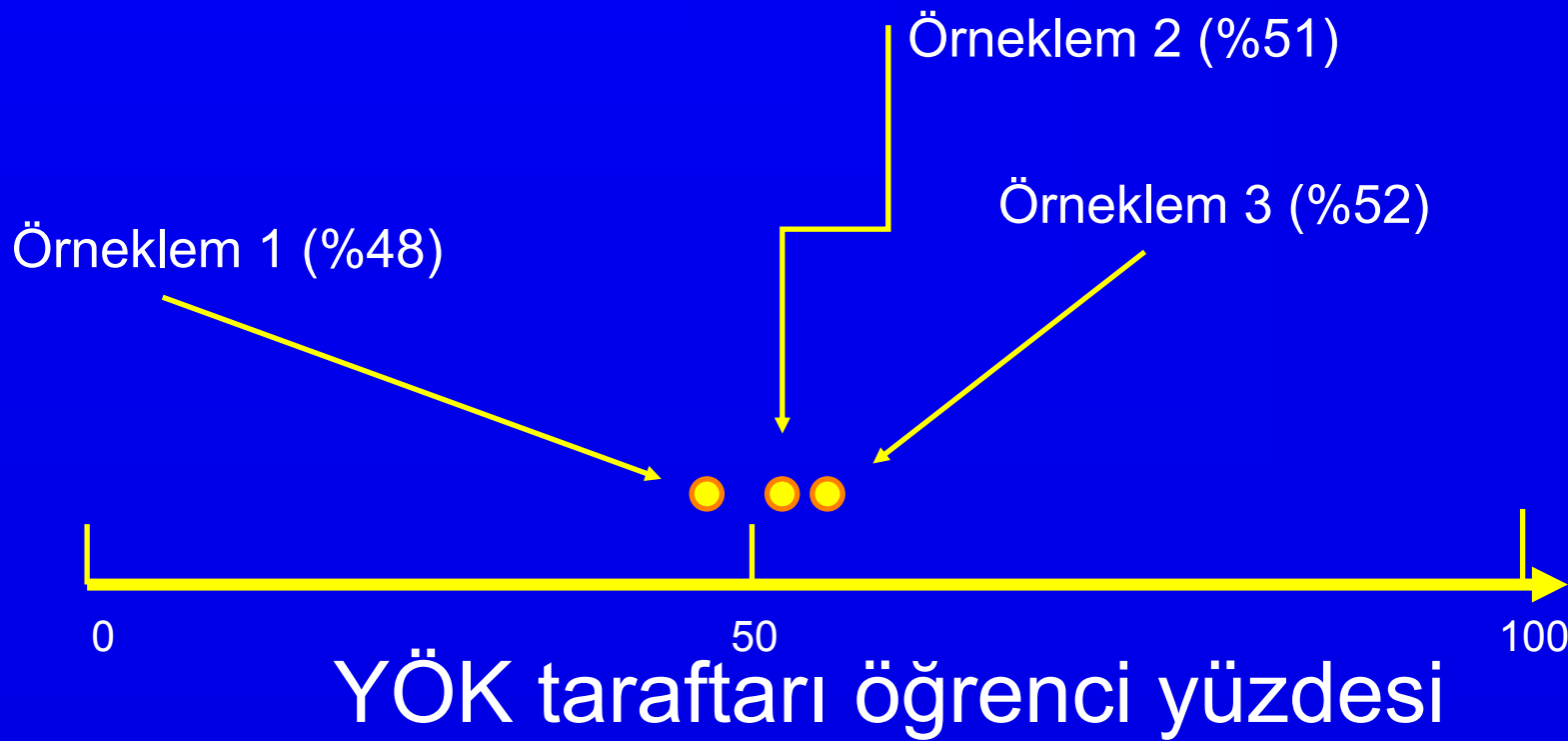


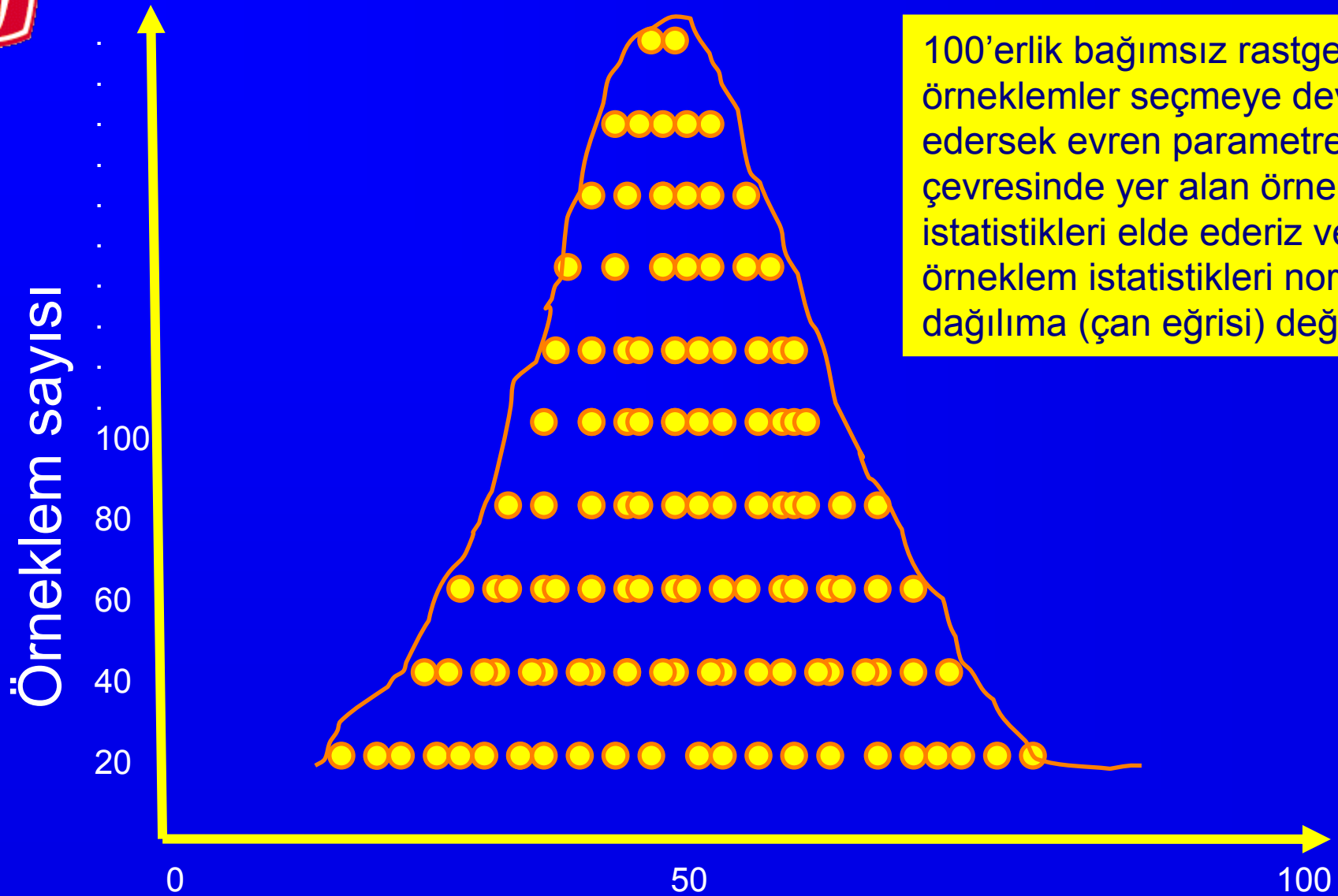
Öğrencilerin YÖK'e karşı tutumu

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

Öğrencilerin YÖK'e karşı tutumu...

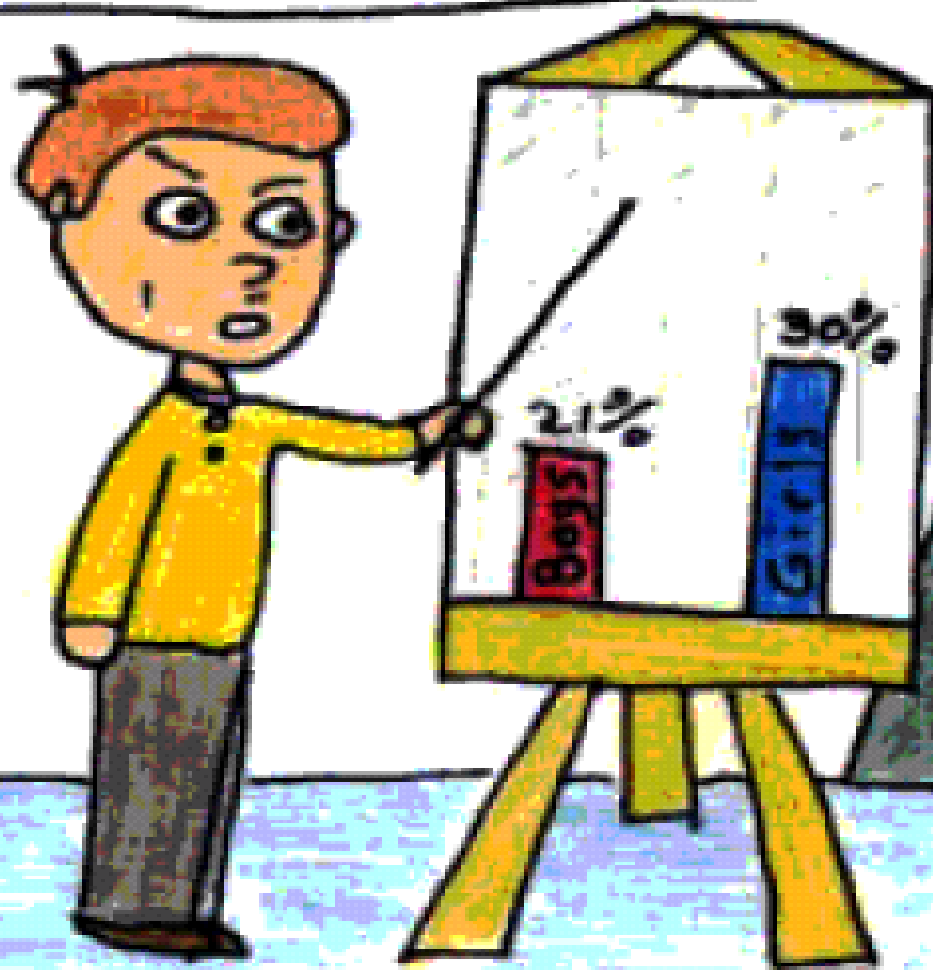
- ➔ Arařtırma evreni: 20.000 HÜ öğrencisi
- ➔ Örneklem çerçevesi: Öğrenci listesi
- ➔ Ögeler: Birey olarak öğrenciler
- ➔ Değişken: Öğrencinin YÖK'e karşı tutumu (ikili değişken)
- ➔ Her öğrenciye bir numara verelim
- ➔ Rastgele 100 öğrenci seçelim





100'erlik bağımsız rastgele örneklem seçmeye devam edersek evren parametresi çevresinde yer alan örneklem istatistikleri elde ederiz ve örneklem istatistikleri normal dağılıma (çan eğrisi) değişir

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



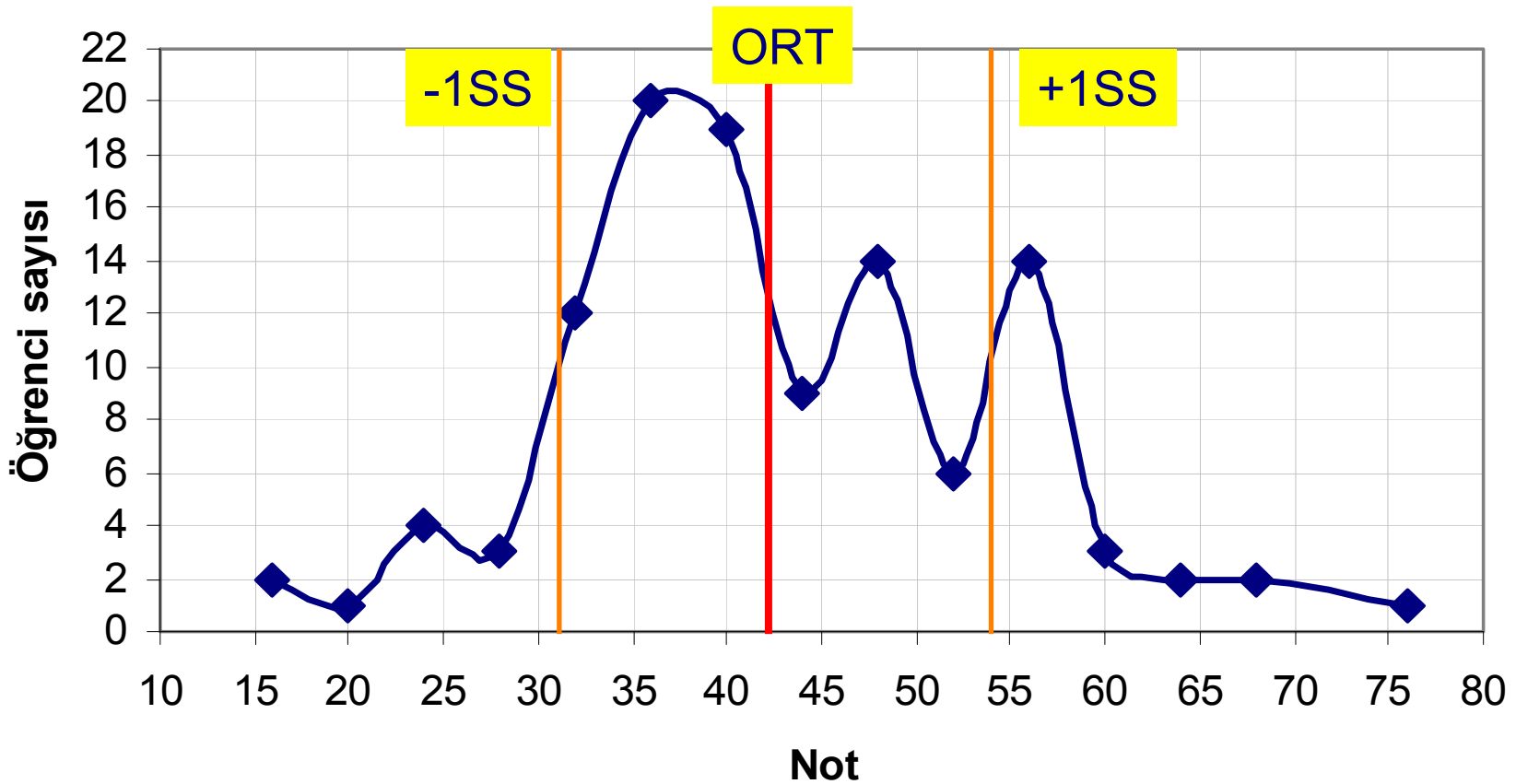
Örnekleme Dağılımı

- ➔ 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 artırılırsa daha isabetli örneklem ortalaması tutturabilirsiniz.

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



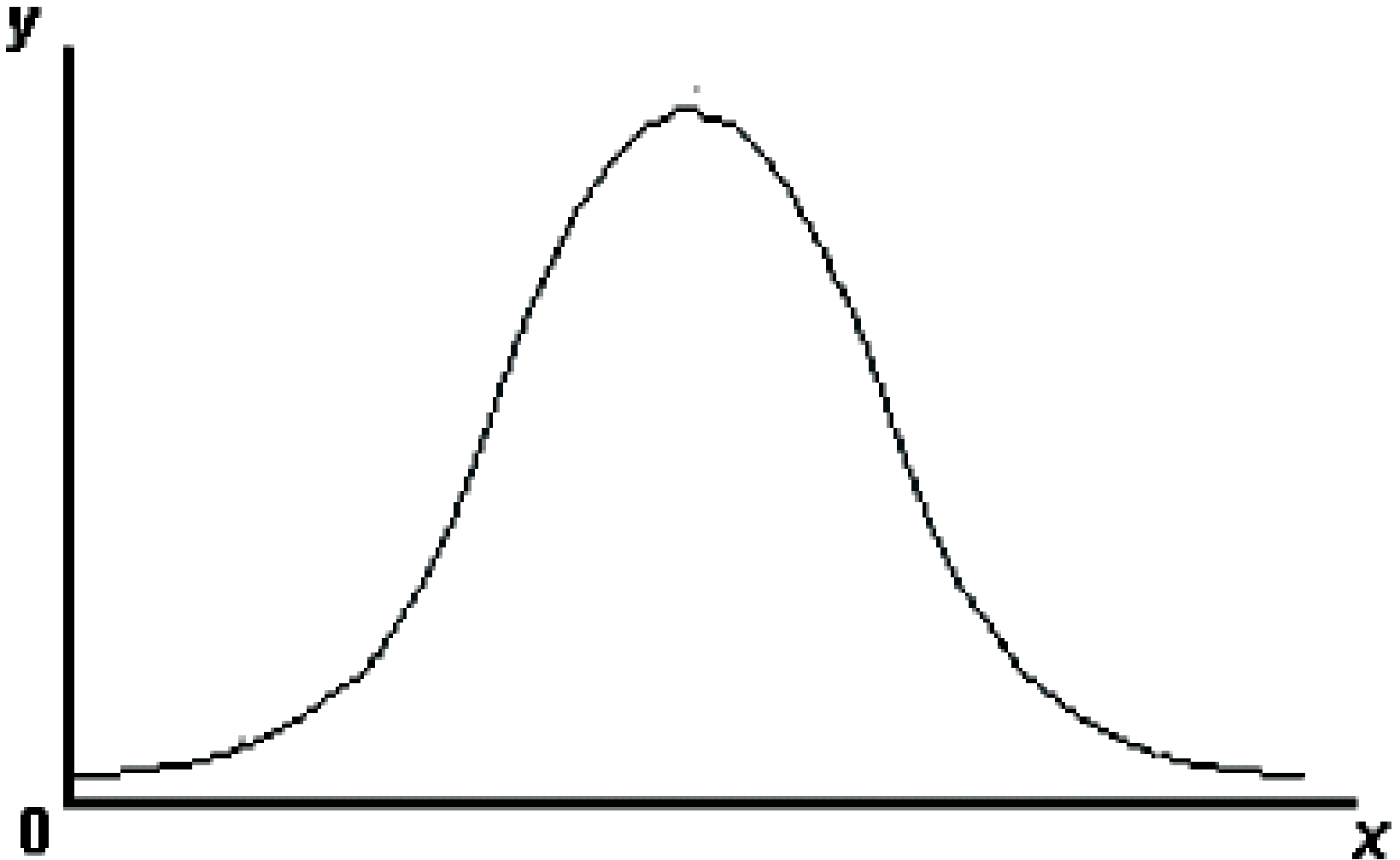
BBY 207 Sınav Sonuçları



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

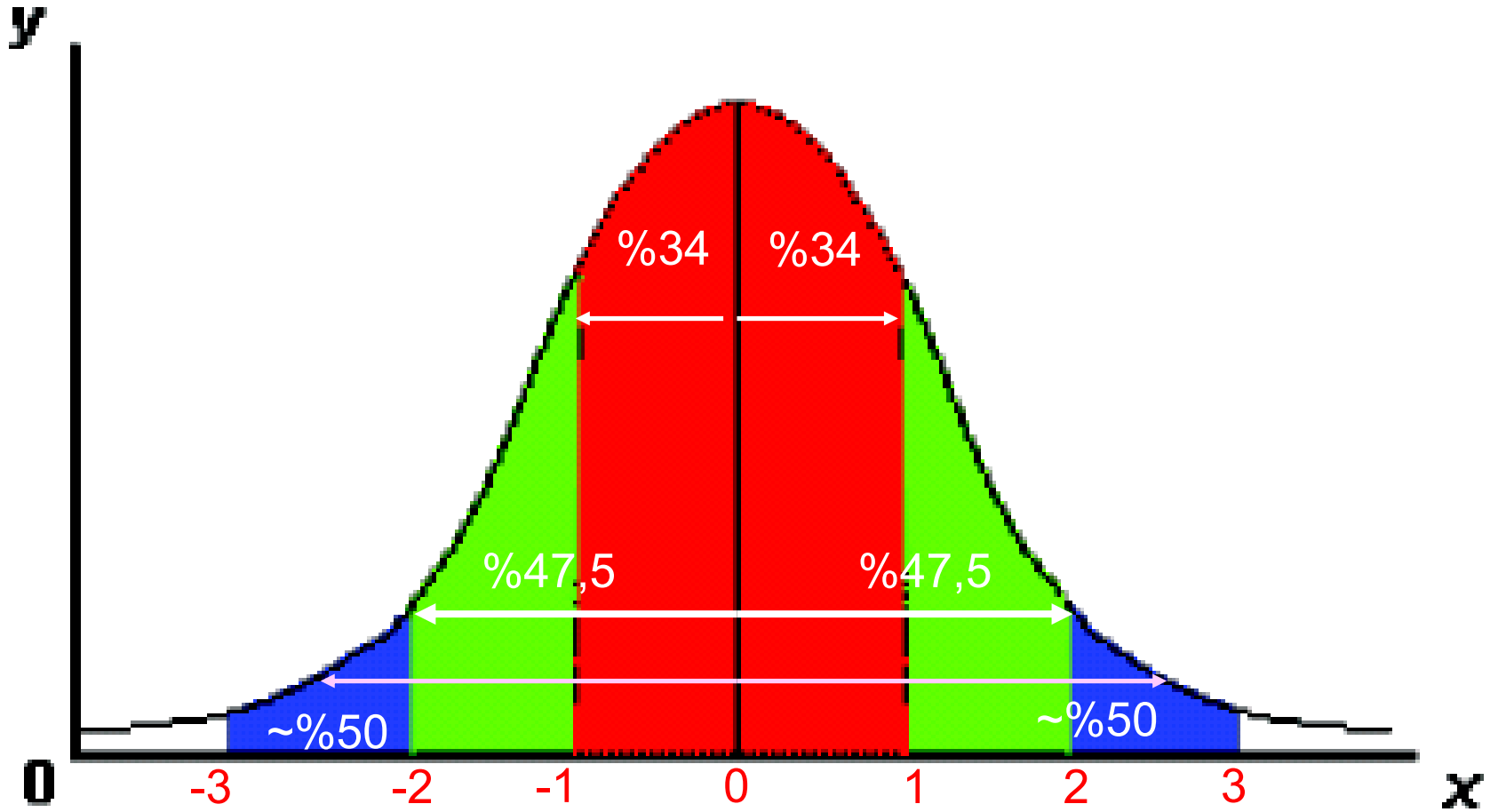


Normal Dağılım





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)



Örnekleme Dağılımı I

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 artırılırsa daha isabetli örneklem ortalaması tutturabilirsiniz.



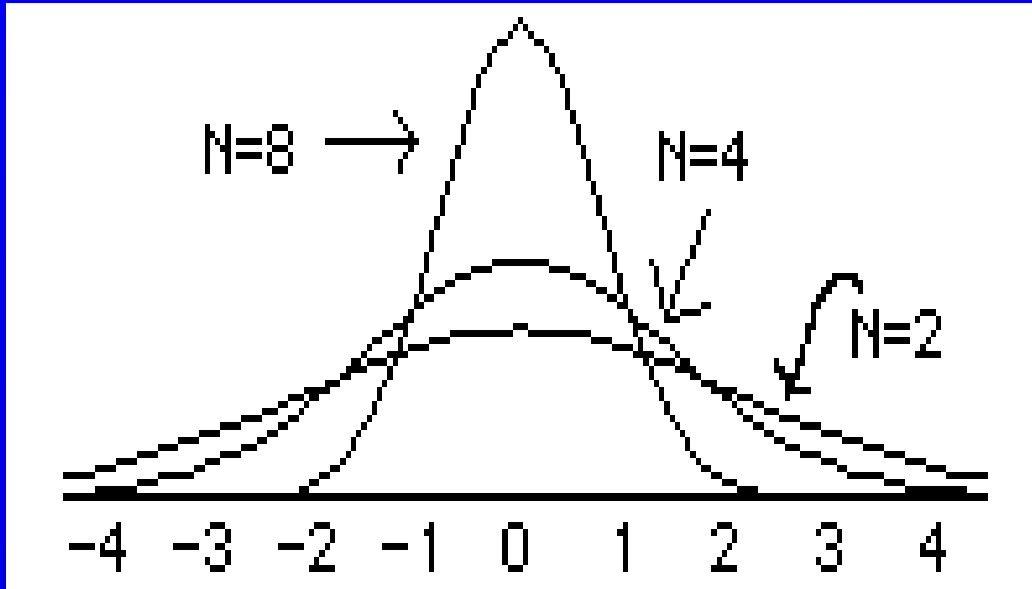
Örneklem Dağılımı II

Örneklem büyüklüğü arttıkça standart hata azalır.

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

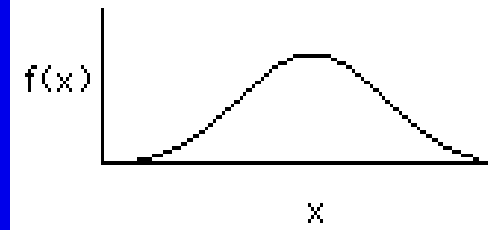
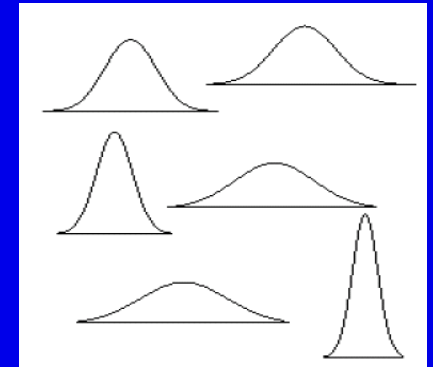
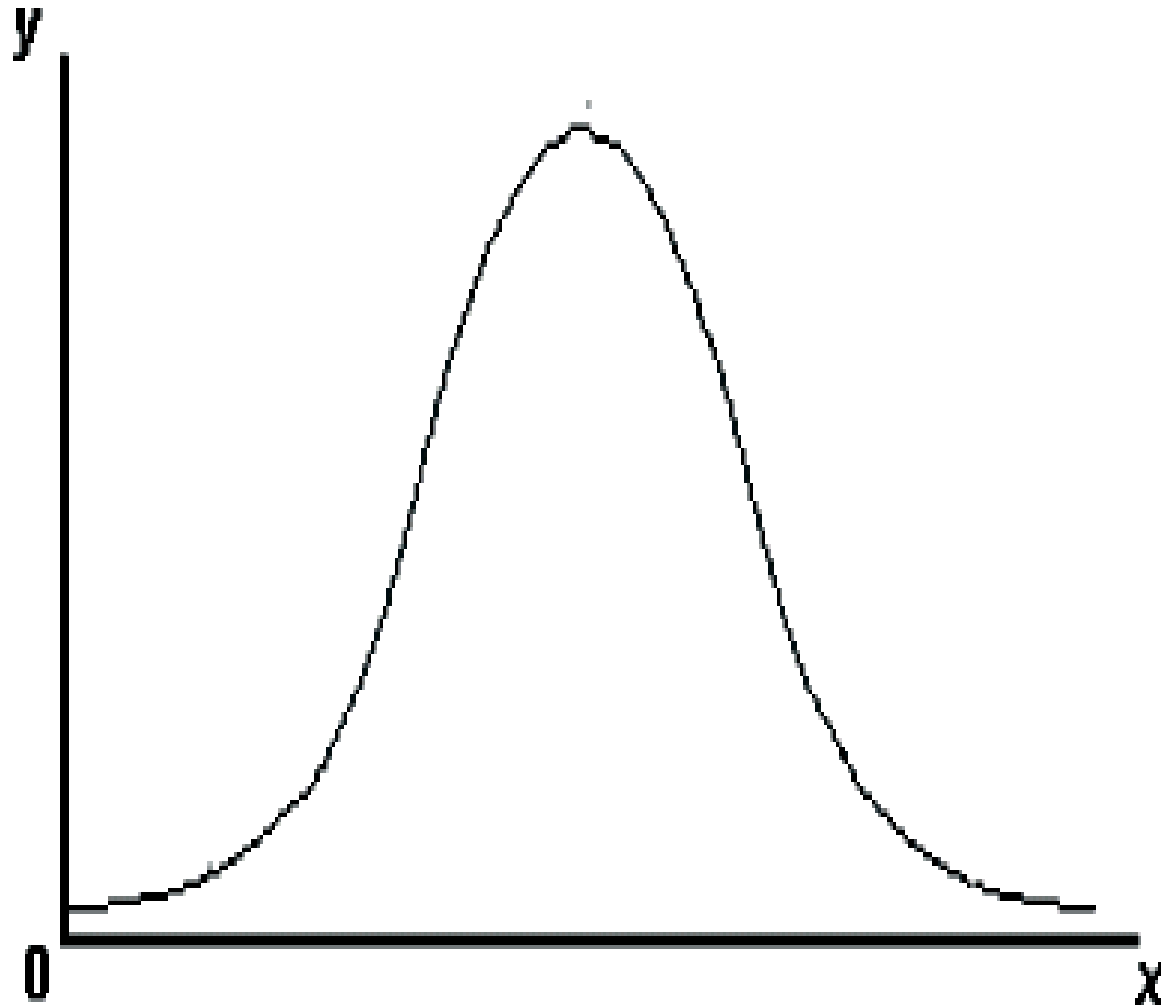
Örneklemin ortalaması μ , SS'si $\frac{\sigma}{\sqrt{N}}$ olur (N = örneklem büyüklüğü)

Örneklemin standart sapması ortalamasının standart hatası olarak bilinir.





Normal Dağılım



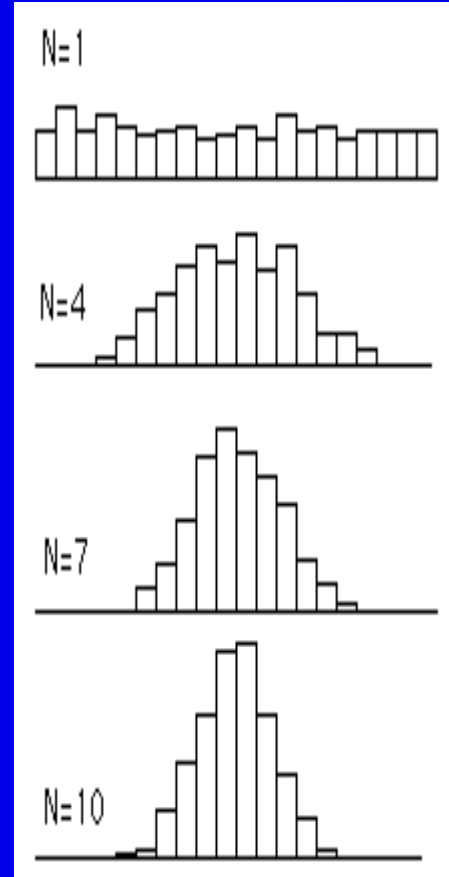
$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

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



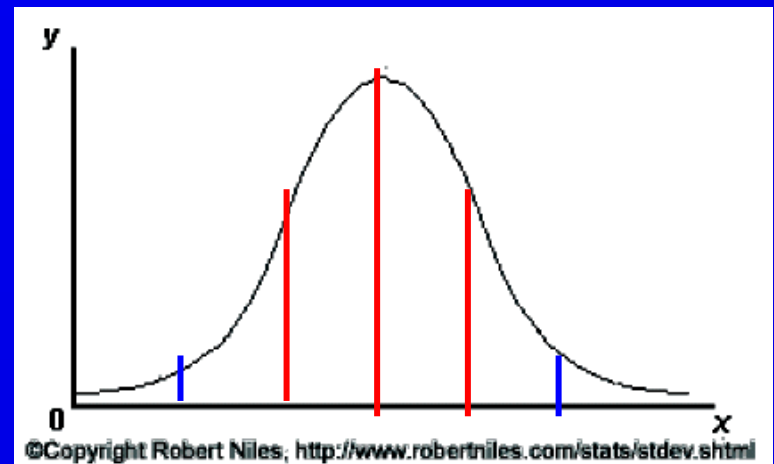
???

- ➔ 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 örneği I ($n = 100$, $s = 0,05$)

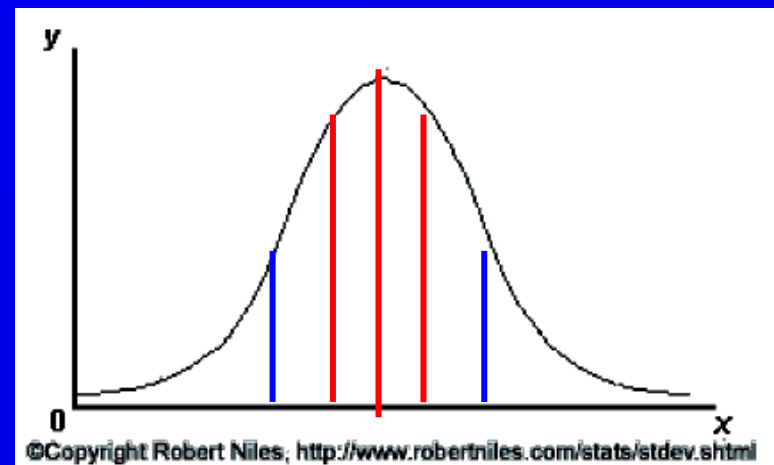
- ➔ %95 güven düzeyinde öğrencilerin %40-%60 arasında (± 2 SH) YÖK taraftarı olduğu söylenebilir
- ➔ %40-%60 güven aralığıdır
- ➔ %68 güven düzeyinde güven aralığı %45-%55 olur
- ➔ Örneklem istatistiğine dayalı tahmin yapılırken hem güven düzeyi hem de güven aralığı belirtilmelidir
- ➔ Standart hata oranına karar verdikten sonra örneklem büyüklüğü saptanabilir





YÖK örneği 2 ($n = 400, s = 0,025$)

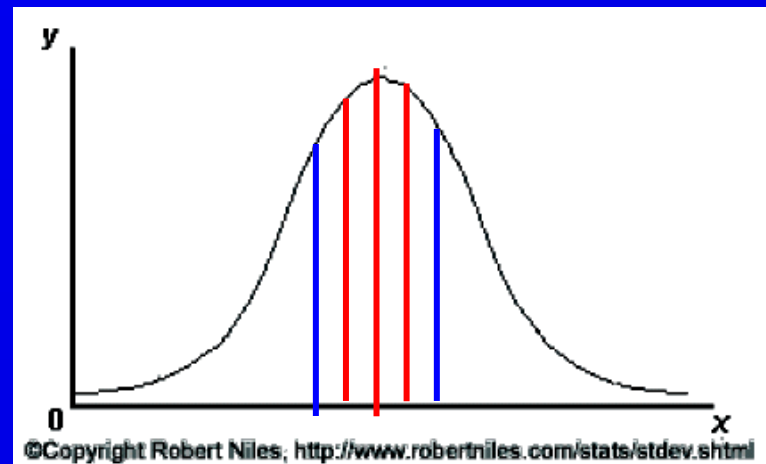
- ➔ %95 güven düzeyinde öğrencilerin %45-%55 arasında ($\pm 2 SH$) YÖK taraftarı olduğu söylenebilir
- ➔ %45-%55 güven aralığıdır
- ➔ %68 güven düzeyinde güven aralığı %47,5-%52,5 olur





YÖK örneği 3 ($n = 1600$, $s = 0,0125$)

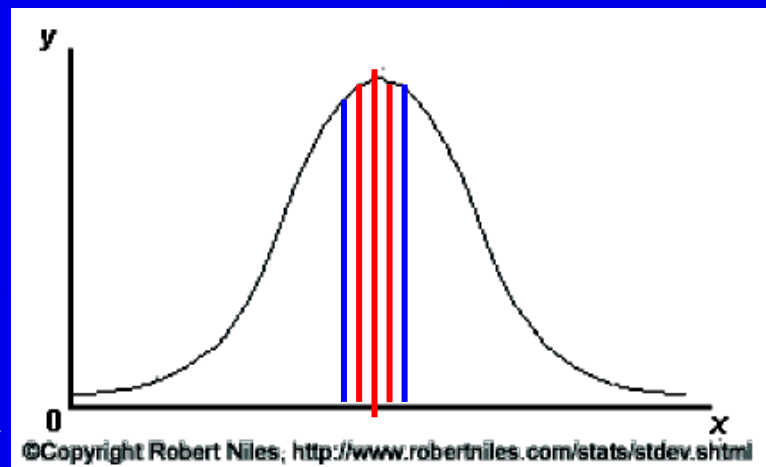
- ➔ %95 güven düzeyinde öğrencilerin %47,5-%52,5 arasında (± 2 SH) YÖK taraftarı olduğu söylenebilir
- ➔ %47,5-%52,5 güven aralığıdır
- ➔ %68 güven düzeyinde güven aralığı %48,75-%51,25 olur





YÖK örneği 4 ($n = 6400$, $s = 0,00625$)

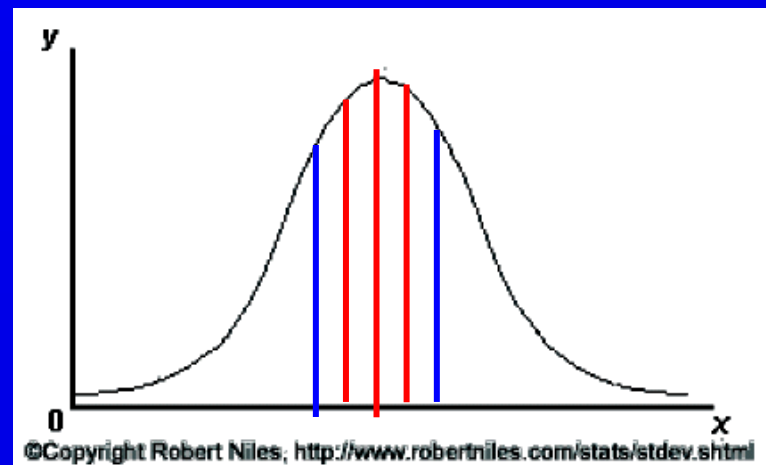
- ➔ %95 güven düzeyinde öğrencilerin %48,75- %51,25 arasında (± 2 SH) YÖK taraftarı olduğu söylenebilir
- ➔ %48,75-%51,25 güven aralığıdır
- ➔ %68 güven düzeyinde güven aralığı %49,375-%52,125 olur





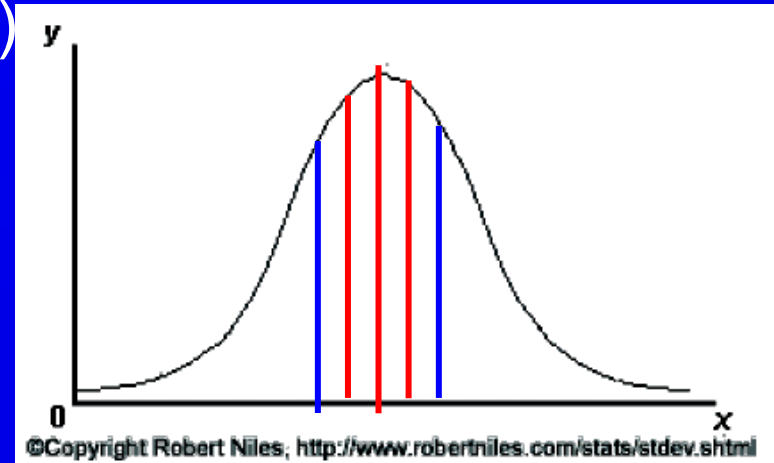
Sigara İçme - Akciğer Kanseri İlişkisi – I

- Örneklem büyüklüğü = 1600
- Örneklem istatistiği: %90
- Standart Hata: %1,25
- Sigara içenlerin %90'ı kansere yakalanıyor
- %95 güven düzeyinde sigara içenlerin %87,5-%92,5'unun (± 2 SH) kansere yakalanacağı söylenebilir
- %87,5-%92,5 güven aralığıdır
- %68 güven düzeyinde güven aralığı %88,75-%91,25 olur



Sigara İçme - Akciğer Kanseri İlişkisi – II

- Örneklem büyüklüğü = 400
- Örneklem istatistiği: %85
- Standart Hata: %2,5
- Sigara içenlerin %85'i kansere yakalanıyor
- %95 güven düzeyinde sigara içenlerin %80-%90'ının (± 2 SH) kansere yakalanacağı söylenebilir
- %80-%90 güven aralığıdır
- %68 güven düzeyinde güven aralığı %82,5-%87,5 olur



Sigara İçme - Akciğer Kanseri İlişkisi – III

Örneklem büyüklüğü = 1600

Örneklem istatistiği: %88

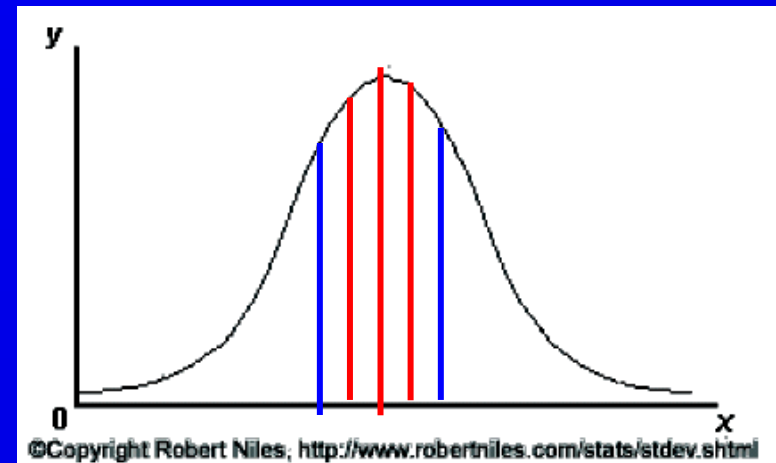
Standart Hata: %1,25

Sigara içenlerin %88'i kansere yakalanıyor

%95 güven düzeyinde sigara içenlerin %85,5-%90,5'unun (± 2 SH) kansere yakalanacağı söylenebilir

%88,5-%90,5 güven aralığıdır

%68 güven düzeyinde güven aralığı %87,25-%89,75 olur

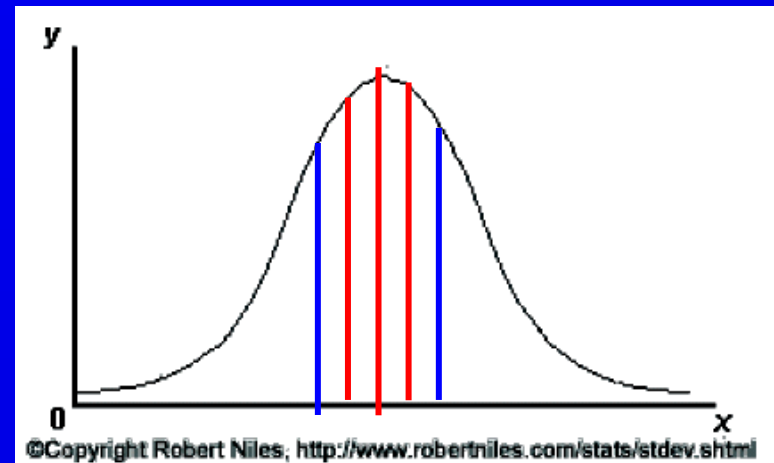


Sigara İçme - Akciğer Kanseri İlişkisi – Sonuç

- Diyelim ki evren parametresini bilmiyorum (yani gerçekte bütün sigara içenlerin kaçta kaçını kansere yakalıyor bilmiyorum)
- Belki hiçbir zaman da bilemeyeceğim.

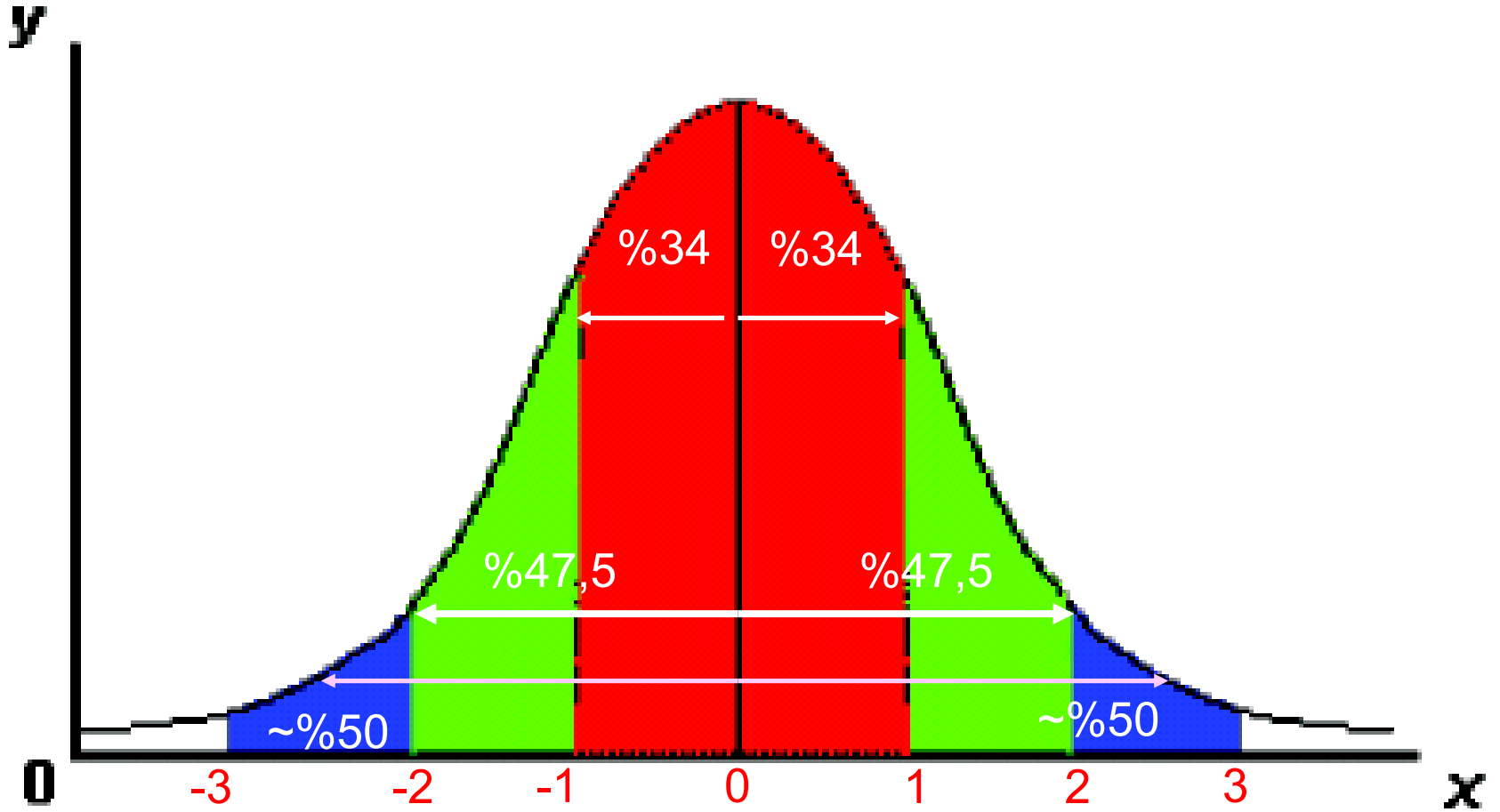
➡ **NE FARKEDER???**

➡ Örneklem istatistiğini evren parametresi olarak kabul etsem ne kadar yanılabilirim?





Normal Dağılım



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



Standart Normal Dağılım

SND aritmetik ortalaması 0,
standart sapması 1 olan bir normal dağılımdır.

Normal dağılımlar yandaki formül kullanılarak SND'ye çevrilebilir:

$$z = \frac{X - \mu}{\sigma}$$

Formülde

X özgün normal dağılımdan bir değer,
 μ özgün dağılımın aritmetik ortalaması,
 σ özgün dağılımın standart sapmasıdır.

SND bazen Z dağılımı olarak da adlandırılır.

Z değeri belirli bir değer aritmetik ortalamadan kaç standart sapma aşağıda ya da yukarıda olduğunu belirlemek için kullanılır.

Örneğin: Notların normal dağıldığı ve sınıf ortalamasının (μ) 50 olduğu bir sınavdan 70 (X) almış olun. Standart sapma (σ) 10 olsun.

Bu durumda sınıf ortalamasından 2 standart sapma daha yüksek not almış olursunuz.

$$z = \frac{70-50}{10} = 2$$



Z tablosu

- ➔ Artı eksi 3.49 arasında deęiřiyor.
- ➔ Bu, teorik evrenin %99.96'sına karşılık geliyor.
- ➔ Z tablosu 1/10'luk aralarla standart sapmayı gösteriyor
- ➔ Örneęin, en üst satır -3.4, -3.41, -3.42 .. SS'yi gösteriyor
- ➔ Arařtırmacılar z tablosundaki birkaç deęerle ilgili. Çünkü çoęu hipotez testlerinde %95 ve %99'luk alanlarla ilgileniyor.



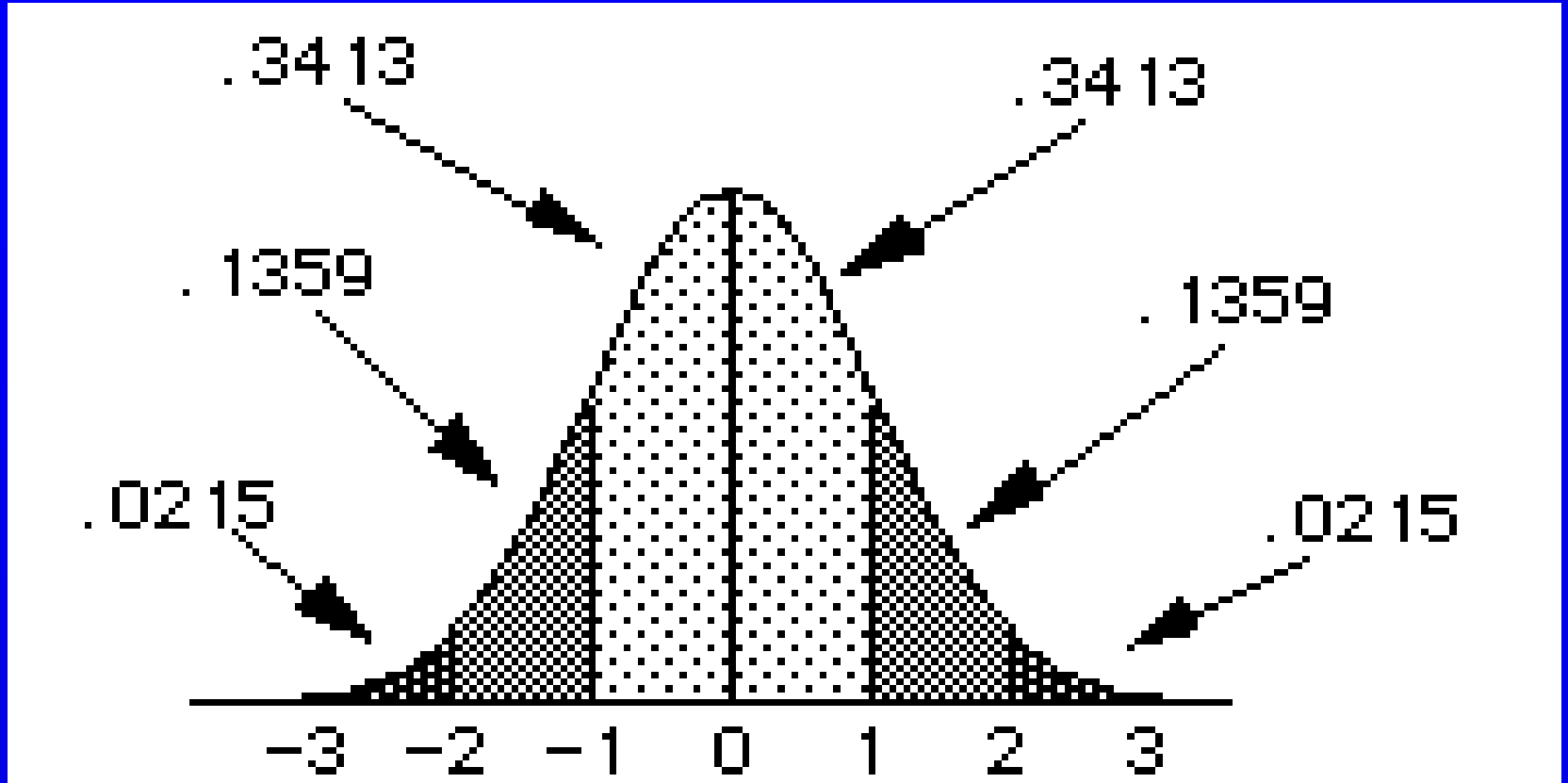
Güven Aralıkları

- ➔ Örneklem istatistikleri belirli bir güven düzeyinde evrene genellenebilir.
- ➔ Çünkü bilinen olasılıklara dayanıyor
- ➔ SND'de ölçümlerin yüzde 68'i ± 1 SS, %96'sı ± 2 SS, sadece %1'i $\pm 2,575$ SS dışında kalıyor
- ➔ Farklı örneklem istatistiklerinin de her birinin farklı SS'leri olabilir (buna standart hata diyoruz)
- ➔ Tek örneklem ortalaması birçok örneklem ortalamasından sadece biri ama güvenle diyebiliriz ki bu ortalama evren parametresine yakın olmalı
- ➔ %95 güven düzeyinde örneklem ortalaması evren parametresinden 1,96, %99 güven düzeyinde 2,575 standart hata uzaklıktadır



$$z = \frac{X - \mu}{\sigma}$$

Formül her zaman ortalaması 0, SS'si 1 olan bir dağılım üretir.
X değerinin alındığı dağılım normal değilse, bu, dönüştürüme de yansır.





Yüzdelerle Çevirme I

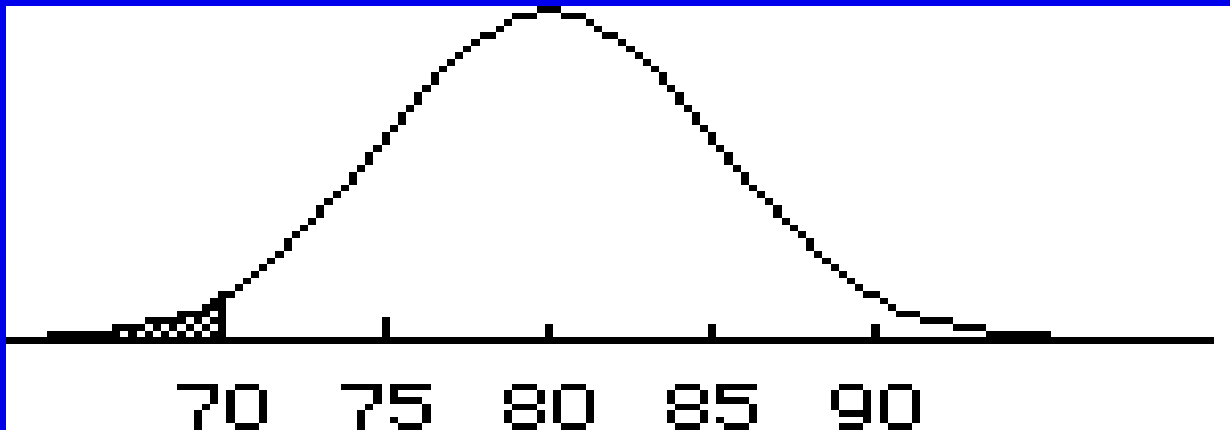
☞ Notların normal dağılım gösterdiği bir sınavdan 70 aldınız. (Ort= 80, SS=5) Sınıftaki yeriniz (yüzde olarak) neresidir?

$$Z = (70 - 80) / 5 = -2.$$

Ortalamanın 2 SS altında.

Sınıftaki öğrencilerin sadece %2.15'i 70 ya da daha altında not almıştır.

$$z = \frac{X - \mu}{\sigma}$$





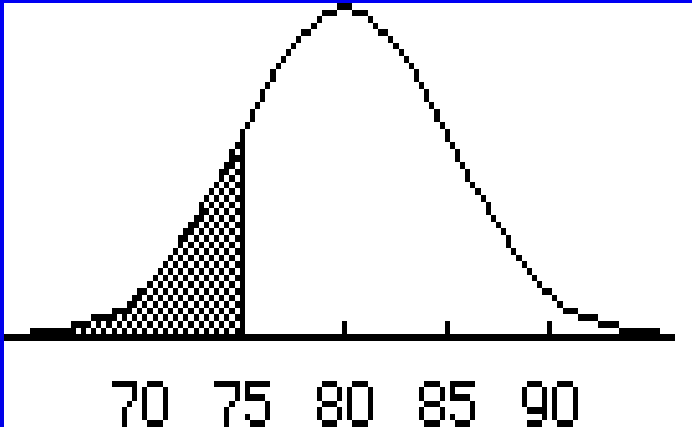
Yüzdelerle Çevirme II

Peki ya aynı sınavdan 75 almış olsaydınız?

Yani ortalamadan 1SS aşağıda?

Yani sınıfın sadece %15.9'u sizinle aynı ya da daha düşük not almış olurdu.

Z tablosu



z	Area from $-\infty$ to z
-3.0	.0013
-2.5	.0062
-2.0	.0227
-1.5	.0668
-1.0	.1587
-0.5	.3085
0.0	.5000
0.5	.6915
1.0	.8413
1.5	.9332
2.0	.9772
2.5	.9938
3.0	.9987



Yüzdelerle Çevirme III

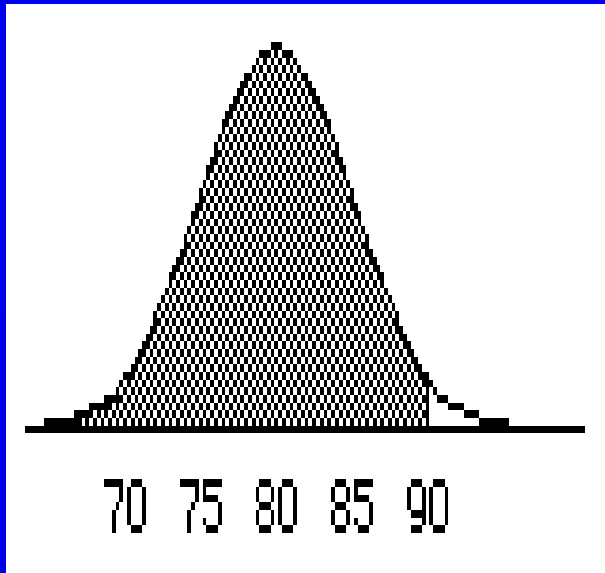
Z tablosu

Peki ya sınıf ortalamasının 2 SS üstünde not almış olsaydınız?

SS= 5 olduğuna göre notunuz $80 + 2*5 = 90$ olacaktı.

Z tablosundan 2 SS'e karşılık gelen yüzde 97.72.

Yani sınıfın üst %2'sindesiniz.



z	Area from $-\infty$ to z
-3.0	.0013
-2.5	.0062
-2.0	.0227
-1.5	.0668
-1.0	.1587
-0.5	.3085
0.0	.5000
0.5	.6915
1.0	.8413
1.5	.9332
2.0	.9772
2.5	.9938
3.0	.9987

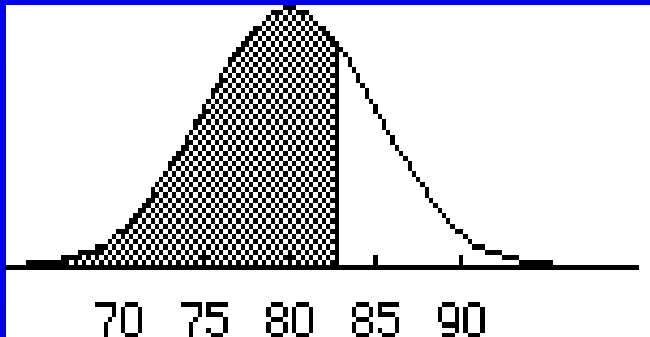


Yüzdelerle Çevirme IV

Peki hangi notu almış olsaydınız %75'lik dilimde olurdunuz?
Doğrudan z tablosu kullanılarak %75'e karşılık gelen z değeri (.7734) bulunur.

SS=5 olduğuna göre, ortalamanın $5 * .7734 = 3.87$ puan üstünde not almamız lazım (yani $80 + 3.87 = 83.87$).

Zaten $z = \frac{X - \mu}{\sigma}$ formülünü $X = \mu + z \sigma$ olarak ifade edebiliriz.
Bu formülle X değerini kolayca bulabiliriz
($X = 80 + .7734 * 5 = 83.87$).



Çan eğrisi altındaki alanın hesabı I

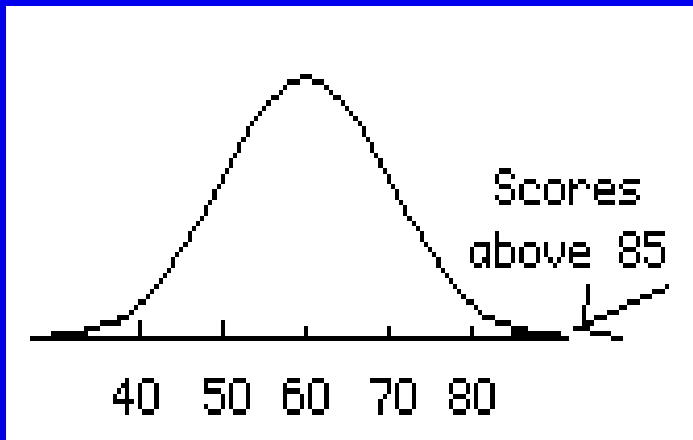
Ort = 60, SS = 10

Notların yüzde kaçı 85 ve üzerindedir?

$$85-60/10=2.5$$

Z tablosundan +2.5 standart sapma .9938'e karşılık geliyor.

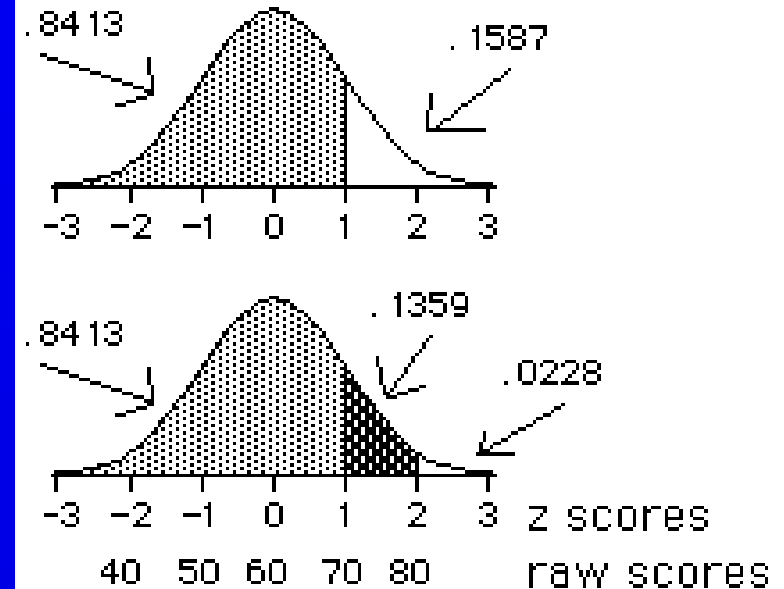
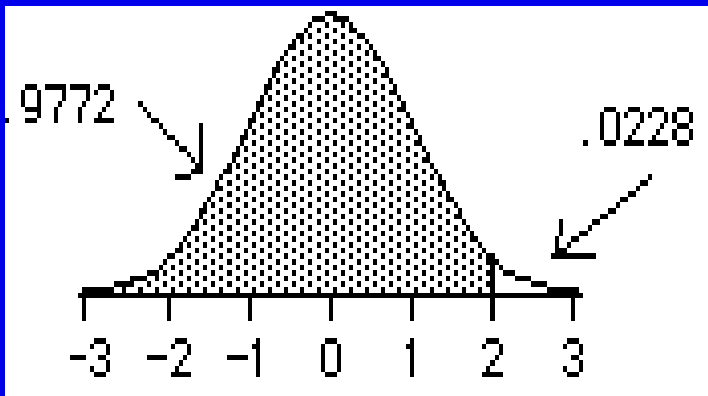
Yani öğrencilerin sadece % 0.62'si (binde 6'sı bu notun üzerinde not almıştır.

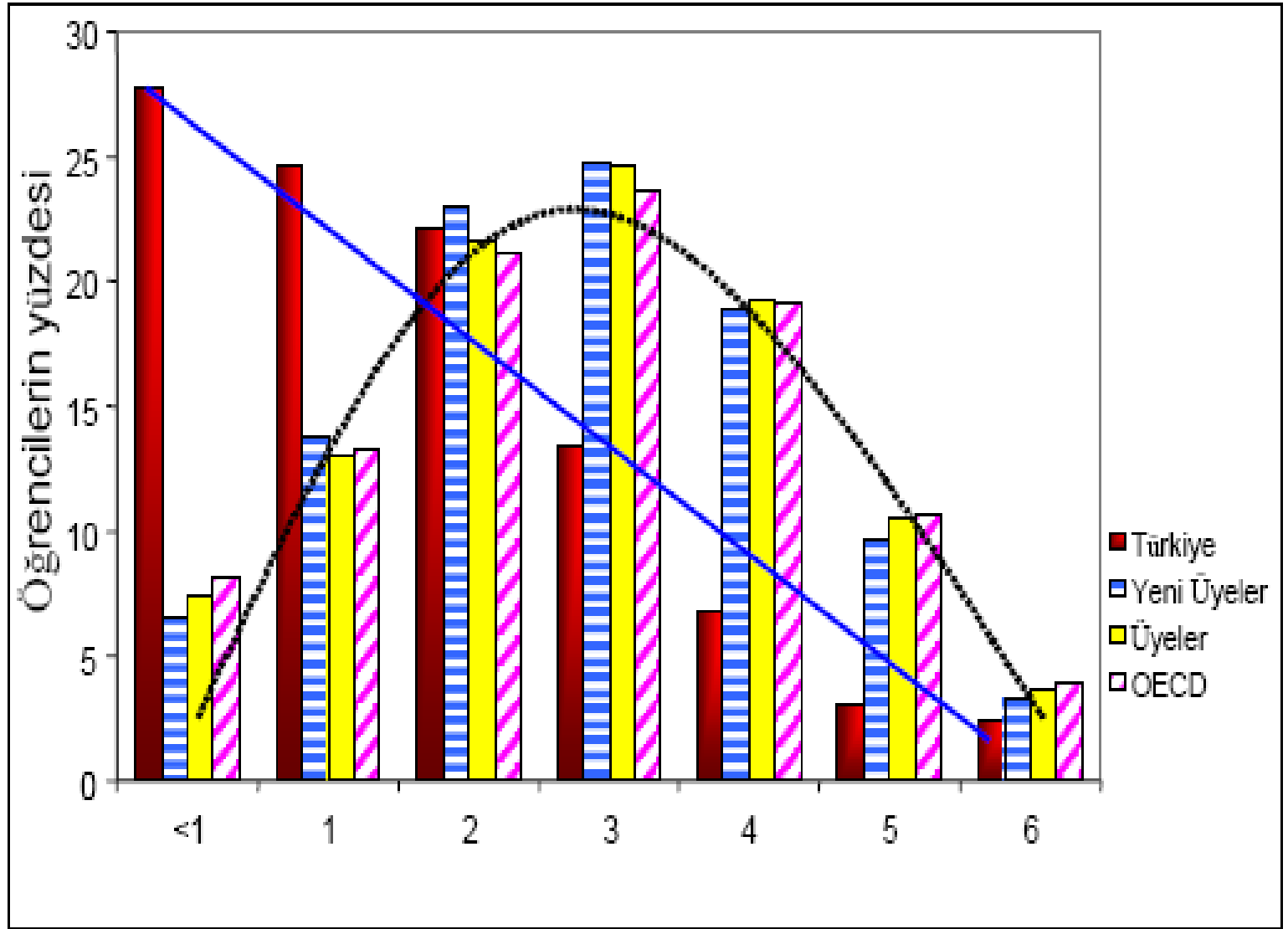




Çan eğrisi altındaki alanın hesabı II

Aynı sınavda 70 ile 80 arasında not alan öğrencilerin oranı nedir? Önce 80 ve daha az alanların oranını, sonra 70 ve daha az alanların oranını bul, birbirinden çıkar, sonuç 70 ile 80 arasında not alanların oranını verir. 80 ortalamamanın 2SS üstünde. Z tablosundan öğrencilerin %97.72'sinin 80 ve daha düşük not aldığını hesaplarız. 70 ortalamamanın 1SS üstünde. Z tablosundan öğrencilerin %84.13'ünün 70 ve daha düşük not aldığını hesaplarız. İkisi arasındaki fark %13.59.





Şekil 1.: PISA 2003, Matematik, 6'lı öğrenme yeterliği ölçeği üzerinden öğrenci performansı¹²²

¹²² PISA 2003, Projesi Ulusal Ön Rapor, MEB, 2004, Andrew Vorkink, Türkiye Eğitim Sektörü Konferansı, 17 Mart 2006, Ankara



Şekil 1'de, AB ve OECD üyeleri ile Türkiye'nin, Matematik için 6 öğrenme düzeyinde öğrenci performansının dağılımı gösterilmiştir. 1.düzye, en alt düzeyde matematik bilgisi performansın, 6.düzye ise en üst performansın ölçüsüdür. Öteki ülkelerde bu düzeyler açısından normal bir dağılım olmasına karşılık yalnız Türkiye için farklı bir dağılım söz konusudur. Bu dağılımın ortaya çıkardığı önemli gerçek, okullarımızın, çok az öğrenciyi iyi eğitirken öğrencilerin çoğunluğunu iyi eğitemediğidir. Arařtırmada yer alan öteki ülkelerin hiç birinde bulunmayan bu durum, eğitim sistemimizin çarpıklığı açısından üzerinde durulması gereken bir göstergedir.