

Örnekleme Mantığı II

Yaşar Tonta

H.Ü. Bilgi ve Belge Yönetimi Bölümü

tonta@hacettepe.edu.tr

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

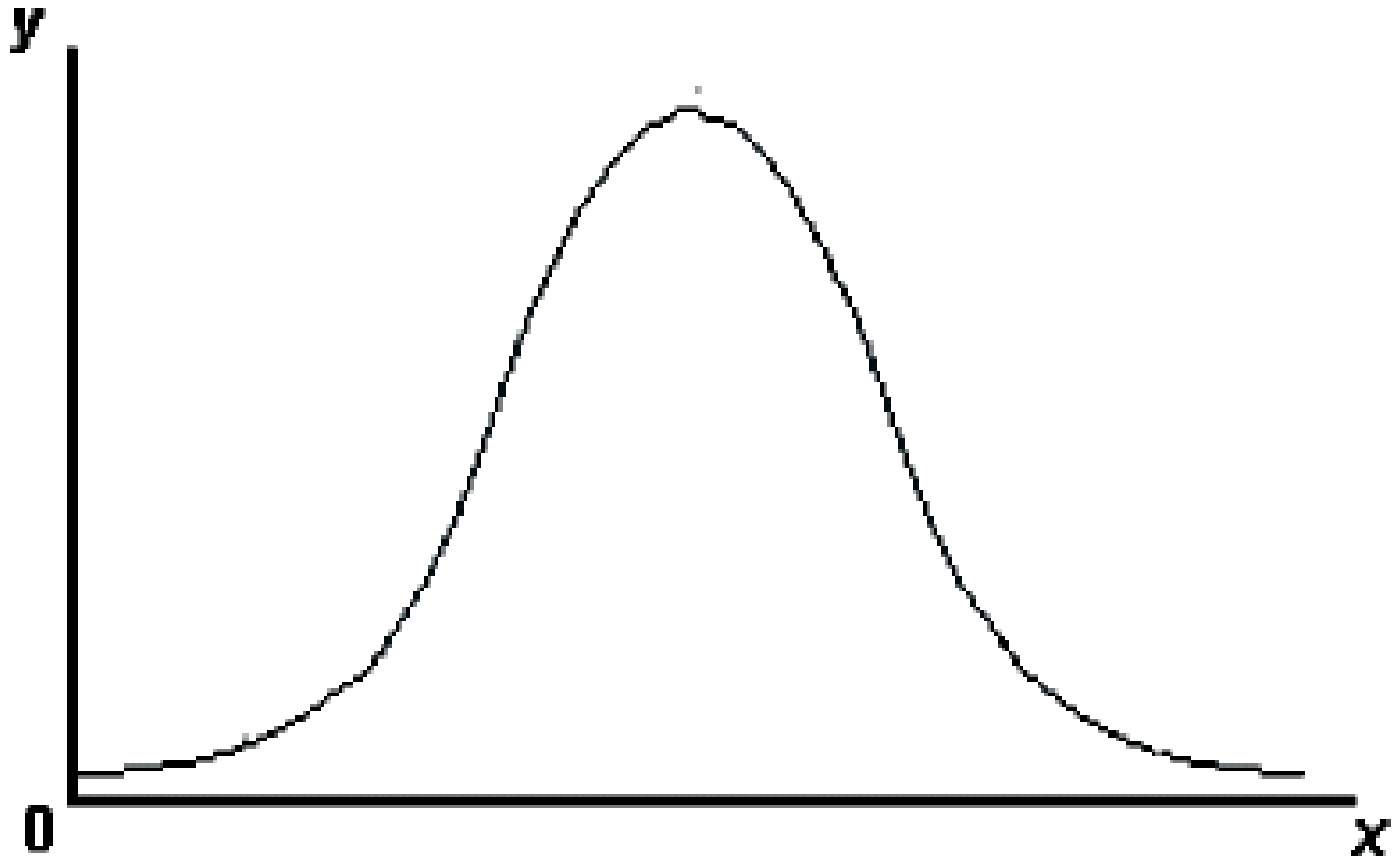
Not

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

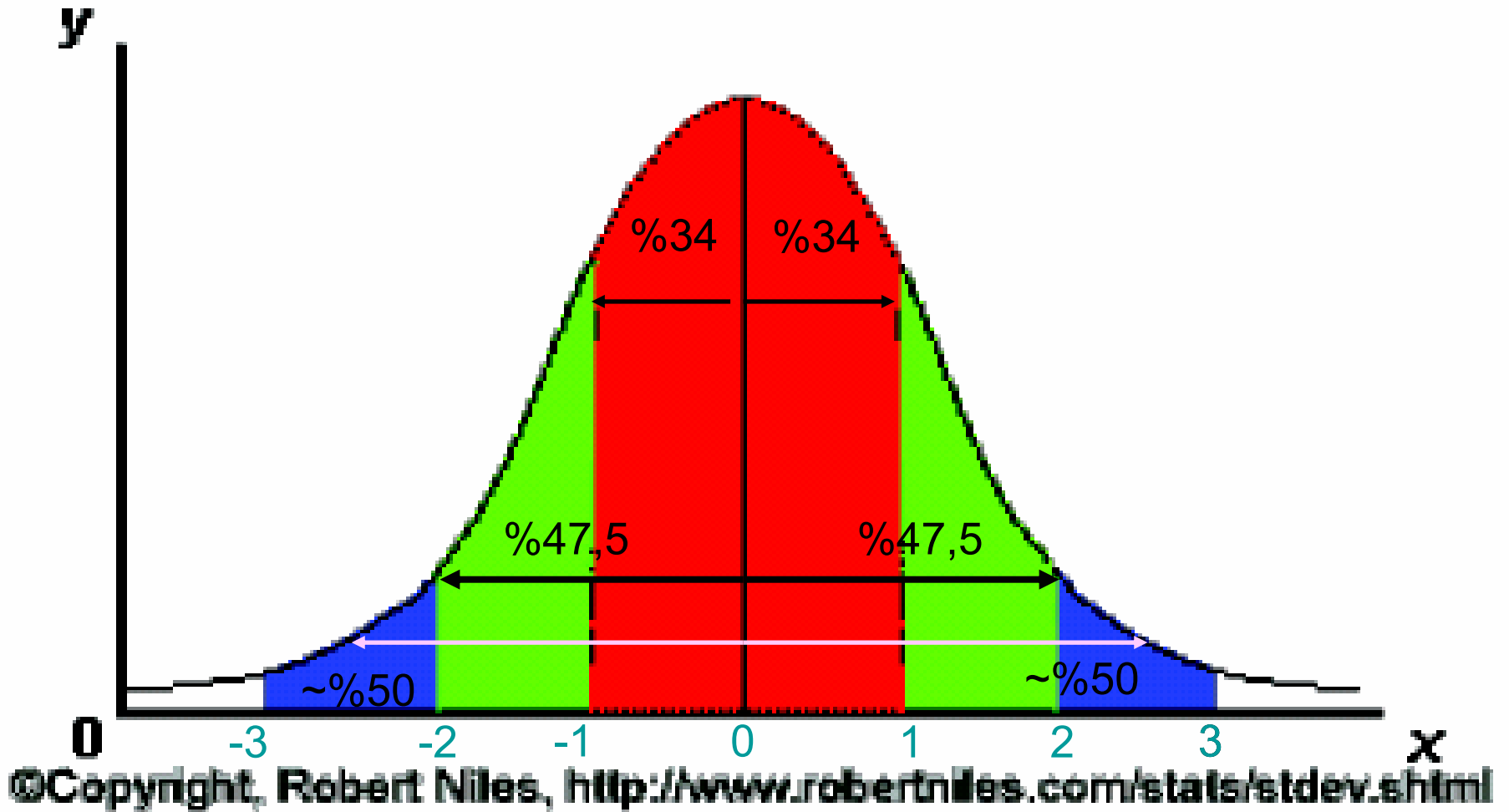
Plan

- Örneklem mantığı (devam)
- Verilerin sunumu

Normal Dağılım



Standart Sapma



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 örneği

- Varsayalım örnekleme'deki 100 öğrencinin yarısı YÖK taraftarı, yarısı değil
- 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 istatistiği 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)

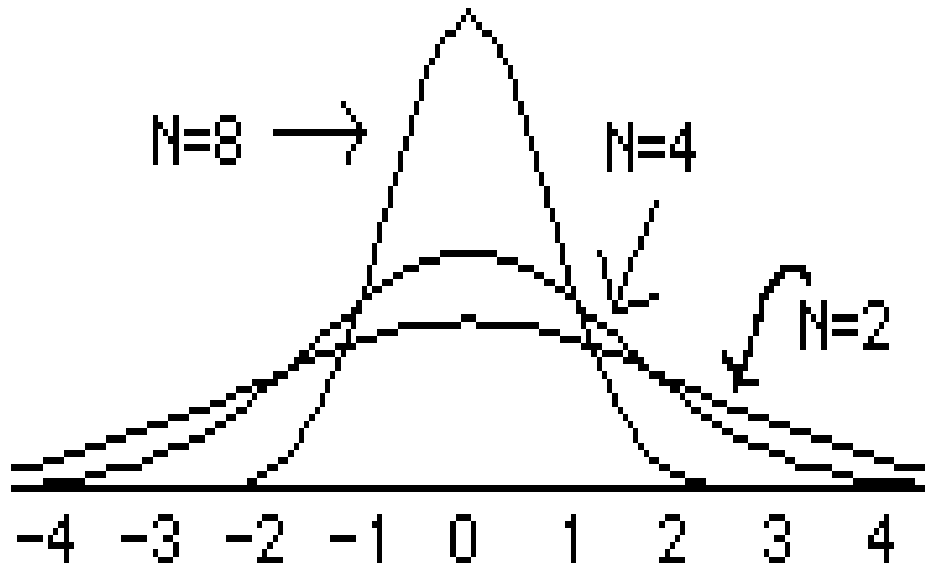
Örneklem Dağılımı II

Örneklem büyüklüğü arttıkça standart hata azalır.

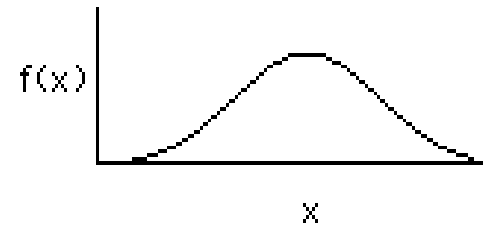
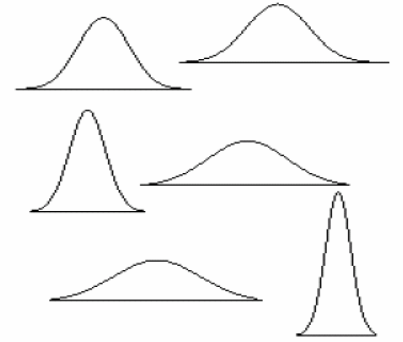
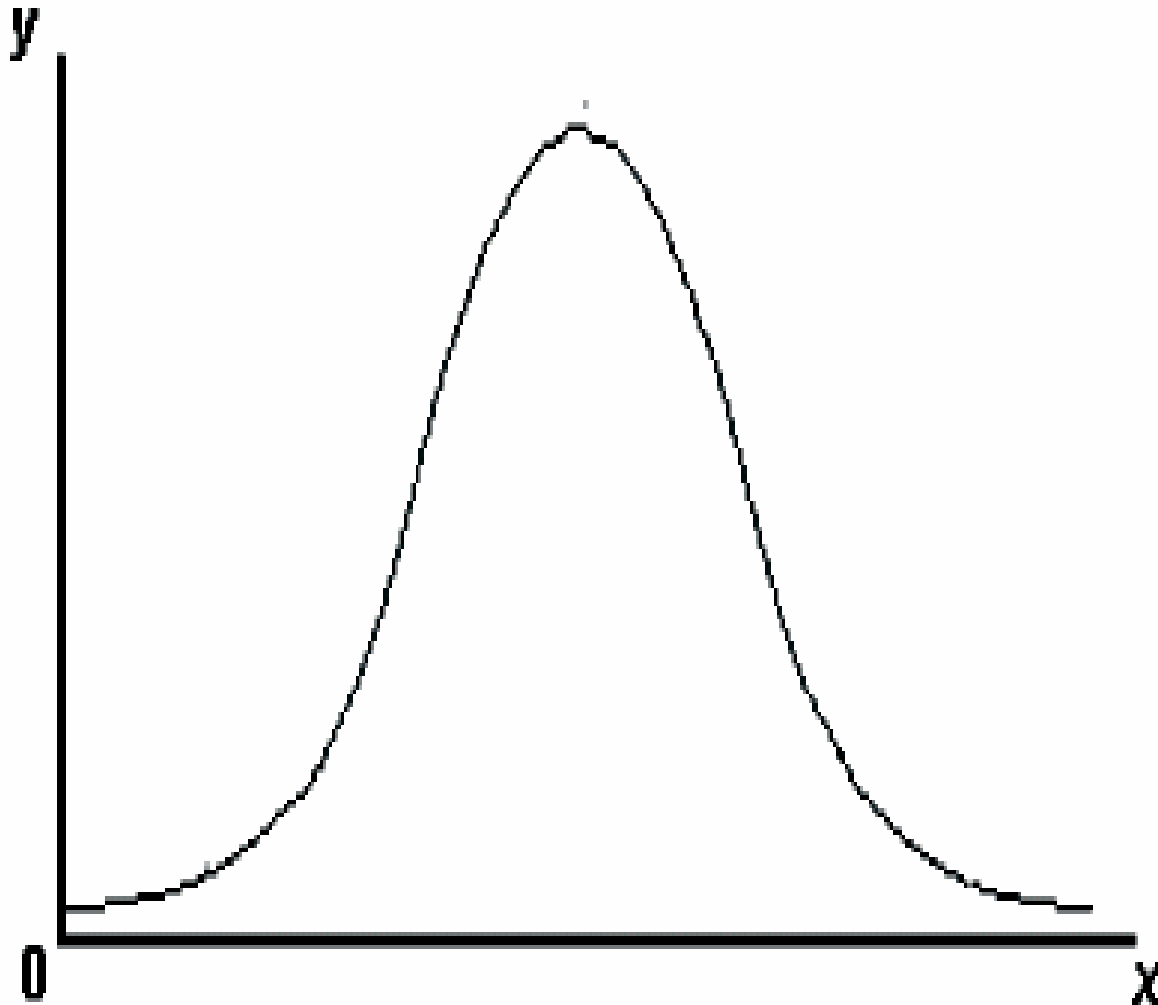
Ortalaması μ , SS'si σ olan bir evrenden bir örneklem seçerseniz,

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

Örneklemin standart sapması ortalamanın standart hatası olarak bilinir.



Normal Dağılım

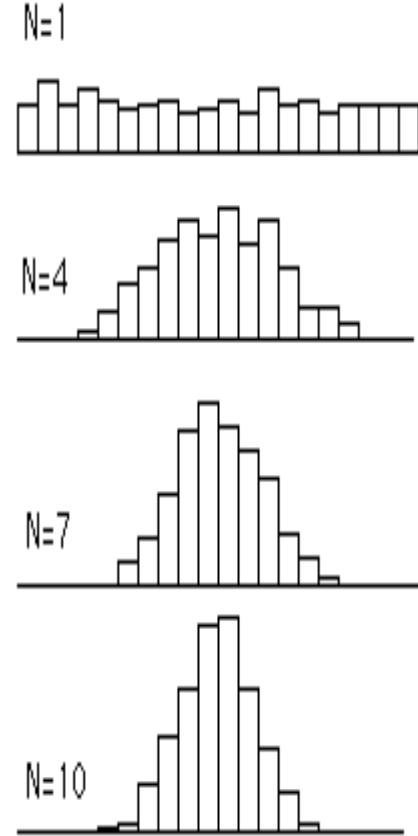


$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

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

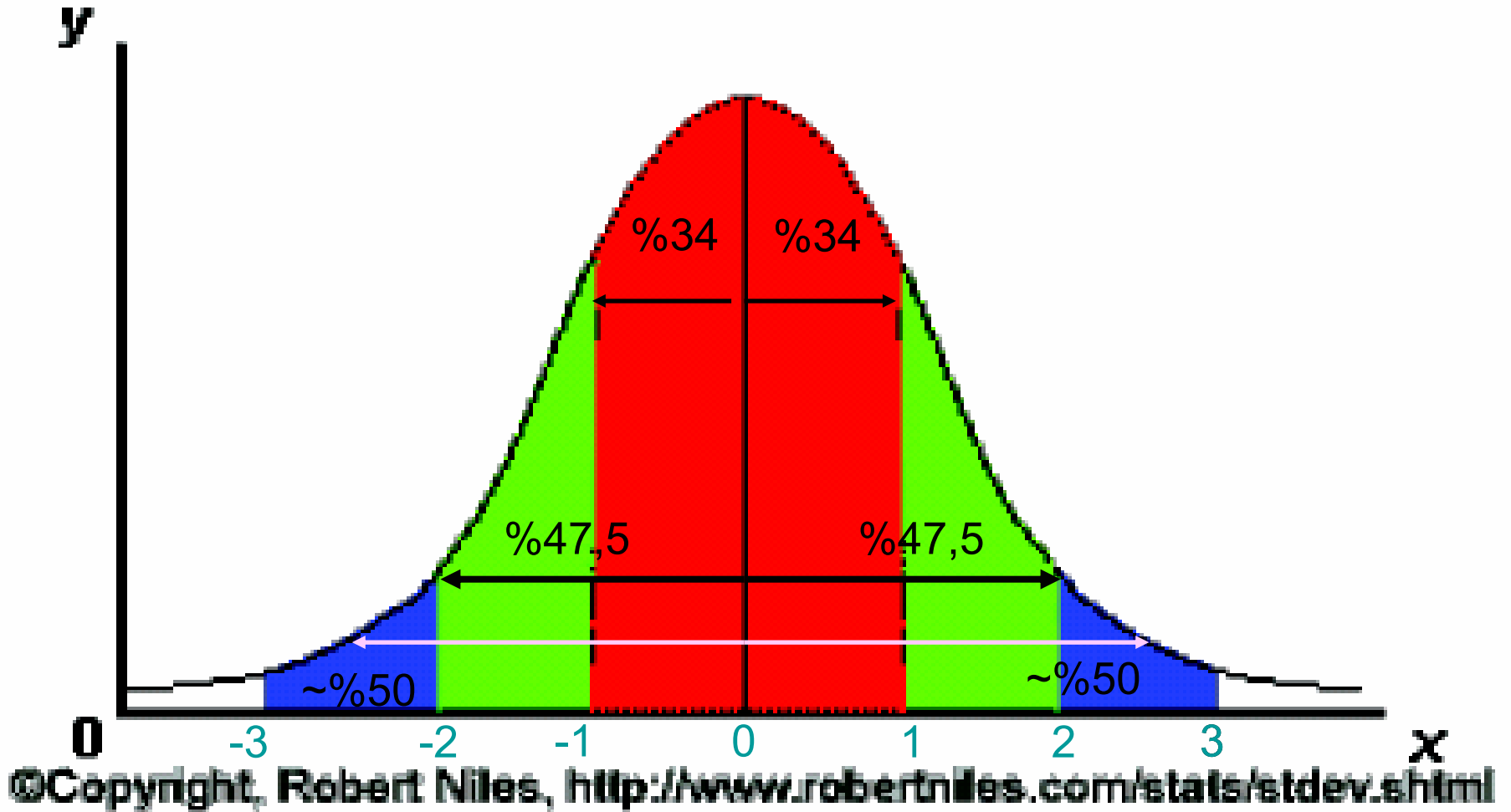
???

- Tek bir rastgele örneklem seçerek elde edilen istatistik evren parametresinin $1SH$ 'lık sınırlar içinde olma olasılığı %68.
- Buna güven düzeyi deniyor
- Yani %68 güvenle örneklem tahmini evren parametresinin $1SH$ içindedir
- Ya da %95 güvenle örneklem tahmini evren parametresinin $2H$ 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

Normal Dağılım



Standart Normal Dağılım

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

Normal dağılımlar aşağıdaki formül kullanılarak SN'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.

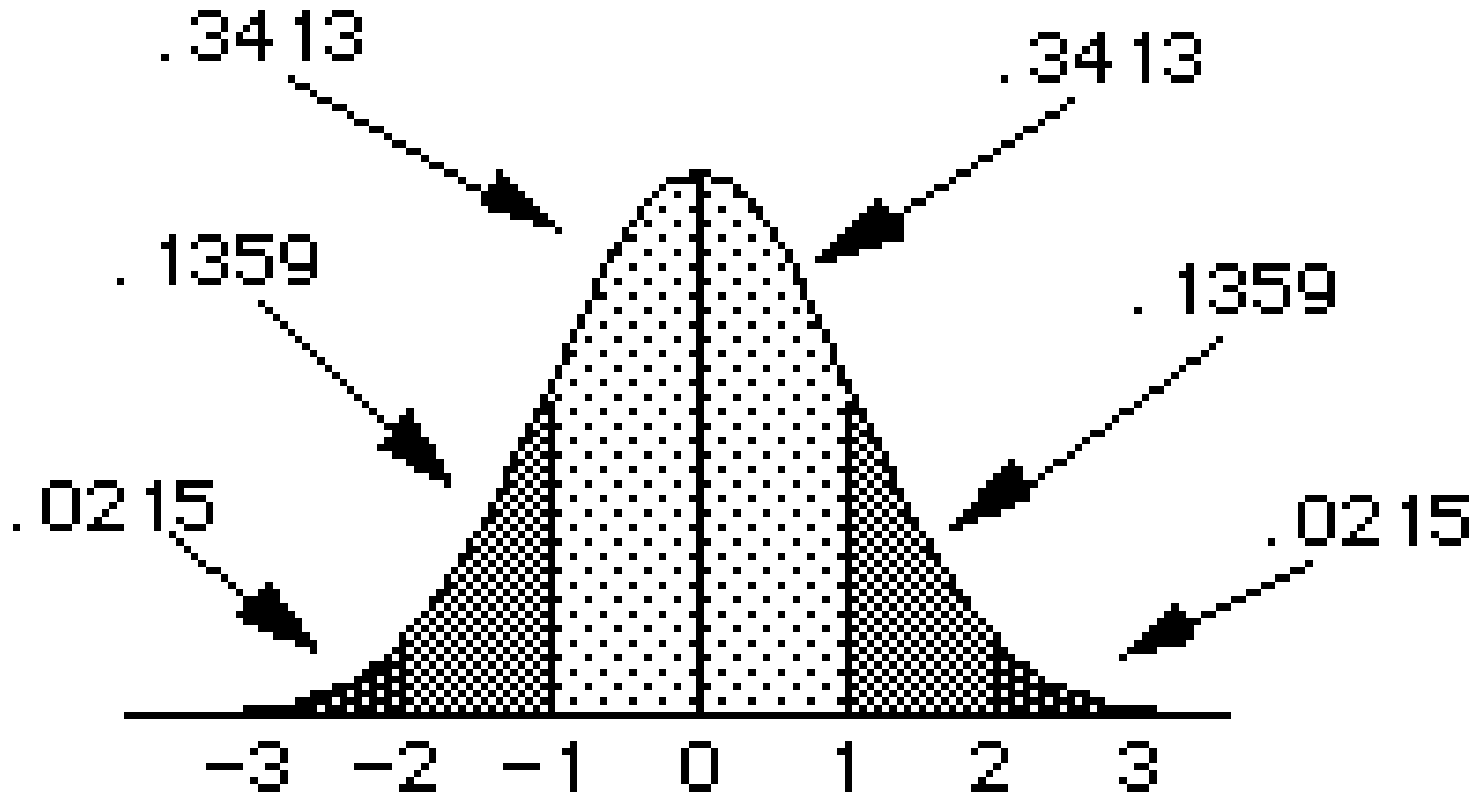
Örneğin: Notların normal dağıldığı ve sınıf ortalamasının (μ) 50 olduğu bir sınavdan 70 (X) almış olun. Standart sapma (σ) 10 olsun.

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

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

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

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



Yüzdelere Çevirme I

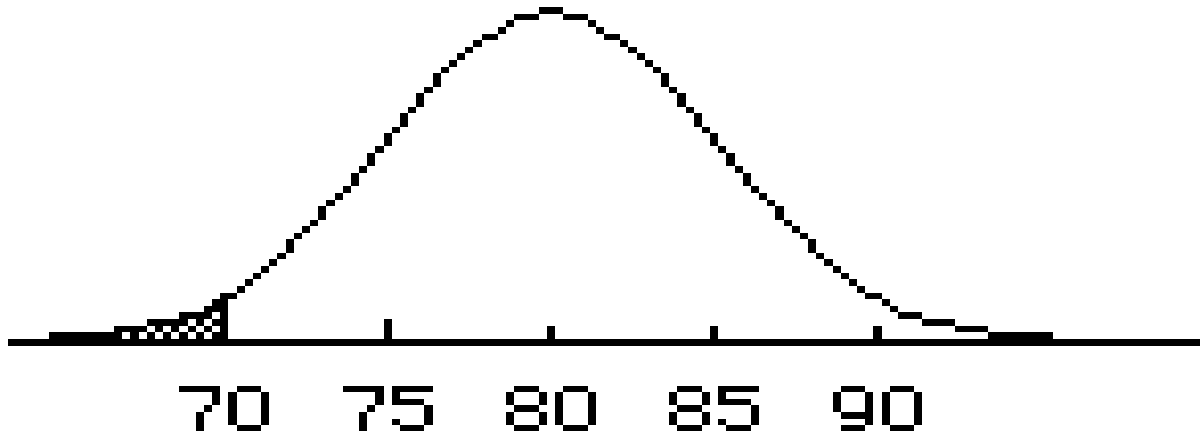
- Notların normal dağılım gösterdiği bir sınavdan 70 aldınız. (Ort=80, SS=5) Sınıftaki yeriniz (yüzde olarak) neresidir?

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

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

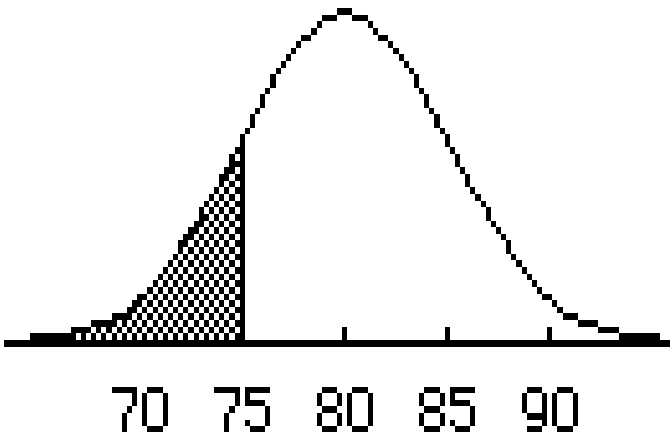
Ortalamanın 2 SS altında.

Sınıftaki öğrencilerin sadece %2.3'ü 70 ya da daha altında not almıştır.



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.

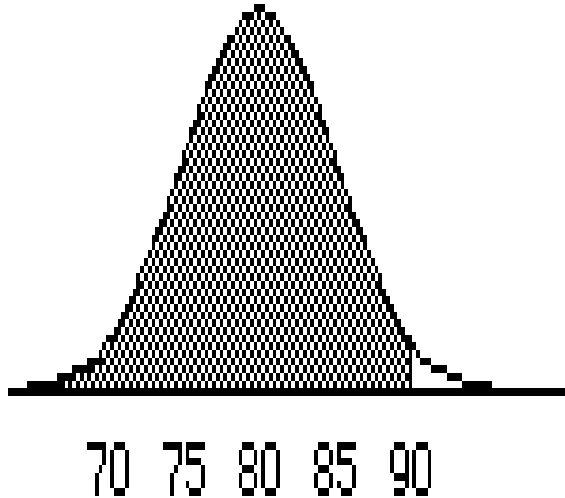


Z tablosu

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

Yüzdelere Çevirme III

- 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.

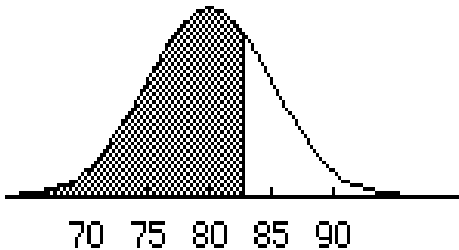


Z tablosu

z	Area from -∞ 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

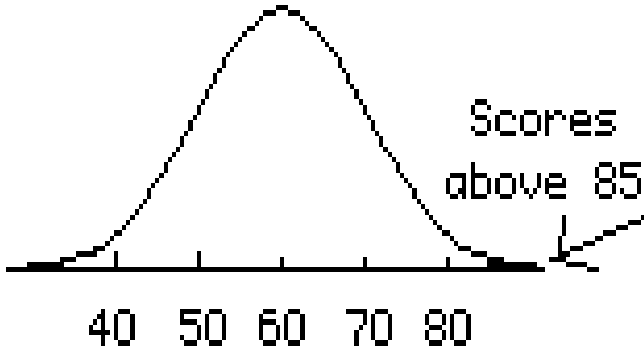
Yüzdelerle Çevirme IV

- Peki hangi notu almış olsaydınız %75'lik dilimde olurduunuz?
- Doğrudan z tablosu kullanılarak %75'e karşılık gelen z değeri (.7734) blunur.
- $SS=5$ 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 ($X = 80 + .7734 * 5 = 83.87$).



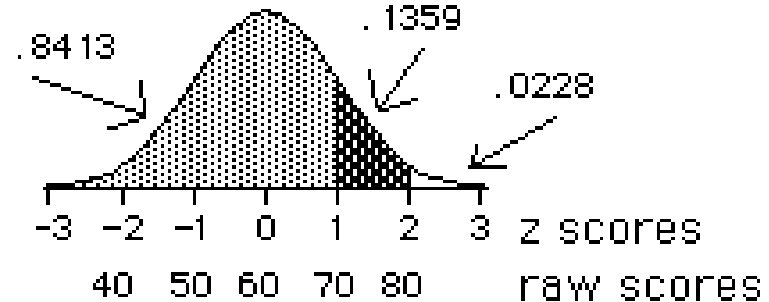
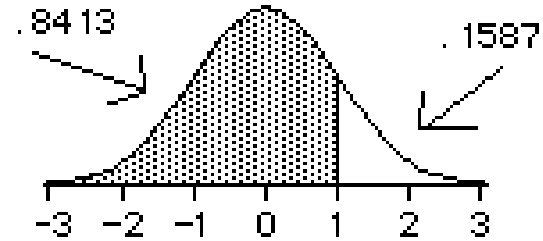
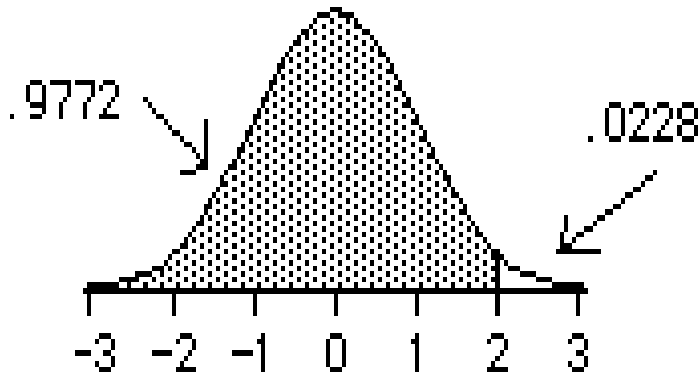
Çan eğrisi altındaki alan hesabı I

- 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 alan 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.



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 değerler 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

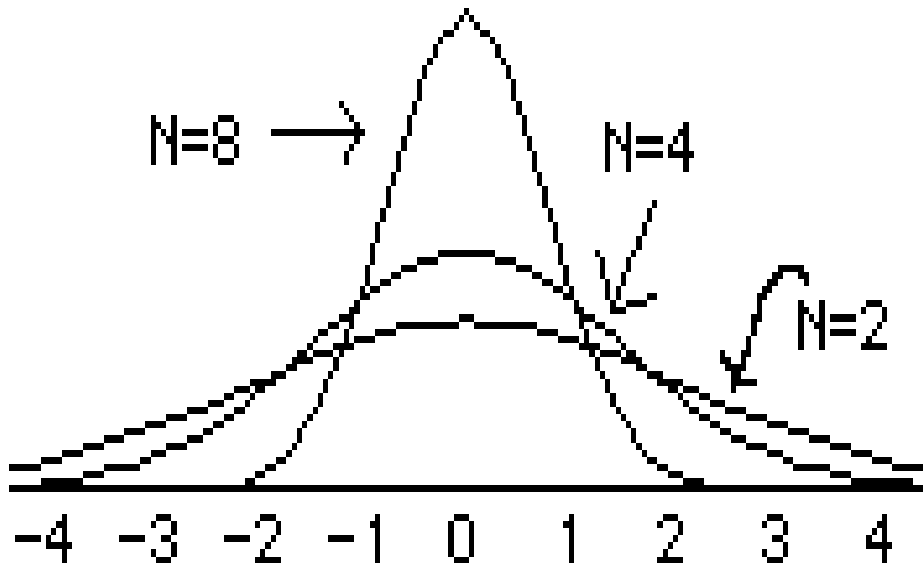
Örnekleme Dağılımı II

Örnekleme büyüklüğü arttıkça standart hata azalır.

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

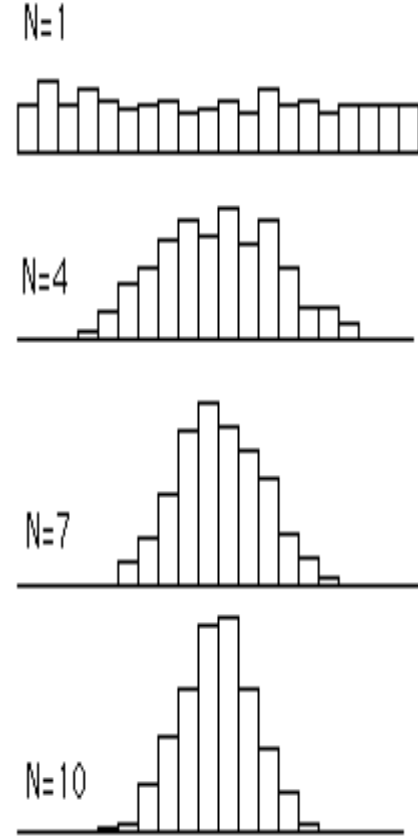
Örneklemin ortalaması μ , SS'si $\frac{\sigma}{\sqrt{N}}$ olur (N = örneklem büyüklüğü)

Örneklemin standart sapması ortalamasının standart hatası olarak bilinir.



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



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

Veri Analizi ve Sunumu

Kullanıcı ihtiyaçlarının karşılanması

Cevap	Sıklık	Geçerli Yüzde	Birikimli Yüzde
1,00	7	25,0	25,0
2,00	19	67,9	92,9
3,00	2	7,1	100,0
Toplam	28	100,0	
Ortalama 1,821			

Çalış, Asuman. (1998). "Üniversite kütüphanelerinde stratejik planlama: bir araştırma,"
Türk Kütüphaneciliği 12(3): 201-230.

- "Kullanıcıların materyal isteklerini ne ölçüde karşılayabiliyorsunuz?"
- "1,00", -Tamamen "2,00" Kısmen "3,00" – Hiç
- Yedi üniversite kütüphanesi tamamen, 19'u kısmen karşılayabilmekte, ikisi ise hiç karşılayamamaktadır.
- Aritmetik ortalamaya (1.821) dikkat!!
- Sınıflama ölçeği kullanılmış, yanıtlar kodlanmış- kodlu veriler birbiriyle çarpılmış ve elde edilen rakam toplam yanıt sayısına bölünmüş!
- $((1,00 \times 7) + (2,00 \times 19) + (3,00 \times 2) = 51/28 = 1.821)$.
- Ya veriler 1, 2, 3 yerine A, B, C diye kodlansaydı?!
- Sınıflama ölçeği ile elde edilen veriler için tabloda verilen sıklıklar ve yüzdeler yeterlidir.

	Hizmetler	Hizmet Veriliyor		Hizmet Verilmiyor		Ort.
		Küt.sayı.		Küt.sayı.		
		Sıklık	Yüzde	Sıklık	Yüzde	
Kullanıcısına verilen hizmet	Multimedya	1	3,6	27	96,4	1,964
	Masaüstü yayıncılık	2	7,1	26	92,9	1,929
	Halkla ilişkiler	2	7,1	26	92,9	1,929
	Kullanıcı Hizmeti	6	21,4	2	78,6	1,786
	Elektronik dergiler	6	21,4	2	78,6	1,786
	OPAC	12	42,9	16	57,1	1,571
	İnternet	15	53,6	13	46,4	1,464
	CD-ROM	20	71,4	8	28,6	1,286
	Kart kataloğu tarama	21	75	7	25	1,25
	Kütüphaneler arası ödünç verme	23	82,1	5	17,9	1,179
	Müracaat hizmeti	25	89,3	3	10,7	1,107
	Kütüphaneler arası işbirliği	26	53,6	13	46,4	1,464

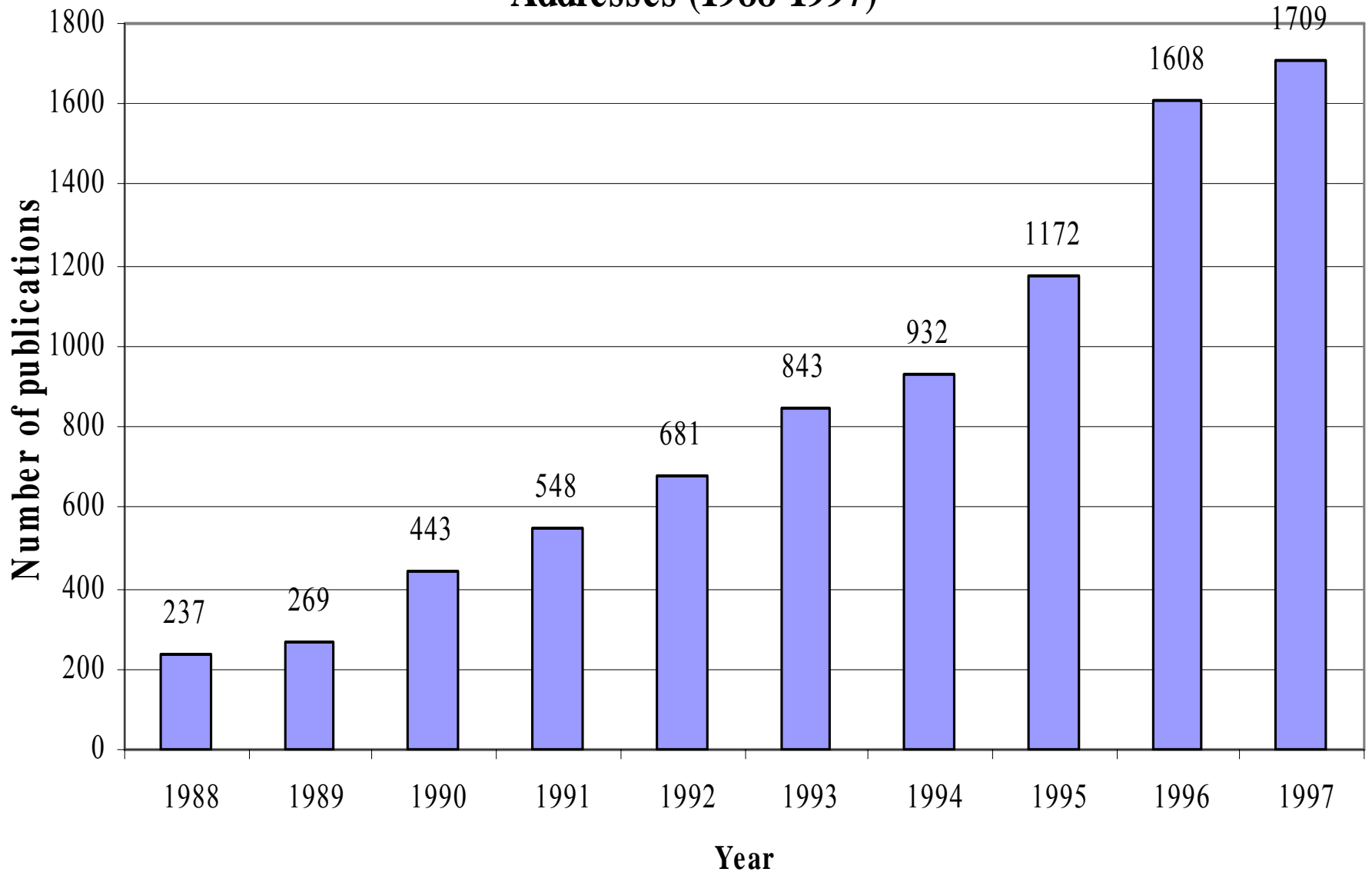
Tablo 3. Kullanıcıya Verilen Hizmetler (N=28)

Verilen Hizmetler	Kütüphane	
	Sayısı	(%)
Multimedya	1	3,6
Masaüstü yayıncılık	2	7,1
...
Kart katalođu tarama	21	75,0
Kütüphaneler arası ödünç verme	23	82,1
Müracaat hizmeti	25	89,3
Kütüphanelerarası işbirliđi	26	53,6

Kullanıcılara verilen CD-ROM hizmetleri ile “pazar payını artırmak ve rekabet avantajı sağlamak” arasında istatistiksel açıdan anlamlı bir ilişki olup olmadığının test edildiği Tablo 4

V188 (soru.30.7)					
V134 1(soru 18.3)	20	2,4000	1,046		
,234					
V134 2	8	3,7500	1,282		
,453					
Anlam farkı = -1,3500					
Varyansların Eritliđi: F=1,160 P= ,291					
t-testi için Anlamlılık Derecesi					
95%					
Varyanslar	t-deđer	df	2-T.S.		
Erit	-2,90	26	,008	,466	(-2,308; -
,392)					
Erit olmayan	-2,65	10,94	,023	,510	(-2,473; -,227)

**Figure 2. Number of Biomedical Publications With Turkish
Addresses (1988-1997)**

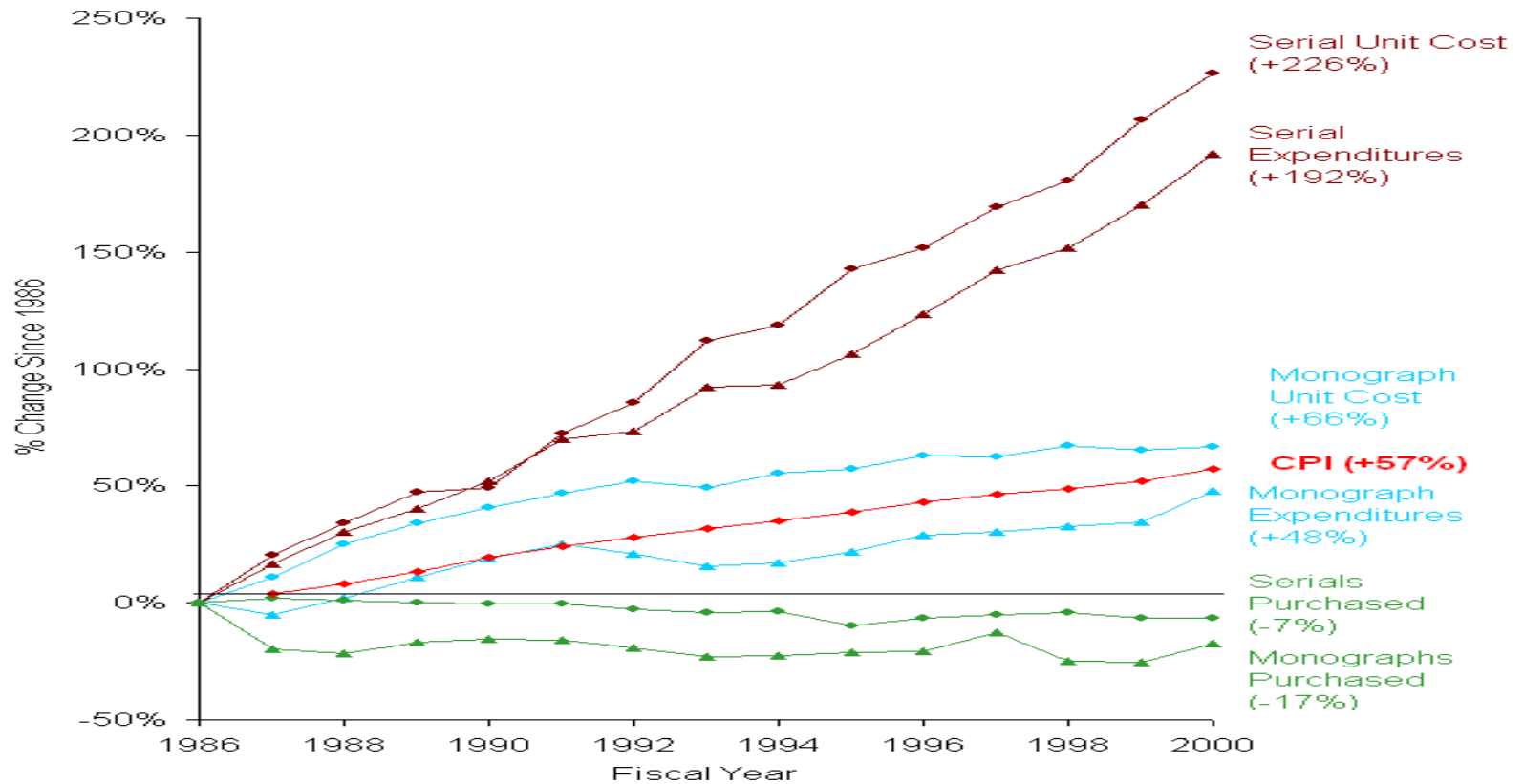


Storage Costs



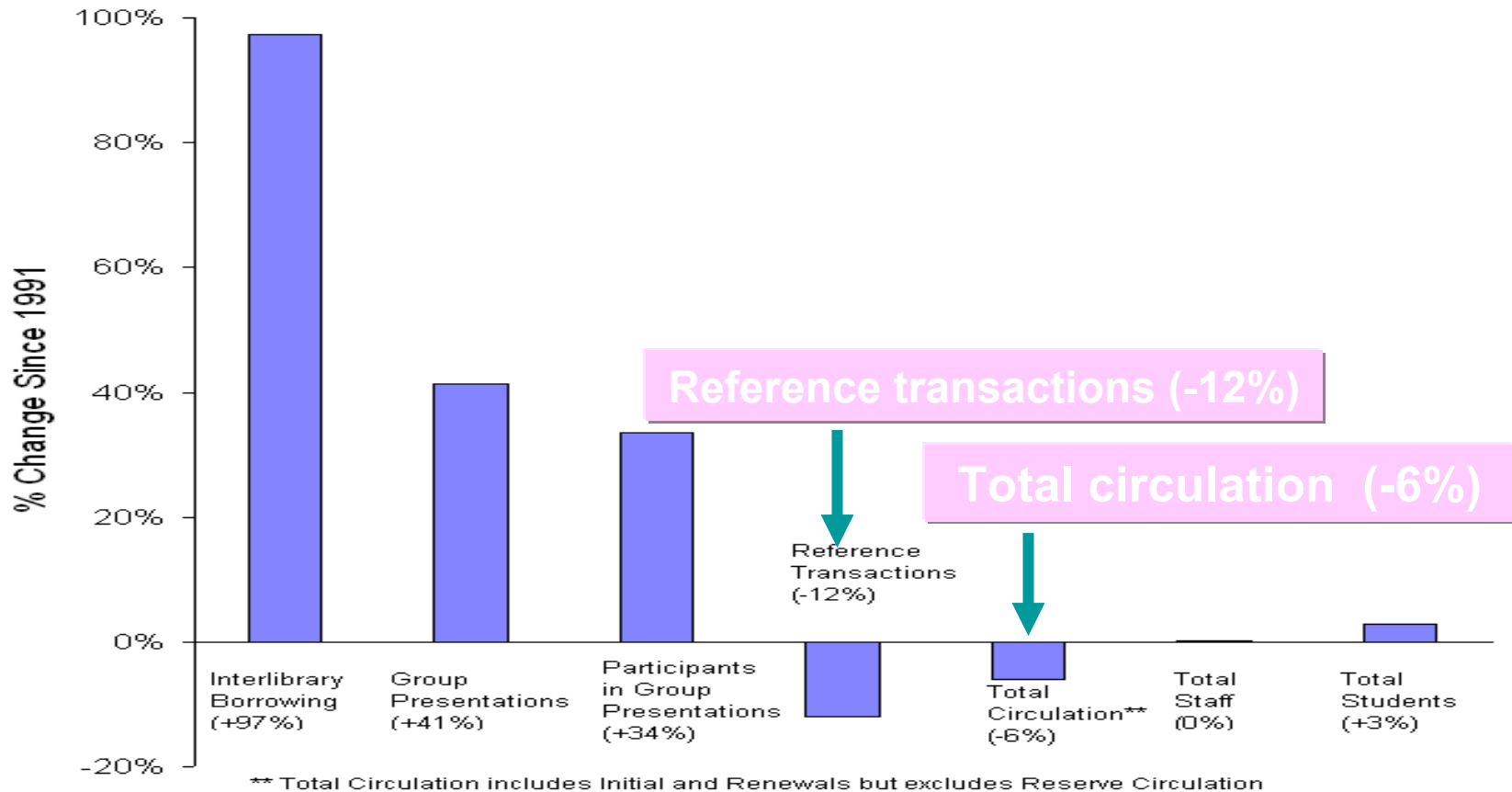
Increasing Costs

Graph 2
Monograph and Serial Costs
in ARL Libraries, 1986-2000



Impact of Remote Access

Graph 1
Service Trends in ARL Libraries, 1991-2000



Tablo 1. Sunuřların Deęerlendirilmesi

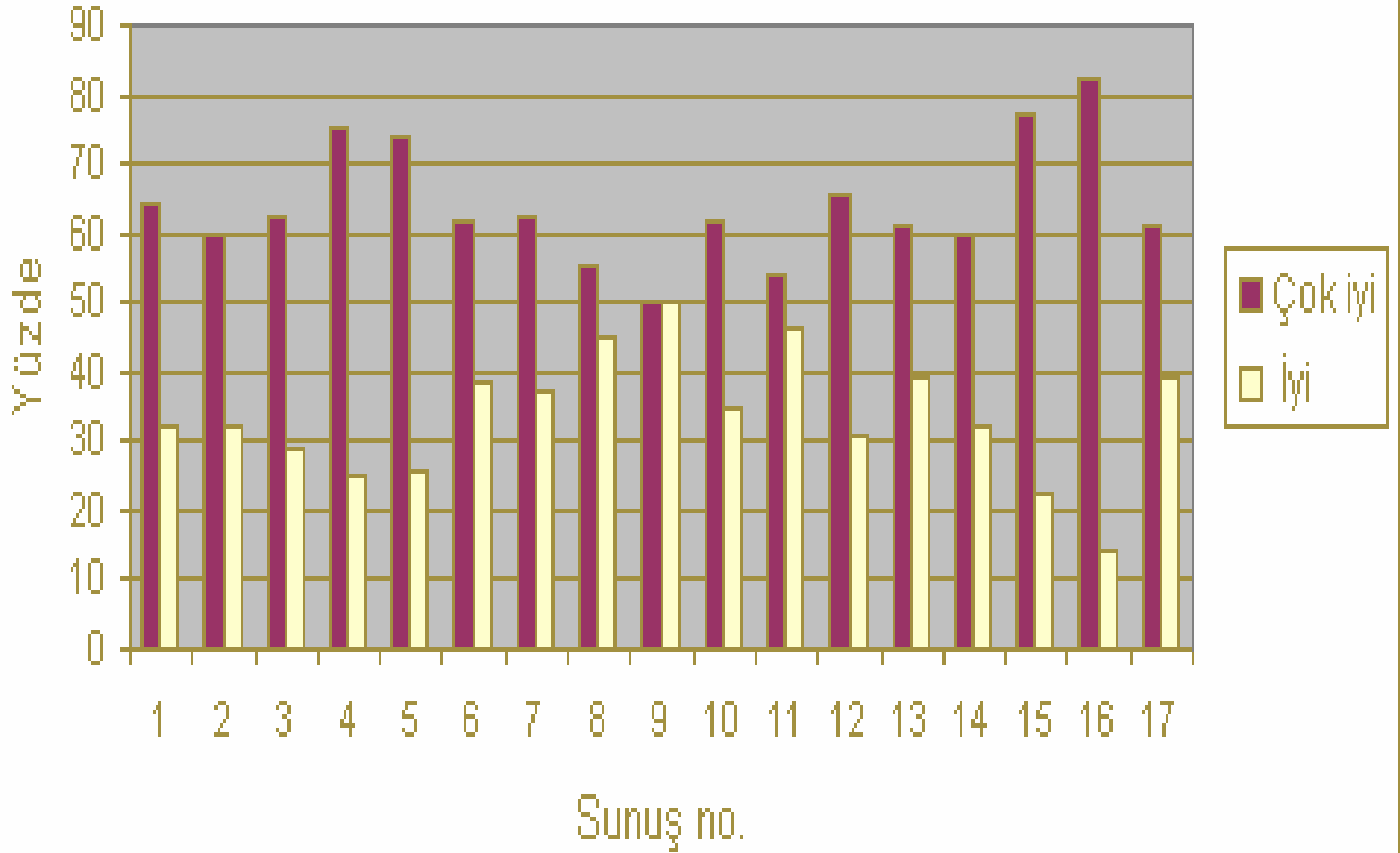
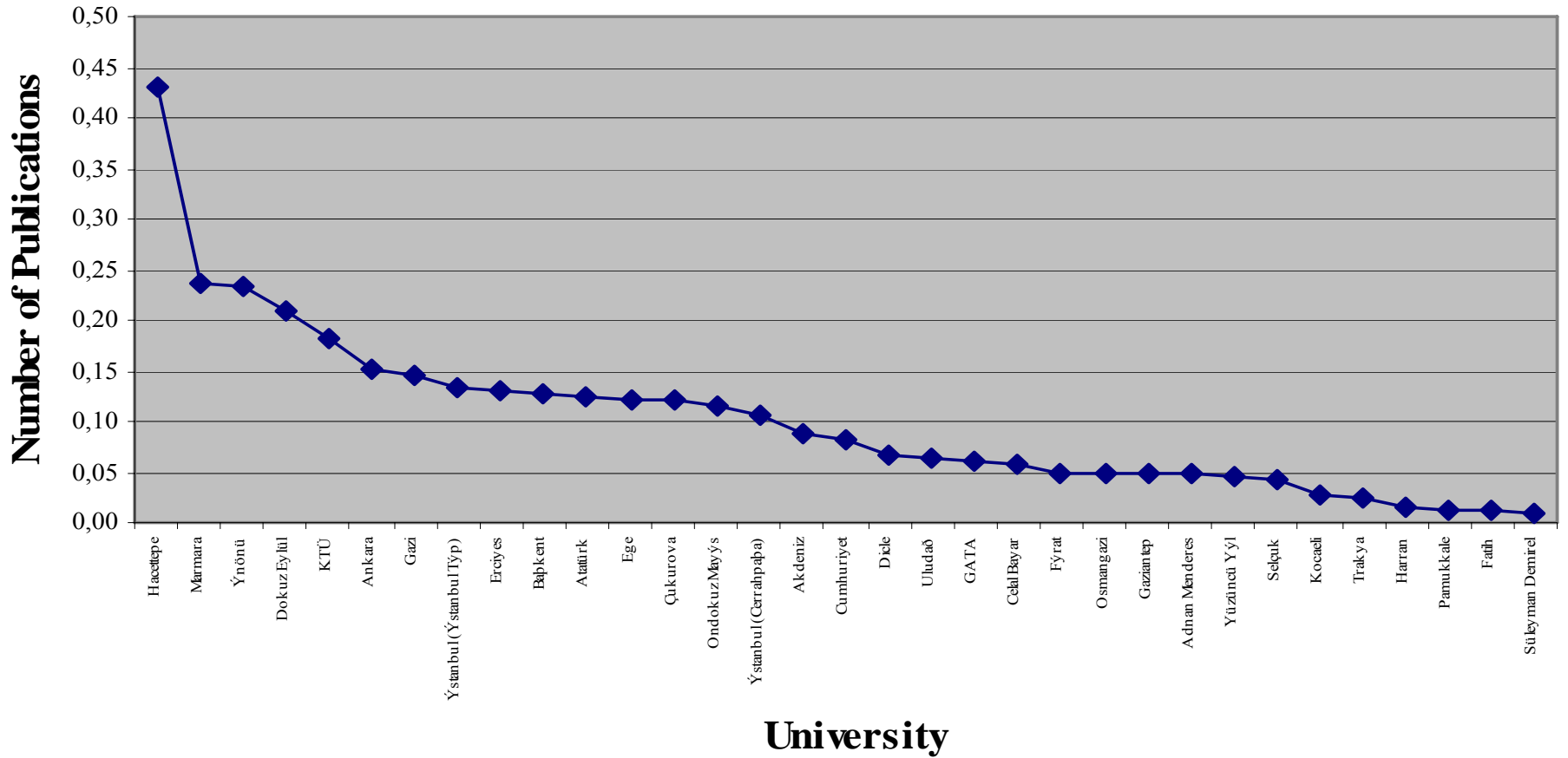
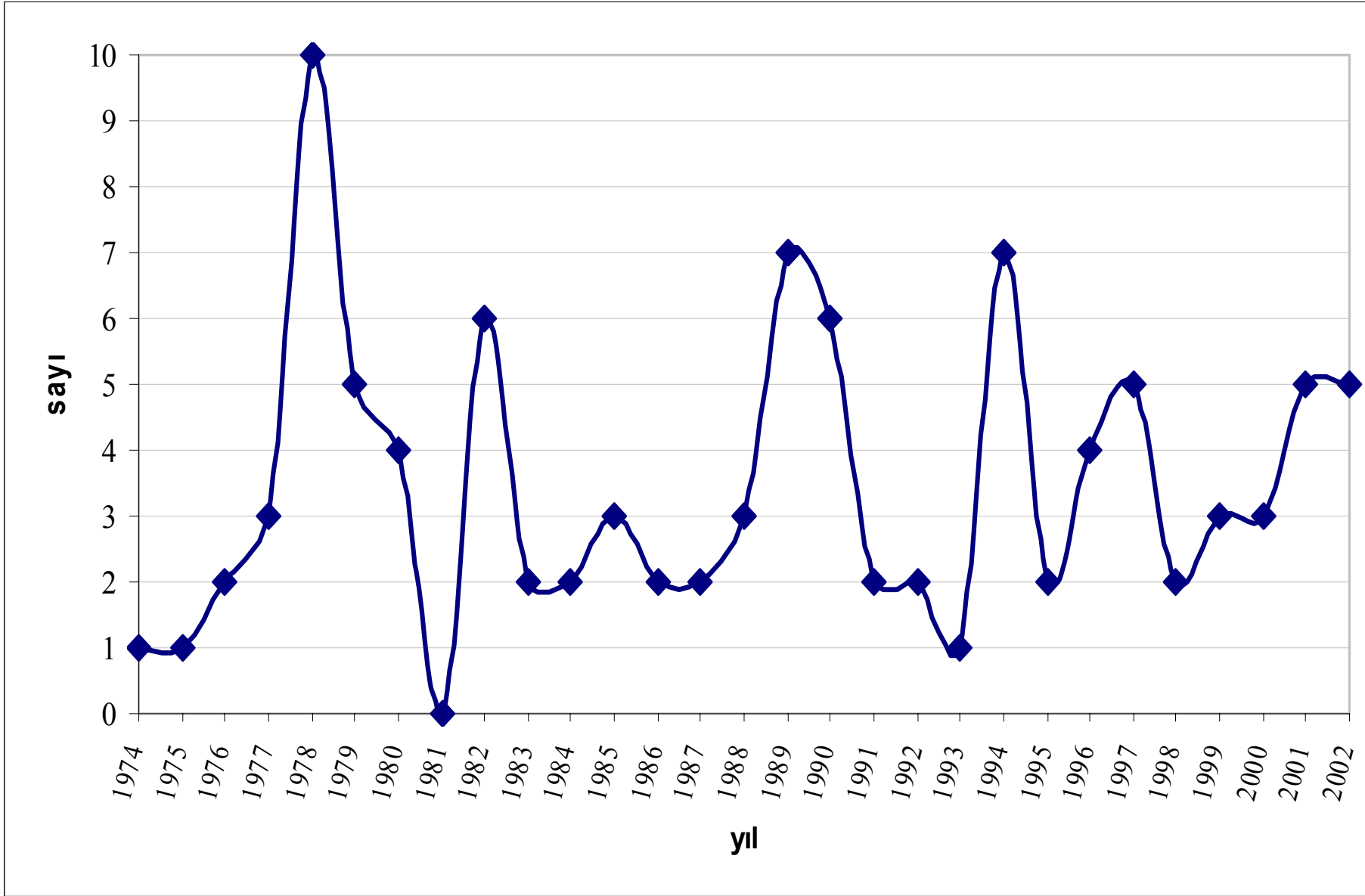
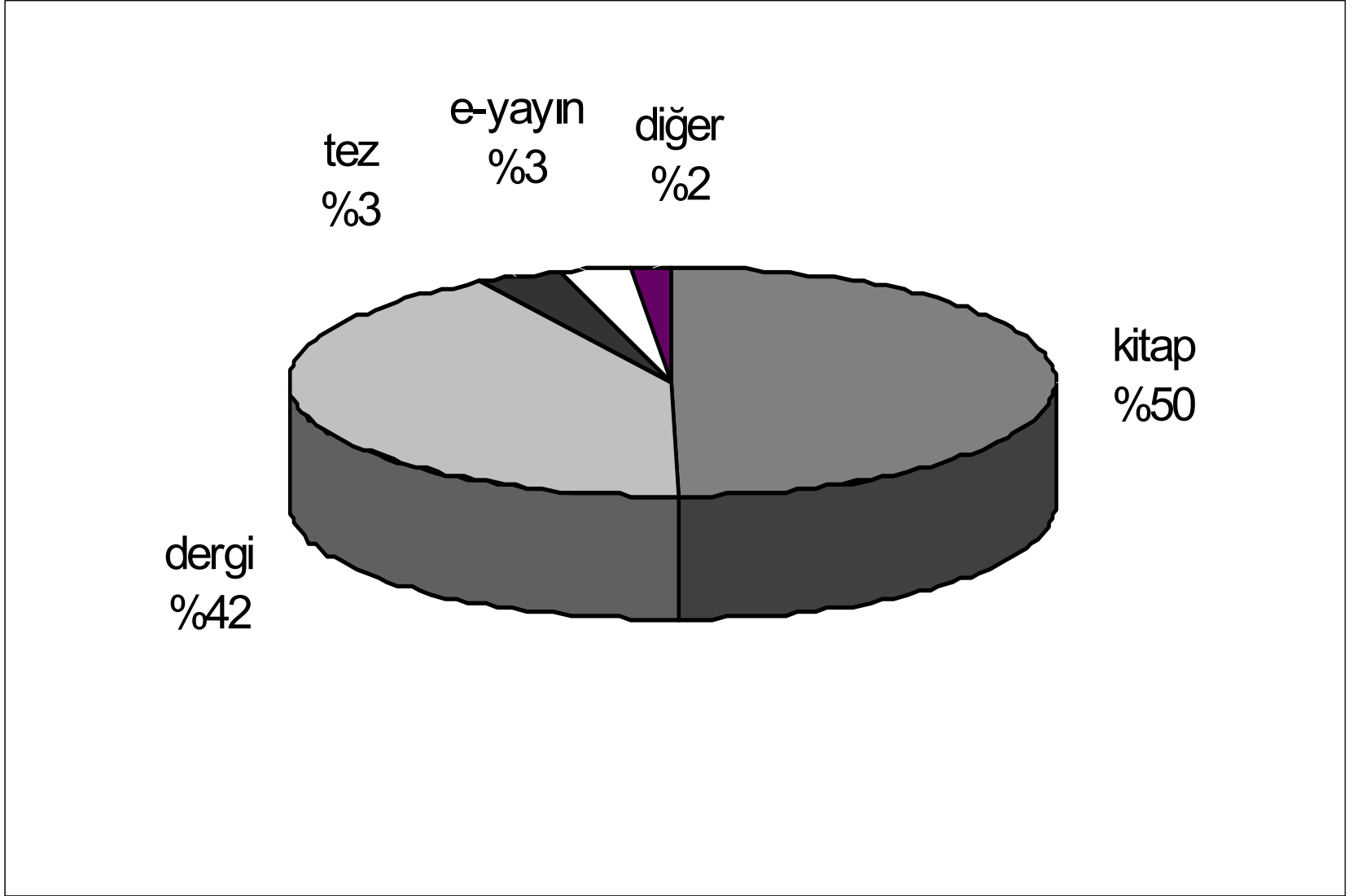


Figure 6. Average Number of Publications Per Faculty in Turkish Medical Schools





Şekil 1. H.Ü. Kütüphanecilik Bölümünde Yıllar itibariyle üretilen tez sayıları



Şekil 2. H.Ü. Kütüphanecilik Bölümü tezlerinde yapılan atıfların yayın türlerine göre dağılımı

Tablo 3. Tez türüne göre atıf yapılan kaynakların dağılımı

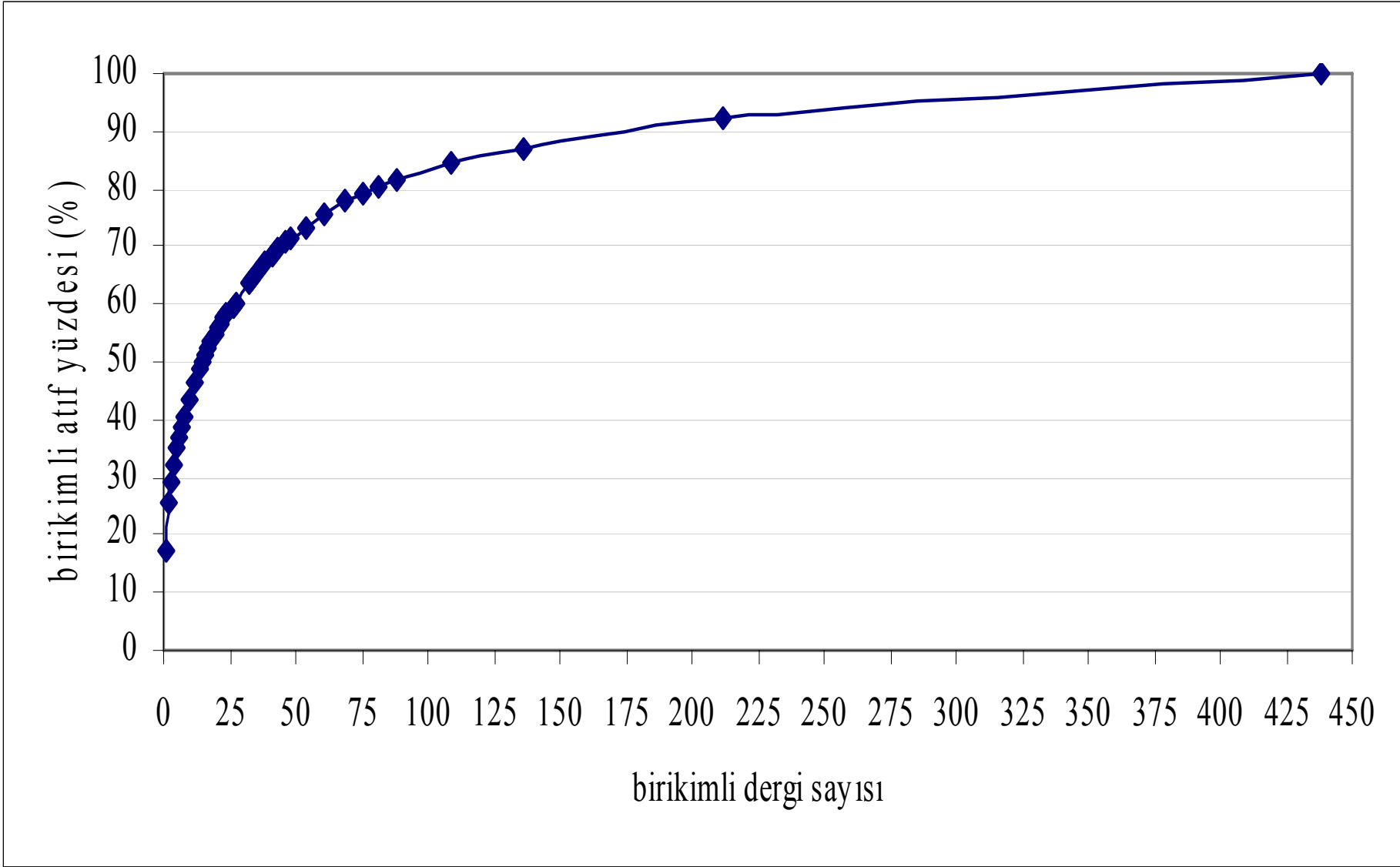
	yüksek lisans		doktora		toplam	
	N	%	N	%	N	%
dergi	1576	38,4	1369	47,0	2945	42,0
kitap	2127	51,8	1371	47,0	3498	49,8
e-yayı n	179	4,4	38	1,3	217	3,1
tez	139	3,4	102	3,5	241	3,4
diğer	84	2,0	34	1,2	118	1,7
toplam	4105	100,0	2914	100,0	7109	100,0

Table 3. Number of Publications Generated by Researchers Based in the Three Largest Cities in Turkey

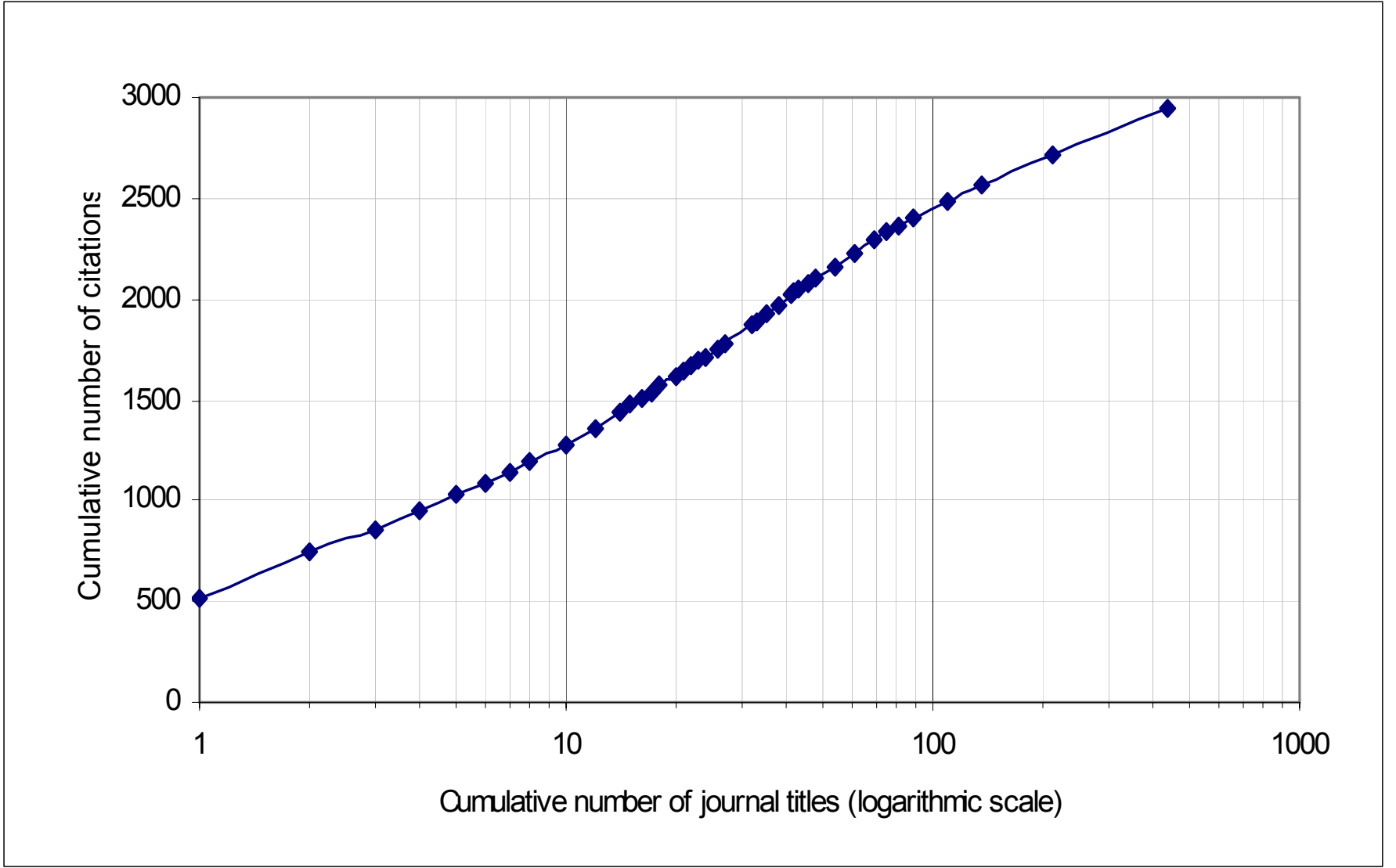
<i>City/Univers ity</i>	<i>Number of Publicatio ns</i>	<i>Share Within Total (%)</i>
Ankara		
Hacettepe	1.718	23,1
Ankara	773	10,4
Gazi	450	6,1
GATA	240	3,2
ODTÜ	135	1,8
Başkent	51	0,7
Bilkent	5	0,1
Fatih	1	0,0
Total	3.373	45,4
İstanbul		
İstanbul	946	12,7
Marmara	394	5,3
Boğaziçi	39	0,5
İTÜ	9	0,1
YTÜ	3	0,0
Total	1.391	18,6
İzmir		
Ege	507	6,8
Dokuz Eylül	261	3,5
İzmir YTE	2	0,0
Total	770	10,4
Other Provinces		
Total	1.893	25,5
Grand Total	7.427	100,0

Tablo 8. Atıfların dergi sayısına göre dağılımı

dergi sayısı	atıf sayısı	toplam atıf sayısı	birikimli dergi sayısı	birikimli atıf yüzdesi
1	512	512	1	17,4
1	236	236	2	25,4
1	102	102	3	28,9
1	99	99	4	32,2
1	79	79	5	34,9
1	60	60	6	36,9
1	54	54	7	38,8
1	48	48	8	40,4
2	43	86	10	43,3
2	42	84	12	46,2
2	40	80	14	48,9
1	35	35	15	50,1
1	33	33	16	51,2
1	31	31	17	52,3
1	30	30	18	53,3
2	26	52	20	55,0
1	25	25	21	55,9
1	24	24	22	56,7
1	23	23	23	57,5
1	22	22	24	58,2
2	21	42	26	59,7
1	20	20	27	60,3
5	19	95	32	63,6
1	18	18	33	64,2
2	17	34	35	65,3
3	16	48	38	67,0
3	15	45	41	68,5
1	14	14	42	69,0
1	13	13	43	69,4
3	12	36	46	70,6
2	11	22	48	71,4
6	10	60	54	73,4
7	9	63	61	75,6
8	8	64	69	77,7
6	7	42	75	79,2
6	6	36	81	80,4
7	5	35	88	81,6
21	4	84	109	84,4
27	3	81	136	87,2
76	2	152	212	92,3
226	1	226	438	100,0
Top.438		2945		



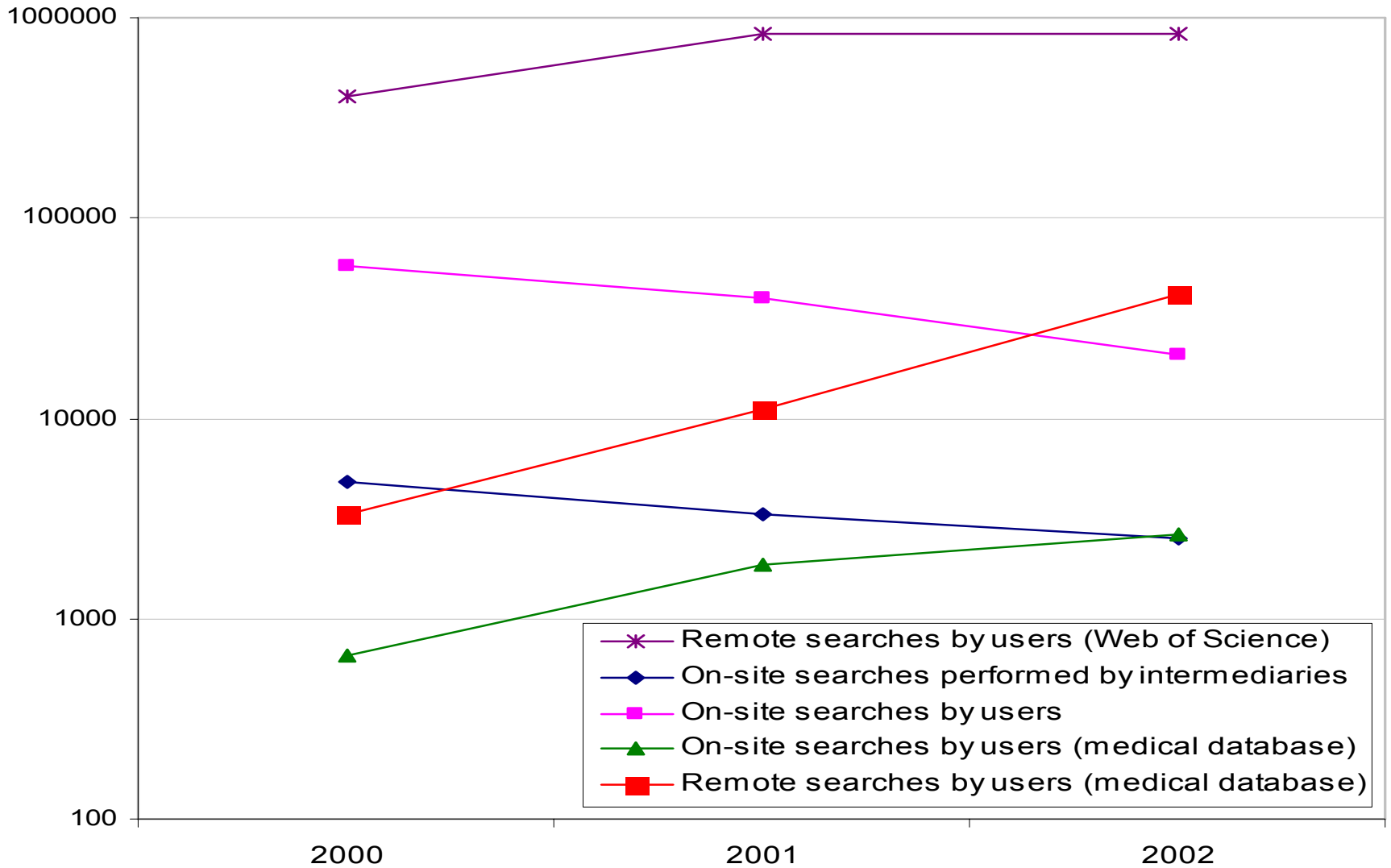
Şekil 3. Atıfların dergi sayısına göre dağılımı



Şekil 3. Atıfların dergi sayısına göre dağılımı (logaritmik)

logarithmic
scale

**Fig. 1. Turkish Academic Network and Information Center (ULAKBİM)
Use Statistics**



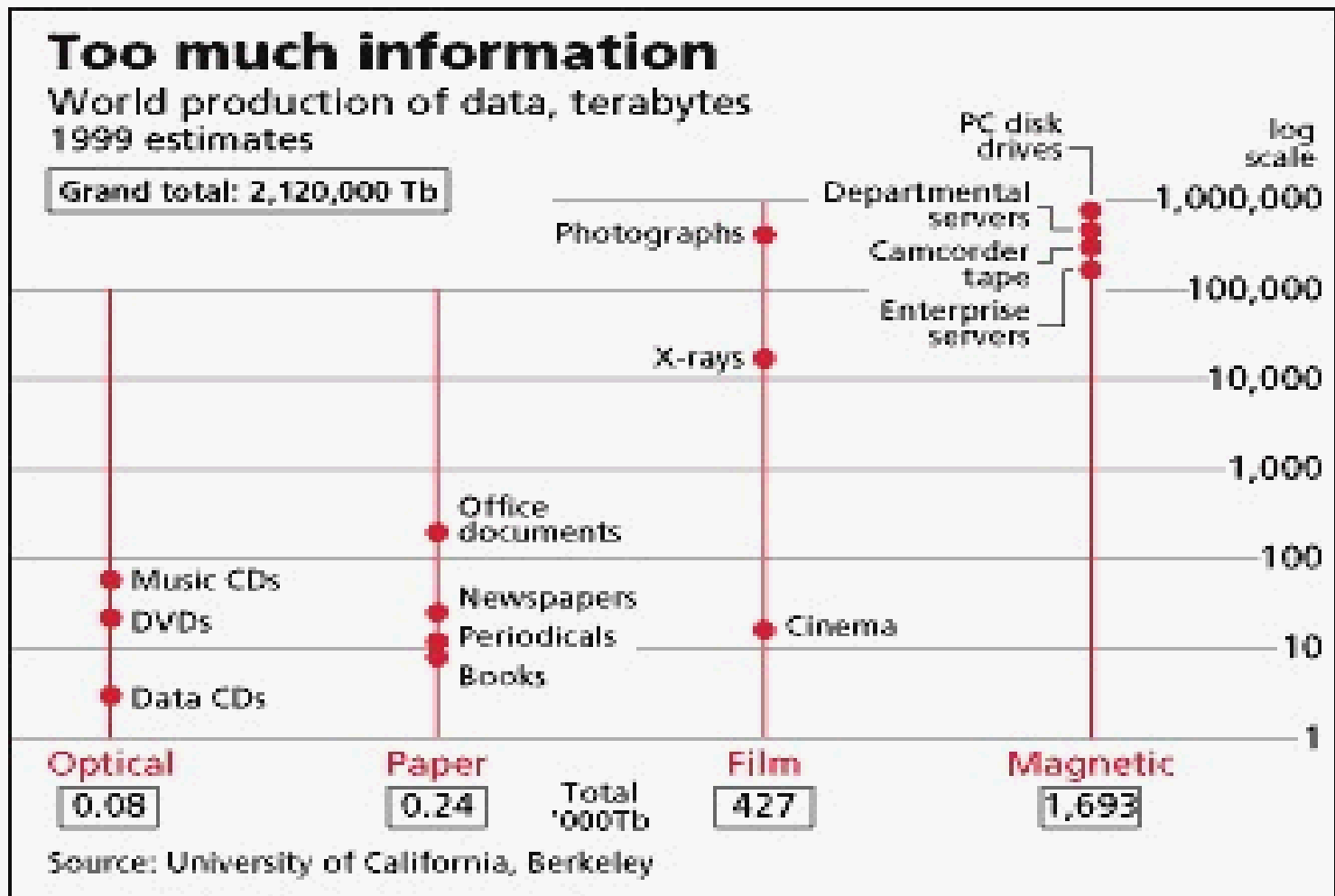


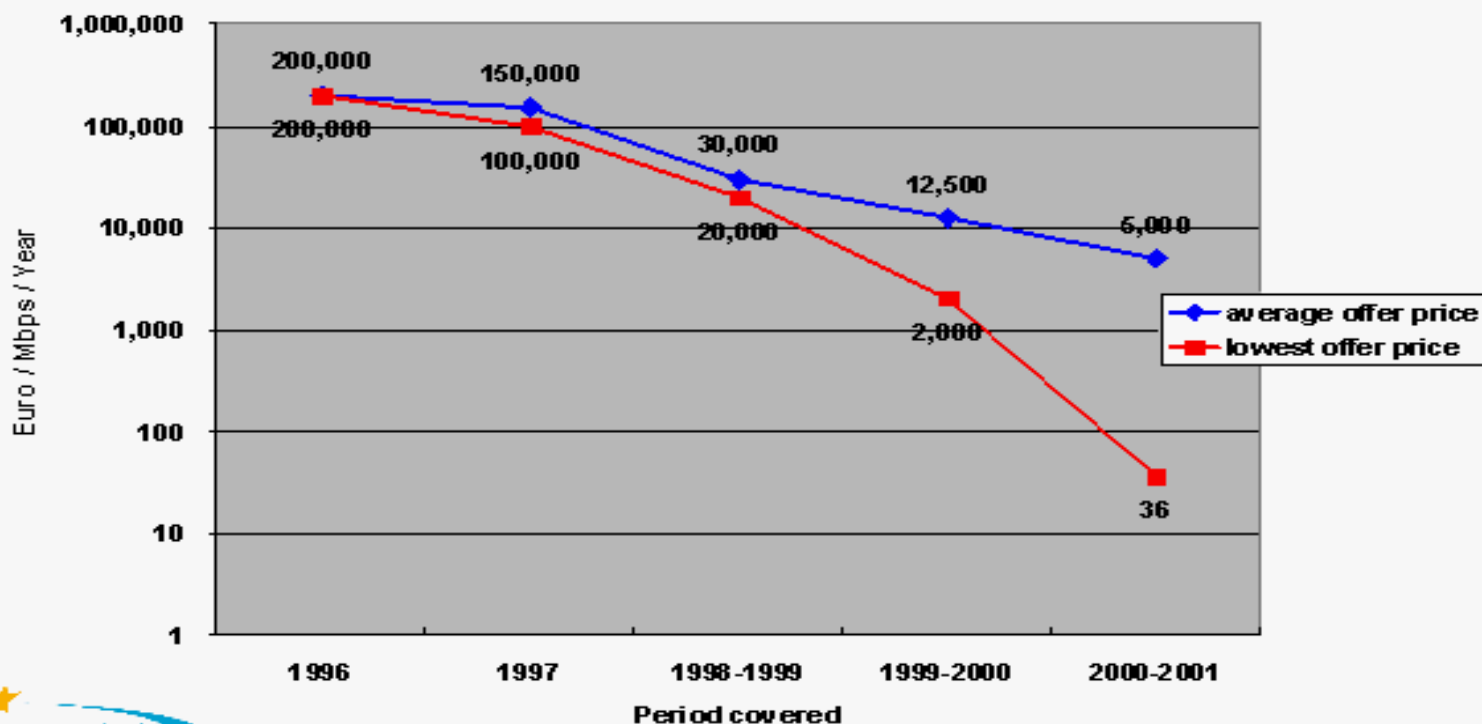
Fig. 1. World-wide information production (1999)

Source: Lyman and Varian (2000). Available:

<http://www.sims.berkeley.edu/research/projects/how-much-info/charts/charts.html>

Transmission Costs

Price Development of International Bandwidth (logarithmic scale)



GEANT - A Giant Step - Vincent.Berkhout@dante.org.uk (12/16)