

Konum ve Dağılım Ölçüleri

BBY606 Araştırma Yöntemleri
Güleda Doğan

Konum ölçüleri

- Merkezi eğilim ölçüleri
- Verilerin ortalamaya göre olan gruplanması nasıl? Yakın, uzak?
- Sıklık dağılımlarının karşılaştırılması

- Aritmetik ortalama
- Ortanca
- Tepe değeri

Dağılım ölçüleri

- Değişim ölçüleri
- Verilerin değişkenliği nasıl?
- Dağılım genişliği
- Standart sapma

En önemlileri

- Aritmetik ortalama
- Standart sapma
- Çıkarsamalı istatistik

Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Reports

Descriptive Statistics

Tables

Compare Means

General Linear Model

Generalized Linear Models

Mixed Models

Correlate

Regression

Loglinear

Neural Networks

Classify

Dimension Reduction

Scale

Nonparametric Tests

Forecasting

Survival

Multiple Response

Missing Value Analysis...

Multiple Imputation

Frequencies...

Descriptives...

Explore...

Crosstabs...

Ratio...

P-P Plots...

Q-Q Plots...

	cinsiyet	var	var	var
ksek	Kadın			
ksek	Erkek			
ksek	Kadın			
ksek	Erkek			

Frequencies

Variable(s):

Final Notu [Final]

Öğrencinin Adı [Ad]
Öğrencinin Soyadı [...]
Uygulama Şubesi [...]
Ara Sınav Notu [Aras...]
Arasınav_gruplu [Ar...]
Öğrencinin cinsiyeti ...

☐ Display frequency tables

Statistics...

Charts...

Format...

Bootstrap...

OK

Paste

Reset

Cancel

Help

Frequencies

Variable(s):
Final Notu [Final]

☐ Display frequency tables

OK Paste Reset Cancel Help

60,0	Ortalamadan yüksek	Kadın	
86,0	Ortalamadan düşük	Kadın	
88,0	Ortalamadan düşük	Erkek	
97,0	Ortalamadan düşük	Kadın	
86,0	Ortalamadan düşük	Kadın	
87,0	Ortalamadan düşük	Kadın	

Frequencies: Statistics

Percentile Values

☐ Quartiles
☐ Cut points for: 10 equal groups
☐ Percentile(s):

Add
Change
Remove

Central Tendency

☒ Mean
☒ Median
☒ Mode
☐ Sum

☐ Values are group midpoints

Dispersion

☐ Std. deviation ☐ Minimum
☐ Variance ☐ Maximum
☐ Range ☐ S.E. mean

Distribution

☐ Skewness
☐ Kurtosis

Continue Cancel Help

Mean = Aritmetik ortalama, Median = Ortanca, Mode = Tepe değeri

Konum ölçüleri

Statistics

Final Notu

N	Valid	91
	Missing	0
Mean		53,220
Median		57,000
Mode		54,0

Aritmetik ortalama

Ortanca

Tepe değeri

Aritmetik ortalama

- Anlaşılması ve hesaplanması kolay
- En çok kullanılan merkezi eğilim ölçüsü
 - ✓ İstatistiksel hesaplamalar için temel oluşturması

Aritmetik ortalama

- Dağılımdaki tüm değerler hesaplamaya dahil

19, 19, 19, 19, 19, 19, 19, 19, 19, 19, 20, 20, 20, 21, 21, 21, 21, 21, 22, 22, 22

- Tüm değerlerin toplamı / Veri sayısı
- $(19+19+19+ \dots +22+22+22) / 20 = 20,10$

	yas
1	19
2	19
3	19
4	19
5	19
6	19
7	19
8	19
9	19
10	20
11	20
12	20
13	21
14	21
15	21
16	21
17	21
18	22
19	22
20	22
21	

Variable(s):

Öğrencinin Yaşı [yas]

☒ Display frequency tables

OK Paste Reset Cancel Help

Statistics...

Percentile Values

☐ Quartiles

☐ Cut points for: 10 equal groups

☐ Percentile(s):

Add Change Remove

Central Tendency

☒ Mean

☐ Median

☐ Mode

☐ Sum

☐ Values are group midpoints

Dispersion

☐ Std. deviation ☐ Minimum

☐ Variance ☐ Maximum

☐ Range ☐ S.E. mean

Distribution

☐ Skewness

☐ Kurtosis

$$(19+19+19+ \dots +22+22+22) / 20 = 20,10$$

Statistics

Öğrencinin Yaşı

N	Valid	20
	Missing	0
Mean		20,10

Aritmetik ortalama

- Dağılımdaki tüm değerler hesaplamaya dahil
 - ✓ Çok büyük ve çok küçük değerlerden etkilenme
- Çok yüksek not alan bir kişinin sınıfın not ortalamasını yükseltmesi gibi



26 :

	yas
1	19
2	19
3	19
4	19
5	19
6	19
7	19
8	19
9	19
10	20
11	20
12	20
13	21
14	21
15	21
16	21
17	21
18	22
19	22
20	22
21	55

Variable(s):

Öğrencinin Yaşı [yas]

☒ Display frequency tables

OK Paste Reset Cancel Help

Statistics...

Frequencies: Statistics

Percentile Values

☐ Quartiles

☐ Cut points for: 10 equal groups

☐ Percentile(s):

Add Change Remove

Central Tendency

☒ Mean

☐ Median

☐ Mode

☐ Sum

☐ Values are group midpoints

Dispersion

☐ Std. deviation ☐ Minimum

☐ Variance ☐ Maximum

☐ Range ☐ S.E. mean

Distribution

☐ Skewness

☐ Kurtosis

$$(19+19+19+ \dots +22+22+22+55) / 21 = 21,76$$

Statistics

Öğrencinin Yaşı

N	Valid	21
	Missing	0
Mean		21,76

Aritmetik ortalama

- Kesikli değişkenler için kullanımı pek uygun değil
 - ✓ Ailedeki çocuk sayısı 2,3
 - ✓ Öğrencilerin derse gelmediği gün sayısı 4,4
 - ✓ Kullanılması gerekiyorsa yuvarlanmalı!
- Evrenden alınacak farklı örneklemeler için en az değişecek merkezi eğilim ölçüsü

Tepe değeri

- Bir dağılımda en sık görülen değer
- Farklı her bir değerın sıklığını bilmek gerek!
- 21 kişilik sınıftaki öğrencilerin yaşları
 - ✓ 19 yaşındaki öğrenci sayısı
 - ✓ 20 yaşındaki öğrenci sayısı
 - ✓ ...

Statistics

Öğrencinin Yaşı

N	Valid	21
	Missing	0
Mode		19

Öğrencinin Yaşı

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	19	9	42,9	42,9	42,9
	20	3	14,3	14,3	57,1
	21	5	23,8	23,8	81,0
	22	3	14,3	14,3	95,2
	55	1	4,8	4,8	100,0
	Total	21	100,0	100,0	

Tepe değeri

- Uç değerlerden etkilenmez
- Kesikli veriler için en uygun merkezi eğilim ölçüsü
 - ✓ Ailedeki çocuk sayısı
 - ✓ Aritmetik ortalama = 2,3 (0,3 çocuk mu?)
 - ✓ Tepe değeri = 2

Tepe değeri

- Birden çok tepe değeri söz konusuysa
 - ✓ Merkezi eğilim ölçüsü olarak aritmetik ortalama ya da ortancayı kullanmak daha doğru

Statistics

Öğrencinin Yaşı

N	Valid	23
	Missing	0
Mode		19 ^a

a. Multiple modes exist. The smallest value is shown

Öğrencinin Yaşı

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	19	9	39,1	39,1	39,1
	20	2	8,7	8,7	47,8
	21	3	13,0	13,0	60,9
	22	9	39,1	39,1	100,0
Total		23	100,0	100,0	

Tepe değeri

- Merkezi eğilim/konum ölçmede her zaman çok kesin bir değer olmadığı için ileri hesaplamalarda kullanımı az
- Tekstil sektöründe kullanımı
 - ✓ En çok satılan bedenden en fazla üretmek

Ortanca

- Veriler küçükten büyüğe ya da büyükten küçüğe doğru sıralandığında en ortada kalan değer
- Veri setinin alt ve üst limitlerinin net olarak bilinmediği durumlarda yararlı

	yas
1	19
2	19
3	19
4	19
5	19
6	19
7	19
8	19
9	19
10	20
11	20
12	21
13	21
14	21
15	22
16	22
17	22
18	22
19	22
20	22
21	22
22	22
23	22

Statistics

Öğrencinin Yaşı

N	Valid	23
	Missing	0
Median		21,00

OK

Paste

Change

Remove

Dispersion

- ☐ Std. deviation ☐ Minimum
☐ Variance ☐ Maximum
☐ Range ☐ S.E. mean

Continue

Cancel

Help

Central Tendency

- ☐ Mean
☒ Median
☐ Mode
☐ Sum

☐ Values are group midpoints

Distribution

- ☐ Skewness
☐ Kurtosis

Veri sayısı çift olursa ?

	yas	
1	19	
2	19	
3	19	
4	19	
5	19	
6	19	
7	19	
8	19	
9	19	
10	20	
11	20	
12	21	
13	21	
14	21	
15	22	
16	22	
17	22	
18	22	
19	22	
20	22	
21	22	
22	22	
23	22	
24	23	

Veri sayısı çift olursa ?

$$(21 + 21) / 2$$

Ortanca

- Alanımız ile ilgili belli 7 kitabın Bilgi ve Belge Yönetimi Bölümü öğrencileri tarafından kütüphaneden ödünç alınma sayıları
- 5, 7, 10, 8, 6, 11, 13
- Ortanca?
 - ✓ Küçükten büyüğe sırala: 5, 6, 7, 8, 10, 11, 13
 - ✓ 7 tane veri: Tek sayı
 - ✓ En ortadaki değer: 4. değer = 8

Ortanca

- En önemli avantajı
 - ✓ Uç / aykırı değerlerden etkilenmez
 - ✓ Uç /aykırı değer: Dağılımın geneline göre çok büyük ve çok küçük değerler

Statistics

Öğrencinin Yaşı

N	Valid	21
	Missing	0
Median		20,00

	yas
1	19
2	19
3	19
4	19
5	19
6	19
7	19
8	19
9	19
10	20
11	20
12	20
13	21
14	21
15	21
16	21
17	21
18	22
19	22
20	22
21	55

Ortanca

- Dağılımdaki tüm değerleri değil en ortadaki bir ya da iki değeri dikkate alması
- Verilerin dağılımı geniş ise uygun merkezi eğilim/konum ölçüsü olmayabilir

Aritmetik, Ortanca ve Tepe Değeri

- Hangi konum ölçüsü hangi değişken türü için uygun?

	Tepe değeri	Ortanca	Aritmetik ortalama
Sınıflama	+	-	-
Sıralama	+	+	-
Aralıklı/Oranlı	+	+	+

Dağılım ölçüleri

- Değişim ölçüleri
- Verilerin yayılışını/dağılımını ölçmek için kullanılan ölçüler
- Dağılım genişliği
- Çeyrek değerler
- Standart sapma
- ...

Dağılım ölçüleri

- Konum ölçüleri: Sıklık dağılımlarının karşılaştırılması
- Çok yönlü karşılaştırmalar için konum ölçüleri yeterli değil
- Örneğin, 2 farklı sınıfın matematik sınav ortalamaları aynı (60), notların dağılımı farklı olabilir
 - ✓ İlk sınıfta en düşük puan 40, en yüksek puan 80
 - ✓ İkinci sınıfta en düşük puan 10, en yüksek puan 90

Dağılım genişliği

- Bir dağılımdaki en büyük değer – en küçük değer
- Bir fast-food zincirinde çalışan kişilerin yıllık maaşları: 3.000, 4.000, 7.000, 16.000, 20.000, 30.000, 38.000, 53.000, 61.000, 88.000 TL
- Dağılım genişliği: $88.000 - 3.000 = 85.000$
- Aritmetik ortalamayı hesaplayın
- Ortalama yıllık maaş, 85.000 TL'lik bir dağılım genişliği ile 32.000 TL

Dağılım genişliği

- 357 kişi 3.000, 301 kişi 4.000, 154 kişi 7.000, 21 kişi 16.000, 11 kişi 20.000, 6 kişi 30.000, 4 kişi 38.000, 3 kişi 53.000, 2 kişi 61.000, 1 kişi 88.000 TL
- Aritmetik ortalamayı hesaplayın
- Dağılım genişliği değişir mi?
- Ortalama yıllık maaş, 85.000 TL'lik bir dağılım genişliği ile 4.896 TL

Dağılım genişliği

- Hesaplanması oldukça kolay
- Dağılım genişliği konum ölçülerinden (aritmetik ortalama, ortanca, tepe değeri) biriyle birlikte verilirse dağılımın değişkenliği hakkında daha net resmi görmek mümkün
- Herhangi bir konum ölçüsünün alacağı değer daima dağılım sınırları içinde

Analyze Direct Marketing Graphs Utilities Add-ons Window Help

Reports

Descriptive Statistics

Tables

Compare Means

General Linear Model

Generalized Linear Models

Mixed Models

Correlate

Regression

Loglinear

Neural Networks

Classify

Dimension Reduction

Scale

Nonparametric Tests

Forecasting

Survival

Multiple Response

Missing Value Analysis...

123 Frequencies...

Descriptives...

Explore...

Crosstabs...

Ratio...

P-P Plots...

Q-Q Plots...

Referanslar Latin a...
Yayın politikası var ...
Elektronik formatta ...
Hakemlik türü bilgi...
Yayın kurulu kaç kiş...
Editörlerin coğrafik ...
Yazarların coğrafik ...
Tüm dergi içeriği in...
İngilizce web sayfa...

☒ Display frequency tables

Ülke	Yıldakaçsayıcıkar yor	Kaçyılndanbericikey or	Kaçyılndar arenwebde cut
ia	12	1948	1

Frequencies

Variable(s):

Derginin Scopustan at...

Statistics...

Charts...

Format...

Bootstrap...

OK

Paste

Reset

Cancel

Help

Frequencies

Variable(s):

Derginin Scopustan at...

Statistics...

Statistics

Derginin Scopustan atıf alma dı

N	Valid	203
	Missing	0
Range		21 49
Minimum		0
Maximum		21 49

Çeyrek değerler

- Ortanca: Dağılımı iki eşit parçaya böler
- Bir dağılımı dört eşit parçaya bölen değerler
 - Birinci çeyrek - İlk %25: Q1
 - İkinci çeyrek - İkinci %25 (%50- ortanca): Q2
 - Üçüncü çeyrek - Üçüncü %25 (%75): Q3

Çeyrek değerler

- Çeyrek değerler genişliği: $Q3 - Q1$
- Çeyrek değer genişliği ile ölçülen dağılımın ortada kalan %50 lik kısmının dağılımı
 - ✓ Dağılım genişliğine benzer
- En alttaki %25'lik kısım ile en üstteki %25'lik kısmın ihmal edilmesi
- Uç değerlerin dağılıma etkisini azaltma

Ondalıklar

- Dağılımı 10 eşit parçaya ayırır
- 4.ondalık: Deneklerin %40'ı kendisinden küçük değerli, %60'ı kendisinden büyük değerli olan değer
- 8.ondalık: Deneklerin %80'i kendisinden küçük değerli, %20'si kendisinden büyük değerli olan değer

Yüzdelikler

- Dağılımı 100 eşit parçaya böler
- Çeyrek değerler ve ondalıklar da dahil
- 45. yüzdelik: Deneklerin %45'i kendisinden küçük değerli, %55'i kendisinden büyük değerli olan değerler
- 40. yüzdelik değer = 4. ondalık değer
- 25. yüzdelik değer = 1. çeyrek değer

Yüzdelikler

- Yüzdelik $\neq$ Yüzde
- %40: Deneklerin 40/100'ü
- 40. yüzdelik: Deneklerin 40/100'ü kendisinden küçük değerli, 60/100'ü kendisinden büyük değerli olan nokta
- Ortanca, tepe değeri, çeyrek değerler, ondalıklar ve yüzdeliklerin hesaplanmasında verilerin bazıları işlem dahilinde, hepsi değil

Statistics

Derginin Scopus'tan atf alma durumu

N

Percent

Çeyrek değer genişliği nedir?

40

13.00

60

51.40

80

178.40

Continue

Cancel

Help

Varyans

- Bir dağılımı tanımlayan iki değer: ortalama ve varyans
- Dağılımdaki deneklerin ortalamadan ayrılışlarının ortalama ölçüsü
- Kitle varyansı: σ^2
- Örneklem varyansı: S^2

Standart sapma

- En çok kullanılan dağılım ölçüsü
- Önemli bir tanımlayıcı istatistik
- Standart sapma dağılımdaki değerlerin her birini dikkate alır
 - ✓ Aritmetik ortalamada olduğu gibi

Standart sapma

- Diğer dağılım ölçüleri dağılımdaki tüm değerleri dikkate almaz
 - ✓ Dağılım genişliği 2 değeri dikkate alır: en büyük ve en küçük değer
 - ✓ Çeyrek değerler genişliği dağılımın en ortadaki %50 lik kısmını dikkate alır

Standart sapma

- Anlamak önemli
 - ✓ Örnekleme kuramındaki kullanımı
- Simetrik ve tek tepeli dağılımlar için bir değişim ölçüsü
- Verilerin ortalama etrafındaki dağılışı
 - Aritmetik ortalama
 - Dağılım ne kadar genişse, standart sapma o kadar büyük

Standart sapma

- Dağılımdaki deneklerin ortalamadan ayrılışlarının karesel ortalaması
- Varyansın pozitif karekökü
- Kitle standart sapması: σ
- Örneklem standart sapması: S

Standart sapma

- Örn: 5 hocanın ödünç aldıkları kitap sayıları
✓ 15, 25, 26, 20, 14
- Aritmetik ortalama = $(15+25+26+20+14) / 5 = 20$
- Ortalamadan sapmalar: -5, 5, 6, 0, -6
- Sapmaların kareleri: 25, 25, 36, 0, 36
- Bunların toplamı: 122
- $122/5 = 22,4$
- 22,4'ün karekökü 4,9
- Standart sapma 4,9 (5 kitap)

$$S = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{n-1}}$$

Statistics

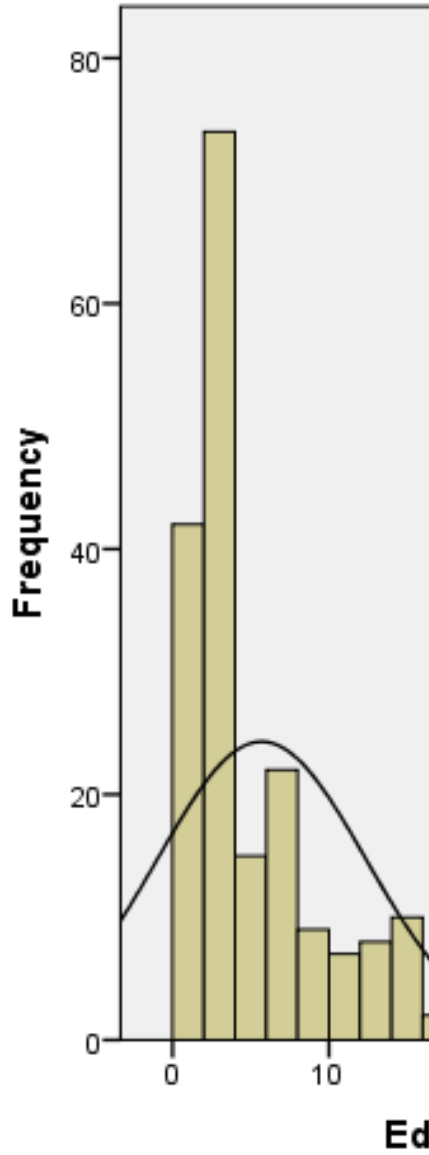
		Derginin Scopustan atıf alma durumu - total cites 3 years	Yayın kurulu kaç kişiden oluşuyor?
N	Valid	203	203
	Missing	0	0
Std. Deviation		277.373	40.511

nomical Society of Japan
and Development

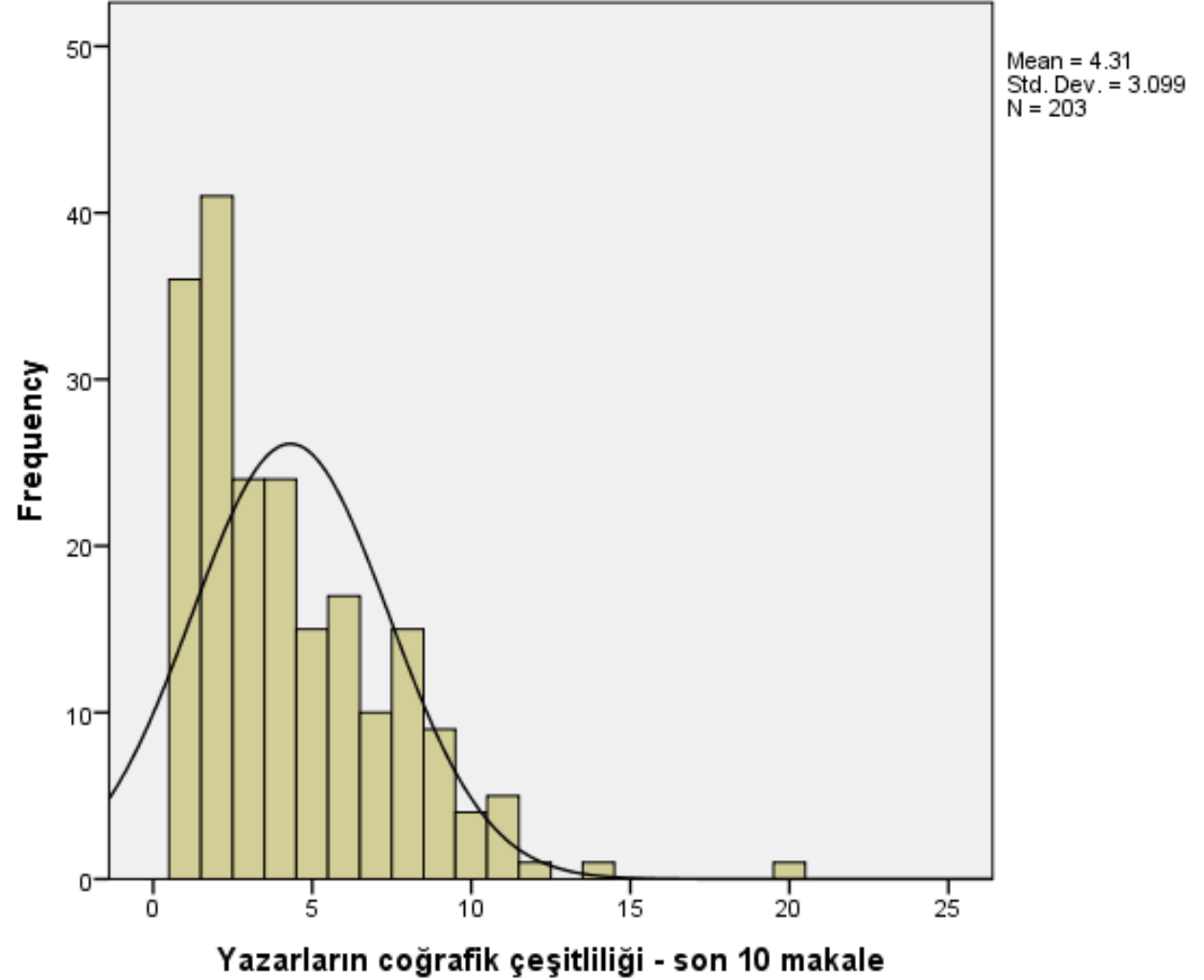
cal Engineering

☐ Variance	☐ Maximum	☐ Kurtosis
☐ Range	☐ S.E. mean	
Continue Cancel Help		

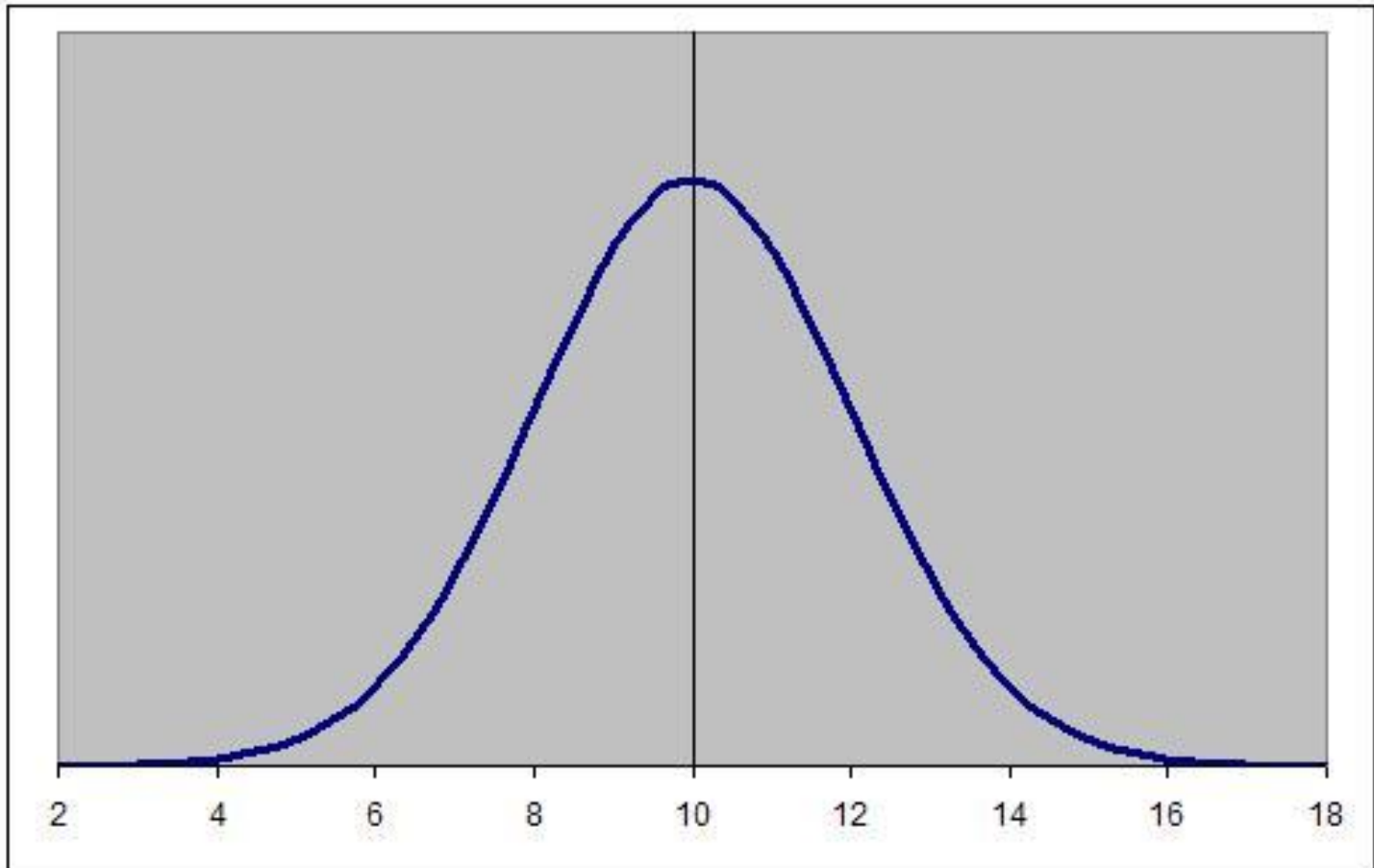
Editörlerin coğrafik çeşitliliği



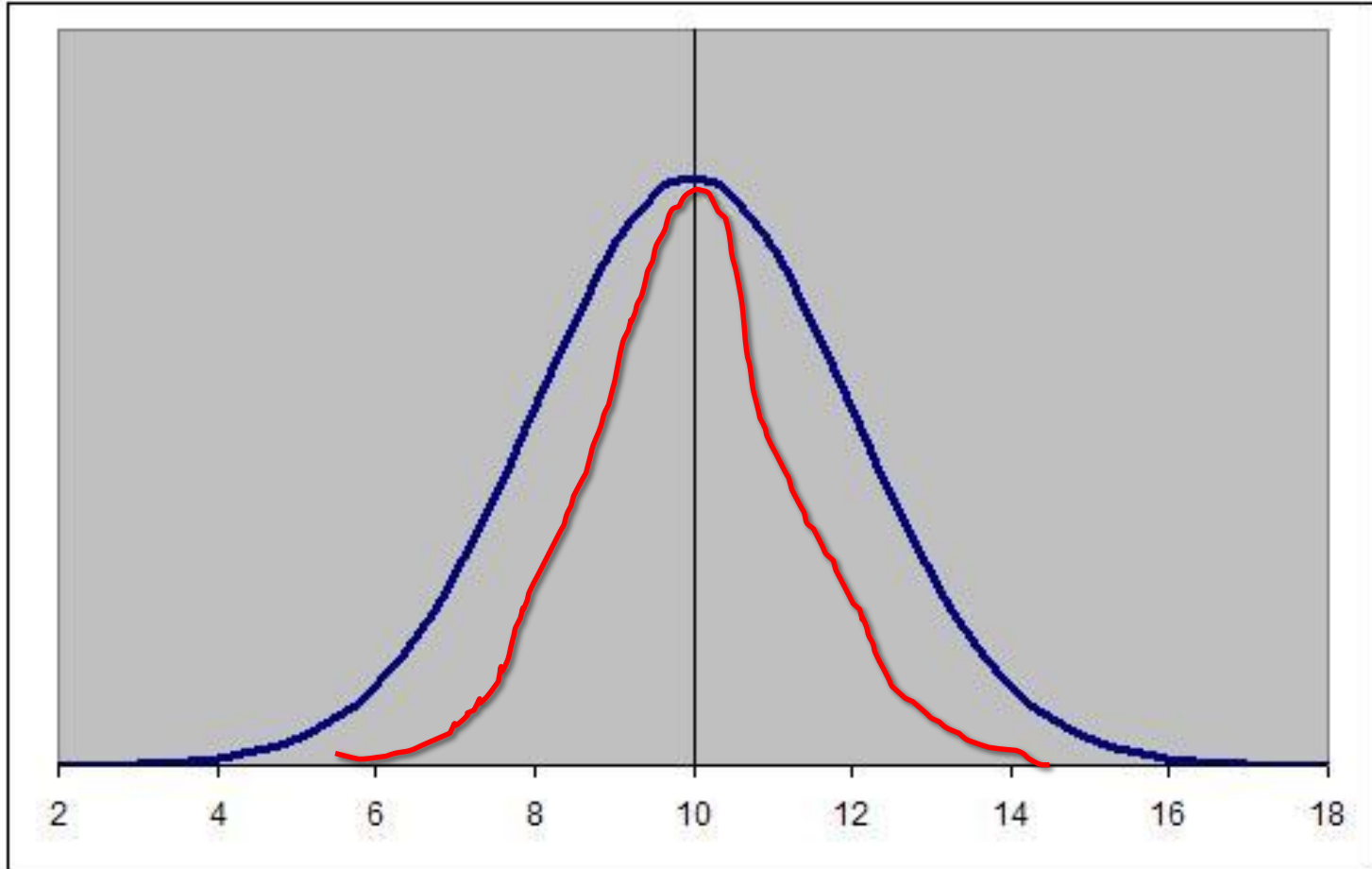
Yazarların coğrafik çeşitliliği - son 10 makale



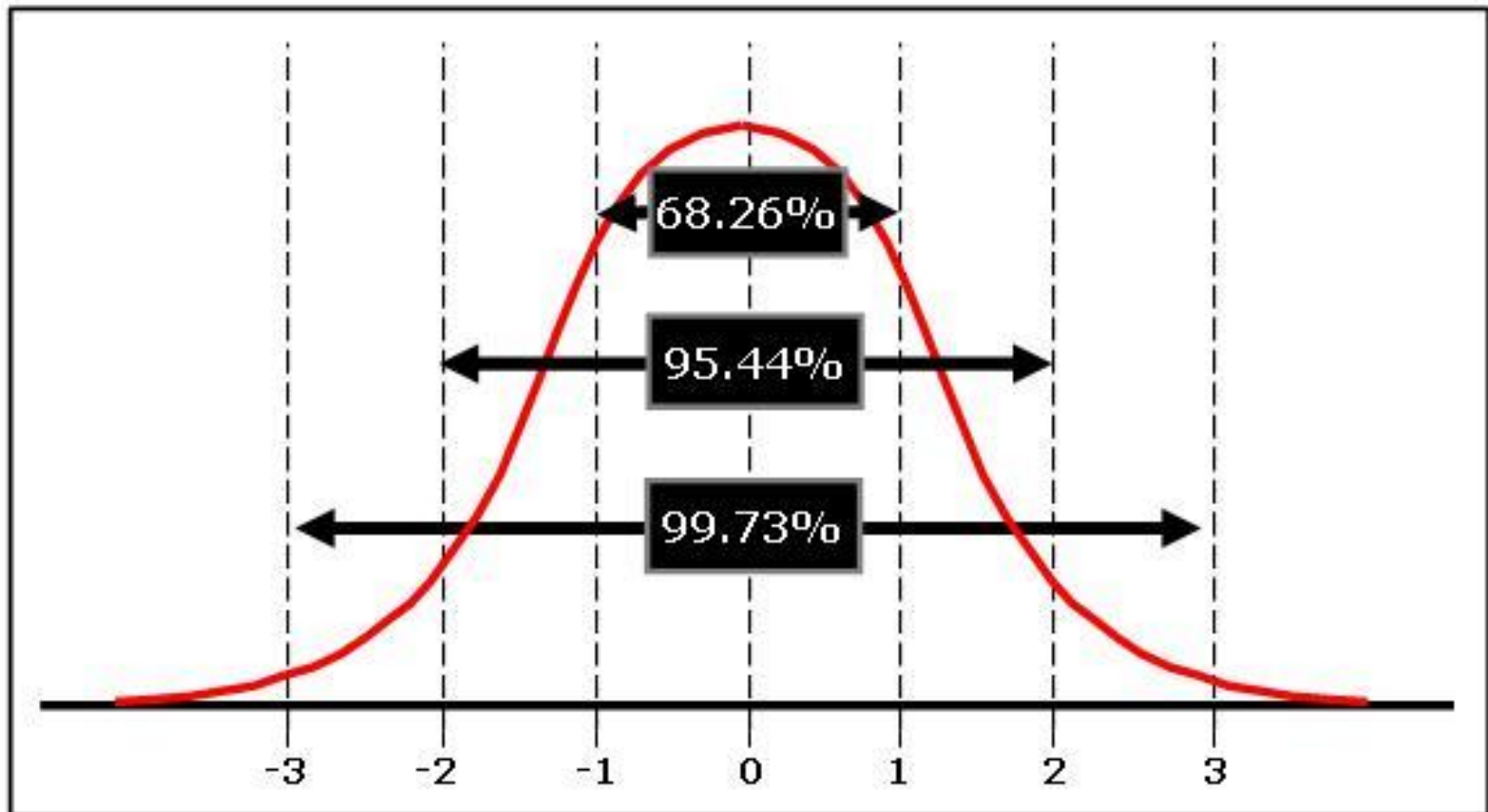
Normal dağılım



Hangisinin standart sapması daha büyük? Ortalamalar ile ilgili ne söylersiniz?



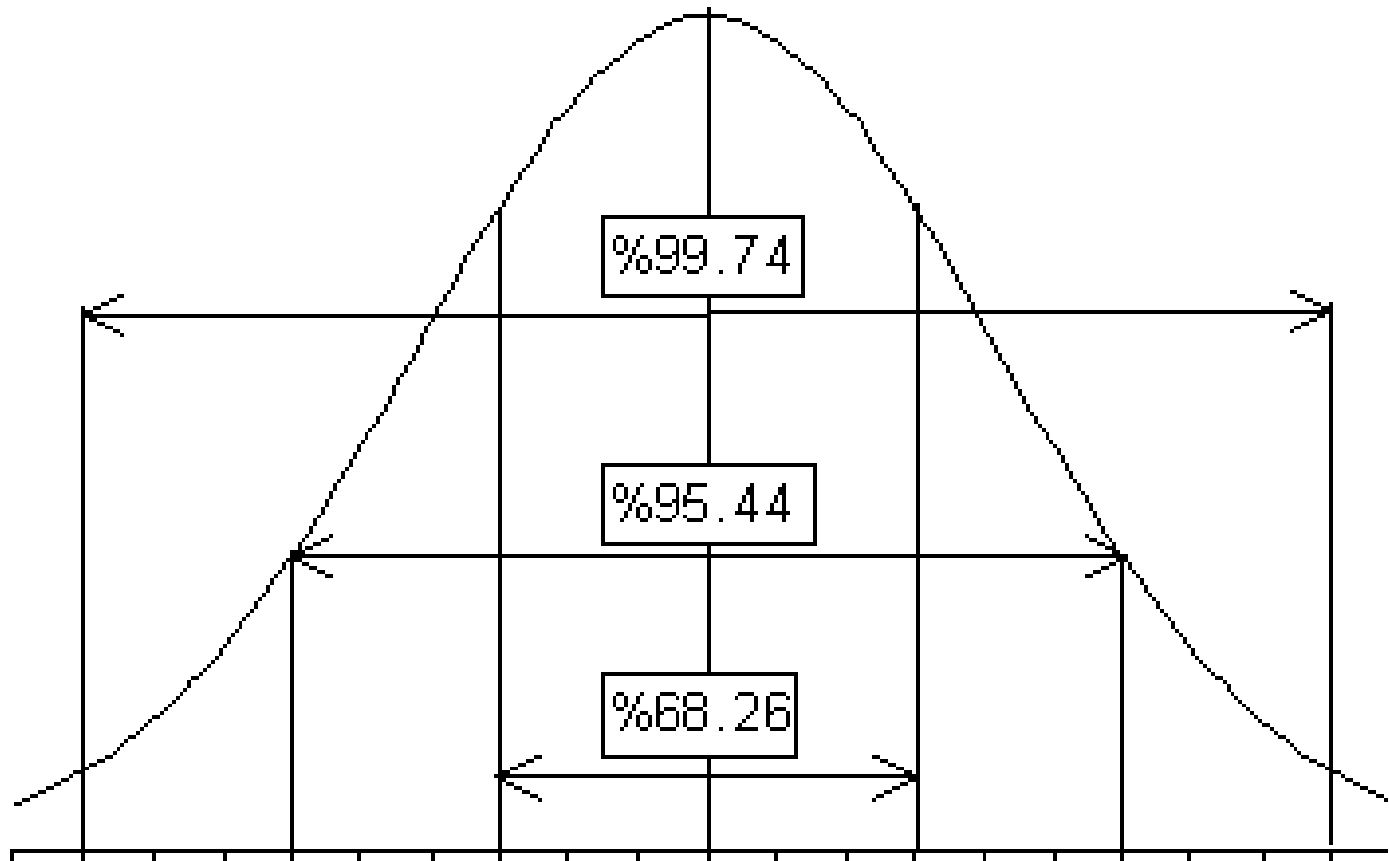
Standart normal dağılım



Standart normal dağılım

- Deneklerin %68,26'sı (%68)
 - ✓ Ortalama ± 1 standart sapma sınırları içinde
- Deneklerin %95,44'ü (%95)
 - ✓ Ortalama ± 2 standart sapma sınırları içinde
- Deneklerin %99,74'ü (%99)
 - ✓ Ortalama ± 3 standart sapma sınırları içinde
- Deneklerin %100'ü
 - ✓ Ortalama ± 4 standart sapma sınırları içinde

Standart normal dağılım



Standart hata

- Normal dağılım gösteren bir kitleden çekilen örneklemelerin ortalamalarının gösterdiği dağılışın standart sapması
- Örneklem ortalamalarının standart hatası
- Standart hata = $S / \sqrt{n}$
- Örneklem ortalaması standart hatası ile birlikte verilir
- Örneklemden hesaplanan her istatistik değerin kendi standart hatası var

Standart hata

- 30 hastanın kandaki alyuvar sayıları ortalaması 5,398 ve standart sapması 0,3928'dir. Alyuvar sayılarının ortalamasının standart hatası nedir?
- Standart hata = $0,3928/\sqrt{30} = 0,0717$
- $5,398 \pm 0,0717 = (5,3263; 5,4697)$
- Alınacak farklı örneklemlerden bulunacak örneklem ortalamalarının değişim aralığı

Statistics

		Derginin Scopustan atıf alma durumu - total cites 3 years	Yayın kurulu kaç kişiden oluşuyor?
N	Valid	203	203
	Missing	0	0
Mean		138.46	27.92
Std. Error of Mean		19.468	2.843

and Development

cal Engineering

Engineering

Continue

Cancel

Help

Değişim katsayısı

- $(\text{Standart sapma} / \text{Ortalama}) \cdot 100$
- Standart sapmanın ortalamaya yüzdesi
- Denekler arasındaki değişimin azlığı ya da çokluğu hakkında bilgi
- % olarak gösterim
- Ortalama yerine ortanca kullanılıyorsa değişim katsayısı = $(\text{Çeyrek sapma} / \text{Ortanca}) \cdot 100$

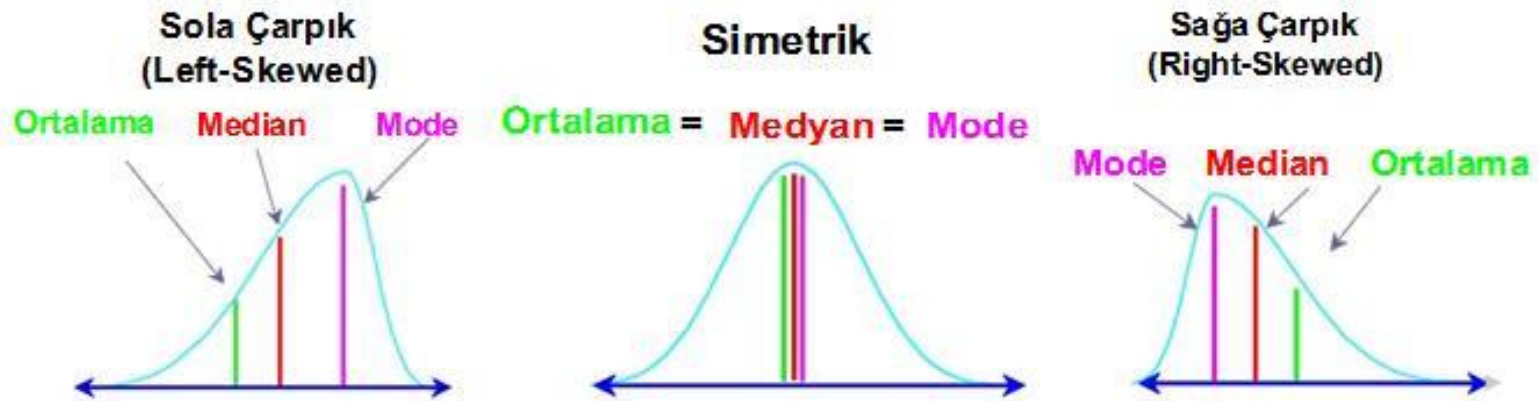
Scopustan atıf alma durumu adlı değişken için her iki formülasyona göre değişim katsayısını bulup yorumlayın

Çarpıklık katsayısı

- Çarpıklık: Ortanca, tepe değeri ve aritmetik ortalama arasındaki bağıntı
- Çarpıklık katsayısı 0 ise dağılım ortalamaya göre simetrik
- Çarpıklık katsayısı 0'dan küçük (negatif değerli) ise dağılım – yöne eğimli, sola çarpık
- Çarpıklık katsayısı 0'dan büyük (pozitif değerli) ise dağılım + yöne eğimli, sağa çarpık

Aritmetik, Ortanca ve Tepe Değeri

- Her zaman eşit değil



- Sağa çarpık (pozitif yönlü, + yöne eğimli)
 - ✓ Aritmetik ortalama > Ortanca > Tepe Değeri
- Sola çarpık (negatif yönlü, - yöne eğimli)
 - ✓ Aritmetik ortalama < Ortanca < Tepe Değeri

Basıklık katsayısı

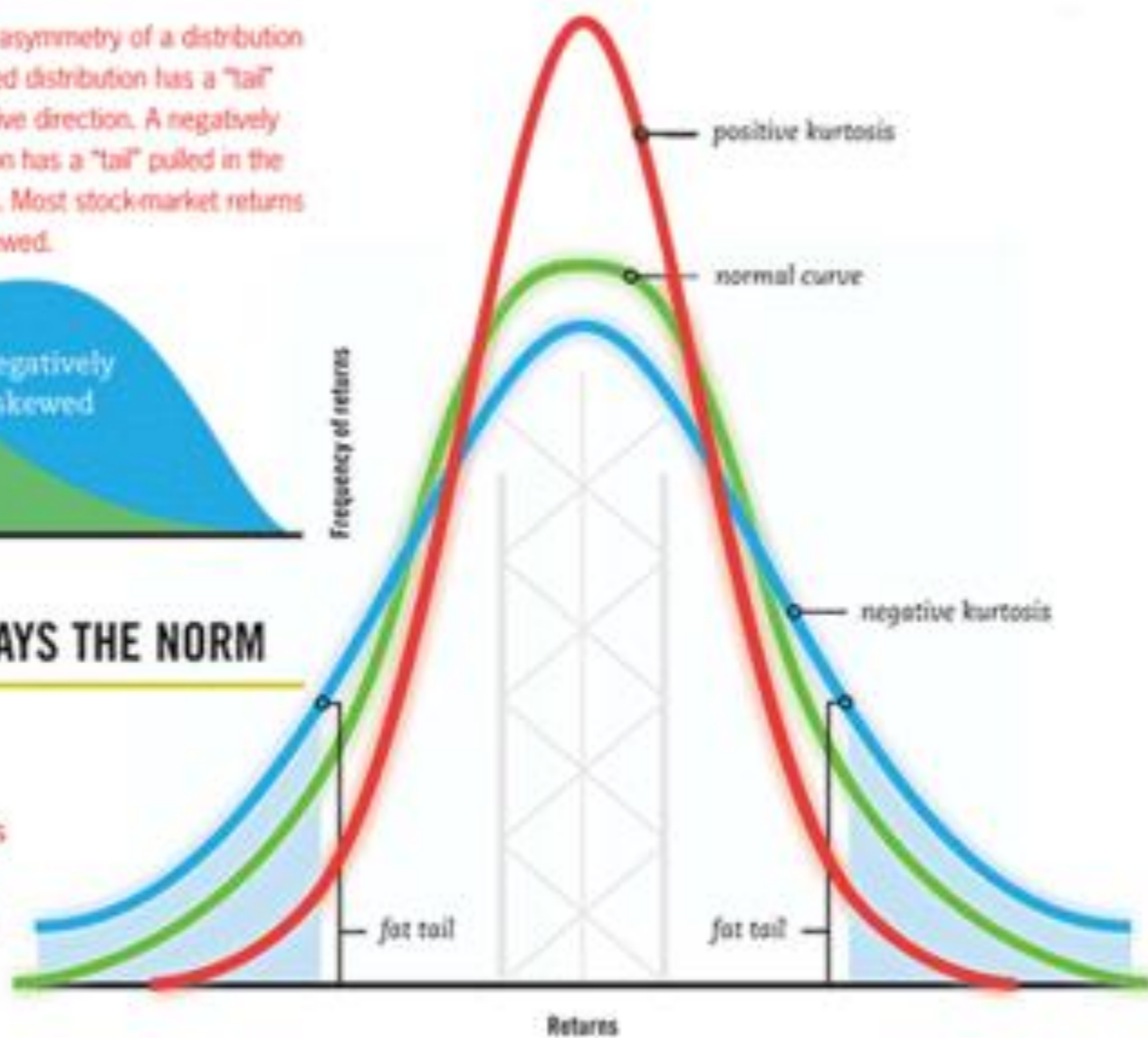
- Verilerin gösterdiği dağılımın standart normal dağılıma göre yüksekliği
- Sivri dağılım / basık dağılım
- Basıklık katsayısı 0 ise dağılımın yüksekliği standart normal dağılıma uygun, aynı
- Basıklık katsayısı 0'dan küçük (negatif değerli) ise dağılım basık
- Basıklık katsayısı 0'dan büyük (pozitif değerli) ise dağılım sivri

Skewness is the asymmetry of a distribution. A positively skewed distribution has a "tail" pulled in the positive direction. A negatively skewed distribution has a "tail" pulled in the negative direction. Most stock-market returns are negatively skewed.



NORMAL NOT ALWAYS THE NORM

Kurtosis refers to how peaked the curve is: steeper means positive kurtosis and flatter means negative kurtosis. Fat tails occur when there are more outsize returns on the downside or upside, or both, than the normal curve suggests.



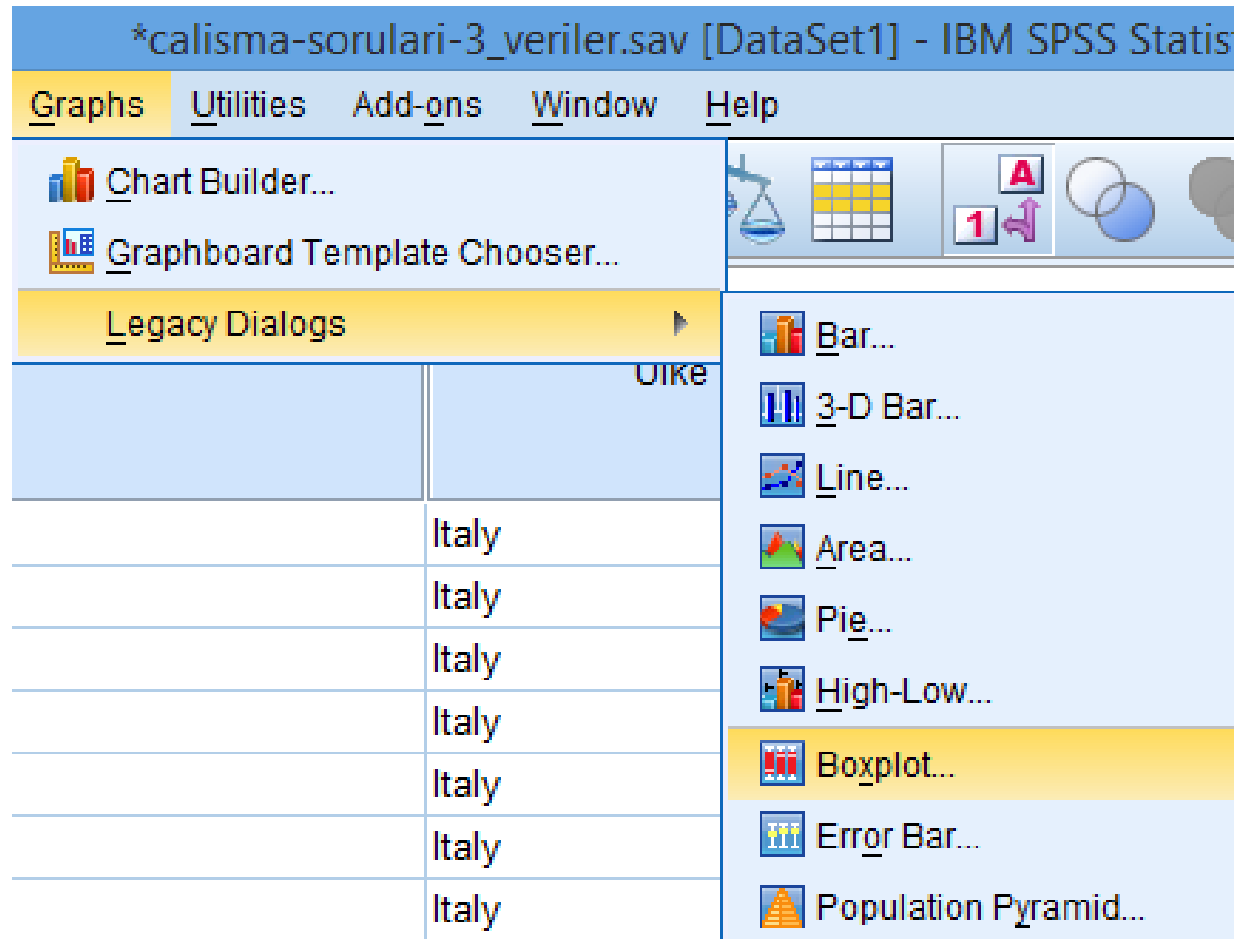
Çarpıklık / Basıklık katsayısı

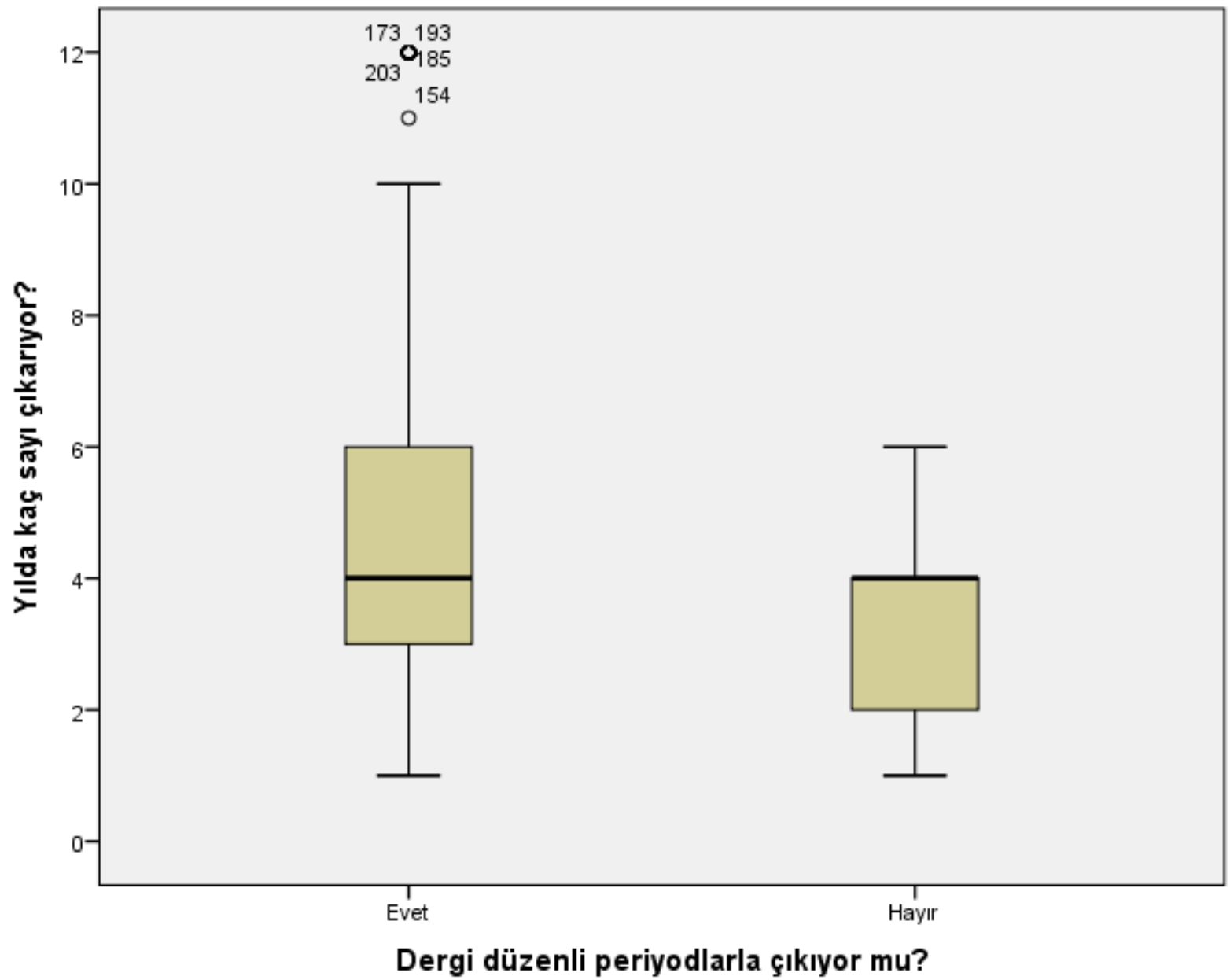
- Birlikte verilir
- İkisi de 0 değerini alıyor ise dağılım standart normal dağılıma uygun dağılmakta

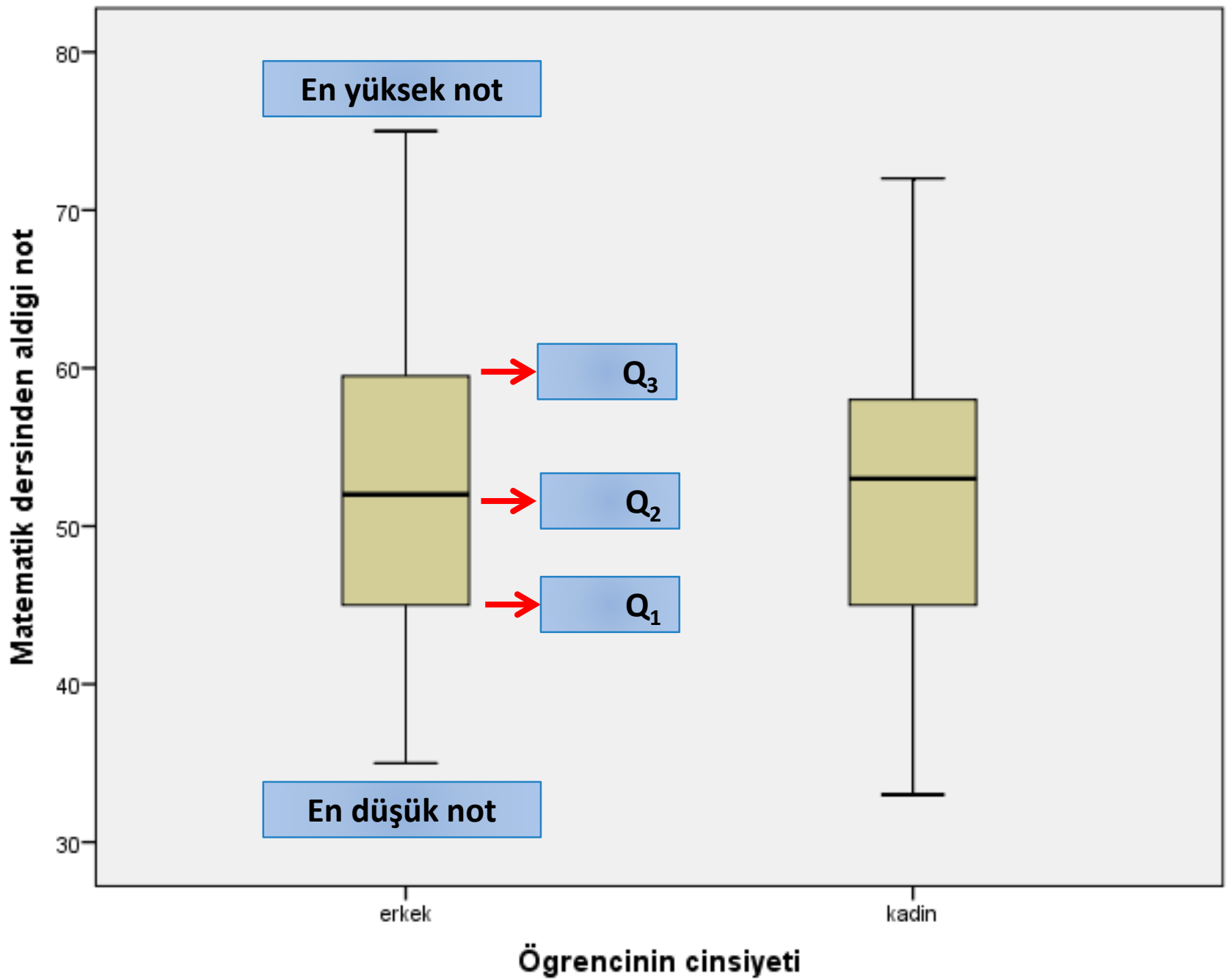
Statistics

		Editörlerin coğrafik çeşitliliği	Yazarların coğrafik çeşitliliği - son 10 makale
N	Valid	203	203
	Missing	0	0
Skewness		2.809	1.270
Std. Error of Skewness		.171	.171
Kurtosis		13.002	2.513
Std. Error of Kurtosis		.340	.340

Boxplot







Boxplot

- Ortanca çizgisi kutucuğun merkezine yakınsa
 - ✓ Dağılımın ortadaki %50'lik kısmı simetrik
- Ortanca çizgisi birinci çeyrek ya da üçüncü çeyreğe yakınsa
 - ✓ Dağılım çarpık
- Kuyrukların uzunluğu
 - ✓ Eşitse, simetrik dağılım
 - ✓ Eşit değilse, çarpık dağılım
- Sapan değer varsa

Konum ve Dağılım Ölçüleri

BBY606 Araştırma Yöntemleri
Güleda Doğan