



HACETTEPE
ÜNİVERSİTESİ
İSTATİSTİK BÖLÜMÜ

İST155 İSTATİSTİĞE GİRİŞ I

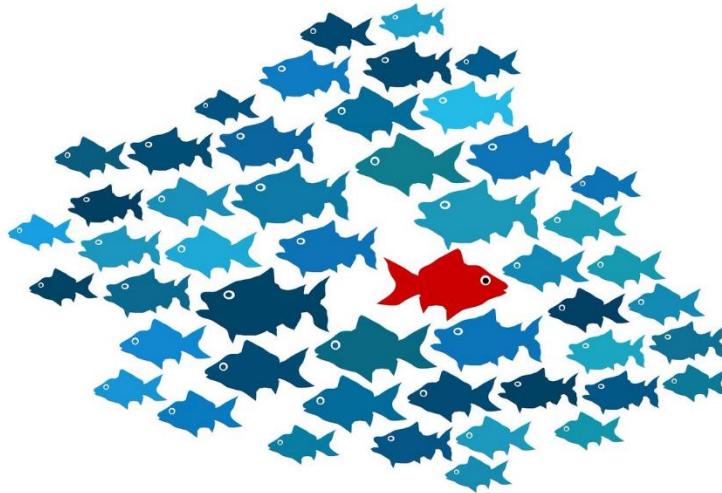
DERS 10
Aykırı Değer İncelemesi

**Ders sorumluları: Prof.Dr.Serpil AKTAŞ ALTUNAY (01 Şubesi)
Doç.Dr. Ayten YİĞİTER (02 Şubesi)**

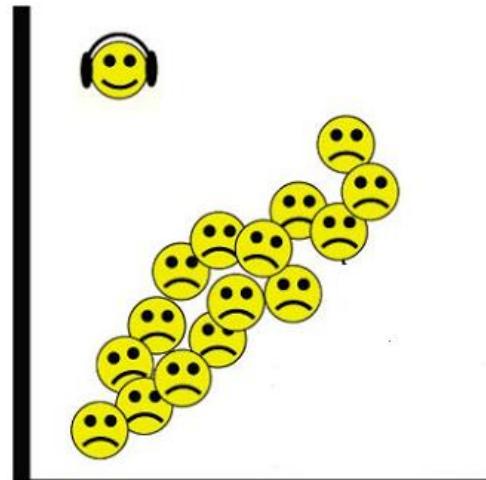
AYKIRI DEĞER İNCELEMESİ

Aykırı/Sapan değer (Outlier): Dağılımdaki gözlem değerlerinden farklılık gösteren uç değer ya da uç değerlerdir.

Aykırı değerin hesaplamalara katılması ya da katılmaması ortalamada ya da diğer istatistiklerde farklı sonuçlara götürebilir. Kitledeki uç değerlerin örneklemeye çıkması olasılığı çok düşüktür. Örneklemde görülen aykırı değerin pek çok sebebi olabilir.



Kaynak: <https://medium.com/datarunner/outlier-analiz-y%C3%B6ntemleri-8ec0e63551f8>



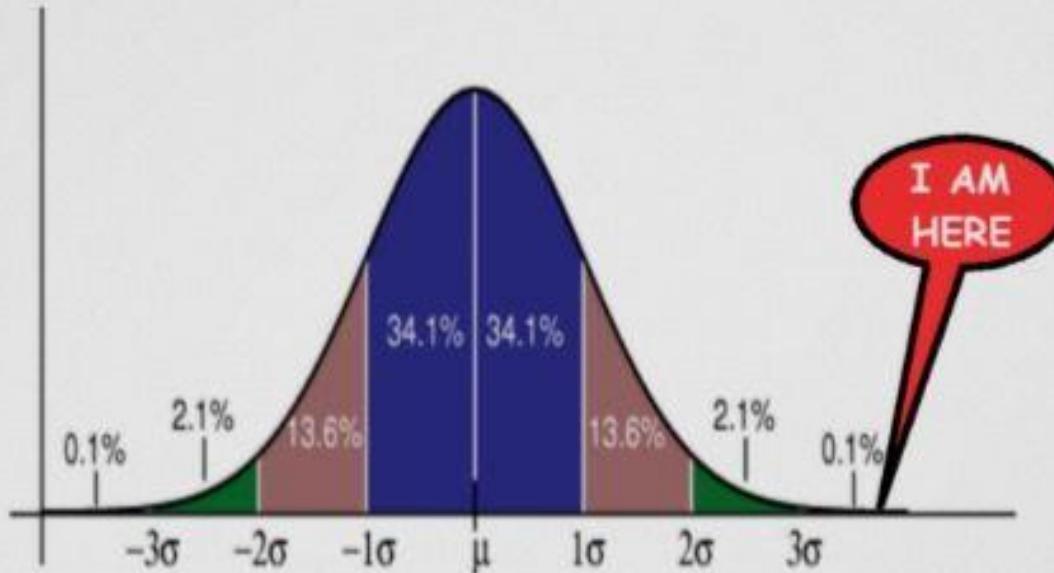
Aykırı Değer Neden Ortaya Çıkar?

- Örnekleme hatası
- Ölçüm hatası
- Veri işleme hatası (operatör hatası)
- Rastgele
-

Aykırı Değer Nasıl Tespit Edilir?

- Kutu Grafiği (Box-Plot) ile ,
- $\bar{X} \mp 2S$ ya da $\bar{X} \mp 3S$ sınırlarına göre,
- Kontrol kartları ile,
- $\left| \frac{X_i - \bar{X}}{S} \right| \geq 3$ (Z skoru) ile,
- ...

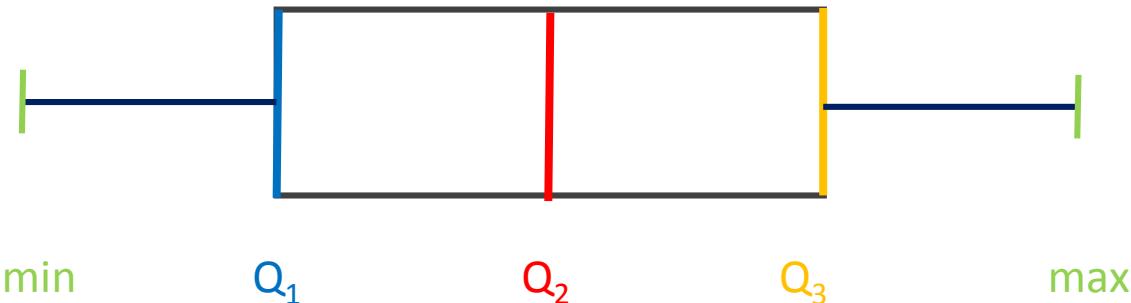
STATISTICAL OUTLIER



INSIDE

© WORDS & UNWORDS

KUTU GRAFİĞİ (BOX PLOT)



Kutu çizimi verilerin çeyrek değerlerine dayalı bir çizimdir.

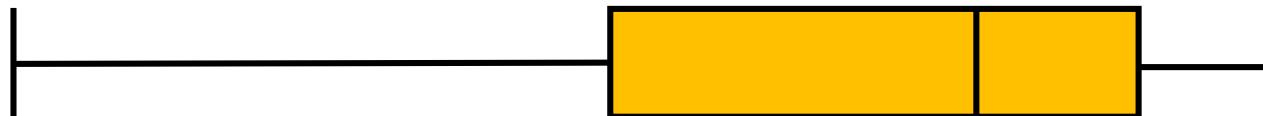
Adım 1: Q_1 , Q_2 , Q_3 çeyrek değerleri hesaplanır.

Adım 2: $d=Q_3-Q_1$ hesaplanarak, bu uzunluk kutunun uzun kenarı olacak biçimde yatay bir kutu çizilir.

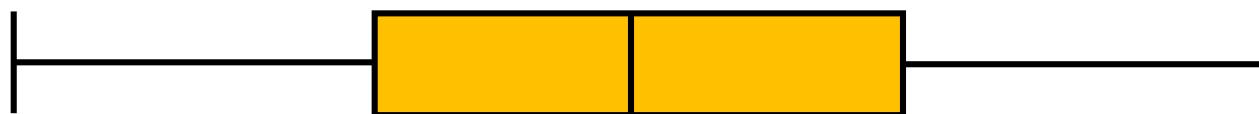
Adım 3: Ortancanın yeri, Q_2-Q_1 farkı ve Q_3-Q_2 farkına göre kutu içinde dikey çizgi ile belirtilir.

Adım 4: Kuyruklar, veri kümesindeki en küçük (min) ve en büyük (max) değerlerden oluşur. Bu değerlerin aykırı değer olmaması gereklidir.

Sola çarpık kutu çizimi:



Simetrik kutu çizimi:



Sağça çarpık kutu çizimi



Örnek 1 : Bir online alışveriş sitesinde, saat 10.00-10.15 arası alışveriş yapan 76 kişinin yaptıkları harcama için kutu grafiğini çiziniz.

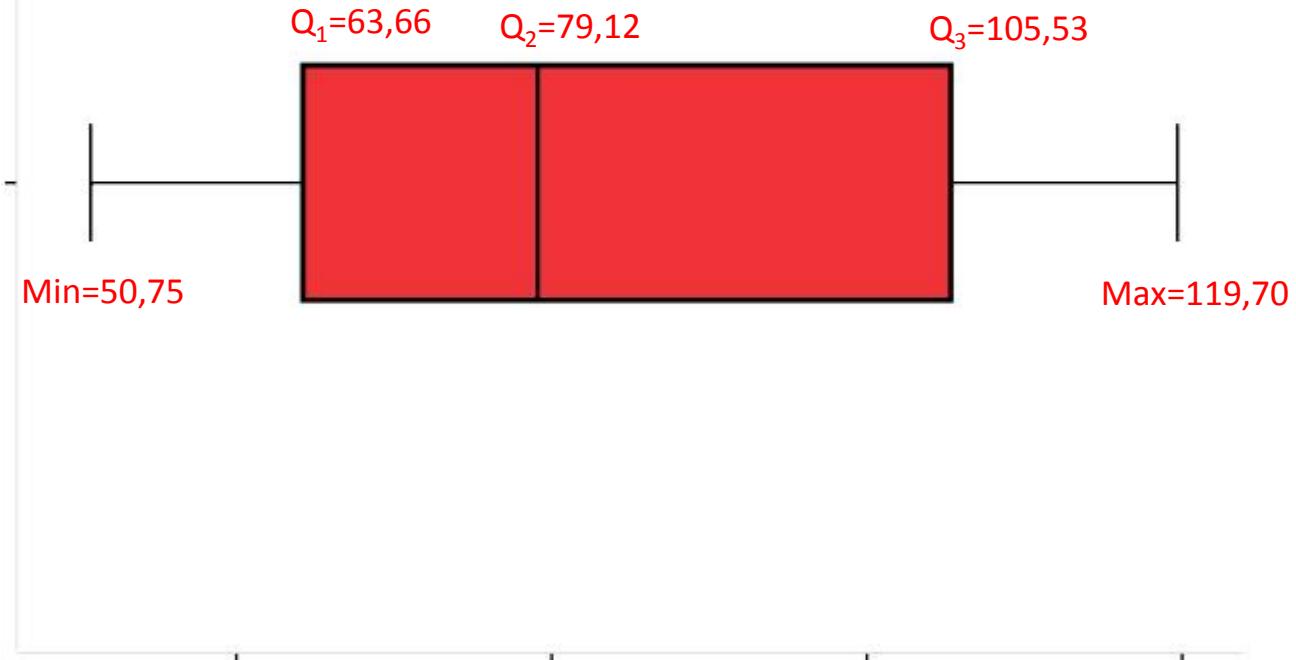
$$Q_2 = 79,12$$

$$Q_1 = 63,66$$

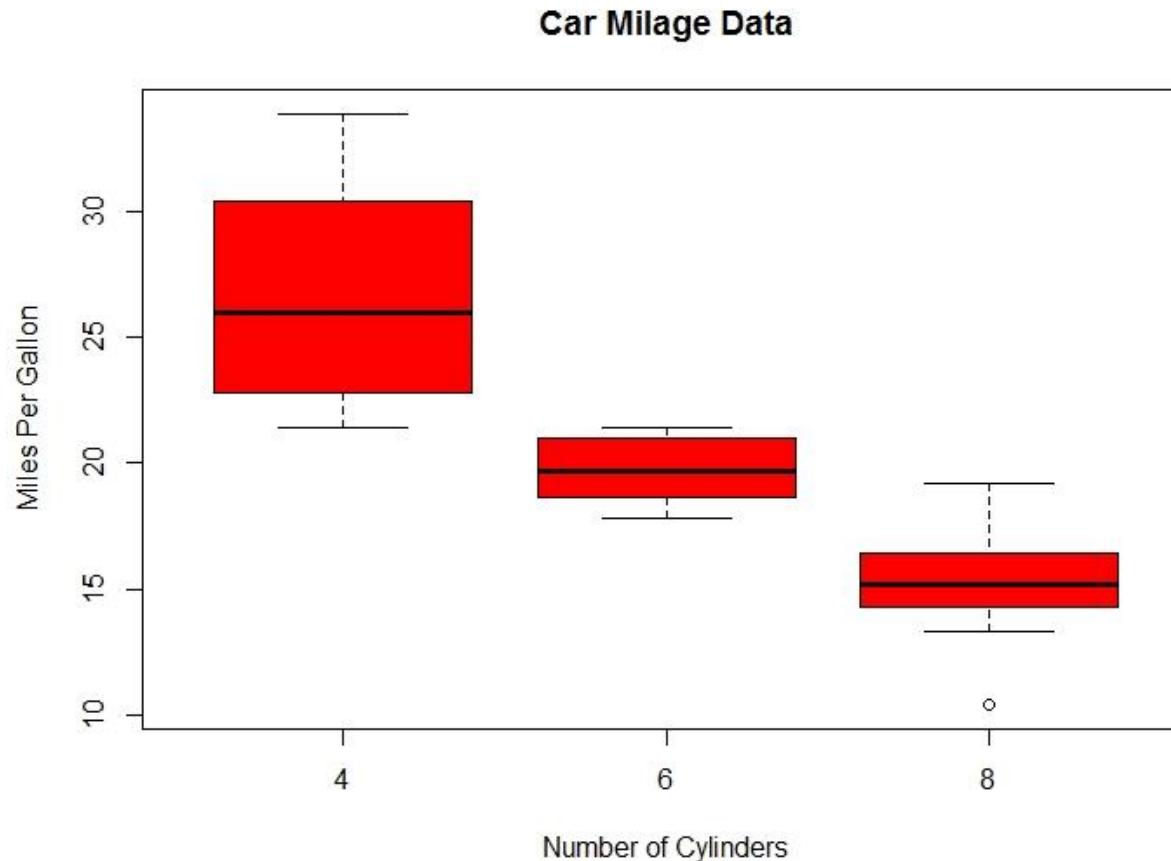
$$Q_3 = 105,53$$

$$\text{Min} = 50,75$$

$$\text{Max} = 119,70$$



Birden fazla değişken için kutu grafiği bir arada çizilebilir.



<https://www.statmethods.net/graphs/boxplot.html>

1. Kutu Grafiği (Box-Whisker Plot) ile aykırı değer incelemesi

Kutu grafiği üzerinde aykırı değer incelemesi için aşağıdaki adımlar izlenir.

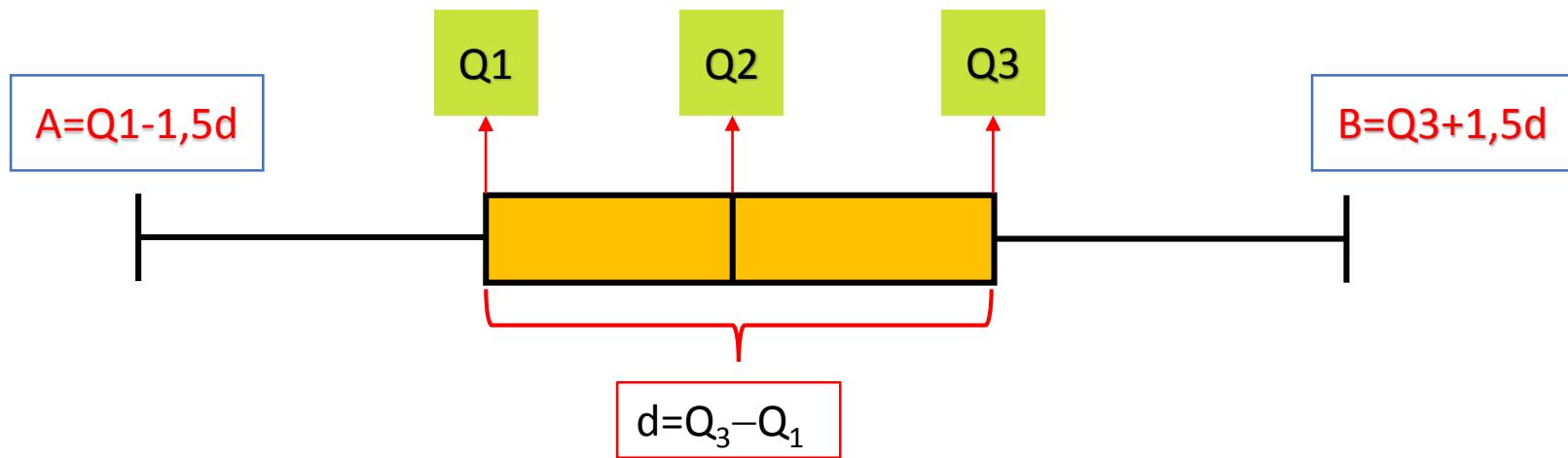
Adım 1: $d=Q_3-Q_1$ hesaplanarak kutunun her iki ucundan alt sınır $A=Q_1-1,5d$ ve üst sınır $B=Q_3+1,5d$ uzunlukları şüpheli gözlem sınırları olarak belirlenmiş olur (A ve B 'ye aynı zamanda CUTOFF denir).

Adım 2: Alt sınır A 'dan daha küçük ve üst sınır B 'den büyük değerler aykırı değer olarak adlandırılır.

$A=Q_1-1.5(Q_3-Q_1)$ 'den küçük gözlemler aykırı değerdir.

$B=Q_3+1.5 (Q_3-Q_1)$ 'den büyük gözlemler aykırı değerdir.

Aynı zamanda «Box-Whisker Plot» olarak da bilinir.



Örnek 2: Bir bölgede 9 gün boyunca metre kare başına düşen yağmur miktarı verilmiştir. 9.gün aykırı değer midir?

i.	1.	2.	3.	4.	5.	6.	7.	8.	9.
X _i	2	5	8	10	11	14	17	20	50

$$i = \frac{n+1}{4} = 2,5$$

$$Q_1 = \frac{x_2 + x_3}{2} = \frac{5 + 8}{2} = 6,5$$

$$i = \frac{n+1}{2} = 5$$

$$Q_2 = x_5 = 11$$

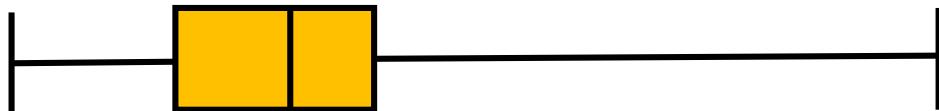
$$i = \frac{3(n+1)}{4} = 7,5$$

$$Q_3 = \frac{x_7 + x_8}{2} = \frac{17 + 20}{2} = 18,5$$

$$d = Q_3 - Q_1 = 18,5 - 6,5 = 12$$

$$A = Q_1 - 1,5d = 6,5 - 1,5 \times 12 = -11,5$$

$$B = Q_3 + 1,5d = 18,5 + 1,5 \times 12 = 36,5$$



9

9.Gözlem B değerini aştiği için aykırı değerdir.

2. Ortalama \pm Standart Sapma bağıntısı ile aykırı değer incelemesi

$$\bar{X} \mp 2S \quad \text{ya da} \quad \bar{X} \mp 3S$$

Sınırları dışında kalan gözlemler aykırı değerdir.

Örnek 3: Bir bölgede 9 gün boyunca metre kare başına düşen yağmur miktarı verilmiştir. $\bar{X} \mp 2S$ ve $\bar{X} \mp 3S$ bağıntılarından yararlanarak aykırı değer incelemesi yapınız.

i.	1.	2.	3.	4.	5.	6.	7.	8.	9.
x_i	2	5	8	10	11	14	17	20	50

$$\bar{X} = 15,22$$

$$S = 14,2$$

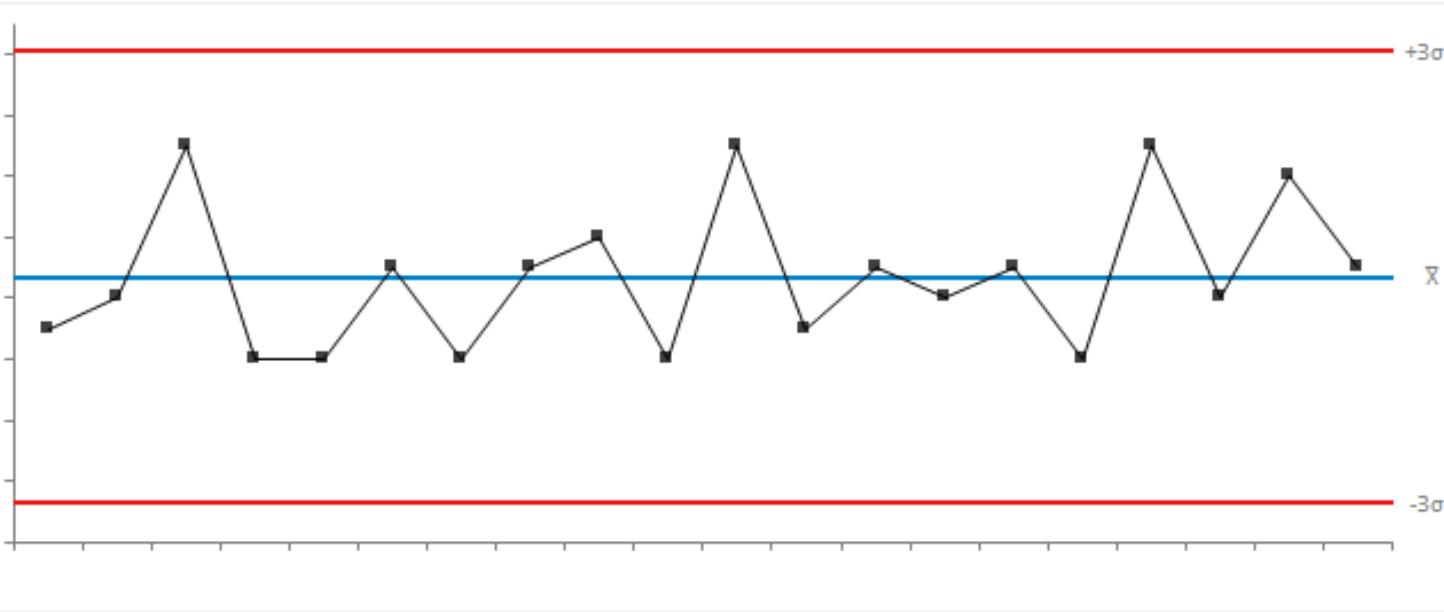
$$\bar{X} \mp 2S \Rightarrow (15,22 - 2 \times 14,2 ; 15,22 + 2 \times 14,2) = (-13; 43,6)$$

$\bar{X} \mp 2S$ sınırlına göre 9. gözlem değeri (50) aykırı değer olur.

$$\bar{X} \mp 3S \Rightarrow (15,22 - 3 \times 14,2 ; 15,22 + 3 \times 14,2) = (-27; 57,8)$$

$\bar{X} \mp 3S$ sınırlına göre, aykırı değer yoktur denir.

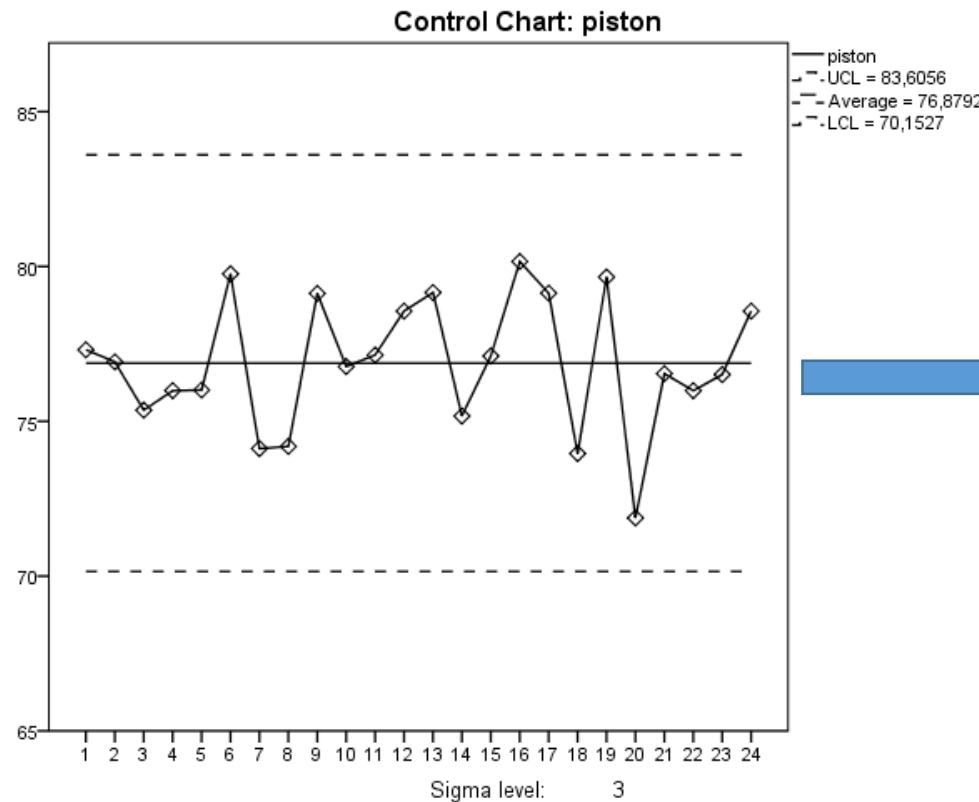
3. Kontrol Kartları ile aykırı değer incelemesi



$\bar{X} \mp 3S$ sınırları dışında kalan gözlemler aykırı değer yani «hatalı üretim» dir.
6 Sigma yaklaşımında(sürecinde) bu değer $\bar{X} \mp 6S$ olarak kabul edilir.

Örnek 4: Otomobil pistonu üreten bir fabrikada, piston çaplarının kalite kontrolünün sağlanması için üretimdeki bir partiden $n=24$ birimlik bir örneklem alınıyor. Kontrol kartını çizerek süreç kontrol altında mıdır gösteriniz.

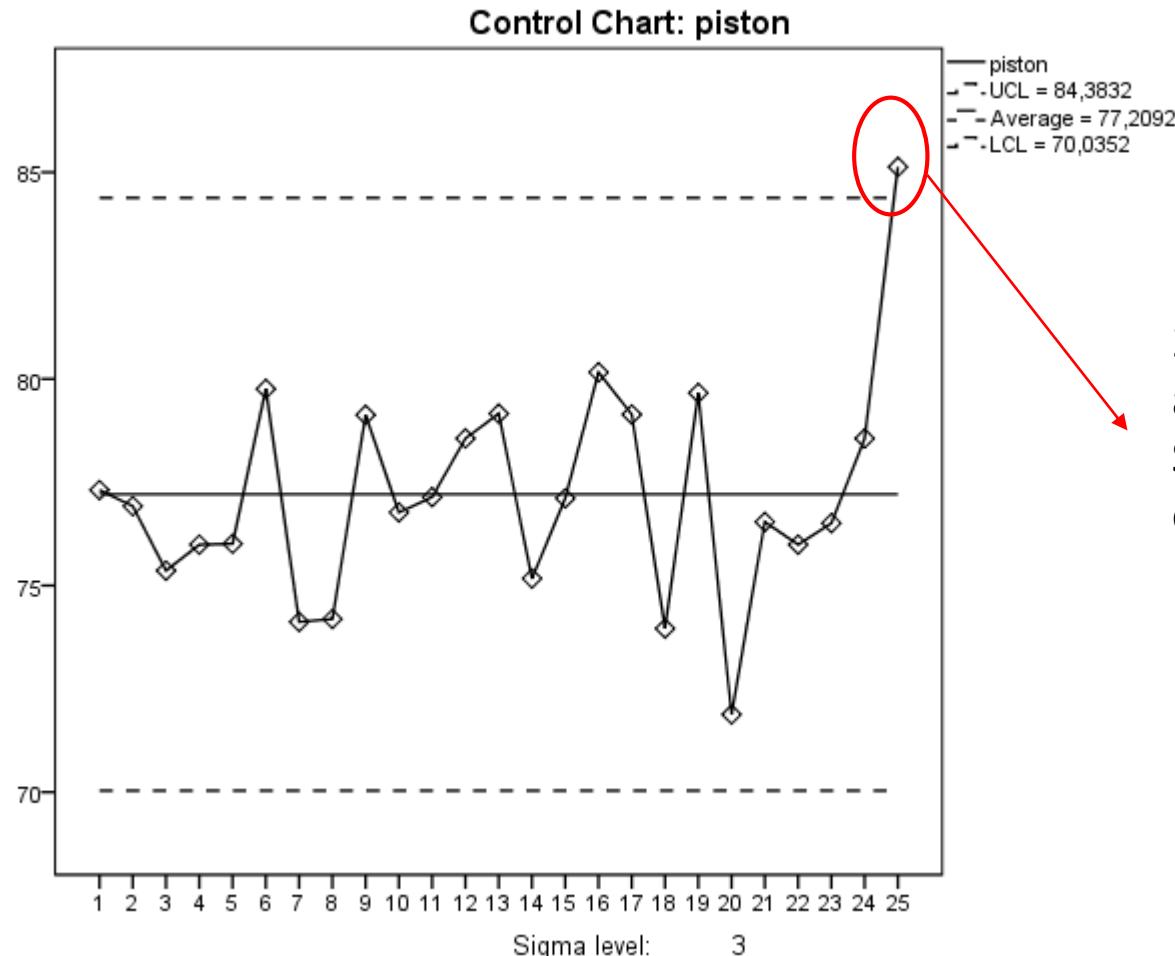
Piston çapı	
77,31	79,16
76,92	75,17
75,36	77,11
75,99	80,16
76,01	79,14
79,76	73,96
74,12	79,66
74,19	71,88
79,13	76,54
76,77	75,99
77,14	76,51
78,56	78,56



Aykırı değer yok...
Süreç kontrol altında

Örnek 5: Süreç devam ederken 25. gözlem değeri 85,13 olarak kaydedilmiştir.
Bu değer aykırı değer midir?

Piston Çapı
77,31
76,92
75,36
75,99
76,01
79,76
74,12
74,19
79,13
76,77
77,14
78,56
79,16
75,17
77,11
80,16
79,14
73,96
79,66
71,88
76,54
75,99
76,51
78,56
85,13



25. gözlem
aykırı değerdir.
Süreç kontrol
edilmelidir.

4. Z skoru ile aykırı değer incelemesi

$\left| \frac{X_i - \bar{X}}{S} \right| \geq 3$ olan değerler aykırı değerdir. $i=1,2,\dots,n$

Örnek 6: Bir bölgede 9 gün boyunca metre kare başına düşen yağmur miktarı verilmiştir. Aykırı değer var mı inceleyiniz.

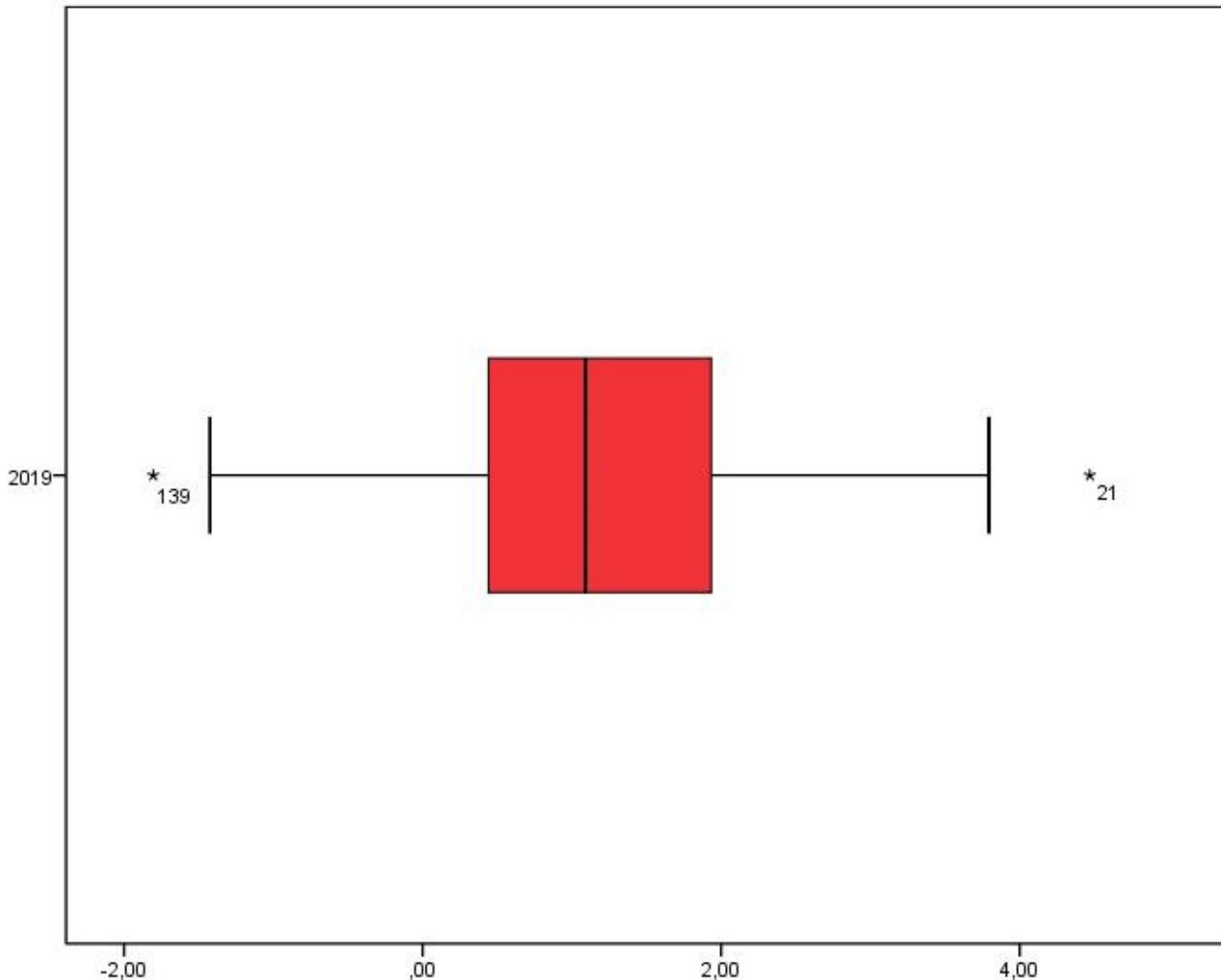
i.	1.	2.	3.	4.	5.	6.	7.	8.	9.
X _i	2	5	8	10	11	14	17	20	50

$$\bar{X} = 15,22$$
$$S = 14,2$$

X _i	$\left \frac{X_i - \bar{X}}{S} \right $
50	2.449296
20	0.33662
17	0.125352
14	0.08592
11	0.29718
10	0.36761
8	0.50845
5	0.71972
2	0.93099

Aykırı değer yoktur denir.

Örnek 7: Dünya Bankası verilerinden 237 ülkenin 2019 yılı nüfus artış hızları alınarak aykırı değer incelemesi yapılmıştır. (Kaynak:<https://data.worldbank.org/indicator/SP.POP.GROW>)

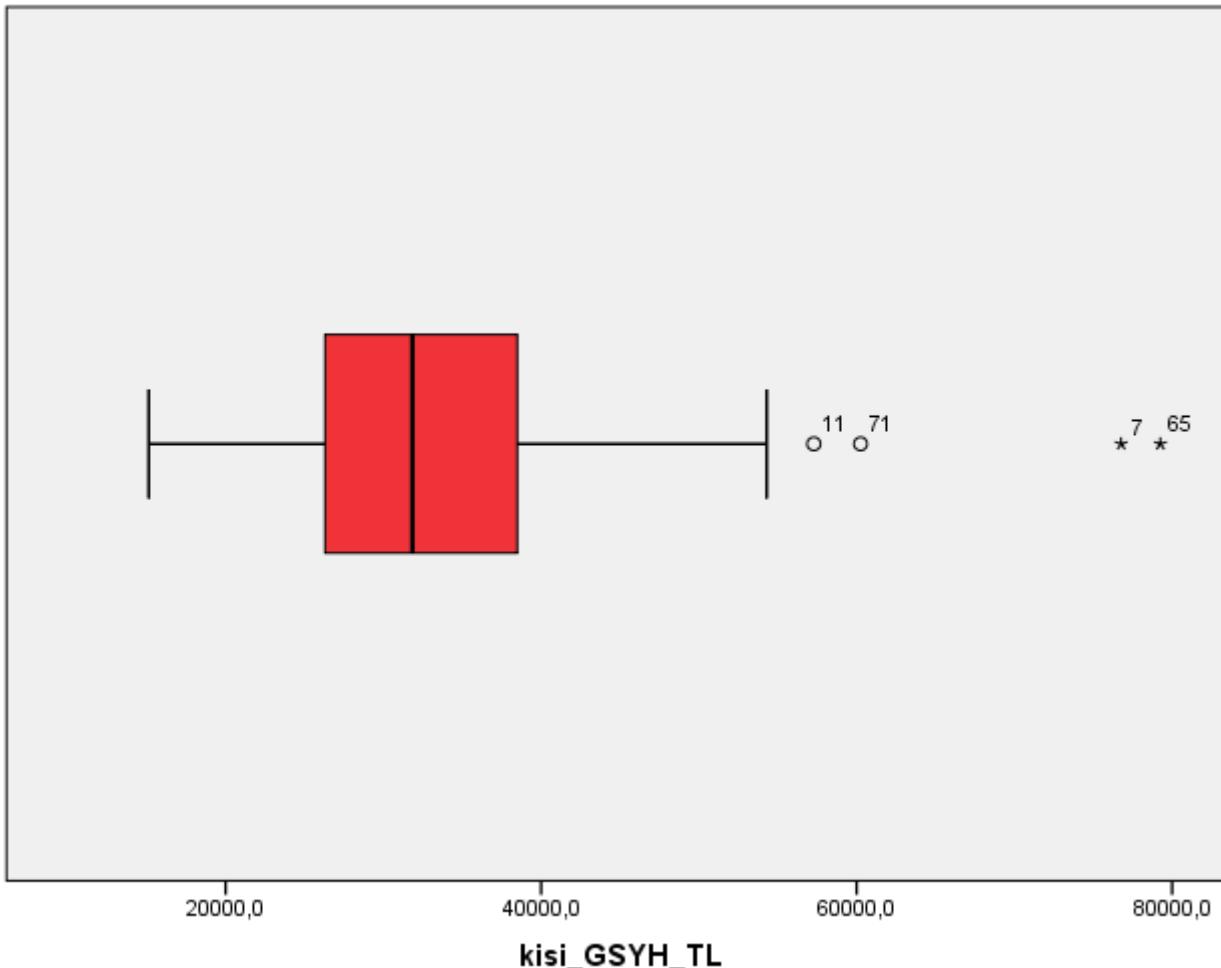


21.gözlem:
Bahreyn
Nüfus artış
hızı=4,469

139.gözlem:
Moldova
Nüfus artış
hızı=-1,805

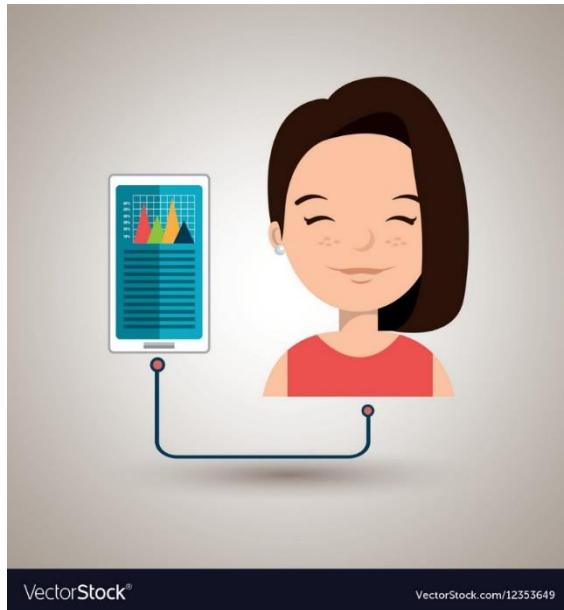
Türkiye ?
Nüfus artış
hızı=1,339

Örnek 8: Türkiye'deki illerin 2018 yılı kişi başı düşen GSYH ilişkin kutu grafiği.
(Kaynak:<https://biruni.tuik.gov.tr/ilgosterge/?locale=tr,13.12.2020>)



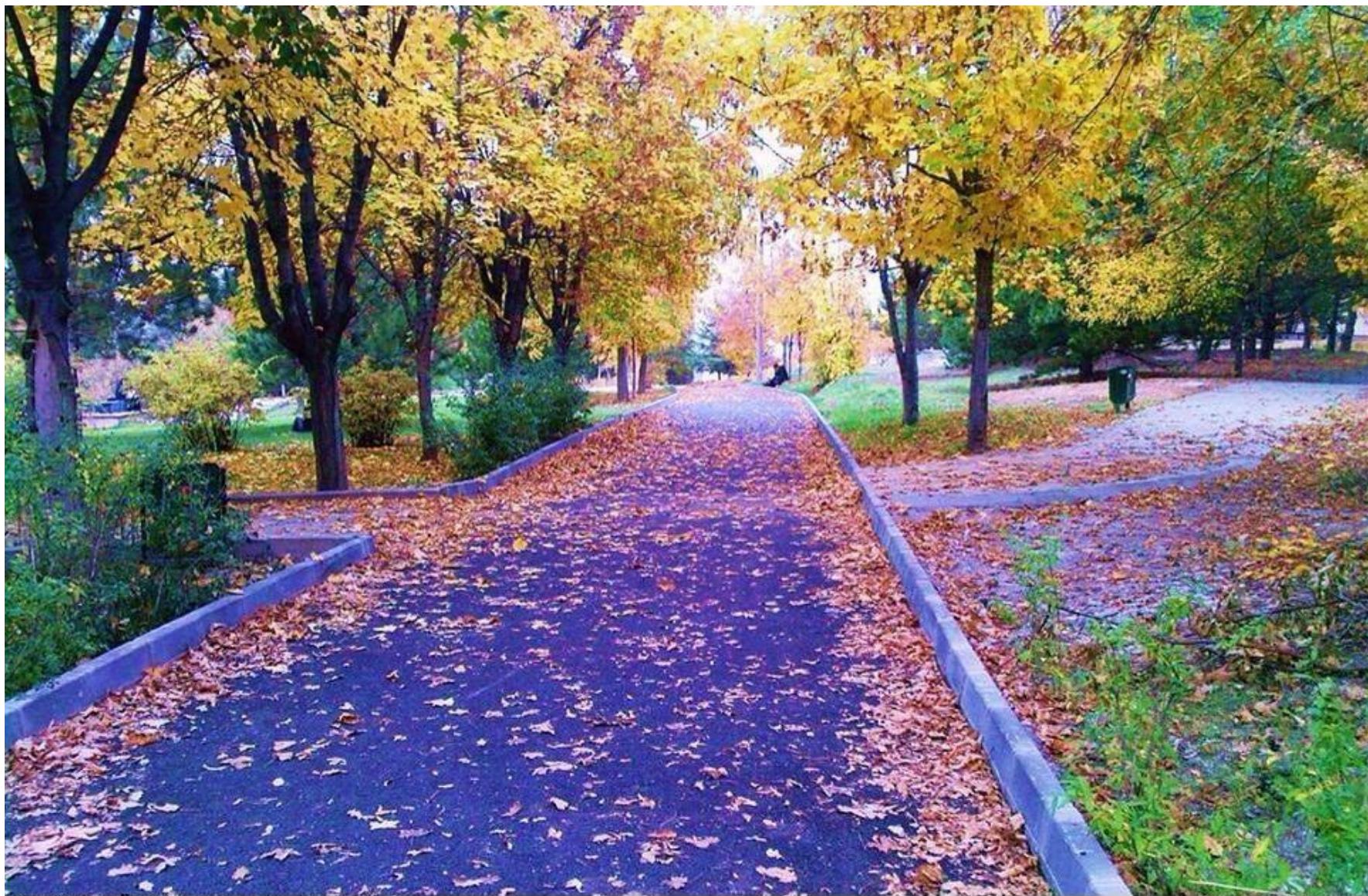
7. Gözlem
değeri :İstanbul

65. Gözlem
değeri :Kocaeli



Bir sonraki derste «İki değişkenli veriler» incelenecek.





Fotoğraf: Özlem Temuçin

sayfahacettepe



06.01.2020, A.Yigiter

KAYNAKLAR

- 1-) H.Demirhan, C.Hamurkaroğlu ,“İstatistiksel Yöntemlere Giriş”, H.Ü.Yayınları, 2011.
- 2-) Serpil Cula, Zehra Muluk, “Temel İstatistik Yöntemler”, Başkent Üniversitesi yayınları,2006.
- 3-) Levent Özbek, Esin Köksal Babacan, “İstatistiğe Giriş”, TÜBİTAK e-kitap.
- 4-) Birdal Şenoğlu, Mehmet Yılmaz, Sibel Açık Kemaloğlu, İstatistiğe Giriş, TÜBİTAK e-kitap.