

IST2084/ IST104.1/ IST104.2
Biyostatistik
5. Hafta

Doç. Dr. Fatih KIZILASLAN

<http://mimoza.marmara.edu.tr/~fatih.kizilaslan/>

Çeyreklikler (Quartiles)

- Bir verideki gözlemleri küçükten büyüğe doğru sıraladığımızda sıralanmış verileri dört eşit parçaya bölen değerlere çeyrek değerler denir.

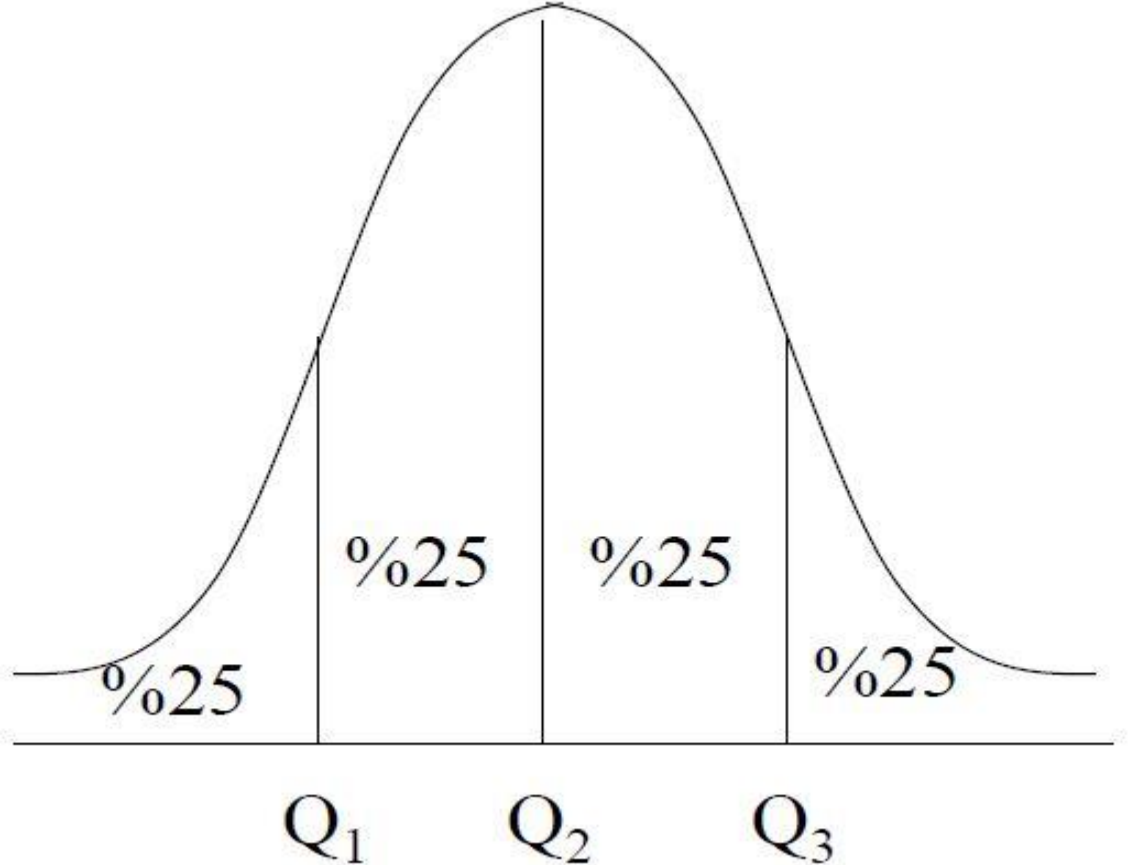
- Verinin ilk %25 lik kısmını içinde

bulunduran birinci çeyreklik Q_1 .

İkinci çeyreklik Q_2 ortanca olup veriyi ortadan ikiye bölen değerdir.

Üçüncü çeyreklik Q_3 verinin %75 lik kısmını içinde bulunduran çeyrekliktir.

Q_3 değerinin solunda verilerin %75lik kısmı bulunur.



• Çeyrekliklerin hesaplanması:

Veri küçükten büyüğe doğru sıralanır.

Birinci çeyreklik sıralanmış veride sıra numarası $\frac{(n+1)}{4}$ olan gözlem değeridir yani $Q_1 = x_{\left(\frac{n+1}{4}\right)}$.

İkinci çeyreklik sıralanmış veride sıra numarası $\frac{(n+1)}{2}$ olan gözlem değeridir yani $Q_2 = x_{\left(\frac{n+1}{2}\right)}$.

Üçüncü çeyreklik sıralanmış veride sıra numarası $\frac{3(n+1)}{4}$ olan gözlem değeridir yani $Q_3 = x_{\left(\frac{3(n+1)}{4}\right)}$.

Eğer bu sıra numaraları tam sayı değil ise çeyrekliğin değeri interpolasyon ile belirlenir (ortanca da yaptığımızı benzer olarak).

• **Örnek 1:** Küçük tansiyon verisi 65, 85, 75, 80, 90, 65, 80, 95, 85, 80 için sıralanmış veri 65, 65, 75, 80, 80, 80, 85, 85, 90, 95 olur.

Bu veride $n=10$ olduğundan medyan (ortanca) $(10+1)/2=5.5$ gözlem olduğundan $x_{Medyan} = Q_2 = \frac{x_{(5)} + x_{(6)}}{2} = \frac{80+80}{2} = 80$ dir.

Birinci çeyreklik için sıra numarası $(10+1)/4=2.75$ olduğundan 2.75 gözlemi 2. gözlemi ve 2 ile 3. gözlemlerin farkını $2.75-2=0.75$ birimlik ağırlıklandırarak

$$Q_1 = x_{(2)} + (x_{(3)} - x_{(2)})0.75 = 65 + (75 - 65)0.75 = 72.5$$

bulunur.

Üçüncü çeyreklikte benzer olarak $3(11/4)=8.25$ olduğundan

$$Q_3 = x_{(8)} + (x_{(9)} - x_{(8)})0.25 = 85 + (90 - 85)0.25 = 86.25$$

bulunur.

Soru 1: Büyük tansiyon 125, 140, 130, 136, 150, 135, 134, 155, 140, 145 verisi için çarpıklık ve basıklık katsayıları ile çeyreklikleri bulunuz.

Örnek 2: 32 tür peynirin içerdiği su değerleri (gr/100 gr) sıralı olarak aşağıda verilmiştir.

18	34	36	37	39	40	41	41
44	45	46	46	47	49	51	53
57	58	62	65	70	72	73	77
78	79	80	82	84	84	85	94

Bu veri çeyreklikleri bulalım. $n=32$ olduğundan Q_1 $(32+1)/4=8.25$ gözlem olduğundan $Q_1 = x_{(8)} + (x_{(9)} - x_{(8)})0.25 = 41 + (44 - 41)0.25 = 41.75$

n çift olduğundan $Q_2 = \frac{x_{(16)} + x_{(17)}}{2} = \frac{53 + 57}{2} = 55$ ve $3(32+1)/4=24.75$

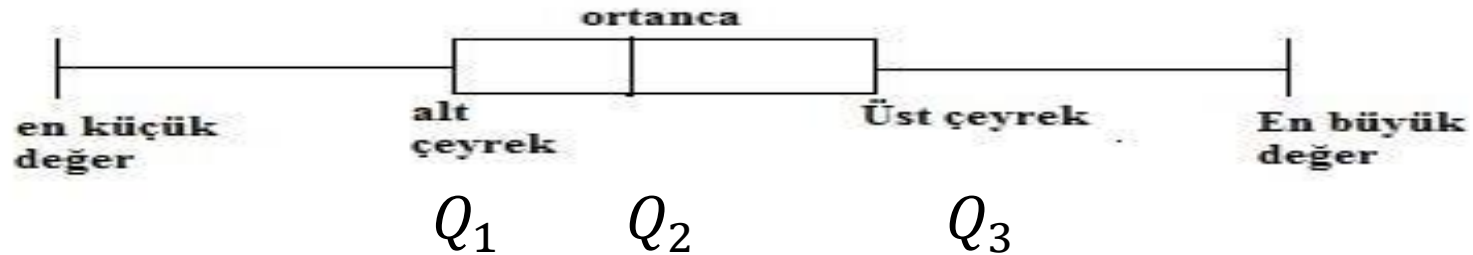
$$Q_3 = x_{(24)} + (x_{(25)} - x_{(24)})0.75 = 77 + (78 - 77)0.75 = 77.75$$

Kutu Grafiđi (Box plot)

- Kutu grafiđi eyreklikler kullanılarak verinin yapısını ortaya ıkarmak iin kullanılan basit bir grafiksel yntemdir. Grafiđin izimi:

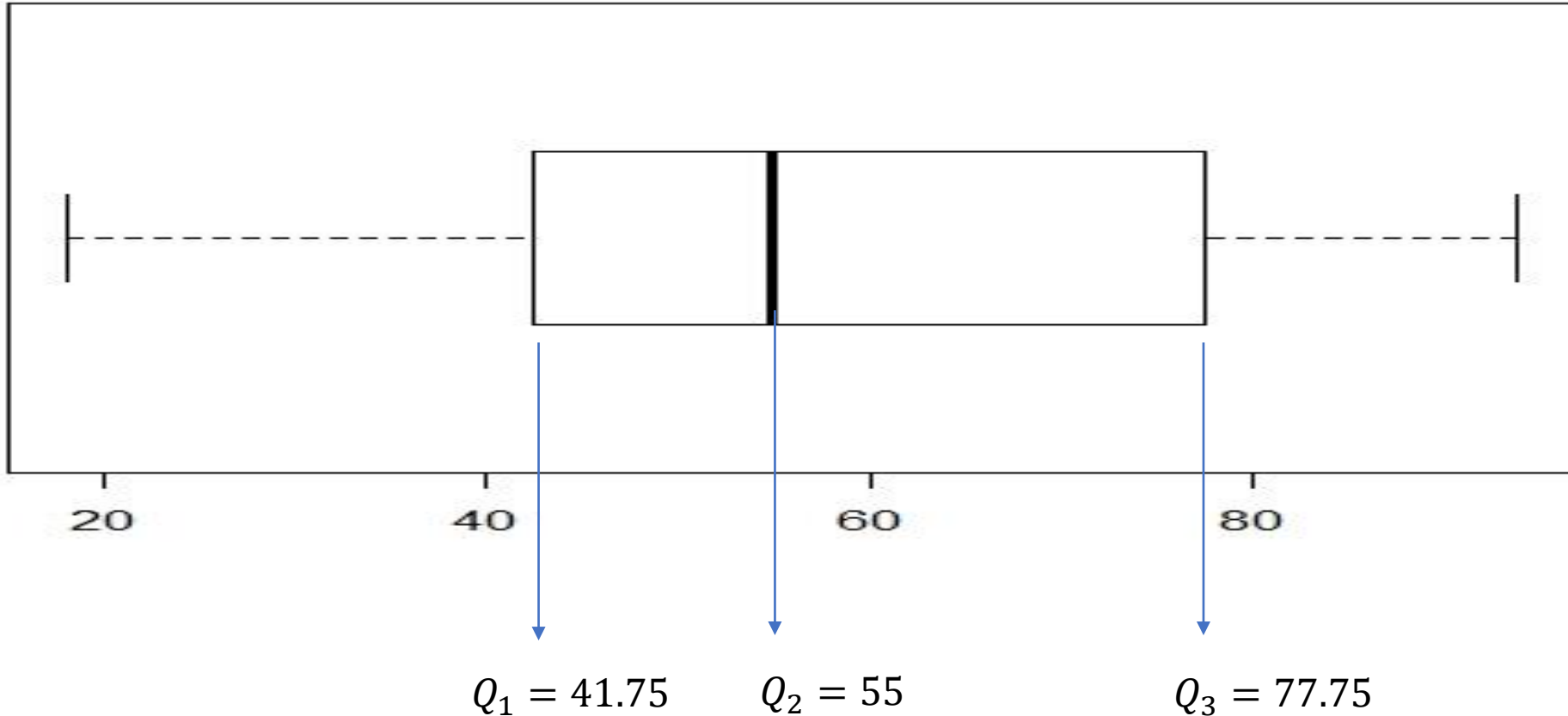
1. Verideki gzlemlerin en kk ve en byk deđerleri ile Q_1 , Q_2 ve Q_3 eyrek deđerleri sayı dođrusu zerinde iřaretlenir.
2. Q_1 ile Q_3 deđerleri arasındaki gzlemler bir kutu ile gsterilir.
3. Kutunun iki ucundan ıkarılan yatay dođrular en kk ve en byk gzlem deđerlerine kadar uzatılır.

Bylece ařađıdaki gibi bir grafik elde edilir.



- Örnek 2 de verilen su verisi için kutu grafiği aşağıdaki gibi olur. Bu veride en küçük değer=18, en büyük değer=94, $Q_1 = 41.75$, $Q_2 = 55$, $Q_3 = 77.75$ dir.

Su verisi için kutu grafiği



- **Aykırı Değer:** Veri kümesi içinde diğerlerinden çok farklı olan gözlem değerlerine aykırı değer (outliers) denir. Aykırı değerler verideki diğer gözlemlerden çok küçük ya da çok büyük olan değerler gibi düşünülebilir.
- Aykırı değerleri çeyreklikleri kullanarak belirleyebiliriz. $d = Q_3 - Q_1$ olmak üzere veri kümesi içinde $(Q_1 - 1.5d)$ değerinden **küçük** ve $(Q_3 + 1.5d)$ değerinden **büyük** gözlemler var ise bu gözlemler **aykırı değer** olarak adlandırılır.
- Örnek 2 de verilen su verisi için $d = Q_3 - Q_1 = 77.75 - 41.75 = 36$,
 $Q_1 - 1.5d = 41.75 - 54 = -12.75$ ve
 $Q_3 + 1.5d = 77.75 + 54 = 131.75$ bulunur. -12.75 den küçük ve 131.75 den büyük herhangi bir gözlem olmadığından bu veride aykırı değer yoktur.

- **Örnek 3:** Otomobillerdeki silindir sayısının benzin tüketimi üzerindeki etkisi olup olmadığı araştırılmak isteniyor. Bu nedenle farklı silindir sayılarına sahip araçların 1lt benzin ile aldıkları yol km olarak aşağıdaki gibi elde edilmiştir.

Silindir sayısı	Alınan yol (km)
8	18.7, 14.3, 16.4, 17.3, 15.2, 10.4, 21.4, 10.4, 21.5, 15.5, 15.2, 13.3, 30.4, 19.7
6	21, 21, 21.4, 18.1, 19.2, 17.8, 15.8
4	22.8, 24.4, 22.8, 14.7, 32.4, 30.4, 33.9, 19.2, 27.3, 26, 15

8 silindir için kutu grafiğini oluşturalım.

• **Çözüm:** 8 silindir için $n=14$, en küçük gözlem= 10.4 , en büyük gözlem= 30.4

$$Q_1 = x_{(3)} + (x_{(4)} - x_{(3)})0.75 = 13.3 + (14.3 - 13.3)0.75 = 14.05$$

$$Q_2 = \frac{x_{(7)} + x_{(8)}}{2} = \frac{15.5 + 16.4}{2} = 15.95$$

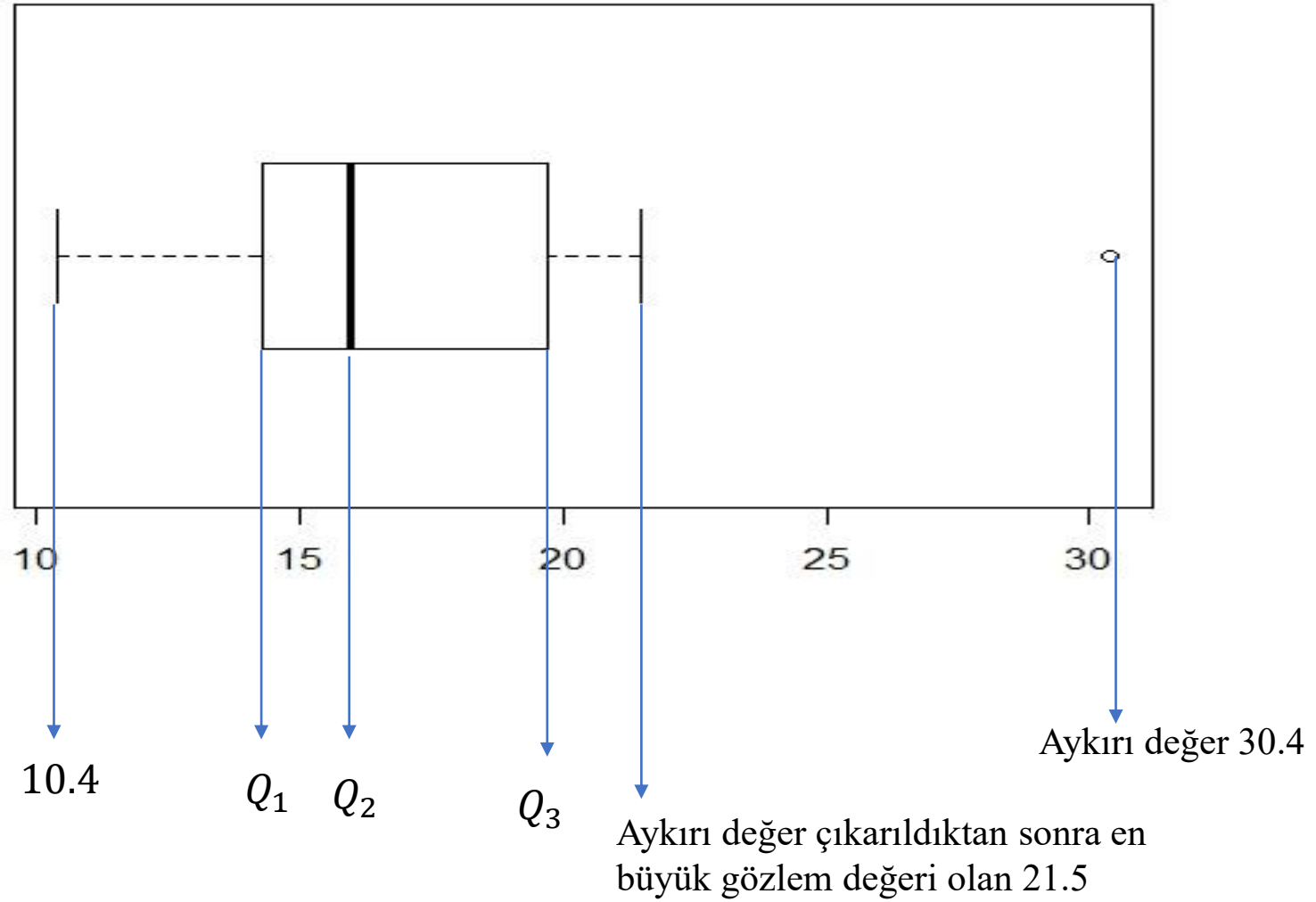
$$Q_3 = x_{(11)} + (x_{(12)} - x_{(11)})0.25 = 19.7 + (21.4 - 19.7)0.25 = 20.125$$

$$d = Q_3 - Q_1 = 20.125 - 14.05 = 6.075,$$

$$Q_1 - 1.5d = 14.05 - (1.5 * 6.075) = 4.9375 \text{ ve}$$

$Q_3 + 1.5d = 20.125 + (1.5 * 6.075) = 29.2375$ bulunur. Böylece, 30.4 gözlem değeri 29.2375 den büyük olduğu için 30.4 aykırı değerdir. Bu durumda kutu grafiği aşağıdaki gibi olur.

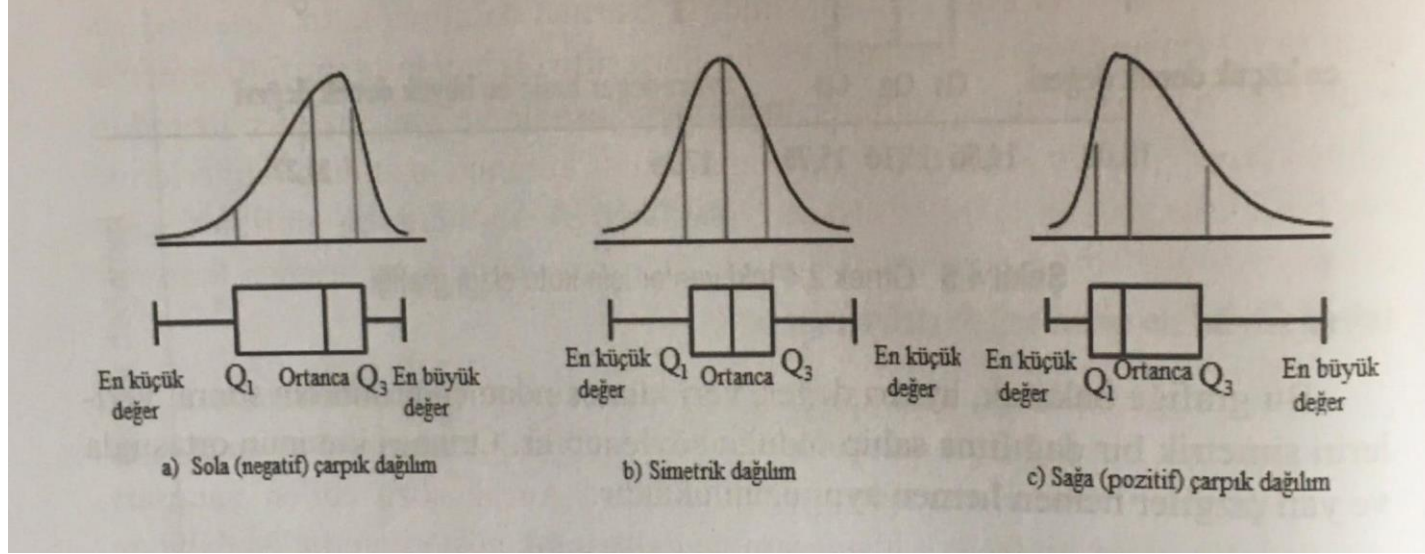
8 Silindir için katedilen yol



Sorular: Örnek 3 de verilen veriler için

1. Her bir silindir grubu için alınan yol değerlerinin ortalaması ve standart sapmasını bulunuz.
2. Her bir silindir grubu için değişim katsayılarını bularak bu verileri homojenlik bakımından karşılaştırınız.
3. 4 ve 6 silindir grupları için çeyreklikleri bularak kutu grafiklerini çiziniz.

- Kutu grafiği çarpıklık ve basıklık hakkında da bilgiler vermektedir.
- Medyan çizgisi Q_1 değerine yaklaştığında dağılım sağa çarpık, Q_3 değerine yaklaştığında ise dağılım sola çarpıktır, ortada olur ise simetriktir.



- Kutunun genişliği çizginin genişliğine yaklaştığında dağılım basık aksi durumda ise dağılımın sivri olduğu gibi bir yorum yapılabilir.

