İstatistik 1
BÖLÜM 4

Sayısal Verilerin Özetlenmesi: Gruplandırılmış Veriler

Ortalama-Grup Data

- En yaygın kullanılan merkezi eğilim ölçüsüdür.
- Veri setinde aşırı uçlar varsa bu ölçü etkilenir.

Populasyon: $$\mu = \frac{\sum_{i=1}^{K} f_i m_i}{N}$$

Örnek: $$\bar{x} = \frac{\sum_{i=1}^{K} f_i m_i}{n}$$

Medyan-Grup Data

$$\text{Medyan} = L + \frac{n/2 - F}{f_m} c$$

L: medyan sınıfında yer alan en düşük değer

n: gözlem sayası

F: medyan sınıfına kadar olan sıklıkların toplamı

f_m: medyan sınıfına ait sıklık

c: sınıf aralığının genişliği

Mod-Grup Data

$$Mod = L + \frac{d_1}{d_1 + d_2} c$$

L: Mod sınıfının en düşük değeri

d_1: Mod sınıfı ile bir önceki sınıf arasındaki sınıf sıklıkğı farkı

d_2: Mod sınıfı ile bir sonraki sınıf arasındaki sınıf sıklığı farkı

c: sınıf aralığı genişliği

Varyans-Grup Data

Örnek için varyans formülü:

$$S^2 = \frac{\sum_{i=1}^{K} f_i (m_i - \bar{x})^2}{n-1} \text{ veya } \frac{\sum_{i=1}^{K} f_i m_i^2 - n\bar{x}^2}{n-1}$$

Populasyon için varyans formülü:

$$\sigma^2 = \frac{\sum_{i=1}^{K} f_i (m_i - \mu)^2}{N} \text{ veya } \frac{\sum_{i=1}^{K} f_i m_i^2}{N} - \mu^2$$

Ortalama Mutlak Sapma (OMS) Grup Data

Populasyon için OMS $$\frac{\sum_{i=1}^{K} f_i \left| m_i - \mu \right|}{N}$$

Örnek için OMS

$$OMS = \frac{\sum_{i=1}^{n} f_i \left| m_i - \bar{x} \right|}{n} = \frac{500.71}{51} = 9.81$$

Populasyon Standart Sapmasının açıklanması

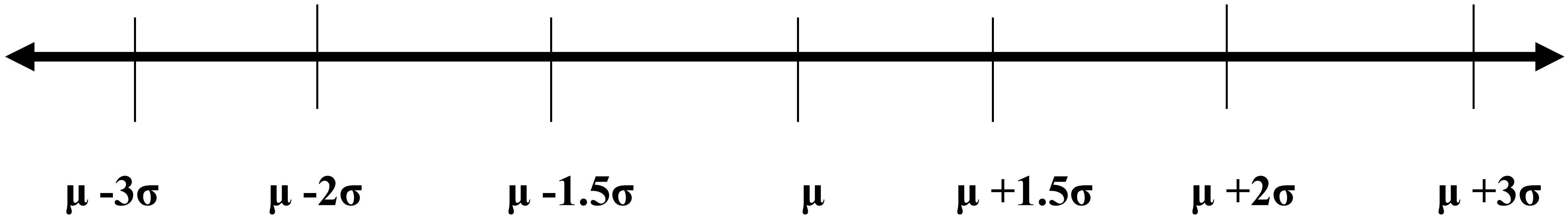
Chebychev (çebişev biçiminde telaffuz edilir) Kuralı
Ortalaması μ ve standart sapması σ olan bir

populasyon üyelerinin en az % 100 (1 -$(\frac{1}{M^2})$' lik

bölümü, M > 1 olması koşulu ile ortalamanın M standart sapması etrafında toplanır.

Populasyon Standart Sapmasının açıklanması

Pratikte Chebychev kuralının nasıl işlemekte olduğunu şu şekilde açıklayabiliriz:



Populasyon Standart Sapmasının açıklanması

Ampirik Kuralı

Ampirik kuralı, Başparmak Kuralı yada Gaus Kuralı olarak da bilinmektedir. Ampirik Kuralı'na göre gözlem değerleri normal dağılma sahip veya çan eğrisi biçiminde ise bu değerlerin yaklaşık olarak

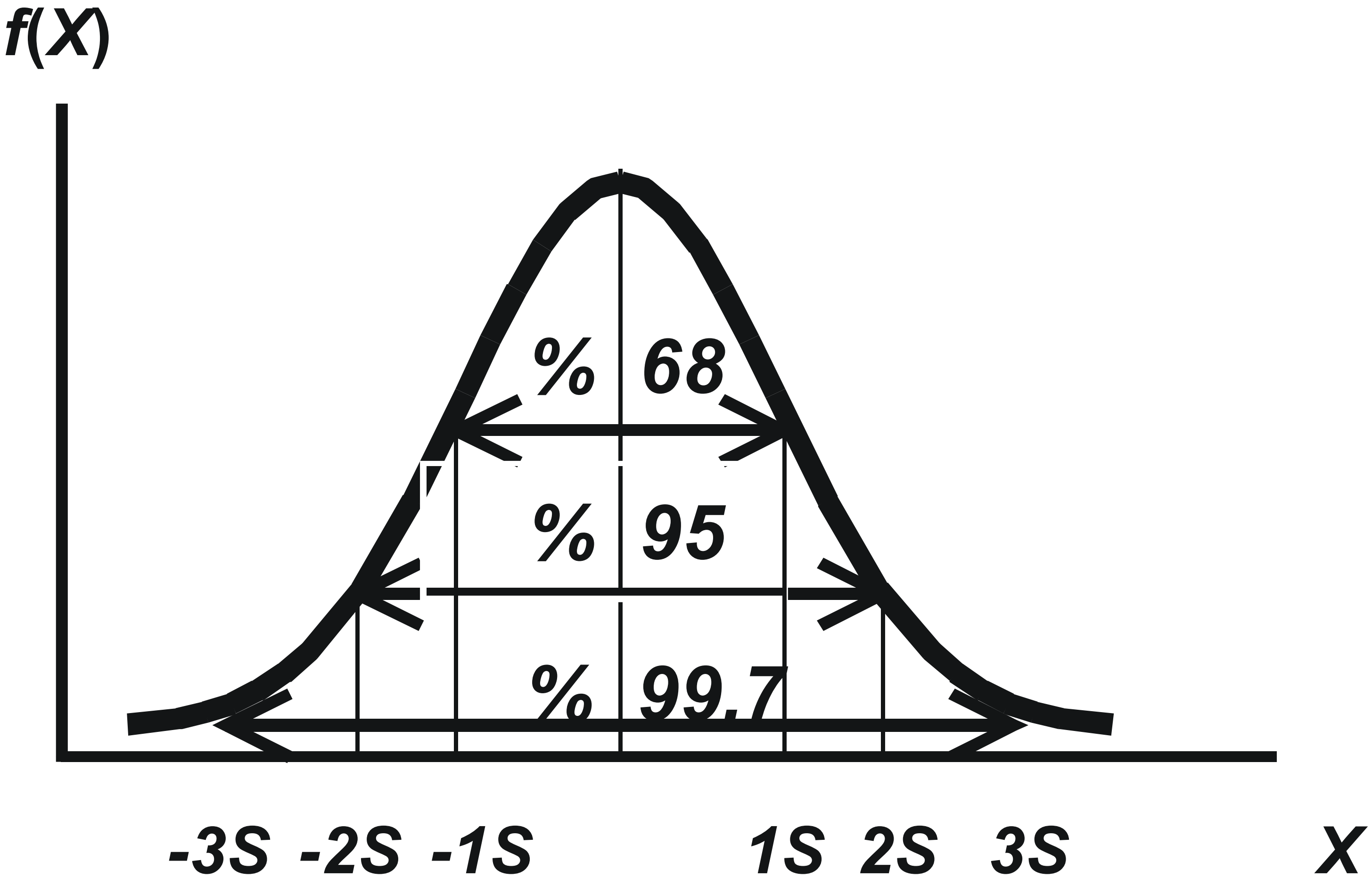
Populasyon Standart Sapmasının açıklanması

% 68’i($\bar{x}$ -S $\bar{x}$ +S) aralığında

% 95’i $\bar{x}$ (-2S $\bar{x}$ +2S) aralığında

%100’ü $\bar{x}$(-3S $\bar{x}$ +3S) aralığında

Populasyon Standart Sapmasının açıklanması



13	31	22	5	3	6
2	19	17	23	22	17
31	13	31	2	1	5
8	9	7	10	17	6
19	3	19	17	22	23
22	31	13	2	5	22
5	22	13	9	13	13
22	25	13	13	3	2
1	2	17			

Sınıf Sıklığı

Sınıf Aralığı			
1 – 5	m_i	14 f_i	f_i m_i
6 – 10	m_i 3	7	42
11 – 15	8	8	56
16 – 20	13	8	104
21 – 25	18	10	144
26 – 31	23	4	230
	28.5		114

$$\bar{x} = \frac{\sum_{i=1}^{K} f_i m_i}{n} = \frac{690}{51} = 13.53$$

Sınıf Aralığı	m_i	Sınıf Sıklığı f_i	$f_i m_i^2$
1 – 5	3	14	126
6 – 10	8	7	448
11 – 15	13	8	1352
16 – 20	18	8	2592
21 – 25	23	10	5290
26 – 31	28.5	4	3249

Varyans:

$$S^2 = \frac{\sum_{i=1}^{K} f_i m_i^2 - n\bar{x}^2}{n-1} = \frac{13{,}057 - (51)*(13.53)}{50} = 74.41$$

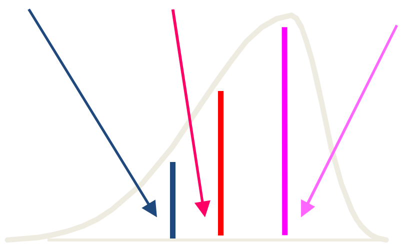
Standart sapma:

$$S = \sqrt{74.41} = 8.62$$

Dağılımların Şekilleri

- Datanın dağılımını tanımlar
- Şekil Ölçüleri
  - Simetrik yada Çarpık

Sola-Çarpık



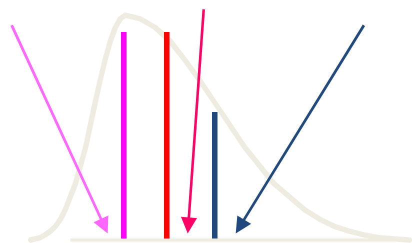
Ortalama < Medyan < Mod

Simetrik

Ortalama = Medyan =Mod

Sağa-Çarpık

Mod < Medyan < Ortalama



Çarpıklık (Skewness):Sınıflandırılmamış data

Çarpıklık (Skewness) Pearson Skewness katsayısı formülü ile hesaplanır

Populasyon için Pearson Skewness katsayısı formülü: Sk = $$\frac{3(\mu - medyan)}{\sigma}$$

Örnek için Pearson Skewness katsayısı formülü:

$$Sk = \frac{3(\overline{x} - medyan)}{S}$$

Çarpıklık (Skewness):Sınıflandırılmış data

Populasyon için Pearson Skewness katsayısı formülü:

$$\text{Sk} = \frac{\sum_{i=1}^{N} f_i (x-\mu)^3}{\sigma^3}$$

Örnek için Pearson Skewness katsayısı formülü:

$$\text{Sk} = \frac{\sum_{i=1}^{n} f_i (x-\bar{x})^3}{S^3}$$

Tepeleme Katsayısı (Kurtosis): Gruplandırılmamış Data

Populasyon için Tepeleme =

$$\frac{(x-\mu)^4}{\sigma^4}$$

Örnek için Tepeleme = $$\frac{(x-\bar{x})^4}{S^4}$$

Tepeleme Katsayısı (Kurtosis): Gruplandırılmış Data

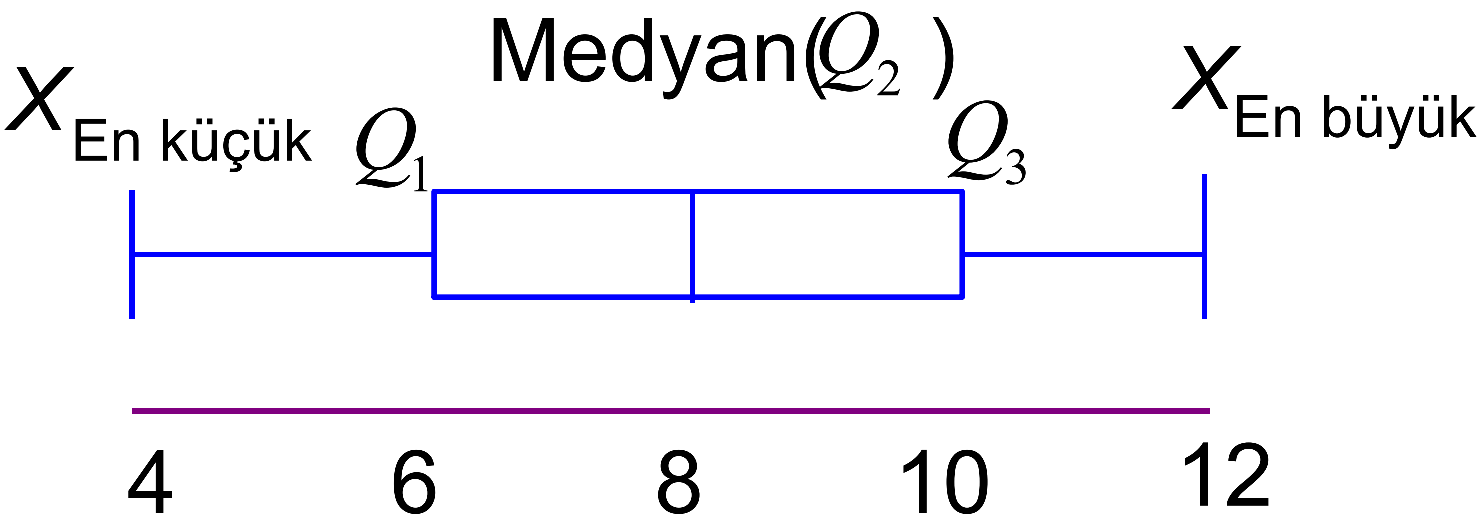
Populasyon için Tepeleme =

$$\frac{\sum_{i=1}^{N} f_i (x-\mu)^4}{\sigma^4}$$

Örnek için Tepeleme = $$\frac{\sum_{i=1}^{n} f_i (x-\bar{x})^4}{S^4}$$

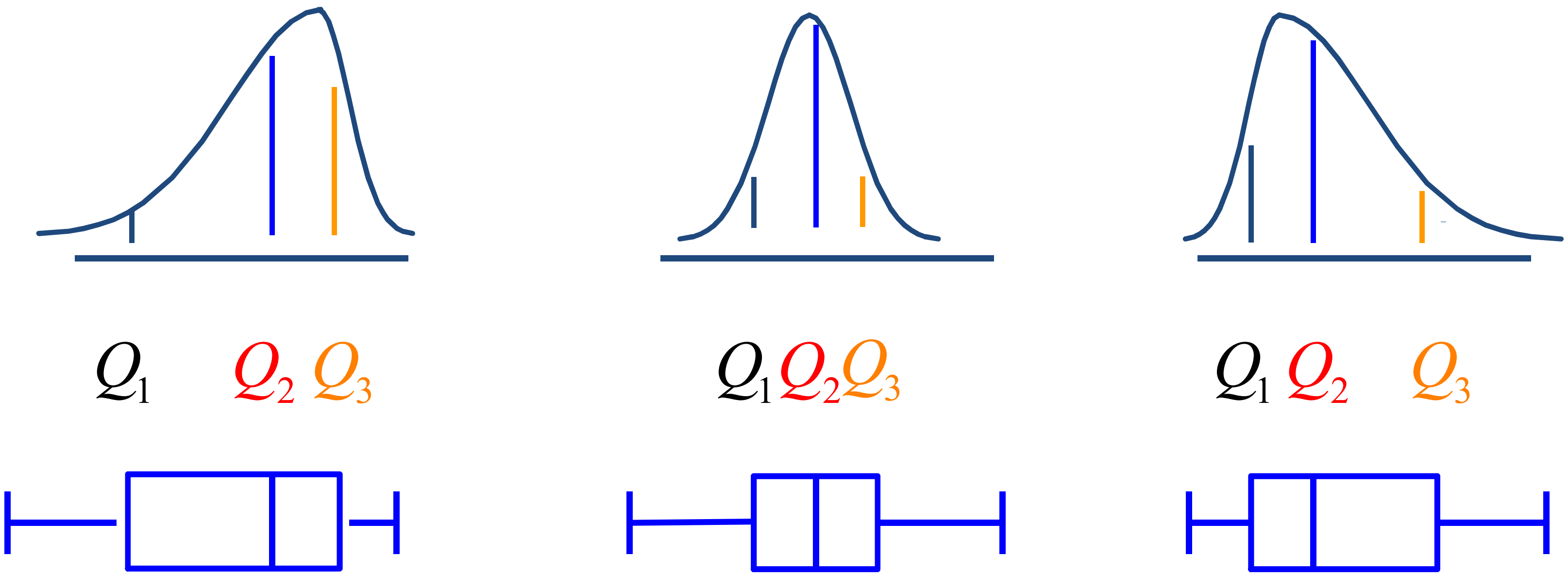
Box ve Whisker Grafiği

- Box-ve-Whisker grafiği
- Datanın grafiksel özetlenmesi: 5-adet sayısal özetleme



Dağılım Şekilleri ve Box-and-Whisker Grafikleri

Sola-Çarpık **Simetrik** **Sağa-Çarpık**



Korelasyon Katsayısı

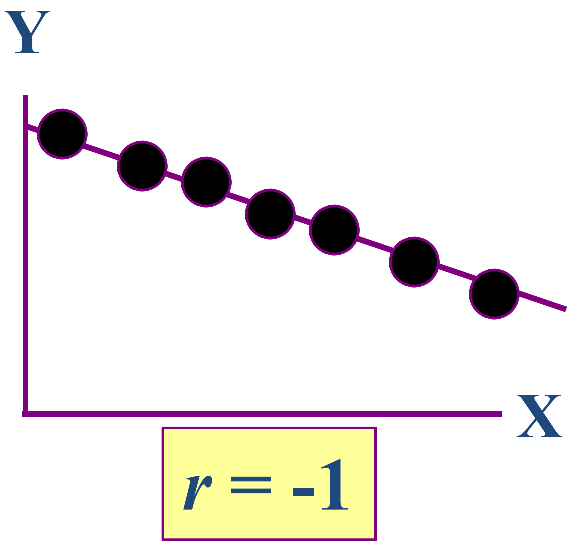
- İki adet sayılabilen değişken arasındaki ilişkiyi ölçümlemede kullanılır ve r ile gösterilir.

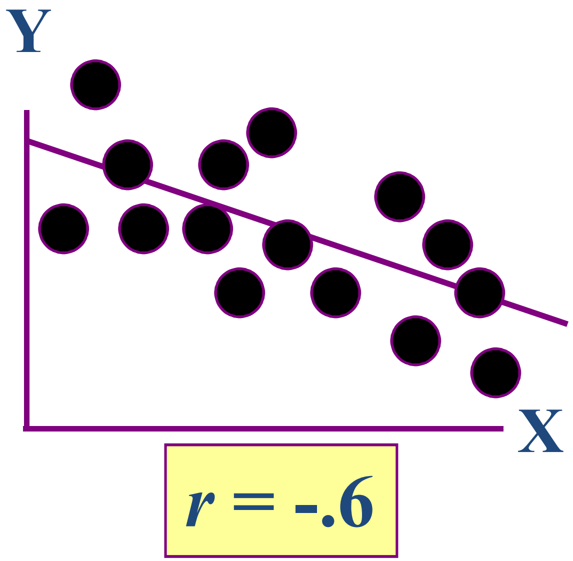
$$r = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}}$$

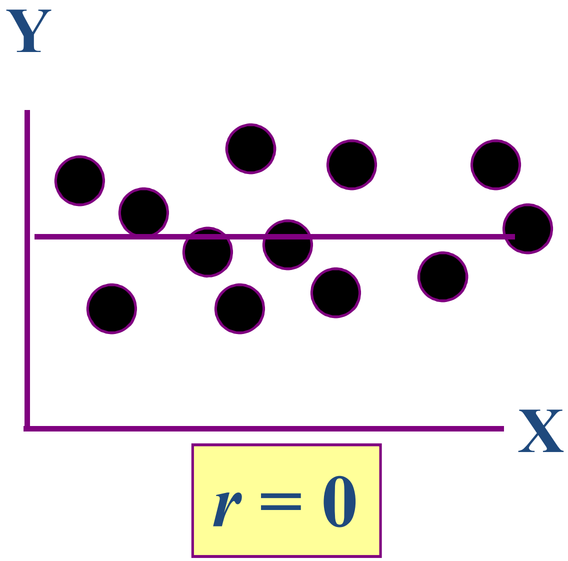
Korelasyon Katsayısının Özellikleri

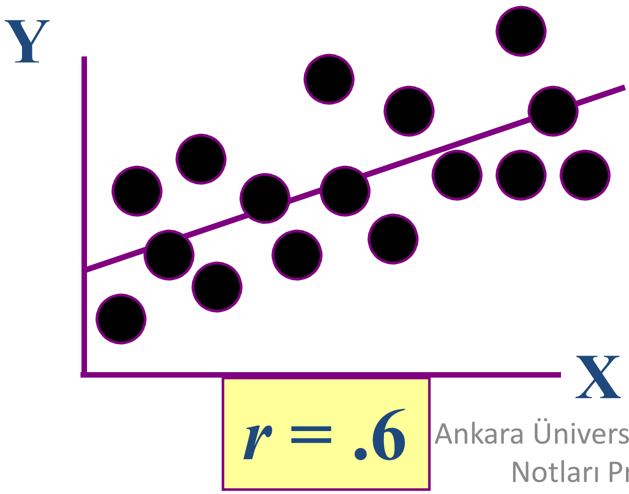
- -1 ile 1 arasında bir değer alır
- -1 mükemmel negatif, 1 ise mükemmel pozitif ilişkiyi gösterir.
- 0 yakın olduğunda ise ilişki çok zayıftır.

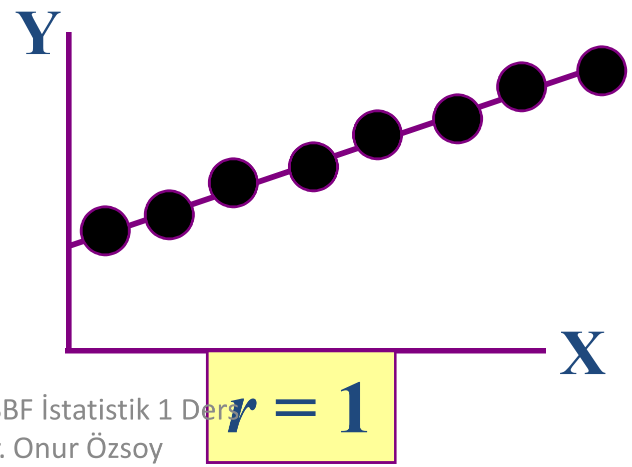
Farklı Korelasyonlar











Sayısal Özetleme Yöntemlerinde Dikkat Edilmesi Gereken Önemli Noktalar

- Data analizi objektif olmalı
  - Data seti ile ilgili bilgileri objektif bir şekilde yansıtacak türden özetleme yöntemleri kullanılmalı
- Yapılacak yorumların, adil tarafsız ve anlaşılır olması gerekli

Etiksel Noktalar

Sayısal özetleme yöntemleri ile sonuçların:

- Tamamı rapor edilmeli: İyi yada kötü
- Adil, tarafsız anlaşılır bir şekilde rapor edilmesi gerekir
- Gerçekleri çarpıtıcı nitelikte yöntemlerden sakınılmalı