- Yaygınlık ölçüleri, bir dağılımdaki değerlerin farklılıklarının ölçüsünü tanımlar.
- Şöyle ki, bir dağılımdaki değerlerin ortalama değere olan uzaklıkları farklılıklar gösterir. Bu farklılıkların derecesi dağılımın yaygınlığı kavramını oluşturur.
- İki dağılım aynı ortalama, ortanca ya da tepe değerine sahipken yaygınlıkları farklı olabilir.

- Şimdi aşağıda verilen 5'er gözlemli iki dağılımı inceleyelim. Bu iki dağılımın aritmetik ortalama, ortanca ve tepe değerleri aynı olup 6 'ya eşittir.

Tablo 2.16. Ortalama Ölçüleri Aynı Olan İki Dağılım

Dağılım I		Dağılım II	
2		6	
6	$\overline{X}$ =6	5	$\overline{X}$ =6
1	Ortanca=6	6	Ortanca=6
15	Tepe D.=6	7	Tepe D.=6
6		6	

- Ancak dağılım I'deki değerlerin aritmetik ortalamaya uzaklığı, dağılım II'ye göre daha yaygındır. Bir dağılımdaki değerler aritmetik ortalamadan uzaklaştıkça dağılımın yaygınlığı artar.
- Dolayısıyla, bir dağılımı tanımlamakta sadece tablo ve grafiklerden ya da ortalama ölçülerinden yararlanmak, incelenen dağılım hakkında yeterli bilgiyi sağlamamaktadır.

- Dağılımların yaygınlığı hakkında bilgi veren ve en çok kullanılan ölçü;
- **Standart sapma (standart deviation)** ve onun karesi olan **varyans (variance)** 'tır.

- **Değişim katsayısı**na da sıklıkla başvurulur (coefficient of variation).
- **Dağılım aralığı** (Range),
- **Çeyreklikler arası dağılım aralığı** (interquartiale range-IQR)
- **Çeyrek sapma**(semi-interquartiale range) bu amaçla yaygın kullanılan diğer ölçülerdir.

- Dağılım aralığı en basit yaygınlık ölçüsüdür.
- **Dağılımdaki en büyük değerden en küçük değerin çıkartılması** ile bulunur. Dağılım aralığının göreceli olarak daha küçük olduğu dağılımlarda, değişkenliğin (yaygınlığının) daha az olduğu düşünülür.

Tablo 2.16. Ortalama Ölçüleri Aynı Olan İki Dağılım

Dağılım I		*Dağılım II*	
2		6	
6	$\bar{X}=6$	5	$\bar{X}=6$
1	Ortanca=6	6	Ortanca=6
15	Tepe D.=6	7	Tepe D.=6
6		6	

- Daha yaygın olan I. Dağılımın dağılım aralığı R=15-1=14 iken, II. dağılım için bu değer R=7-5=2'dir.

- Ancak, dağılım aralığı dağılımdaki diğer değerlerden oldukça farklı değer(ler) olan aşırı değer(ler)den etkilenmesi ve sadece iki gözleme ilişkin değeri dikkate alması nedeniyle **kaba bir değişkenlik ölçüsüdür.**
- Aşağıdaki iki dağılımı inceleyelim ;

Dağılım A : 0 0 2 3 3 4 5 6 8 9 10 10 10

Dağılım B : 0 4 5 5 5 5 5 6 6 6 6 6 10

- *A* Dağılımı daha yaygın olmakla birlikte *B* Dağılımı ile aynı dağılım aralığına (genişliğe) sahiptir.
- *B* Dağılımında, en büyük ve en küçük değer çıkartıldığında, dağılım aralığı 10'dan 2'ye düşmektedir.
- Ayrıca, gözlemlerin çoğunun en büyük ya da en küçük değere yakın olduğu durumlarda da gerçek değişkenlik hakkında bilgi vermez.
- Dağılım aralığı daha ileri düzeydeki istatistiksel hesaplamalarda da pek kullanılmaz.

- **Verilerin kesikli ya da sürekli sayısal veri tipi**nde olduğu durumlarda eğer dağılımlar çarpıksa (dolayısıyla, ortalama ölçüsü olarak ortanca kullanılıyorsa) ya da **veri sıralı bir veri** (1.,2., .. ,10., ..gibi) ise **yaygınlık ölçüsü olarak çeyrekler arası dağılım aralığından sıklıkla yararlanılır**.
- Özellikle, araştırıcıların **uçtaki değerlerden çok ortadaki değerlerle ilgilendiği durumlarda** kullanılır.
- Çeyrekler arası dağılım aralığı, 75. yüzdelik değerinden 25. yüzdelik değerinin çıkartılması ile bulunur.
- Dağılım aralığı aşırı uç değerlerden etkilendiği halde *çeyrekler* arası dağılım aralığı aşırı uç değerlerden etkilenmez.

- Örneğin
  - 5 7 9 12 26 33 34 35 38 42 70
- Q1 Medyan Q3

- Çeyrekler Arası Dağılım: Q3-Q1 = 38-9 = 29
- Yüzücülerin *MaxV02* değerlerinin 75. yüzdeliği 62 *ml/kgldk* ve 25. yüzdeliği 53 *ml/kg/dk* olarak bulunur. Buradan, bu dağılımın çeyrekler arası dağılım aralığı 62-53=9 *ml/kg/dk* olarak bulunur.

- 25. ve 75. yüzdelikler arasındaki mesafenin yarısı; bu yüzdeliklerle ortanca arasındaki mesafenin ortalama bir ölçüsü olup bir yaygınlık ölçüsü olarak kullanılabilir.
- 25. ve 75. yüzdelikler arasındaki mesafenin yarısına (ya da çeyreklikler arası genişliğinin yarısına) ***çeyrek sapma*** ya da **yarı *çeyreklikler* arası dağılım aralığı *(semi-interquartile range)*** adı verilir.
- Yüzücülerin *MaxV02* değerlerinin 75. yüzdeliği 62 *ml/kgldk* ve 25. yüzdeliği 53 *ml/kg/dk* olarak bulunur. Buradan, bu dağılımın çeyrekler arası dağılım aralığı 62-53=9 *ml/kg/dk* olarak bulunmuştu.
- çeyrek sapma ise 9/2=4.5 olarak bulunur.

- Dağılımdaki değerlerin sadece yüzde ellisi dikkate alınarak hesaplandığından dağılımın tümünün gösterdiği değişim hakkında tam bir bilgi vermez.

- Özet olarak, çeyrek sapma, **"ortalama ölçüsü olarak ortancanın kullanıldığı durumlarda"** sıklıkla kullanılan bir yaygınlık ölçüsüdür. Özellikle de, aşırı değerlerin **"dağılımın sadece bir tarafında olduğu"** durumlarda kullanılması gerekir.
- Yukarıda belirtilen üç yaygınlık ölçüsünün hesaplanışında verilerin tümü kullanılmadığından, bu ölçüler daha ileri düzeydeki istatistiksel hesaplamalarda pek kullanılmazlar.
- **Bu nedenle, verilerin tümünü kullanan ve istatistiksel açıdan daha güçlü yaygınlık ölçülerine gereksinim vardır.**

- Bir dağılımdaki yaygınlığı belirlemek için bir diğer yaklaşım, dağılımdaki değerlerin aritmetik ortalamaya olan uzaklıklarının (sapmalarının) ortalamasının bulunmasıdır.
- Standart sapma dağılımdaki tüm verileri dikkate alır ve **“dağılımdaki tüm değerlerin aritmetik ortalamaya olan uzaklıklarının ortalama bir göstergesi”** olarak tanımlanır.
- Standart sapma **büyüdükçe dağılımın yaygınlığı artar**. Dağılımdaki değerlerin birbirine eşit olduğu durumlarda (örneğin, 7, 7, 7, 7, ..) yaygınlık yoktur ve standart sapma sıfıra eşittir.

- Örneklem değerleri için standart sapma genellikle *s* ya da S ile gösterilir. **Standart sapmanın birimi incelenen değişkenin birimi ile aynıdır.**
- **Dağılımdaki tüm değerleri dikkate almakla birlikte, standart sapmanın çarpık dağılımlarda (ortalama ölçüsü olarak aritmetik ortalamanın kullanılmadığı) bir yaygınlık ölçüsü olarak kullanılmaması önerilmektedir.**
- Standart sapma, sınıflandırılmış ve sınıflandırılmamış veriler için ayrı formüllerle hesaplanır.

- Sınıflandırılmamış verilerde standart sapma formülü aşağıda verilmiştir.

$$S = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}}$$

- Örnek olarak Tablo 2.16 'da verilen ve ortalama ölçüleri aynı olan verilerin standart sapmalarını hesaplayalım.

Tablo 2.16. Ortalama Ölçüleri Aynı Olan İki Dağılım

Dağılım I		*Dağılım II*	
2		6	
6	$\overline{X}=6$	5	$\overline{X}=6$
1	Ortanca=6	6	Ortanca=6
15	Tepe D.=6	7	Tepe D.=6
6		6	

$$S=\sqrt{\frac{\sum_{i=1}^{n} x_i^2-\frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}}$$

- Örnek olarak Tablo 2.16 'da verilen ve ortalama ölçüleri aynı olan verilerin standart sapmalarını hesaplayalım.

I. Dağılım için :

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{5} x_i^2 = 2^2 + 6^2 + 1^2 + 15^2 + 6^2 = 302$$

$$\left(\sum_{i=1}^{n} x_i\right)^2 = \left(\sum_{i=1}^{5} x_i\right)^2 = (2+6+1+15+6)^2 = (30)^2 = 900$$

olarak elde edilir ve (2.8) yardımıyla standart sapma,

$$S = \sqrt{\frac{302 - \frac{(30)^2}{5}}{5-1}} = 5.5227$$

- Örnek olarak Tablo 2.16 'da verilen ve ortalama ölçüleri aynı olan verilerin standart sapmalarını hesaplayalım.
- Buna göre; bu dağılımdaki değerler aritmetik ortalama etrafında ortalama ± 5.5227 birimlik değişkenliğe sahiptir.
- Birbirine yakın değerlere sahip olan Dağılım II için ise standart sapma benzer işlemlerle S=0.7071 birim olarak bulunur.
- Sonuçlardan anlaşılacağı gibi, ikinci dağılımın yaygınlığı birinciye göre oldukça düşüktür.

- Sınıflandırılmış verilerde standart sapmayı bulabilmek amacıyla geliştirilmiş çeşitli formüller vardır. Bunlardan biri aşağıda verilmiştir (2.9)

$$S = \sqrt{\frac{\sum_{i=1}^{k} f_i s_i^2 - \frac{\left(\sum_{i=1}^{k} f_i s_i\right)^2}{n}}{n-1}}$$

- Burada; *k:* Sınıf sayısı *f; : i.* Sınıfın frekansı, *si : i.* Sınıfın sınıf değeri' dir.

- Yüzücülerin *Max V02* ölçümlerine ilişkin standart sapmayı sınıflandırılmış veriler yardımıyla bulabilmek için gerekli hazırlıklar Tablo 2.17' da verilmiştir.

MaxVO₂	f_i	s_i	$f_i\ s_i$
40-44	2	42	84
45-49	3	47	141
50-54	8	52	416
55-59	13	57	741
60-64	8	62	496
65-69	4	67	268
70-74	2	72	144
Toplam	**40**		**2290**

$$\bar{x} = \frac{2290}{40} = 57.25\ ml/kg/dk$$

$$S = \sqrt{\frac{\sum_{i=1}^{k} f_i s_i^2 - \frac{\left(\sum_{i=1}^{k} f_i s_i\right)^2}{n}}{n-1}}$$

- Yüzücülerin *Max V02* ölçümlerine ilişkin standart sapmayı sınıflandırılmış veriler yardımıyla bulabilmek için gerekli hazırlıklar Tablo 2.17' da verilmiştir.

$MaxVO_2$	f_i	s_i	$f_i\ s_i$
40-44	2	42	84
45-49	3	47	141
50-54	8	52	416
55-59	13	57	741
60-64	8	62	496
65-69	4	67	268
70-74	2	72	144
Toplam	**40**		**2290**

$$\bar{x} = \frac{2290}{40} = 57.25\ ml/kg/dk$$

$$S = \sqrt{\frac{133100 - \frac{(2290)^2}{40}}{39}} = 7.15\ ml/kg/dk$$

Bazen sınıflandırılmamış ve sınıflandırılmış verilerden elde edilen standart sapmalar az da olsa farklı çıkabilir. Bu fark verileri sınıflandırmaktan kaynaklanmaktadır.

- *Birden* fazla gruba ilişkin standart sapmalar elimizde var ise, toplam denek sayısı için standart sapma ham verilere gerek kalmadan hesaplanabilir. Bu amaçla, ağırlıklı standart sapmadan yararlanılır.

$$\text{Ağırlıklı } S = \sqrt{\frac{\sum_{i=1}^{k}\left[(n_i - 1)S_i^2 + (n_i \times \bar{x}_i^2)\right] - N \times \bar{\bar{x}}^2}{N-1}} \qquad (2.10)$$

ile verilir. Burada, k: *Grup sayısı,* S_i: *i. grubun standart sapması,* $\bar{x}_i$: *i. grubun ortalaması,* $\bar{\bar{x}}$: *Ağırlıklı Ortalama,* N : *Toplam denek sayısı*'dır.

Örneğin Tablo 2.16'da verdiğimiz iki farklı dağılıma (iki farklı gruba) ilişkin ortalamalar; $\bar{x}_1 = 6$ ve $\bar{x}_2 = 6$, standart sapmalar; $S_1 = 5.5227$ ve $S_2 = 0.7071$ olarak bulunmuştu. Grupların denek sayıları eşit olduğu için *N=10* için ortalama (ağırlıklı ortalama) $\bar{\bar{x}} = 6$ olacaktır. Buna göre (2.10)'dan ağırlıklı standart sapma (*N=10* için standart sapma);

$$S = \sqrt{\frac{\left[(5-1)\,5.5227^2 + 5\times 6^2\right] + \left[(5-1)\,0.7071^2 + 5\times 6^2\right] - 10\times 6^2}{10-1}} = 3.7118$$

Tablo 2.16. Ortalama Ölçüleri Aynı Olan İki Dağılım

Dağılım I		*Dağılım II*	
2		6	
6	$\bar{x}$ =6	5	$\bar{x}$ =6
1	Ortanca=6	6	Ortanca=6
15	Tepe D.=6	7	Tepe D.=6
6		6	

$$\text{Ağırlıklı } S = \sqrt{\frac{\sum_{i=1}^{k}\left[(n_i-1)S_i^2 + (n_i \times \bar{x}_i^2)\right] - N \times \bar{\bar{x}}^2}{N-1}} \qquad (2.1($$

- Standart sapma bir dağılımın yaygınlığını gösteren ölçülerden birisidir. Ancak standart sapmanın büyüklüğüne bakarak bir dağılımın yaygınlığı konusunda yargıya varmak güçtür.

- Örneğin, bir dağılımın standart sapması 6.3 ise bu değerin büyük, küçük ya da normal bir sapma olduğu konusunda bir yargıya varılamaz.

- Diğer taraftan, iki ya da daha fazla dağılımın yaygınlığını karşılaştırmak istediğimizde de standart sapmayı doğrudan kullanamayız.

- Bu nedenlerle, dağılımların göreceli değişkenliklerine gereksinim duyulur.

- Değişim katsayısı (coefficient of variation), standart sapmanın ortalama etrafında yüzde kaçlık bir değişim gösterdiği konusunda bilgi verir. Değişim katsayısı, DK;

$$DK = \frac{S}{\bar{x}} \times 100 \qquad (2.11)$$

- Değişim katsayısı ölçüm biriminden bağımsızdır. DK ’nın sıfıra yaklaşması dağılımın yaygınlığının azaldığını gösterirken, DK’ nın %25’ in üzerinde olması incelenen dağılımın oldukça yaygın bir dağılım olduğunu gösterir.

- Tablo 2.16'da Dağılım *1 (S1=5.5227)* için DK:
- *DK=(5.5227/6) x 100= % 92* iken,
- Dağılım *II* için *(S2=0.707116) için DK:*
- *DK=(0.707116/6) x 100 = % 11.8* olarak bulunur.
- Buna göre, Dağılım *II* ortalama etrafında % 11.8 'lik bir değişim gösterirken Dağılım *I,* % 92 'lik bir değişim gösterir.

- Dolayısıyla Dağılım *I,* Dağılım *II‘ ye* göre çok yaygındır.
- **Değişim katsayısı, özellikle ortalamaları farklı olan dağılımların yaygınlıkları hakkında bilgi sahibi olmak ve bir karşılaştırma yapmak amacı ile kullanılır.**

- Standart sapmanın karesine varyans denir.
- **Varyans** dağılım ölçüsüdür,Elimizdeki veri setindeki değerlerin ortalamaya göre dağılımını(ve değişimi) gösterir.
- En yaygın olan dağilim ölçülerinden varyans, gözlem değerimiz ile ortalama arasındaki farka(sapmaya) dayanır. **Sapma nedir** ? Bir serinin(yada bir setteki) herhangi bir değişkenin (değeri) ile ortalaması arasındaki farktır.

$$V = S^2 = 7.15^2 = 51.12 \quad (ml/kg/dk)^2$$

- Tablo 1.1.A 'daki *Max V02* ölçümlerinin varyansı olarak bulunur. Görüldüğü gibi, varyansın birimi kareseldir.
- Bu nedenle varyans, istatistiksel hesaplamalar dışında, **yaygınlık ölçüsü olarak veriyi tanımlamakta pek kullanılmaz.**

- Dağılımların yapısı hakkında görsel bilgi etmek amacıyla kullanılan çeşitli grafikler üzerinde daha önce durulmuş ve bu grafiklerin kullanım yerleri hakkında bilgi verilmişti.

- Simetrik ve tek tepeli dağılımları tanımlamak amacıyla kullanılan bir diğer basit grafik yöntemi ***ortalama± standart sapma*** grafiğidir. Bu grafik, genellikle, *aritmetik ortalamaya 1 ya da 2 standart* sapmalık uzaklıkların çizgi ile birleştirilmesi ile çizilir.

- Örneğin, Tablo 2.8'deki Veri Gurubu 2, tek tepeli simetrik bir yapıdadır ve ortalaması 66.8, standart sapması ise 7.2 'dir.
- Buna göre, ortalama± 1 x (standart sapma) grafiği Grafik 2 .12 'de verilmiştir. Bu grafikte,
- Alt uç=66.8-1 x 7.2= 59.6
- Üst uç=66.8+1x7.2=74.0 olarak elde edilir.

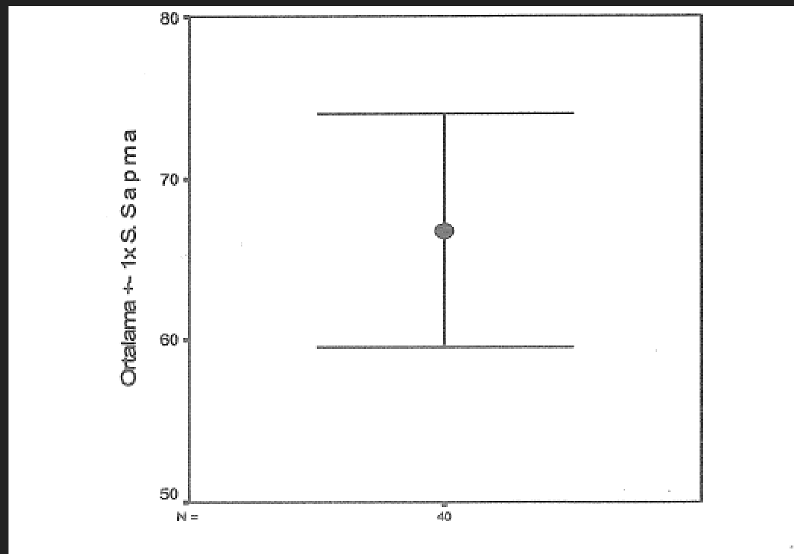
Tablo 2.8. Üç Örnek Veri Grubu (devam)

Veri Grubu 2.

73	72	65	74	69	64	60	67	78	66
54	50	72	77	67	68	67	57	70	66
83	76	68	76	69	57	60	66	62	71
65	67	56	58	77	62	61	67	62	72

Veri Grubu 3.

69	72	53	76	81	72	67	78	76	79
78	57	58	82	77	70	67	68	73	72
76	63	56	62	65	71	65	75	77	74
50	64	62	71	77	77	61	76	78	77



Grafik 2.12. Tablo 2.8'deki Veri Gurubu 2 İçin Ortalama ± 1 × (Standart Sapma) Grafiği

- Bazen dağılımların negatif ya da pozitif çarpık ve/ya da çok yaygın (standart sapmanın büyük) olması 1-2 değer nedeniyle ortaya çıkabilmektedir. Bu değerlere aşırı değer denir.
- Örneğin, 18-20 yaşlarındaki bireyler üzerinde yapılan bir çalışmada, boy uzunluğu 142 cm olan bir birey boy uzunluğu açısından bu yaş grubunun genel eğiliminin dışında olacaktır. Aşırı değer(ler), dağılımdaki diğer değerlerden çok farklı olan az sayıdaki değer olup ölçüm hatalarına, yanlış kaydetmeye, ölçüm aracının bozulmasına, .. vb bağlı olarak ortaya çıkabilir(ler).
- Aşırı değerler, bireysel farklılıklar nedeniyle ortaya çıkan gerçek değerler de olabilmektedir.

- Aşırı değerler sadece genel eğilime göre büyük değerli gözlemler değildir. Genel eğilime göre küçük değerli gözlemler de aşırı değer olabilir.
- Diğer bir deyişle, aşırı gözlemler dağılımın her iki tarafında da olabilir.
- Bir dağılımdaki aşırı değerlerin belirlenmesinde kutu-çizgi grafiklerine sıklıkla başvurulur.
- Bu yaklaşımda, kutunun alt ve üst sınırlarından (sırasıyla 25. ve 75. yüzdelikler) bir buçuk ile üç kutu boyu arası uzakta yer alan gözlemler aşırı değer olarak nitelendirililer ve istatistiksel yazılımlarda "o" ile gösterilirler.
- Kutunun alt ve üst sınırlarından üç kutu boyu ve daha fazla uzak olan gözlemler ise *çok* aşırı değer olarak nitelendirililer ve grafik üzerinde "*" ile belirtilirler.

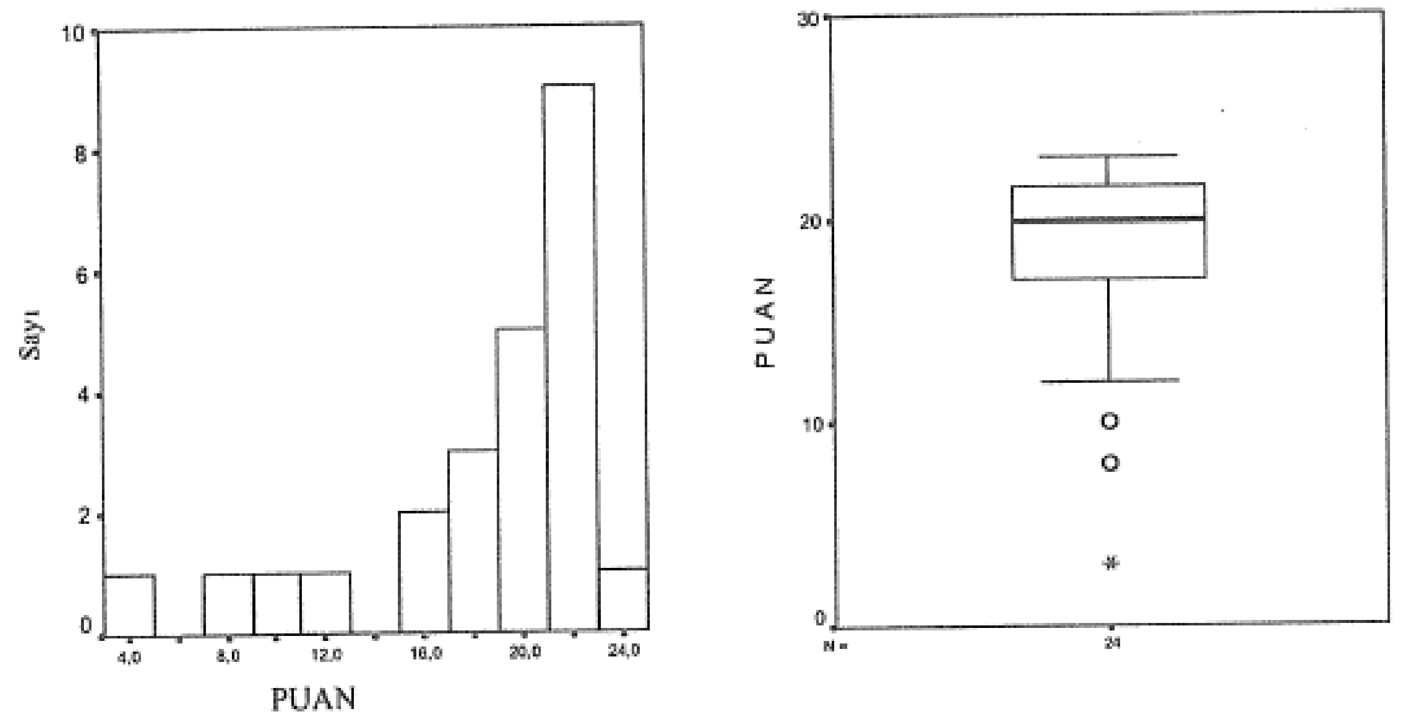
- Bir örnek olması açısından, 24 küçük yaş sporcusunun genel kültür düzeyi puanları (30 üzerinden) Tablo da verilmiştir. Bu veriye ilişkin histogram ve kutu çizgi grafiği ise Grafikte verilmiştir.

Tablo 2. 18. Genel Kültür Düzeyi Verisi

3	8	10	12	15	16	18	18	18.6	19.2	19.2	19.3	20.4	20.5
21	21	21.1	21.6	21.6	22	22	22.5	22.8	23				

- Grafikteki histogramdan görüldüğü gibi dağılım negatif çarpıktır.
- Çarpıklık, küçük değerli birkaç gözlem nedeniyle ortaya çıkmaktadır. Benzer yoruma, kutu-çizgi grafiği aracılığı ile de gidilebilir.
- Buna göre, kutunun üst sınırının en büyük değere yakın iken ortanca da kutunun üst sınırına yakındır.
- Ayrıca, bu dağılımda iki tane aşırı ve bir tane de çok aşırı gözlem vardır (not: grafikler SPSS 'de çizilmiştir).

18-21.6



Grafik 2.13. Genel Kültür Düzeyi Verisi Grafikleri

- Spor Bilimlerinde Uygulamalı İstatistik“Prof. Dr. Reha Alpar”
- http://ekonomiturk.blogspot.com.tr/2010/04/varyansstandart-sapma-nedir.html (erişim: 20.10.2015)
- Olasılık ve İstatistik “Prof. Dr. Fikri Akdeniz”
- İstatistik, Schaum Serisi “Çeviri: Prof. Dr. Salih Çelebioğlu”