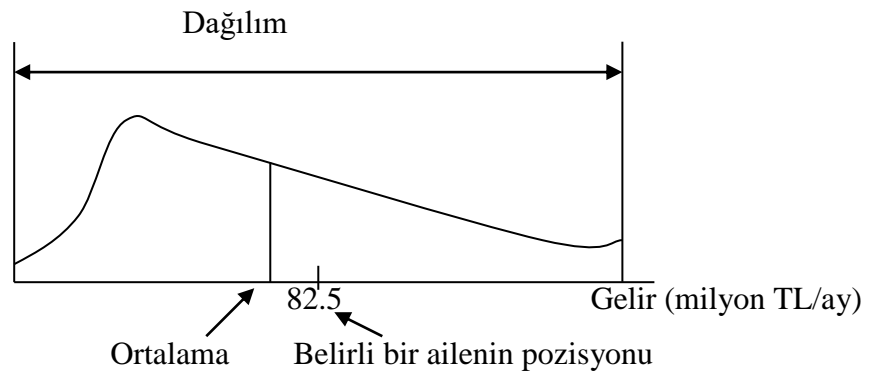


4. TANIMLAYICI İSTATİSTİK

Tanımlayıcı istatistik, belirli bir konuda karar alıcıların elinde bulunan işlenmemiş (ham) veya işlenmiş (sınıflandırılmış veya grafiklendirilmiş) verilerin temel özelliklerinin belirlenmesinde kullanılmaktadır. Sahip olunan verilerin hangi tipik değer etrafında toplandığı veya verilerin dağılımının nasıl olduğu araştırmacıların en fazla merak ettiği konulardır. Örneğin kırsal kesimde yaşayan ailelerin ortalama aylık gelirlerini araştıran bir kişi kırsal kesimde yaşayan bir ailenin ortalama olarak aylık ne kadar gelir elde ettiğini, ailelerin aylık gelirinin dağılımının nasıl olduğunu ve belirli bir ailenin aylık gelirinin bu dağılım içerisinde tam olarak hangi noktada yer aldığını öğrenmek isteyebilir. Şekil 4.1’de kırsal kesimde yaşayan tipik bir ailenin ortalama aylık geliri, kırsal kesimde yaşayan ailelerinin aylık gelirlerine ait dağılım ve belirli bir ailenin aylık gelirinin bu dağılım içindeki yeri gösterilmiştir.. İşte bütün bunları belirlemek için bazı tanımlayıcı ve özetleyici istatistik ölçülerini kullanmak gerekmektedir. Tanımlayıcı ve özetleyici ölçümler (1) merkezi eğilim ölçüleri, (2) dağılım ölçüleri ve (3) pozisyon ölçüleri olmak üzere üç grup altında incelenmektedir. Bu bölümde elimizde bulunan veri setini tanımlayıp özetlemeye yarayacak olan bu üç grup ölçümler anlatılacaktır.



Şekil 4.1. Tanımlayıcı ve Özetleyici Ölçümler

4.1. Merkezi Eğilim Ölçüleri

Merkezi eğilim ölçüleri veri setindeki verilerin özelliklerini sayısal bir değer ile özetlemede kullanılan en yaygın ölçülerdir. Merkezi eğilim ölçüleri veri setindeki verileri yansıtabilecek tipik değeri, yani veri setinin tam merkezinde bulunan değeri ifade etmektedir. Tarımsal ürünlerin ortalama fiyatları, tarım işletmelerinin sahip oldukları ortalama arazi büyüklüğü, ortalama aile büyüklüğü, tarım işletmelerinin sahip olduğu ortalama sermaye ve tüketicilerin tarım ürünlerine olan tercih durumu vb gibi ifadeler tarım ekonomisi alanında merkezi eğilim ölçülerinin kullanımına örnek olarak verilebilir. Çizelge 4.1’de çalışan kişilerin cinsiyet itibarıyla özellikleri verilmiştir. Bu çizelgede verilen değerler tipik değerlerdir, yani çalışan erkek ve bayanlar için tam merkezde bulunan değerlerdir.

En çok kullanılan merkezi eğilim ölçüleri aritmetik ortalama, ortanca değer (medyan), tepe değeri (mod), geometrik ortalama ve harmonik ortalamadır.

Çizelge 4.1. Çalışan Erkek ve Bayanların bazı Özellikleri

Özellik	Bayan	Erkek
Yaş	44	52
Maaş (milyon TL/ay)	350	600
Deneyim (yıl)	12	17
Evlilik oranı (%)	69	91
Çocuk sayısı	1	3
Haftalık çalışma süresi (saat)	56	56

4.1.1. Aritmetik ortalama

En yaygın kullanılan merkezi eğilim ölçüsüdür. Genellikle sadece ortalama kavramı kullanıldığında da aritmetik ortalama anlaşılmaktadır. Aritmetik ortalamanın diğer merkezi eğilim ölçülerine göre çok kolay hesaplanması, bütün gözlemlere dayanması, hesaplanmış bir değer olması ve cebirsel işlemlere çok uygun olması gibi bazı üstünlükleri bulunmakla beraber bazı dezavantajlara da sahiptir. Aritmetik ortalamanın en önemli dezavantajı veri setinde bulunan çok küçük veya çok büyük (extrem) değerlerden çok etkilenmesidir. Örneğin Samsun ilinin Çarşamba ilçesinde bulunan 5 tarım işletmesinin yıllık gelirlerinin sırasıyla 6.78, 9.27, 9.38, 6.47 ve 75.09 milyar TL olduğunu varsayalım. Söz konusu tarım işletmelerin dört tanesi 10 milyar TL’den daha az gelir elde ederken, bir tanesi diğerlerinin yaklaşık 7 katı fazla gelir elde etmektedir. İşte bu 75 milyar TL’lik veri, ilgilendiğimiz veri seti içinde extrem

bir deęerdir. Őimdi bu ok byk deęerin aritmetik ortalamaya etkisini grelim. ok byk olan bu deęer aritmetik ortalamaya dahil edildięinde tarım iřletmelerinin ortalama geliri 21.9 milyar TL olacak iken, dahil edilmedięinde 7.97 milyar TL olacaktır. Yani ok byk veya ok kk deęerler ieren veri setinde aritmetik ortalamanın hesaplanmasında ok dikkatli olmak gerekmektedir. İkinci dezavantaj ise, aritmetik ortalamanın simetrik olmayan daęılımlar iin iyi bir merkezi eęilim ls olmamasıdır. Yani aritmetik ortalama simetrik daęılım gsteren veriler iin uygun bir merkezi eęilim lsdr. Oysa Tarım ekonomisi alanında ve birok bilim dalında elde edilen veriler oęu zaman simetrik daęılım gstermemektedir. Bu sebeple aritmetik ortalama kullanılmadan nce elde bulunan verilerin daęılımının simetrik olup olmadıęı belirlenmelidir. Őayet aritmetik ortalama populasyondan hesaplanıyorsa μ , rnekten hesaplanıyorsa \bar{x} ile gsterilmektedir. Aritmetik ortalamanın sahip olduęu  temel zellik ařaęıda belirtilmiřtir:

1. Veri setinde bulunan btn gzlemlerin veri setinin aritmetik ortalamasından sapmalarının toplamı sıfırdır.

$$\sum_{i=1}^n (X_i - \bar{x}) = 0$$

2. Aritmetik ortalamadan sapmaların karelerinin toplamı en kktr.

$$\sum_{i=1}^n (X_i - \bar{x})^2 = \min$$

3. Aritmetik ortalamanın standart hatası, dięer merkezi eęilim llerinin standart hatasından daha kktr.

Aritmetik ortalama gruplandırılmıř ve gruplandırılmamıř verilerde farklı řekillerde hesaplanmaktadır.

1. Sınıflandırılmamıř verilerde aritmetik ortalamanın hesaplanması

Sınıflandırılmamıř verilerde aritmetik ortalama hesaplanırken veri setinde bulunan btn gzlemlerin toplamı alınıp, gzlem sayısına oranlanmaktadır. Sınıflandırılmamıř verilerde popülasyon ve rnek iin aritmetik ortalamanın hesaplanmasında kullanılan formller ařaęıda verilmiřtir:

$$\mu = \frac{\sum x}{N} \quad \text{Populasyonun aritmetik ortalaması}$$

$$\bar{x} = \frac{\sum x}{n} \quad \text{Örneğin aritmetik ortalaması}$$

Örneğin bir firmada toplam 8 kişinin çalıştığını varsayalım. Bu 8 kişinin aylık maaşları (milyon TL/ay) aşağıda verilmiştir:

450 530 825 370 615 480 910 560

Bu firmada çalışan işçilerin ortalama maaşı merak ediliyorsa, populasyonun içinde bulunan bütün elemanların maaşı bilindiği için populasyon için aritmetik ortalama hesaplanacaktır.

Burada $\sum x = 450 + 530 + 825 + 370 + 615 + 480 + 910 + 560 = 4740$ ve $N=8$ 'dir.

Populasyona ait aritmetik ortalama ise $\mu = \sum x / N$ (4740/8) 592.5 milyon TL'dir.

Ancak bu firmada çalışan 3 kişinin aylık maaşları bilinseydi, bu durumda populasyondan çekilmiş bir örnek için aritmetik ortalama hesaplanacaktır. Bu durumda $\sum x = 530 + 480 + 910 = 1920$, $n=3$ ve örnek için aritmetik ortalama $1920/3=640$ milyon TL olarak hesaplanacaktır.

2. Sınıflandırılmış verilerde aritmetik ortalamanın hesaplanması

Eğer elimizde sınıflandırılmış yani frekans dağılımı oluşturulmuş veriler var ise her bir sınıfın sınıf orta noktasından yararlanarak aritmetik ortalamayı üç farklı yöntemle hesaplamak mümkündür.

1. yöntem: Bu yöntemde her bir sınıfın orta noktası ait olduğu sınıfın frekansı ile çarpılmakta ve bu çarpımların toplamı toplam gözlem sayısına oranlanmaktadır. Bu yöntemle hesaplanan aritmetik ortalama, ham verilerden hesaplanan aritmetik ortalama değerinden farklılık göstermektedir. Bunun sebebi, hesaplamada esas alınan sınıf orta noktasının, sınıfın hakiki ortalamasına tam olarak eşit olmamasıdır. Fakat veri seti boyunca çok yüksek ve çok düşük orta noktalar birbirini telafi edeceklerinden aritmetik ortalamalar arasındaki fark çok büyük olmaz. Eğer sınıf genişlikleri çok iyi tespit edilmişse sınıflandırılmış ve sınıflandırılmamış verilerden hesaplanan ortalamalar birbirine çok yakın olur. Üçüncü bölümde Çorum ili Kızılırmak Havzasında yer alan tarım işletmelerinin mülk arazisi büyüklükleri için oluşturulan frekans dağılım tablosundan yararlanarak, bu yöntemle aritmetik ortalamayı hesaplayalım.

Çizelge 4.2. Sınıflandırılmış Verilerde Aritmetik Ortalamanın Hesaplanması (1.Yöntem)

Sınıflar (dekar)	Frekans (f)	Sınıf orta noktası (o)	Çarpım (o x f)	Hesaplama
1-11	18	6.0	108	$\bar{x} = \frac{\sum of}{\sum f}$
12-22	32	17.0	544	
23-33	16	28.0	448	
34- 44	12	39.0	468	
45-55	6	50.0	300	$\bar{x} = \frac{2544}{94}$
56-66	6	61.0	366	
67-77	2	72.0	144	$\bar{x} = 27.06$
78-88	2	83.0	166	
Toplam	94		2544	

2. yöntem: Bu yöntemde hesaplamaya bir düzeltme faktörü ilave edilmektedir. Yani önce aritmetik ortalama tahmini olarak tespit edilmekte, daha sonra bu ortalama düzeltme faktörü ile kesin değerine kavuşturulmaktadır. Bu yöntemde yapılacak ilk iş, aritmetik ortalamanın içinde bulunduğu sınıfı belirlemektir. Bu sınıf orijin kabul edilmekte, ve bu sınıfın sınıf orta noktası tahmini aritmetik ortalamayı temsil etmektedir. Aritmetik ortalamayı içeren sınıfı belirlemede kullanılan en yaygın yol en büyük frekansa sahip sınıfın orta noktasını almaktır. Tahmini aritmetik ortalamayı yansıtan sınıfın orta noktasına “indi orijin” adı verilmektedir. Daha sonra diğer sınıfların orta noktalarından indi orijin çıkartılmak suretiyle “sapmalar serisi” elde edilmektedir. Bunu takiben her bir sınıfın sapması frekans ile çarpılmakta ve bu çarpımların toplamı toplam frekans sayısına bölünerek düzeltme faktörü hesaplanmaktadır. Sonuç olarak indi orijin ile düzeltme faktörü toplandığında gerçek aritmetik ortalama değeri hesaplanmaktadır. Aşağıda bu yöntemle aritmetik ortalama hesaplarken kullanılan eşitlikler verilmiştir:

$$\bar{x} = A + d$$

$$d = \frac{\sum fb}{f}$$

Eşitliklerde \bar{x} aritmetik ortalamayı, A indi orijini, b aritmetik ortalamasının içinde olduğu tahmin edilen sınıfın sınıf orta noktasından sapmaları ve d düzeltme faktörünü ifade etmektedir. Aritmetik ortalamanın bu yöntemle hesaplanmasını görmek amacıyla Çorum İli

Kızılırmak Havzası tarım işletmelerinin mülk arazisine ait frekans dağılımı kullanılmış ve hesaplama Çizelge 4.3’de gösterilmiştir.

Çizelge 4.3. Sınıflandırılmış Verilerde Aritmetik Ortalamanın Hesaplanması (2.Yöntem)

Sınıflar (dekar)	Frekans (f)	Sınıf orta noktası	Sapma (b)	Çarpım (fb)	Hesaplama
1-11	18	6.0	-22	-396	$\bar{x} = A + d$ $d = \frac{\sum fb}{f}$ <p>A=28 da d= -88/94 d=-0.94</p> $\bar{x} = 28 - 0.94 = 27.06$
12-22	32	17.0	-11	-352	
23-33	16	28.0 (A)	0	0	
34- 44	12	39.0	11	132	
45-55	6	50.0	22	132	
56-66	6	61.0	33	198	
67-77	2	72.0	44	88	
78-88	2	83.0	55	110	
Toplam	94			-88	

2. yöntem: Bu yöntemde aritmetik ortalama hesaplanırken içinde aritmetik ortalamayı ihtiva ettiği tahmin edilen sınıfın orta noktasından yani indi orijinden sınıf genişliği sapmalarını hesaplamaya dayanmaktadır. İndi orijinden yukarıda ve aşağıda bulunan sınıfların artı ve eksi yönde ne kadar uzakta oldukları tespit edilir. Örneğin indi orijinden üç sınıf aşağıda olan sınıfın sapması -3’dür. Daha sonra bu sapmalar her bir sınıfın frekansı ile çarpılarak toplamaları alınır ve toplam frekans sayısına bölünür. Elde edilen sonuç sınıf genişliği ile çarpılarak bir düzeltme faktörü elde edilir. Bütün bu ara işlemler tamamlandıktan sonra aritmetik ortalama, indi orijinin sınıf orta noktasına düzeltme faktörünün ilavesi ile hesaplanmaktadır. Aşağıda bu yöntemle aritmetik ortalama hesaplamasında kullanılan eşitlikler verilmiştir:

$$\bar{x} = A + dc$$

$$d = \frac{\sum fb}{\sum f}$$

Eşitliklerde A aritmetik ortalamanın içinde olduğu düşünülen sınıfın orta noktasını, b indi orijinden sınıf genişliği olarak sapmaları, d düzeltme faktörünü ve c ise sınıf genişliğini ifade etmektedir.

Şimdi Çorum İli Kızılırmak Havzası tarım işletmelerinin mülk arazisine ait frekans dağılımını kullanarak aritmetik ortalamanın bu yöntemle hesaplanmasını görelim (Çizelge 4.4).

Çizelge 4.4. Sınıflandırılmış Verilerde Aritmetik Ortalamanın Hesaplanması (3.Yöntem)

Sınıflar (dekar)	Frekans (f)	Sınıf orta noktası	Sapma (b)	Çarpım (fb)	Hesaplama
1-11	18	6.0	-2	-36	$\bar{x} = A + dc$ $d = \frac{\sum fb}{\sum f}$
12-22	32	17.0	-1	-32	
23-33	16	28.0 (A)	0	0	
34- 44	12	39.0	1	12	
45-55	6	50.0	2	12	A=28 da
56-66	6	61.0	3	18	d= -8/94
67-77	2	72.0	4	8	d=-0.085
78-88	2	83.0	5	10	c=11
					$\bar{x} = 28 - 0.94 = 27.06$
Toplam	94			-8	

Sınıflandırılmış verilerde aritmetik ortalama hesaplamada kullanılan üç farklı yöntemde de aynı sonuç elde edilmiştir (27.06). Aynı veri setinde sınıflandırma yapılmaksızın aritmetik ortalama hesaplandığında 26.35 olarak bulunmaktadır. Görüldüğü gibi sınıflandırılmış ve sınıflandırılmamış verilerle hesaplanan aritmetik ortalamalar birbirilerine oldukça yakın değerler almaktadır. Bu farklılık sınıflandırılmış verilerde hesaplamaların sınıf orta noktasına dayanarak yapılmasından kaynaklanmaktadır. Eğer sınıf orta noktaları ile sınıfın hakiki ortalaması birbirine eşit olsaydı böyle bir fark olmayacaktı. Bu sebeple frekans dağılım tablosu oluşturulurken sınıf genişliklerinin ve sınıf sayısının çok iyi tespit edilmesi gerekmektedir. Eğer gerektiğinden daha az sınıf yapılp sınıf genişliğin büyük seçilirse aritmetik ortalamalar arasındaki fark daha da artacaktır.

4.1.2. Ortanca değer (medyan)

Aritmetik ortalamadan sonra en önemli merkezi eğilim ölçüsü ortanca değerdir. Ortanca değer büyüklüğüne göre dizilmiş veri setinin tam ortasında bulunan gözlemin değeridir. Bir merkezi eğilim ölçüsü olarak ortanca değer en önemli üstünlüğü veri setinde bulunan extrem değerlerden etkilenmemesidir. Özellikle son derece asimetrik dağılım

gösteren veri setinde çok küçük ve çok büyük gözlemleri dikkate almadığı için ortanca değer aritmetik ortalamadan daha iyi bir merkezi eğilim ölçüsüdür. Örneğin 5 adet tarım işletmesinin arazi büyüklükleri sırasıyla 25, 42, 56, 65, 250 dekadır. Bu veri seti için aritmetik ortalama $(25+42+56+65+250)/5=87.6$ dekadır. Veri setinde bulunan tek bir işletme ortalama değerden daha fazla araziye sahipken, 4 tarım işletmesi bu ortalama değerden daha küçük araziye sahiptir. Yani aritmetik ortalama büyük değerden etkilenerek veri setini temsilden uzaklaşmıştır. Oysa aynı veriler için ortanca değer 56 dekadır ve tarım işletmelerinin yarısı bu değerden daha fazla, diğer yarısı ise daha az araziye sahiptir. Ortanca değer bir diğer üstünlüğü de belirlenmesinin çok kolay olmasıdır. Bu üstünlüklerini yanında ortanca değer hesaplanmış bir değer olmaması, veri setindeki gözlemleri sıralamayı gerektirmesi, aritmetik işlemlere uygulanamaması ve çok fazla gözlemle kullanıldığında yanlış yorumlara yol açması gibi olumsuzlukları da bulunmaktadır. Tarım ekonomisi sahasında genellikle fiyat endekslerinin hazırlanmasında kullanılan, ayrıca sıra (ordinal) verilerde kullanılan en önemli merkezi eğilim ölçüsü olan ortanca değer iki temel özelliği vardır:

1. Ortanca değerden pozitif ve negatif sapmaların adeti eşittir.
2. Aritmetik ortalamadan daha büyük standart hataya sahiptir.

1. Sınıflandırılmamış verilerde ortanca değer hesaplanması

Sınıflandırılmamış verilerde ortanca değer hesaplanan bir değer değil, tayin edilen bir değerdir. Ortanca değer belirlenmesi işlemi oldukça basittir ve iki temel basamağı bulunmaktadır. Yapılacak ilk iş veri setinde bulunan gözlemleri büyükten küçüğe doğru veya küçükten büyüğe doğru sıraya dizmektir. Daha sonra sıraya dizilmiş veri setinde tam ortada bulunan gözlem belirlenir ve bu gözlemin değeri ortanca değerdir. Veri setinde bulunan gözlem sayısı tek ise, veri setinin tam ortasında tek bir gözlem vardır ve buda ortanca değerdir. Ortanca değer pozisyonu $\frac{n+1}{2}$ formülü ile belirlenir. Ancak veri setinde bulunan gözlem sayısı çift ise veri setini tam ortasında tek bir gözlem değil iki gözlem yer alır. Bu durumda ortanca değer veri seti içinde bulunan $\frac{n}{2}$ nci gözlemin değeri ile, $\frac{n+2}{2}$ nci gözlemin değerinin ikiye bölünmesi ile belirlenmektedir.

Örneğin Türkiye'nin 5 büyük ihracatçı firmasının ihracat değerleri 10007, 11548, 9693, 75275 ve 18296 milyon dolardır. Acaba bu veri seti için ortanca değer nedir? Ortanca değeri belirlemek için öncelikle mevcut veriler sıralanmalıdır.

9693 10007 11548 18296 75275

↑
Ortanca değer

Elimizde bulunan veri setinde gözlem sayısı tek olduğu için $\frac{n+1}{2}$ nci yani $6/2=3$. gözlemin değeri elimizde bulunan veri seti için ortanca değerdir. Bu değer 11548'dir. Dikkat edilecek olursa medyan değerinin sağında ve solunda bulunan gözlem sayısı eşittir.

Ancak bazen elimizde bulunan veri setinde toplam gözlem sayısı aşağıdaki gibi çift olur.

13 23 36 50 97 210 234 249 257 275 385 506

↖ ↗
6. gözlem medyan 7. gözlem

Bu durumda medyan $(\frac{n}{2})$ yani $12/2=6$. gözlemin değeri ile $(\frac{n+2}{2})$ $14/2=7$. gözlemin değerinin ortalamasına eşittir. Sonuçta medyan altıncı gözlem olan 210 ile yedinci gözlem olan 234'ün toplamının ikiye bölümü olan 222'ye eşittir.

2. Sınıflandırılmış verilerde ortanca değer hesaplanması

Sınıflandırılmış yani frekans tablosu oluşturulmuş verilerde ortanca değer tespit edilirken öncelikle ortanca değer içinde bulunduğu sınıf belirlenir. Bu işlem yapılırken frekansların yarısının bu sınıfın üstünde, diğer yarısının da altında olmasına dikkat edilir. Daha sonra belirlenen sınıf içinde ortanca değer sınıf genişliğinden interpolasyon yardımıyla bulunur. Ortanca değer hesaplanmasında aşağıdaki formül kullanılır:

$$M_e = I_i + \left(\frac{\frac{N}{2} - f_{-i}}{f_0} \right) \cdot c$$

Formülde M_e ortanca değeri, I_i ortanca değer bulunduğu sınıfın alt sınırını, f_{-i} ortanca değer bulunduğu sınıftan daha düşük değerleri olan sınıfların frekanslarının toplamını f_0 ortanca değer bulunduğu sınıfın frekansını ve c sınıf genişliğini ifade etmektedir. Şimdi Çorum İli Kızılırmak Havzası tarım işletmelerinin mülk arazi büyüklüklerine ait frekans tablosunu kullanarak ortanca değeri belirleyelim (Çizelge 4.5).

Çizelge 4.5. Sınıflandırılmış Verilerde Ortanca Değerin Hesaplanması

Sınıflar (dekar)	Frekans (f)	Hesaplama
1-11	18	$M_e = I_i + \left(\frac{\frac{N}{2} - f_{-i}}{f_0} \right) \cdot c$ $M_e = 12 + \left(\frac{\frac{94}{2} - 18}{32} \right) \cdot 11$ $M_e = 21.97$
12-22	32	
23-33	16	
34- 44	12	
45-55	6	
56-66	6	
67-77	2	
78-88	2	
Toplam	94	

Belirlenen ortanca değer 21.9'dir. Örneğimiz için sınıflandırılmamış veriler için ortanca değer ise 20 olarak bulunmuştur. Görüldüğü gibi bu değer aynı verilerden oluşturulmuş frekans tablosundan hesaplanan ortanca değerden farklıdır, ancak oldukça yakındır. Bu fark sınıf genişliklerinin büyüklüğünden, sınıfların alt ve üst sınırlarından ve dağılımın özelliğinden kaynaklanmaktadır.

4.1.3. Tepe değeri (mod)

Günlük yaşantımızda “Sokaktaki adam”, “tipik işletme” vb ifadeler kullanıldığında farkında olunmadan tepe değerinden (mod) söz edilmektedir. Tepe değeri (mod) Fransızca kökenli bir kelime olup, çok genel veya moda anlama gelmektedir. İstatistikte tepe değeri (mod) veri setinde en fazla tekrar eden gözlem olarak tanımlanmaktadır. Diğer bir ifade ile tepe değeri dağılımın en sık ve en tipik değeridir. Tepe değerinin en önemli üstünlüğü en çok rastlanan değer olduğu için veri setini diğer merkezi eğilim ölçülerinden daha iyi temsil etmesidir. Çok küçük ve çok büyük değerlerin tesirini içermemesi ise tepe değerinin bir diğer üstünlüğüdür. Oldukça güç tanımlanması, az sayıda gözlem olduğunda pratik olarak kullanılamaması ve cebirsel işlemlere uygunluk göstermemesi ise tepe değerinin olumsuz yanlarıdır. Ayrıca her dağılımda tek bir aritmetik ortalama ve ortanca değer olmasına rağmen bazı dağılımlarda birden fazla tepe noktası da söz konusu olabilmektedir. Eğer bir dağılımın tek bir tepe noktası varsa tek tepeli dağılım (unimodal), iki tepe noktası mevcutsa iki tepeli

dağılım (bimodal) ve ikiden fazla tepe değeri varsa çok tepeli dağılım (multimodal) adı verilmektedir.

Tepe değeri aritmetik ortalama ve ortanca değerden farklı olarak sınıflandırılmamış verilerde güçlükle belirlenmektedir. Eğer veri setindeki gözlem sayısı en çok tekrar eden gözlemi tespit edecek kadar küçükse (25-30) tepe değerini sınıflandırma yapmadan da belirlemek mümkündür. Ancak çoğu zaman veri seti daha fazla sayıda gözlem ihtiva etmekte, dolayısıyla sınıflandırma yapmadan tepe değerini belirlemek imkânsız hale gelmektedir. Bu durumda eğer dağılım asimetrik ise aritmetik ortalama, ortanca değer ve tepe değeri arasındaki ilişkiden yararlanarak tepe değerini sınıflandırılmamış verilerde tahmin etmek mümkündür. Asimetrik dağılımlarda ortanca değer aritmetik ortalama ile tepe değeri arasındadır ve yaklaşık olarak aritmetik ortalamadan itibaren 1/3 oranında bir mesafede yer alır. Bu yolla tepe değerinin yaklaşık tahmininde aşağıdaki formül kullanılmaktadır:

$$M_0 = \bar{x} - 3(\bar{x} - M_e)$$

Formülde M_0 tepe değerini, M_e ortanca değeri ve \bar{x} aritmetik ortalamayı ifade etmektedir.

Sınıflandırılmış verilerde tepe değerinin hesaplanması için tepe değerinin içinde olduğu model sınıfın orta noktasından yararlanılmaktadır. Model sınıf en yüksek frekansa sahip sınıf olmaktadır. Eğer tepe değerinin bulunduğu sınıfın üzerindeki sınıfta bulunan frekans sayısı, tepe değerinin bulunduğu sınıfın altında bulunan sınıfın frekansından fazla ise tepe değeri model sınıfın orta noktasından büyük bir değerdir. Tepe değerinin belirlenmesinde aşağıdaki formül kullanılmaktadır:

$$M_0 = I_{-i} + \left(\frac{f_{+i}}{f_{+i} + f_{-i}} \right) \cdot c$$

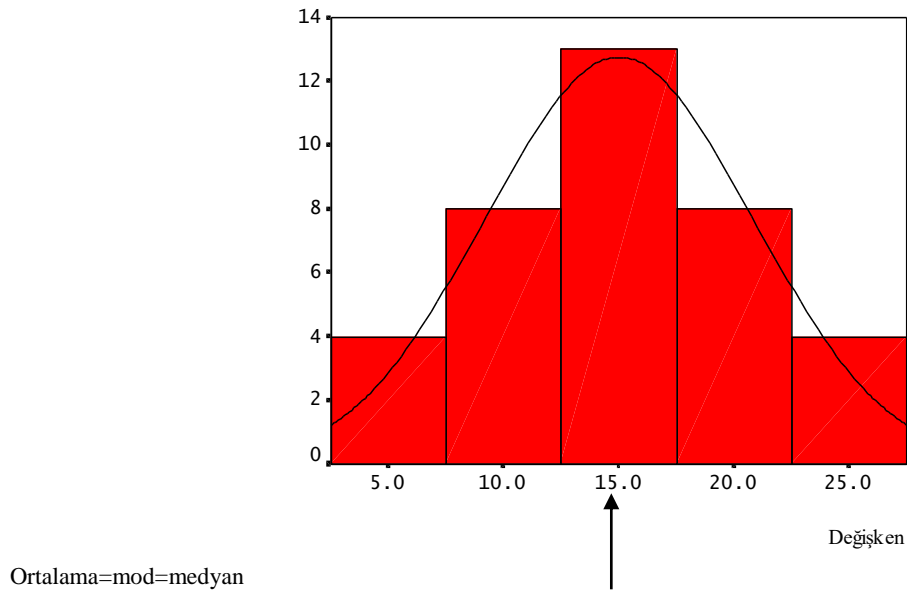
Formülde M_0 tepe değerini, I_{-i} tepe değerinin bulunduğu sınıfın alt sınırını, f_{-i} tepe değerinin bulunduğu sınıftan yukarıda (küçük olan) sınıfın frekansı, f_{+i} tepe değerinin bulunduğu sınıftan aşağıda (büyük olan) sınıfın frekansı ve c sınıf genişliğini ifade etmektedir. Çorum İli Kızılırmak Havzası tarım işletmelerinin mülk arazi büyüklüklerine ait frekans tablosunu kullanarak tepe değerinin bulunması Çizelge 4.6'da gösterilmiştir.

Çizelge 4.6. Sınıflandırılmış Verilerde Tepe Değerinin Hesaplanması

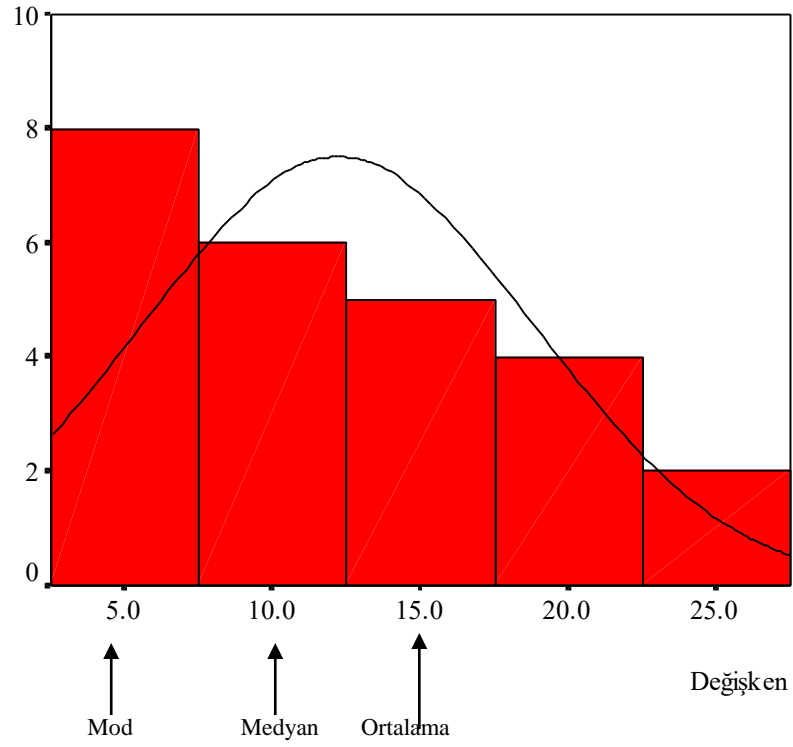
Sınıflar (dekar)	Frekans (f)	Hesaplama
1-11	18	$M_0 = I_{-i} + \left(\frac{f_{+i}}{f_{+i} + f_{-i}} \right) \cdot c$ $M_0 = 12 + \left(\frac{16}{16 + 18} \right) \cdot 11$ $M_0 = 17.18$
12-22	32	
23-33	16	
34- 44	12	
45-55	6	
56-66	6	
67-77	2	
78-88	2	
Toplam	94	

4.1.4. Aritmetik ortalama, ortanca değer (medyan) ve tepe değeri (mod) arasındaki ilişkiler

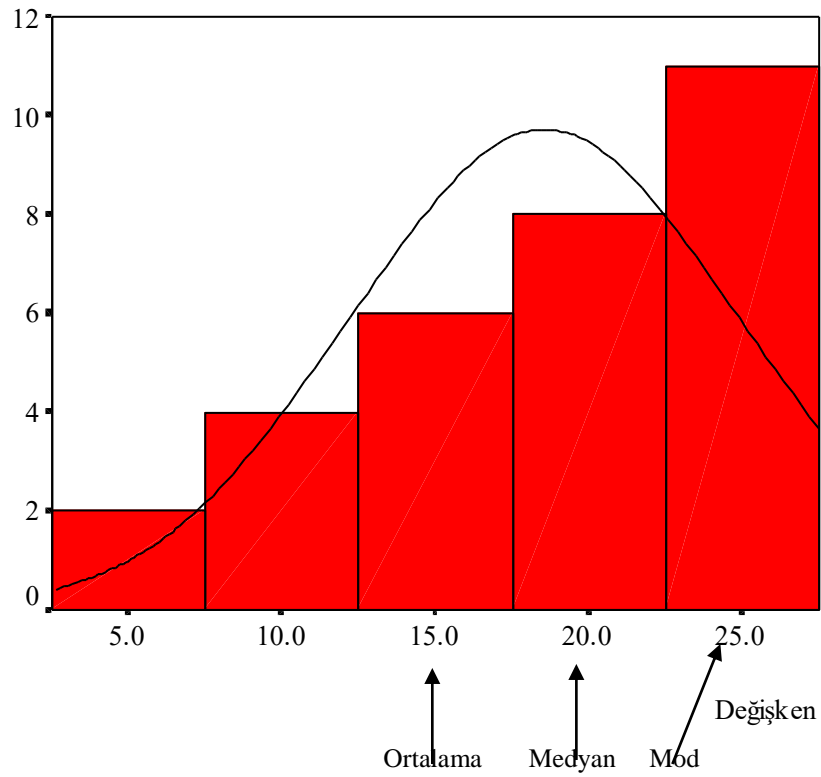
Simetrik bir dağılımda aritmetik ortalama, ortanca değer ve tepe değeri birbirine eşittir. Yani bu üç merkezi eğilim ölçüsü dağılımın tam orta noktasında bulunurlar (Şekil 4.2).



Sağa çarpık bir dağılımda ise aritmetik ortalama, ortanca değerden; ortanca değer ise tepe değerinden büyüktür (Ortalama>ortanca değer>tepe değeri). Eğer dağılım sola çarpıksa tepe değeri ortalama ve ortanca değerden daha büyüktür (tepe değeri>ortanca değer>ortalama).



Şekil 4.2. Sağa çarpık dağılımda aritmetik ortalama, ortanca değer ve tepe değeri



Şekil 4.2. Sola çarpık dağılımda aritmetik ortalama, ortanca değer ve tepe değeri

4.1.5. Geometrik ortalama

Geometrik ortalama veriler arasındaki aritmetik farklar yerine, geometrik farkları dikkate alan merkezi eğilim ölçüsüdür. Veri setinde yer alan bütün gözlemleri kapsayan ve cebirsel işlemlere uygunluk gösteren geometrik ortalama nispi rakamların işlenmesine imkân tanımaktadır. Ancak hesaplanması oldukça güçtür ve genel olarak insanlar tarafından güç anlaşılmaktadır. Geometrik ortalamanın en önemli sakıncası veri setinde hem negatif hem de pozitif değerlerin bulunduğu durumlarda ve veri setinde bir veya birden fazla gözlemin değerinin sıfır olması durumunda hesaplanamamasıdır. Geometrik ortalamanın hesaplanmasında veriler arasındaki geometrik farklar dikkate alınmaktadır. Geometrik ortalamanın hesaplanması aşağıda açıklanmıştır.

Geometrik ortalamanın hesaplanması: Geometrik ortalama hesaplanırken, veri setinde bulunan bütün gözlemlerin çarpımların toplam gözlem sayısı derecesinden kökü alınmaktadır.

$$M_g = \sqrt[n]{\sum_{i=1}^n x_i x_n}$$

Formülde n gözlem sayısını, x gözlem değerini ve M_g geometrik ortalamayı ifade etmektedir. Genellikle gözlem sayısı arttıkça yapılan işlem daha karmaşık hale geldiğinden, hatta imkânsız olduğundan uygulamada bu hesaplama “logaritma” dan yararlanarak yapılmaktadır. Bu yöntemle geometrik ortalama hesaplanırken gözlemlerin logaritmalarının toplamı gözlem sayısına bölünmekte ve elde edilen değer anti logaritması alınmaktadır.

$$\log M_g = \frac{\log x_1 + \log x_2 + \dots + \log x_n}{n} = \frac{\sum \log x}{n}$$

Çorum İli Kızılırmak Havzası tarım işletmelerinin mülk arazi büyüklükleri kullanılarak geometrik ortalama hesaplandığında $\log M_g = \frac{119.11}{94} = 1.27$ 'nin anti logaritması olan 18.50 dekar değeri bulunmaktadır.

Geometrik ortalama veri setinde bulunan küçük gözlemlere daha fazla, büyük gözlemlere daha az önem vermektedir. Bu sebeple dağılım büyük değerlere meyilli olduğu zaman geometrik ortalama tepe değerine daha yakın değer almaktadır ve aritmetik ortalama daha tipiktir. Bunun tersi durumda dağılım küçük değerlere doğru kayarsa aritmetik ortalama, geometrik ortalama daha tipiktir.

Geometrik ortalama bir merkezi eğilim ölçüsü olarak aritmetik ortalamadan daha az önemli olmakla beraber artma eğilimi gösteren değişimlerin açıklanmasında çok faydalı olmaktadır. Faiz, nüfus, tüketim ve gelir artışı gibi verilerin dağılımında geometrik ortalama kullanılmaktadır. Geometrik ortalama özellikle oranlar ile ilgili çalışmalara çok uygun düşmektedir. Örneğin aynı bölgede bulunan 3 tarım işletmesinin 1995 ve 2000 yılında ödedikleri vergiler (milyon TL) aşağıdaki gibi olsun:

İşletme no	1995	2000
1	498	605
2	202	142
3	97	63

Bu işletmelerin ödedikleri vergilerin her iki yıl için aritmetik ve geometrik ortalamaları aşağıdaki gibidir.

	1995	2000
Aritmetik ortalama	266	270
Geometrik ortalama	214	176

İncelenen dönemde birinci tarım işletmesinin ödediği vergi artmış, ancak diğerlerinin ödediği vergi miktarı ise azalmıştır. Aritmetik ortalama incelendiğinde bu işletmenin ödediği vergilerde küçük bir artışı ifade etmektedir. Bu karşılaştırmada aritmetik ortalama birinci işletmenin fazlasıyla tesiri altında kalmıştır. Oysa geometrik ortalama ise ödenen vergi miktarlarında azalmaya işaret etmektedir. Geometrik ortalama işletmelerin ödedikleri vergilerin devre içinde değişme nispetlerine eşit ağırlık verdiği için, aritmetik ortalamaya oranla daha iyi sonuç vermektedir. Bu tarz olaylarda geometrik ortalama çok daha uygun olduğu halde, pratikte daha az kullanılmaktadır. Bunun sebebi insanların geometrik farklardan çok, doğrusal farkları açıklamaya alışmış olmalarıdır.

4.1.6. Harmonik ortalama

Harmonik ortalama nispeten daha az önemli bir merkezi eğilim ölçüsüdür. Harmonik ortalama, verilerin terslerinin aritmetik ortalamasının tersidir. Daha çok oranların ortalamasının alınmasında kullanılır. Harmonik ortalama veri setinde bulunan bütün gözlemlere dayanan ve cebirsel işlemler uygunluk gösteren bir merkezi eğilim ölçüsüdür.

Ancak harmonik ortalamanın hesaplanması güçtür, küçük sayılara gereğinden fazla önem verir ve veri setinde hem negatif hem de pozitif değerler varken veya veri setinde sıfır varsa anlam ifade etmemektedir. Harmonik ortalama d/t oranında d değeri yani incelenen değişkenin değeri sabit, t değeri yani zaman değişkeni iken çok tipiktir ve çok anlamlı olmaktadır. Ters durumda aritmetik ortalama, harmonik ortalamadan daha tipiktir. Harmonik ortalama gerek sınıflandırılmış verilerde gerekse de sınıflandırılmamış verilerde hesaplanabilmektedir.

1. Sınıflandırılmış verilerde harmonik ortalamanın hesaplanması

Sınıflandırılmış verilerde harmonik ortalama hesaplanırken aşağıdaki formülden yararlanılmaktadır.

$$M_h = \frac{\sum f}{\sum \left(\frac{1}{son}\right) f}$$

Formülde “f” frekansı, “son” ise sınıf orta noktasını ifade etmektedir. Çorum ili Kızılırmak havzası tarım işletmelerinin mülk arazisi büyüklüklerine ait frekans tablosu kullanılarak harmonik ortalamanın hesaplanması Çizelge 4’7de verilmiştir. Araştırma alanı için harmonik ortalama 15.58 dekar olarak bulunmuştur. Görüldüğü gibi harmonik ortalama küçük değerlere daha fazla önem verdiği için aritmetik ortalama ve geometrik ortalamadan daha küçük değere sahiptir.

Çizelge 4.7. Sınıflandırılmış Verilerde Harmonik Ortalamanın Hesaplanması

Sınıflar (dekar)	Frekans (f)	Sınıf orta noktası	$1/SON$	Çarpım $(1/SON)f$	Hesaplama
1-11	18	6.0	0.170	3.00	$M_h = \frac{\sum f}{\sum \left(\frac{1}{son}\right) f}$ $M_h = \frac{94}{6.03}$ $M_h = 15.58$
12-22	32	17.0	0.059	1.88	
23-33	16	28.0	0.036	0.57	
34- 44	12	39.0	0.026	0.31	
45-55	6	50.0	0.020	0.12	
56-66	6	61.0	0.016	0.10	
67-77	2	72.0	0.013	0.03	
78-88	2	83.0	0.012	0.02	
Toplam	94			6.03	

1. Sınıflandırılmamış verilerde harmonik ortalamanın hesaplanması

Sınıflandırılmamış verilerde harmonik ortalama hesaplanırken gözlemlerin tersi alınıp, toplam gözlem sayısına oranlanmaktadır. Aşağıda sınıflandırılmamış verilerde harmonik ortalamanın hesaplanmasında kullanılan formül verilmiştir.

$$M_h = \frac{N}{\sum (1/X)}$$

Formülde M_h harmonik ortalamayı, N veri setindeki toplam gözlem sayısını ve X ise her bir gözlemin değerini ifade etmektedir.

Örneğin 3 farklı işçi 1 dekar mısırı tarlasından kaldırılan ürünün koçan yapraklarının soyulması işlemini ayrı ayrı sırasıyla 2, 1.5 ve 1 saatte bitiriyorlarsa, ortak olarak çalıştıklarında aynı işi ne kadar zamanda bitirebileceklerdir. Bu sorunun cevabı ancak harmonik ortalamanın hesaplanması ile verilebilir. Burada harmonik ortalama alınmakla her bir işçi için birim zamanda soyulan koçan sayısı tespit edilmekte ve bunların ortalaması alınmaktadır.

$$M_h = \frac{3}{\frac{1}{2} + \frac{1}{1.5} + \frac{1}{1}} = 1.385$$

Yani üç işçi birlikte çalıştıklarında 1 dekar araziden çıkan mısırların koçan yapraklarını 1.385 saatte tamamlayabileceklerdir.

Bir başka örnek olarak tarım işletmelerinin sahip oldukları mülk arazi büyüklüklerinin, işletme arazisi içindeki paylarının ortalamasını harmonik ortalama yardımıyla alalım. Çizelge 4.8'de 15 tarım işletmesinin sahip olduğu mülk arazisi büyüklüğünün işletme arazisi içindeki payları ve buna dayanarak harmonik ortalamanın hesaplanma tarzı verilmiştir.

Çorum ili Kızılırmak Havzasında yer alan tarım işletmelerinin sahip olduğu sınıflandırılmamış mülk arazi büyüklüklerine ilişkin harmonik ortalama ise $M_h = \frac{N}{\sum (1/X)}$,

$$M_h = \frac{94}{7.60} = 12.36 \text{ olarak bulunmaktadır.}$$

Çizelge 4.8. Sınıflandırılmamış Verilerde Harmonik Ortalamanın Hesaplanması

Mülk arazinin işletme arazisine oranı (%)	Verilerin tersi (1/X)	Hesaplama
86.7	0.01153	$M_h = \frac{N}{\sum (1/X)}$ $M_h = \frac{15}{0.25975}$ $M_h = 57.75$
98.3	0.01017	
46.5	0.02150	
77.4	0.01292	
83.2	0.00120	
72.1	0.01387	
65.4	0.01529	
46.8	0.02137	
97.4	0.01027	
62.6	0.01597	
40.9	0.02445	
27.6	0.03623	
49.8	0.02008	
51.7	0.01934	
67.8	0.01474	
Toplam	0.25975	

4.2. Dağılım Ölçüleri

Merkezi eğilim ölçüleri veri setinde bulunan verilerin ortalama etrafında toplanma özelliklerini yansıtırken, dağılım ölçüleri ise gözlemlerin nasıl dağıldığını göstermektedir. Bir anlamda dağılım ölçüleri veri setinde bulunan gözlemler arasındaki değişimi göstermektedir. Aritmetik ortalama, ortanca değer, tepe değeri, geometrik ortalama ve harmonik ortalama genellikle dağılımın şekli hakkında bilgi vermede yetersiz kalmaktadır. Bu durumlarda değişim genişliği, varyans, standart sapma ve değişim katsayısı gibi dağılım ölçüleri kullanılmaktadır. Aşağıda iki farklı şirkette çalışan işçilerin yaşlarına ait veri setleri verilmiştir.

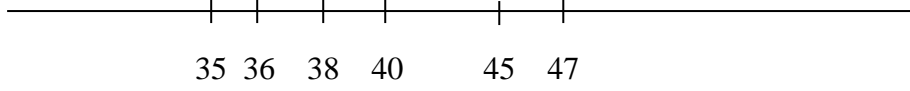
Şirket A : 47 38 35 40 36 45 39

Şirket B : 70 33 18 52 27

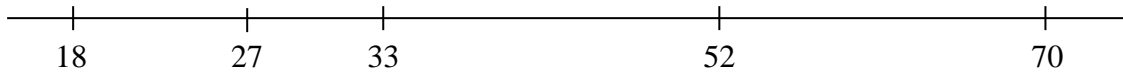
Her iki şirkette de işçilerin ortalama yaşı 40'dır. Bu şirketlerde çalışan her bir işçinin yaşları ve dağılımı bilinmeden sadece merkezi eğilim ölçülerine (aritmetik ortalama, ortanca değer vb) dayanarak her iki şirkette çalışan işçilerin yaşlarının dağılımının aynı olduğunu söylemek mümkündür. Ancak gerçekte B şirketinde çalışanların yaşlarındaki değişim, A şirketine oranla daha fazladır. Dolayısıyla iki şirkette aynı dağılıma sahiptir demek pek doğru bir ifade değildir. Bu sebeple sadece merkezi eğilim ölçülerini kullanmak yerine, merkezi eğilim

ölçüleri ile birlikte dağılım ölçülerini kullanmak veri seti hakkında daha iyi ve doğru bilgiler vermektedir. Şimdi dağılım ölçüleri olan değişim genişliği, varyans, standart sapma ve değişim katsayısını ayrı ayrı inceleyelim.

Şirket A



Şirket B



4.2.1. Değişim genişliği (range)

Değişim genişliği, değişim miktarının en kolay anlaşılan ve en yaygın kullanılan ölçüsüdür. Değişim genişliği veri seti içinde bulunan en büyük gözlem ile en küçük gözlem arasındaki farkı ifade etmektedir. Örneğin Çorum ili Kızılırmak Havzasında yer alan 94 tarım işletmesine ait mülk arazi büyüklükleri incelendiğinde en büyük gözlemin 86 dekar, en küçük gözlemin ise 1 dekar olduğu görülmektedir. Buna göre değişim genişliği $(86-1)$ 85 dekar'dır. Eğer veriler büyüklüklerine göre dizilmiş ise değişim sınırları kolaylıkla tespit edilebilmektedir. Veri setinde bulunan gözlemlere dayanarak frekans tablosu oluşturulmuş ise değişim genişliği, ilk sınıfın orta noktası ile son sınıfın orta noktası arasındaki fark alınarak bulunmaktadır. Çok kolay hesaplanmasına rağmen veri setinde bulunan bütün gözlemleri bir tarafa bırakıp sadece iki tanesine (en küçük ve en büyük) dayanması ve veri setinde bulunan ekstrem değerlerden etkilenmesi değişim genişliğini kullanmanın en önemli sakıncalarıdır.

Değişim genişliği iki farklı veri setini birbirleriyle karşılaştırmada yaygın olarak kullanılmaktadır. Karşılaştırmalarda değişim genişliği kullanılırken 2 farklı veri setinde bulunan gözlem sayıları mutlaka birbirine eşit olmalı ve aynı zamanda aynı birimi taşımalıdır. Veri setinde bulunan gözlem sayıları ve birimleri birbirinden farklı iki serinin değişim genişliği ile karşılaştırılması doğru değildir.

Örneğin iki farklı ilçede 6'şar kişinin gelirlerinin (100 milyon TL/ay) tespit edildiğini ve bunlara ait gözlemlerin aşağıdaki gibi olduğunu varsayalım:

Vezirköprü : 2 3 6 7 8 10

Havza : 5 5 5 6 7 8

İncelenen her iki ilçede de ortalama gelir 600 milyon TL’dir. İlçelerin gelir ortalaması eşit olmakla birlikte gelirin dağılımı birbirinden farklılık göstermektedir. Vezirköprü için gelire ait değişim genişliği $1000 - 200 = 800$ milyon TL, Havza için gelire ait değişim genişliği ise $800 - 500 = 300$ milyon TL’dir. Dikkat edildiğinde Havza ilçesindeki gelir dağılımı, Vezirköprü’den daha iyidir, çünkü değişim genişliği bu ilçede daha küçüktür.

4.2.2. Varyans ve standart sapma

Standart sapma en yaygın kullanılan dağılım ölçüsüdür. Standart sapma veri seti içinde yer alan gözlemlerin ortalamaya ne kadar yakın olduğunu ifade etmektedir. Genellikle küçük standart sapma değeri, ortalamadan sapmaların küçük olduğunu göstermektedir. Ters durumda yani büyük standart sapma söz konusu olduğunda ise, verilerin ortalamadan uzaklaştığı anlaşılmaktadır. Standart sapma aritmetik ortalamadan sapmaların karelerinin toplamının yani varyansın toplam gözlem sayısına oranlanması ve bu değer karekökünü alınmasıyla elde edilmektedir. Özetle standart sapma, varyansın kareköküne eşittir. Eğer varyans populasyon için hesaplanıyorsa σ^2 , örnek için hesaplanıyorsa S^2 ile gösterilmektedir. Buna bağlı olarak popülasyona ait standart sapma σ ile, örneğe ait standart sapma ise S ile gösterilmektedir. Herhangi bir veri seti için hesaplanan varyans ve standart sapma değerleri hiçbir zaman negatif olmazlar. Bu değerler daima pozitifdir. Eğer veri setinde değişim yoksa varyans ve standart sapma “0” a eşittir. Örneğin veri seti içinde yer alan 5 kişinin de yaşı 35 ise, bu durumda değişim yok demektir, yani varyans ve standart sapma sıfırdır. Ayrıca veri seti içinde yer alan gözlemlerin birimi (dekar, TL vb) ne ise, standart sapmanın birimi de aynıdır.

1. Sınıflandırılmamış verilerde varyans ve standart sapmanın hesaplanması

Sınıflandırılmamış verilerde varyans hesaplanmasında kullanılan temel formüller aşağıda verilmiştir.

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N} \quad \text{ve} \quad S^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

Formülde σ^2 populasyon varyansını, S^2 örnek varyansını, $(x - \mu)^2$ ve $(x - \bar{x})^2$ her bir gözlem değerinin ortalamadan sapma miktarını, N toplam gözlem sayısını ifade etmektedir.

Veri seti içinde bulunan gözlemlerin ortalamadan sapma miktarlarının toplamı sıfıra eşittir ($\sum (x - \mu)^2 = 0$ ve $\sum (x - \bar{x})^2 = 0$). Örneğin 4 öğrencinin Tarım Ekonomisi İstatistiği dersinden aldığı notlar 82, 95, 62 ve 92'dir. Bu 4 öğrencinin aldığı notların ortalaması 84'dür.

$$\bar{x} = (82 + 95 + 67 + 92) / 4 = 84$$

Bu ortalamadan sapmaların toplamı Çizelge 4.29'da verilmiştir.

Çizelge 4.29. Öğrencilerin Tarım Ekonomisi İstatistiği Dersinden Aldığı Notlar ve Ortalamadan Sapmalar

x	$(x - \bar{x})^2$
82	$82 - 84 = -2$
95	$95 - 84 = +11$
67	$67 - 84 = -17$
92	$92 - 84 = +8$
	$\sum (x - \bar{x})^2 = 0$

Sınıflandırılmamış verilerde Varyans ve standart sapmanın hesaplanmasında kullanım kolaylığı olan daha pratik aşağıdaki formüller kullanılmaktadır.

$$\sigma^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N} \quad \text{ve} \quad S^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n - 1}$$

$$\sigma = \sqrt{\sigma^2} \quad S = \sqrt{S^2}$$

Bu pratik formüller yardımıyla varyans ve standart sapma hesaplamasının 4 temel basamağı vardır. İlk yapılacak iş gözlem değerlerinin toplamını bulmaktır ($\sum x$). Daha sonra gözlemlerin karelerinin toplamı belirlenir ($\sum x^2$). Üçüncü adımda elde edilen verilerden yararlanarak varyans hesaplanır. Son olarak hesaplanan varyans değerinin karekökü alınarak standart sapma belirlenir. Şimdi Giresun ve Ordu illerinde fındık ihracatı yapan firmaların yıllık fındık ihracat miktarlarını dikkate alarak varyans ve standart sapmayı hesaplayalım (Çizelge 4.30).

Çizelge 4.30. Ordu ve Giresun İllerinde 10 Büyük Firmanın
Fındık İhracat Miktarları (1995)

Firma adı	Fındık ihracat miktarı (ton/iç) (x)	x^2
Yavuz Gıda Ltd. Şti.	7913	62615569
Anadolu AŞ.	5956	35473936
İnternut Ltd. Şti.	5052	25522704
Karayel AŞ	4856	23580736
Furtun Ltd. Şti.	3300	10890000
Durak AŞ.	3000	9000000
Gürsoylar AŞ	2722	7409284
Ornut Ltd. Şti	2700	7290000
Poyraz Ltd.Şti	2380	5664400
Poyraz AŞ.	2340	5475600
	$\sum x = 40219$	$\sum x^2 = 192922229$

Kaynak: Bozoğlu, M. 1999. Doktora Tezi Dökümanları

1. Gözlem değerlerinin toplamının bulunması ($\sum x$) bulunması

Çizelge 4.30’da verilen gözlemler toplamı 40219’dur.

2. Gözlem değerlerinin karelerinin toplamının ($\sum x^2$) bulunması

Çizelge 4.30’da 3. sütunda önce teker teker gözlemlerin karesi alınmış ve daha sonra bunların toplamı olan $\sum x^2 = 192922229$ bulunmuştur.

3. Varyansın hesaplanması

$$S^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{192922229 - \frac{(40219)^2}{10}}{9} = 3462825.88$$

4. Standart sapmanın hesaplanması

$$S = \sqrt{S^2} = \sqrt{3462825.88} = 1861 \text{ ton/iç}$$

1. Sınıflandırılmış verilerde varyans ve standart sapmanın hesaplanması

Sınıflandırılmış verilerde varyans ve standart sapma hesaplamasının iki yolu vardır. Birinci yolda sınıf orta noktalarının aritmetik ortalamadan sapmaları hesaplanır ve bunların kareleri alınır. İkinci yöntemde de aritmetik ortalamadan sapmaların karelerinin toplamı, indi olarak seçilen orijinden sapmaların kareleri toplamından düzeltme faktörünün çıkarılmasına eşittir. Düzeltme faktörü, aritmetik ortalama ile indi olarak seçilen orijinin farkına dayanmaktadır.

1.yol:

$$S^2 = \frac{\sum fx^2}{n-1}$$
$$S = \sqrt{S^2}$$

Formülde S^2 varyansı, S standart sapmayı, $\sum fx^2$ sınıf orta noktalarının aritmetik ortalamadan sapmalarının karelerinin frekans adedi ile çarpımlarının toplamını ve n gözlem sayısını ifade etmektedir. Bu yolla varyans ve standart sapmanın hesaplanması Çizelge 4.31’de verilmiştir.

Çizelge 4.31. Sınıflandırılmış Verilerde Varyans ve Standart Sapmanın Hesaplanması (I.Yol)

Sınıflar (dekar)	Frekans (f)	Sınıf orta noktası	Aritmetik ortalamadan (27.06 da) sapma (x)	Sapmanın karesi (x ²)	Çarpım (fx ²)	Hesaplama
1-11	18	6.0	-21.06	443.52	7983.36	$S^2 = \frac{\sum fx^2}{n-1}$ $S = \sqrt{S^2}$ $S^2 = \frac{33313.28}{93}$ $S = \sqrt{358.21}$ $S = 18.93$ dekar
12-22	32	17.0	-10.06	101.20	3238.40	
23-33	16	28.0	0.94	0.88	14.08	
34- 44	12	39.0	11.94	142.56	1710.72	
45-55	6	50.0	22.94	526.24	3157.44	
56-66	6	61.0	33.94	1151.92	6911.52	
67-77	2	72.0	44.94	2019.60	4039.20	
78-88	2	83.0	55.94	3129.28	6258.56	
Toplam	94			7515.23	33313.28	

2. Yol:

$$S = \sqrt{\frac{\sum f_i d_i^2}{n} - \left(\frac{\sum (f_i d_i)}{n}\right)^2 . c}$$
$$S = \sqrt{S^2}$$

Formülde S^2 varyansı, S standart sapmayı, $\sum f_i d_i^2$ sınıf genişliğindeki indi orijinden sapmalarının karelerinin frekans adedi ile çarpımlarının toplamını, $\sum (f_i d_i)$ sınıf genişliğindeki sapmaların frekans sayısı ile çarpımlarının toplamını, c sınıf genişliğini ve n gözlem sayısını ifade etmektedir. Bu yolla varyans ve standart sapmanın hesaplanması Çizelge 4.32’de verilmiştir.

Çizelge 4.32. Sınıflandırılmış Verilerde Varyans ve Standart Sapmanın Hesaplanması (II.Yol)

Sınıflar (dekar)	Freka ns (f)	Sınıf orta noktası	İndi nokta dan sapma (d_i)	d_i^2	$f_i d_i$	$f_i d_i^2$	Hesaplama
1-11	18	6.0	-2	4	-36	72	$S = \sqrt{\frac{\sum f_i d_i^2}{n} - \left(\frac{\sum (f_i d_i)}{n}\right)^2 . c}$ $S = \sqrt{S^2}$ $S = 13.87$
12-22	32	17.0	-1	1	-32	32	
23-33	16	28.0 (A)	0	0	0	0	
34- 44	12	39.0	1	1	12	12	
45-55	6	50.0	2	4	12	24	
56-66	6	61.0	3	9	18	54	
67-77	2	72.0	4	16	8	32	
78-88	2	83.0	5	25	10	50	
Toplam	94				-8	276	

4.2.3. Değişim katsayısı

Dağılım ölçüleri içerisinde en yaygın kullanılan standart sapmadır ancak standart sapma birimleri farklı olan serilerde kullanılamamakta ve nispi değişimi ortaya koyamamaktadır. Bu sebeple farklı birimlere sahip verileri karşılaştırabilmek ve nispi değişimi ortaya koyabilmek amacıyla “değişim katsayısı” kullanılmaktadır. Değişim katsayısı, standart sapmanın aritmetik ortalamasının yüzdesi olarak ifade edilmiş halidir ve aşağıdaki formüller yardımıyla hesaplanmaktadır.

$$DK = \frac{\sigma}{\mu} \cdot 100 \quad (\text{Populasyon için değişim katsayısı})$$

$$DK = \frac{s}{\bar{x}} \cdot 100 \quad (\text{Örnek için değişim katsayısı})$$

Değişim katsayısı tarım ekonomisi alanında yaygın olarak kullanılmaktadır. Çoğu zaman tarım işletmelerinin ekonomik yapısı incelenirken tarım işletmelerinin arazisinde meydana gelen değişim ile elde edilen üretim miktarında meydana gelen değişim karşılaştırılmaktadır. Bu gibi durumlarda birimler farklı olduğu için standart sapma kullanılamadığından, değişim katsayısı daha kullanışlı olmaktadır. Değişim katsayısının herhangi bir birimi olmadığına ve yüzde olarak ifade edildiğine dikkat edilmelidir.

Örneğin bir şirkette çalışan bütün işçilerin ortalama haftalık geliri 42.35 milyon TL ve bu gelirin standart sapması 3.82 milyon TL’dir. Aynı işçilerin eğitim gördüğü yıl sayısı ise ortalama olarak 15 yıl ve standart sapması 2 yıldır. Acaba bu işçilerin gelirlerindeki değişim mi, yoksa eğitim gördükleri yıl sayısındaki değişim mi daha fazladır?

$$\text{Gelir için değişim katsayısı} \quad DK = \frac{\sigma}{\mu} \cdot 100 = \frac{3.82}{42.35} \cdot 100 = \%9.02$$

$$\text{Eğitim yılı için değişim katsayısı} \quad DK = \frac{\sigma}{\mu} \cdot 100 = \frac{2}{15} \cdot 100 = \%13.33$$

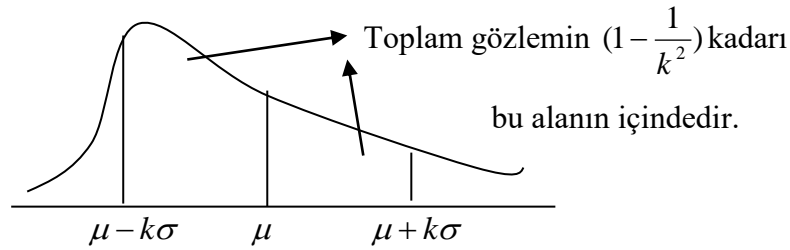
Bu sonuçlara dayanarak eğitim görülen yıl sayısındaki değişimin, gelirdeki değişimden fazla olduğunu söylemek mümkündür.

4.2.4. Standart sapmanın kullanışı

Aritmetik ortalama ve standart sapmadan yararlanarak, ortalama etrafında belirlenen bir aralıkta yer alan gözlemlerin yüzdesini bulmak mümkündür. Bahsedilen iki ölçüden yararlanarak gözlemlerin ortalama etrafında belirli bir aralıkta yer alan gözlemlerin yüzdesi bulunurken “Chebyhsev teoremi” ve “deneysel yaklaşım (empirical rule)” dan yararlanılmaktadır.

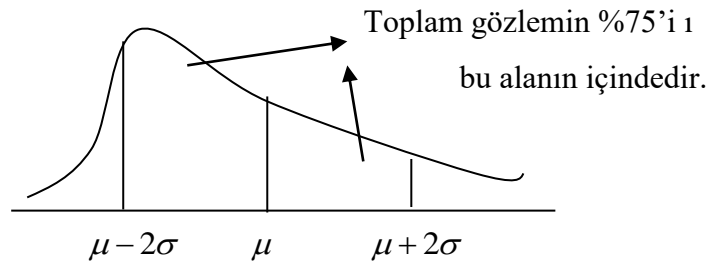
4.2.4.1. Chebyhsev teoremi

Bu teoremden yararlanarak ortalamanın her iki tarafında ortalamaya aynı uzaklıkta bulunan iki nokta arasında kalan eğrinin alanının alt sınırını belirlenmektedir. Bu teoremin en önemli özelliği veri setinde bulunan gözlemlerin dağılımı şekilde ne olursa olsun kullanılabilmesidir. Chebyhsev teoremine göre 1 den büyük her bir “k” sayısı için, ortalamanın k standart sapma kadar altında ve üstünde yer alan iki nokta arasında kalan eğrinin oluşturduğu alan içinde en azından $(1 - \frac{1}{k^2})$ kadar gözlem bulunmaktadır (Şekil 4.2).



Şekil 4.2. Chebyhsev teoremi

Örneğin $k=2$ ise, Chebyhsev teoremine göre toplam gözlemlerin %75'i $[(1 - \frac{1}{k^2}) = 1 - \frac{1}{4} = 1 - 0.25 = 0.75]$ ortalamanın iki standart sapma altında ve üstünde yer alan iki noktanın oluşturduğu alan içerisinde yer almaktadır (Şekil 4.3).



Şekil 4.3. $k=2$ için Chebyhsev teoreminin uygulaması

Bu teoremin $k > 1$ için geçerli olduğu unutulmamalıdır. Çünkü, $k = 1$ olduğunda $(1 - \frac{1}{k^2})$ ın değeri sıfır “0” olmaktadır. Ayrıca $k < 1$ olduğu durumlarda ise $(1 - \frac{1}{k^2})$ negatif değer almaktadır.

Örnek problem:

Ceyhan (2001) tarafından Çorum ili Kızılırmak Havzası tarım işletmelerinde yapılan araştırma sonucunda tarım işletmelerinin tarımsal gelirinin ortalama 1.5 milyar TL ve buna ait standart sapmanın 350 milyon TL olduğu tespit edilmiştir. Acaba 0.45 milyar TL ile 2.55 milyar TL arasında tarımsal gelire sahip olan işletmelerin yüzdesi kaçtır?

Çözüm:

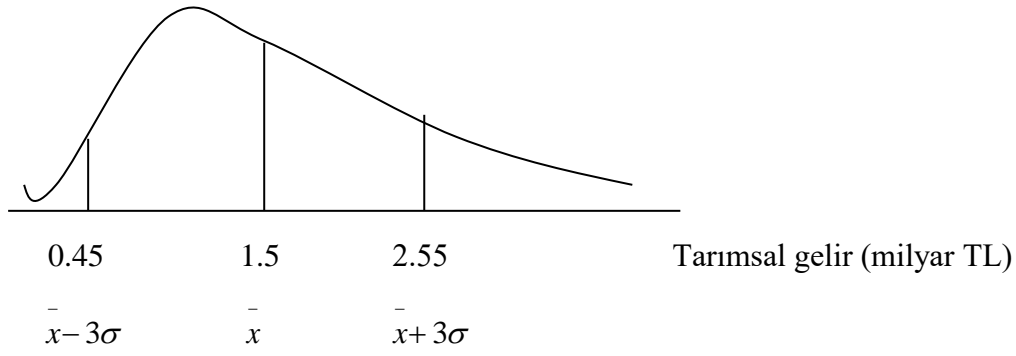
$$\begin{aligned}\bar{x} &= 1.5 \\ S &= 0.35\end{aligned}$$

Burada yapılacak ilk işlem “k” sayısını belirlemektir. Bunun için öncelikle araştırılan alt (0.45 milyar TL) ve üst sınırların (2.55 milyar TL) ortalamadan ne kadar uzakta olduğunu tespit etmek gerekmektedir. Her iki rakam ortalamadan 1050 milyon TL uzaklıkta yer almaktadır. “k” değeri ortalamadan uzaklığın, standart sapma değerine oranlanması ile bulunmaktadır. Örneğimiz için bu $k = 1050 / 350 = 3$ ’dür.

Bundan sonra yapılacak iş Chebyhsev teoreminden yararlanarak belirlenen aralıktaki işletmelerin yüzdesini hesaplamaktır.

$$1 - \frac{1}{k^2} = 1 - \frac{1}{9} = 0.89$$

Yani Çorum İli Kızılırmak havzasında yer alan tarım işletmelerinin %89’u 450 milyon TL ile 2.55 milyar TL arasında tarımsal gelire sahiptir.

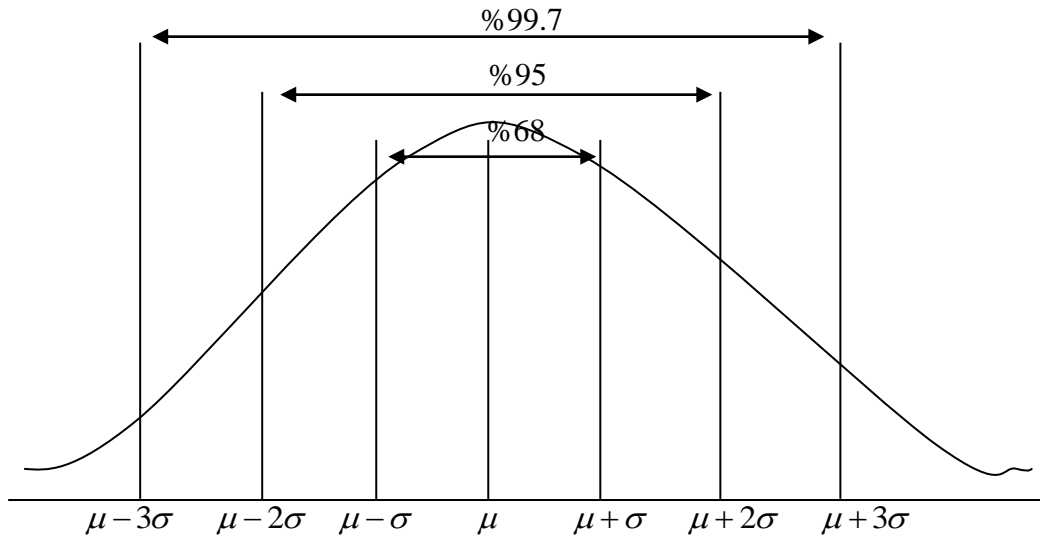


Şekil 4.4. Çorum ili Kızılırmak havzasında tarımsal geliri 0.45-2.55 milyar TL arasında olan işletmelerin yüzdesi

4.2.4.2. Deneysel yaklaşım

Chebyshev teoremi veri setinde bulunan gözlemlerin dağılımı ne olursa olsun kullanılabiliyorken, “deneysel yaklaşım” sadece dağılımı çanı andıran yani normal dağılım gösteren gözlemlerde uygulanabilmektedir. Deneysel yaklaşıma göre normal dağılımda

1. Gözlemlerin %68'i ortalamanın bir standart sapma sağ ve solunda kalan iki nokta arasında kalan alanda yer almaktadır.
2. Gözlemlerin %95'i ortalamanın iki standart sapma sağ ve solunda kalan iki nokta arasında kalan alanda yer almaktadır.
3. Gözlemlerin %99.7'si ortalamanın üç standart sapma sağ ve solunda kalan iki nokta arasında kalan alanda yer almaktadır.



Şekil 4.5. Deneysel yaklaşımın şekille gösterilmesi

Örnek problem:

Ceyhan (2001) tarafından Çorum ili Kızılırmak Havzası tarım işletmelerinde yapılan araştırma sonucunda tarım işletmelerinin tarımsal gelirinin ortalama 1.5 milyar TL ve buna ait standart sapmanın 350 milyon TL olduğu tespit edilmiştir. Verilerin normal dağılım gösterdiğini varsayarsak acaba 0.45 milyar TL ile 2.55 milyar TL arasında tarımsal gelire sahip olan işletmelerin yüzdesi kaçtır?

Çözüm:

$$\bar{x} = 1.5$$

$$S = 0.35$$

Burada yapılacak ilk işlem araştırılan alt (0.45 milyar TL) ve üst sınırların (2.55 milyar TL) ortalamadan ne kadar uzakta olduğunu tespit etmektir. Her iki rakam ortalamadan 1050 milyon TL yani iki standart sapma uzaklıkta yer almaktadır. Bundan sonra yapılacak iş deneysel yaklaşımdan yararlanarak belirlenen aralıktaki işletmelerin yüzdesini hesaplamaktır ki bu %95'dir. Yani Çorum İli Kızılırmak havzasında yer alan tarım işletmelerinin %95'i 450 milyon TL ile 2.55 milyar TL arasında tarımsal gelire sahiptir.

4.3. Pozisyon Ölçüleri

İstatistik analizlerde çok sık kullanılan tanımlayıcı istatistikten birisidir. Herhangi bir gözlemin dağılım içerisindeki durumunu ortaya koyan ölçümlerdir. En yaygın kullanılan pozisyon ölçüleri “çeyrek değişim ölçüleri ve çeyrek değişim ölçüleri arası” ile “yüzdelikler” dir.

4.3.1. Çeyrek değişim ölçüleri

Veri setinde yer alan gözlemlerin oluşturduğu dağılımın belirli noktaları arasındaki gözlem değerlerinin farkını göstermektedir. Çeyrek değişim ölçüleri veri setinin çeyrek mesafelerinde yer alan gözlemleri ifade etmektedir. Birinci çeyrek değişim ölçüsü (Q_1) en küçük gözlemden itibaren, en büyük gözleme doğru $\frac{1}{4}$ mesafede yer alan gözlemin değeridir. Üçüncü çeyrek (Q_3) veri setinin en küçük gözleminden itibaren en büyük gözlemine doğru $\frac{3}{4}$ mesafede yer alan gözlemdir. İkinci çeyrek değişim ölçüsü (Q_2) ortanca değer (medyan) ile

aynı olduğundan sadece ortanca değer olarak anılmaktadır ve veri setinin tam ortasında bulunan gözlemdir. Çeyrek değişim ölçüleri aynı medyan gibi belirlenen bir değerdir, hesaplanan bir değer değildir. Her bir çeyrek değişim ölçüsü veri setinde bulunan gözlemlerin %25'ine sahiptir. Aşağıda çeyrek değişim ölçüleri gösterilmiştir.

%25	%25	%25	%25
-----	-----	-----	-----

Q₁

Q₂

Q₃

Bazen çeyrek değişim ölçüleri arasındaki fark bizi ilgilendirir. Bu sebeple “çeyrek değişim ölçüleri arası” adı verilen bir ölçü kullanılır. Bu ilk çeyrek değişim ölçüsü ile üçüncü çeyrek değişim ölçüsü arasındaki farkı göstermektedir ($QD = Q_3 - Q_1$). Çeyrek değişim ölçüleri arası gözlemlerin medyan etrafında toplanma özelliğini gösterir. Çeyrek değişim ölçüleri arası, bütün değişimin yarısı kadar büyüklükte olduğu zaman, gözlemlerin merkezi bir nokta etrafında toplanma eğiliminde olmadığı anlaşılmaktadır. Bu değer küçüldükçe merkezi bir nokta etrafında toplanma söz konusudur.

Çeyrek değişim ölçüleri sınıflandırılmış ve sınıflandırılmamış verilerde farklı şekillerde hesaplanmaktadır.

1. Sınıflandırılmamış verilerde çeyrek değişim ölçülerinin hesaplanması

Sınıflandırılmamış verilerde yapılacak ilk iş veri setinde bulunan gözlemleri küçükten büyüğe doğru sıralamaktır. Daha sonra veri setinin $\frac{1}{4}$ 'ünü aşağıda ve $\frac{3}{4}$ 'ünü yukarıda bırakan gözlem sayısı tek olduğunda $\frac{N}{4} + \frac{1}{2}$ inci gözlem, gözlem sayısı çift olduğunda $\frac{N}{4}$ üncü gözlem birinci çeyrek değişim ölçüsü olarak tespit edilir. Gözlemlerin $\frac{3}{4}$ 'ünü aşağıda $\frac{1}{4}$ 'ünü yukarıda bırakan gözlem sayısı tek olduğunda $\frac{3N}{4} + \frac{1}{2}$ inci gözlem, gözlem sayısı çift olduğunda $\frac{3N}{4}$ üncü gözlem ise üçüncü çeyrek değişim ölçüsüdür.

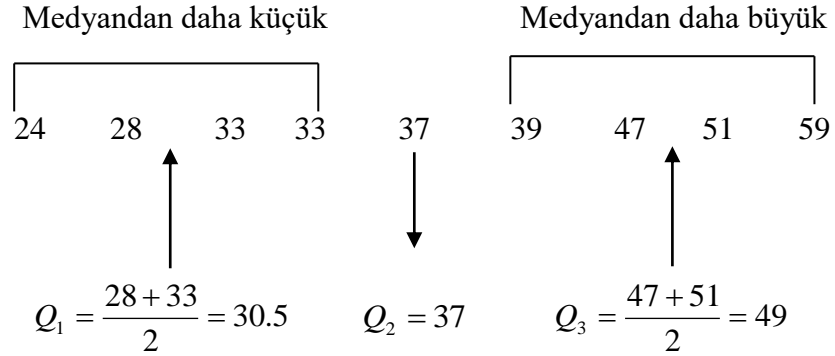
Örnek problem:

Aşağıda bir firmada çalışan 9 kişinin yaşları verilmiştir.

47 28 39 51 33 37 59 24 33

- a) Çeyrek değişim ölçülerini bulunuz. 28 yaş bu yaş dağılımında nerede yer almaktadır (pozisyonu nedir)?
- b) Çeyrek değişim ölçüleri arasını bulunuz?

Çözüm:



28 yaş dağılımında ilk %25'lik dilimde yer alıyor.

b) Çeyrek değişim ölçüleri arası $QD = Q_3 - Q_1 = 49 - 30.5 = 18.5$

Çeyrek değişim ölçüleri arası toplan değişimin $(59-24=34)$ yarısından fazla olduğundan $(18.5>17)$ gözlemlerin medyan etrafında toplanmadığını söyleyebiliriz.

1. Sınıflandırılmış verilerde çeyrek değişim ölçülerinin hesaplanması

Sınıflandırılmamış verilerde çeyrek değişim ölçülerinin belirlenmesinde aşağıdaki eşitlikler kullanılmaktadır:

$$Q_1 = L_{-i} + \left(\frac{\frac{N}{4} - f_{-i}}{f_0} \right) c$$

$$Q_3 = L_{-i} + \left(\frac{\frac{3N}{4} - f_{-i}}{f_0} \right) c$$

Eşitliklerde Q_1 birinci çeyrek değişim ölçüsünü, Q_3 üçüncü çeyrek değişim ölçüsünü, L_{-i} birinci veya üçüncü çeyrek değişim ölçüsünü kapsayan sınıfın alt sınırını, f_{-i} çeyrek değişim ölçüsünün bulunduğu sınıftan yukarıda (küçük olan) olan sınıfların frekanslarının toplamı, f_0 çeyrek değişim ölçülerinin bulunduğu sınıfın frekansını ve c sınıf genişliğini ifade etmektedir

Çorum İli Kızılırmak Havzası tarım işletmelerinin mülk arazi büyüklüklerine ait frekans tablosunu kullanarak çeyrek değişim ölçülerinin bulunması Çizelge 4.33’de gösterilmiştir. Buna göre incelenen 94 tarım işletmesinin %25’i 14 dekarдан daha az mülk araziye sahip iken, %75’i 38 dekarдан daha az mülk araziye sahiptir. İşletmelerin ancak %25’i 38 dekarдан daha fazla mülk araziye sahiptir.

Çizelge 4.32. Sınıflandırılmış Verilerde Çeyrek Değişim Ölçülerinin Bulunması

Sınıflar (dekar)	Frekans (f)	Kümülatif frekans	Q_1	Q_3
1-11	18	18		
12-22	32	50		
23-33	16	66	$Q_1 = L_{-i} + \left(\frac{\frac{N}{4} - f_{-i}}{f_0}\right)c$	$Q_3 = L_{-i} + \left(\frac{\frac{3N}{4} - f_{-i}}{f_0}\right)c$
34- 44	12	78		
45-55	6	84	$Q_1 = 12 + \left(\frac{\frac{94}{4} - 18}{32}\right)11$	$Q_3 = 34 + \left(\frac{\frac{282}{4} - 66}{12}\right)11$
56-66	6	90	$Q_1 = 13.89$	$Q_3 = 38.13$
67-77	2	92		
78-88	2	94		
Toplam	94			

4.3.2. Yüzdelikler ve yüzdelikler arası mesafe

Yüzdelikler, veri setinde bulunan gözlemleri yüz eşit parçaya bölerek gözlemlerin pozisyonunu belirleme de kullanılan bir ölçüdür. Bu ölçümün kullanılabilmesi için öncelikle veri setinde bulunan gözlemlerin küçükten büyüğe doğru sıraya dizilmesi gerekmektedir. Daha sonra belirli bir gözlemin hangi yüzdelik dilimde yer aldığı tespit edilmektedir. Belirli bir yüzdelik dilimde yer alan gözlemin değerini belirlemede aşağıdaki formül kullanılmaktadır:

$$P_k = \frac{kn}{100}$$

Formülde P_k k ıncı yüzdelik dilimde yer alan gözlemin değerini, k yüzdelik dilimi, n gözlem sayısını ifade etmektedir.

Örnek:

Aşağıda 12 kişinin haftalık gelirlerini gösteren veri seti verilmiştir. 62. yüzdelik dilimde yer alan gözlemin değeri nedir?

16 38 18 20 20 18 22 34 7 58 31 19

Çözüm:

Önce mevcut gözlemler büyüklüğüne göre sıraya dizilir.

7 16 18 18 19 20 20 22 31 34 38 58

Daha sonra 62. yüzdilik dilim pozisyonu aşağıdaki gibi hesaplanır.

$$\frac{kn}{100} = \frac{62(12)}{100} = 7.44 . \text{ Gözlem}$$

7.44. gözlemin değeri 7. gözlemin değeri ile 8. gözlemin değerinin ortalaması alınarak bulunmaktadır.

$$P_{62} = \frac{20 + 22}{2} = 21$$

Buna dayanarak çalışanların %62'sinin haftalık gelirinin 21 milyon TL'den daha az, %38'inin ise 21 milyon TL'den daha fazla gelire sahip olduğunu söylemek mümkündür.

Burada örnek olması açısından az sayıda gözlem kullanılmıştır. Dolayısıyla hesaplamaların gerçeği yansıtması düşüktür, yani hata payı içermektedir. Bu sebeple bu ölçümün kullanılması için veri setinin sahip olduğu gözlem sayısı gerekmektedir.

Yüzdelikler arası mesafe ise belirli bir gözlem değerinden daha küçük gözlemlerin sayısının, toplam gözlem sayısına oranını ifade etmektedir ve aşağıdaki formül yardımıyla hesaplanmaktadır.

$$\frac{\sum f_{-i}}{\sum f} . 100$$

Formülde $\sum f_{-i}$ belirli bir gözlem değerinden daha küçük değere sahip gözlemlerin sayısını ve $\sum f$ ise toplam gözlem sayısını ifade etmektedir.

Yüzde dilimlerde örnek olarak kullanılan haftalık gelir serisinde acaba 31 milyon TL'den daha az gelire sahip kişilerin oranı kaçtır?. Bu soruya cevap verebilmek için yüzdelikler arası mesafe kavramı kullanılmaktadır. Cevap $P_{32} = \frac{8}{12} . 100 = \%66.67$ olacaktır.