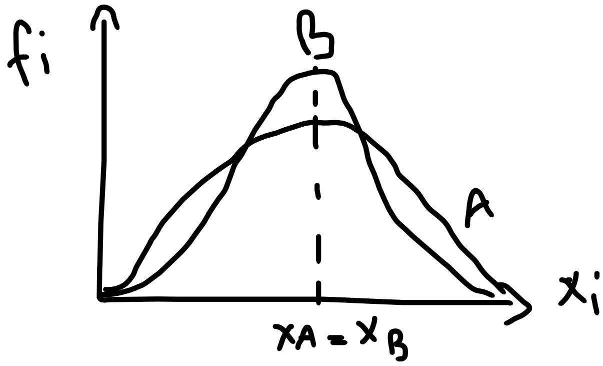
4. DAĞILIŞ ÖLÇÜLERİ

Merkezi eğilim ölçüleri yığının veya yığından elde edilen örneklemin birim değerlerinin etrafında toplandığı birimin değerini belirleyen ölçülerdi. Yayılım\Dağılım\Dağılış ölçüleri ise gözlem değerlerinin birbirlerine göre konumlarını, birbirine yakınlık ve uzaklıklarını yansıtan değerlerdir. Örneğin aynı yaş ortalamasına sahip A ve B gibi iki farklı yığına ait frekans eğrileri aşağıdaki şekilde gösterilsin:



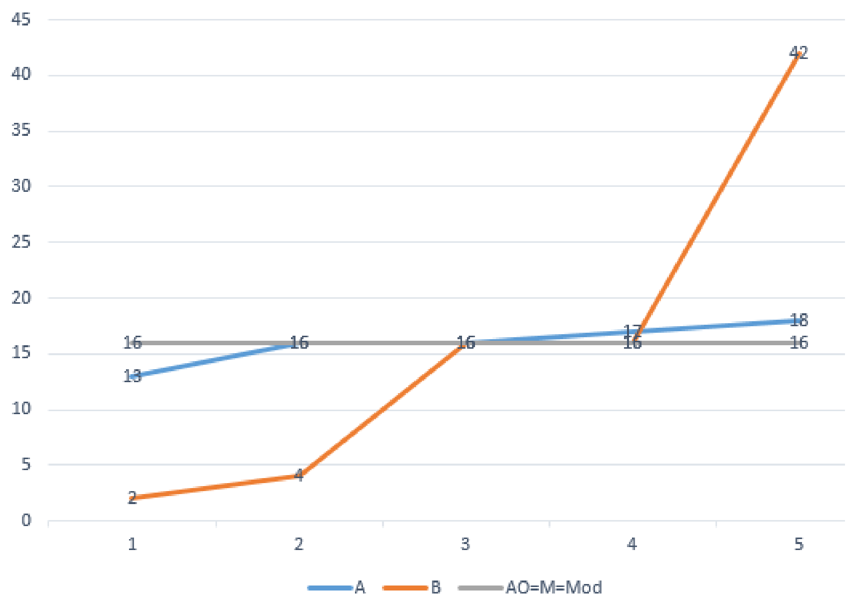
Görüldüğü gibi iki yığının ortalaması eşit olmasına karşın verilerin dağılımı farklıdır. Yayılım\dağılış ölçüleri verilen ortalamadan ve birbirinden ne kadar uzaklaştığının ölçüsüdür.

Bu bölümde Değişim aralığı, Varyans ve Standart sapma, Değişim katsayısı, çeyreklikler anlatılacaktır. Dağılış ölçülerinin önemini daha iyi anlamak için aşağıdaki örneği inceleyelim.

Örnek: A ve B örneklemleri için AO, M ve Mod değerlerini hesaplayınız ve bu veri setlerini karşılaştırınız.

A	13	16	16	17	18
B	2	4	16	16	42

$$\bar{x}_A = M_A = Mod_A = 16$$
$$\bar{x}_B = M_B = Mod_B = 16$$



4.1 Değişim aralığı (Açıklık): Bir veri setinde en büyük (EB) değer ile en küçük değer (EK) arasındaki farktır. Verinin yayılımını gösterir.

$$Değişim\ Aralığı\ (DA) = EB - EK$$

Örnek: $DA_A = 18 - 13 = 5$ ve $DA_B = 42 - 2 = 40$ olup şu halde B veri setinde A veri setine göre yayılıma daha fazladır denir. Yani B de değişkenlik daha fazladır.

Örnek: 100 öğrencinin ağırlığına ilişkin aşağıdaki tablo ile verilen dağılımın değişim aralığını hesaplayalım:

Sınıflar	Frekans
45-50	10
50-55	18
55-60	20
60-65	18
65-70	20
70-75	14

En yüksek değer 75, en küçük değer 45 olduğuna göre DA=75-45=30.

4.2 Varyans ve Standart Sapma: Veri setindeki gözlem değerleri ile aritmetik ortalamanın farklarının karelerinin ortalamasına varyans denir ve varyansın karekökü de standart sapma olarak adlandırılır.

Kitle için varyans: $\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$ olup Standart sapma: $\sigma = \sqrt{\sigma^2}$

Örneklem için varyans: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$ olup Standart sapma: $s = \sqrt{s^2}$

Örnek:

$$s_A^2 = \frac{(13-16)^2 + (16-16)^2 + (16-16)^2 + (17-16)^2 + (18-16)^2}{5-1} = 3.5$$

$$s_B^2 = \frac{(2-16)^2 + (4-16)^2 + (16-16)^2 + (16-16)^2 + (42-16)^2}{5-1} = 254$$

Varyansı büyük olan veri setinde değişkenlik daha fazladır. Bu pek istenilmeyen bir durumdur. Varyansın büyük olması verinin heterojen yapıya sahip olması ve tahminlerde daha fazla sapmalar olması anlamına gelir.

- Eğer veri seti frekans tablosuna sahip ise veri setinin varyansı

$$s^2 = \frac{\sum(x - \bar{x})^2 * f}{n-1}$$

şeklinde hesaplanır.

Örnek: Aşağıda yaşlara ilişkin verilen örneklemin standart sapmasını bulunuz.

Yaş	Kişi sayısı
20	4
22	3
24	2
25	5

$$\bar{x} = \frac{\sum(x * f)}{\sum f} = \frac{319}{14} = 22.79$$

$$s^2 = \frac{\sum(x - \bar{x})^2 * f}{n - 1} = \frac{(20 - 22.79)^2 * 4 + \cdots + (25 - 22.79)^2 * 5}{14 - 1} = 4.64$$

$$s = \sqrt{4.64} = 2.15$$

- Eğer veri seti sınıflandırılmış frekans tablosuna sahip ise y_j sınıf orta değeri olmak üzere

$$s^2 = \frac{\sum(y - \bar{x})^2 * f}{n - 1}$$

şeklinde hesaplanır.

Örnek: Rasgele seçilen 100 öğrencinin istatistik dersinden aldıkları notlara ilişkin sınıflandırılmış frekans tablosuna göre sınıf için ortalama notu ve standart sapmayı hesaplayınız.

Not	**f**		**y**	$\boldsymbol{y * f}$	$(\boldsymbol{y - \bar{x}})$	$(\boldsymbol{y - \bar{x}})^2$	$(\boldsymbol{y - \bar{x}})^2 * \boldsymbol{f}$
0-20	10		10	10*10=100	-39	1521	15210
20-40	15		30	30*15=450	-19	361	5415
40-60	50		50	50*50=2500	1	1	50
60-80	20		70	70*20=1400	21	441	8820
80-100	5		90	90*5=450	41	1681	8405
Toplam	**100**			**4900**			**37900**

$\bar{x} = \frac{\sum(y*f)}{\sum f} = \frac{4900}{100} = 49$ $\quad s^2 = \frac{\sum(y-\bar{x})^2*f}{n-1} = \frac{37900}{100-1} = 382.83$ $\quad s = \sqrt{382.83} = 19.57$

4.3. Değişim Katsayısı (DK): Gözlemlerin birimlerinden arındırılmış bir dağılış ölçüsüdür. Birimleri farklı verilerin karşılaştırılmasında sıklıkla kullanılır. DK küçük olan veri seti değişkenlik bakımından daha homojen olduğu söylenebilir ve bu tercih edilen bir durumdur.

Kitle için: $DK = \frac{\sigma}{\mu}$ olup örneklem için: $DK = \frac{s}{\bar{x}}$

Örnek: $DK_A = \frac{1.87}{16} = 0.12$ olup $DK_B = \frac{15.94}{16} = 0.99$ olarak bulunur. $DK_A < DK_B$ olduğundan A örneklemi B ye göre daha homojen olduğu söylenebilir.

Örnek: Bir grup öğrencinin dönemde devam ettiği derslerden almış olduğu ortalama puanlar ile aylık harcamaları aşağıdaki tablolar ile verilmiştir. Her iki değişkenin merkezi dağılım ölçülerini hesaplayıp yorumlayınız.

Puan Sınıfları	Frekans
$0 \leq x < 20$	5
$20 \leq x < 40$	32
$40 \leq x < 60$	54
$60 \leq x < 80$	5
$80 \leq x \leq 100$	4

Harcama Sınıfları(1000TL)	Frekans
$2 \leq x < 3$	7
$3 \leq x < 4$	30
$4 \leq x < 5$	50
$5 \leq x < 6$	10
$6 \leq x \leq 7$	3

Puan Sınıfları	Frekans	Sınıf orta değeri(y_j)	$y_j f_j$	$(y_j - \bar{x})^2 f_j$
$0 \leq x < 20$	5	10	50	5780
$20 \leq x < 40$	32	30	960	6272
$40 \leq x < 60$	54	50	2700	1944
$60 \leq x < 80$	5	70	350	3380
$80 \leq x \leq 100$	4	90	360	8464

$$\bar{x} = \frac{\sum(y * f)}{\sum f} = 44,2, s^2 = \frac{\sum(y - \bar{x})^2 * f}{n-1} = 261,01, s = 16.15$$

$$DK_1 = \frac{s}{\bar{x}} = \frac{16.15}{44,2} = 0,36$$

Harcama Sınıfları(1000TL)	Frekans	Sınıf orta değeri(y_j)	$y_j f_j$	$(y_j - \bar{x})^2 f_j$
$2 \leq x < 3$	7	2,5	17,5	202,3
$3 \leq x < 4$	30	3,5	105	147
$4 \leq x < 5$	50	4,5	225	45
$5 \leq x < 6$	10	5,5	55	169
$6 \leq x \leq 7$	3	6,5	19,5	158,7

$$\bar{x} = \frac{\sum(y*f)}{\sum f} = 4{,}22, s^2 = \frac{\sum(y-\bar{x})^2 * f}{n-1} = 7{,}29\ , s = 2{,}7$$

$$DK_2 = \frac{s}{\bar{x}} = \frac{2{,}7}{4{,}22} = 0{,}64$$

$DK_2 > DK_1$ olduğundan birinci örneklem daha homojendir.

4.4. Çeyreklik (Kartil) Aralığı (KA): Veri setinin bir yayılma ölçüsü olup aykırı değerlerin tespit edilmesinde kullanılır.

$KA = Q_3 - Q_1$ ile hesaplanır. $(Q_1 - 1.5 * KA)$ ve $(Q_3 + 1.5 * KA)$ sınırlarının dışına düşen gözlem aykırı değer olarak değerlendirilir.

Örnek: 11 gözleme sahip veri setinin çeyreklikleri $Q_1 = 101$,$Q_2 = M = 108$ ve $Q_3 = 120$ ise $KA = Q_3 - Q_1 = 120 - 101 = 19$ olup aykırı değer sınırları $(101 - 1.5 * 19) = 72.5$ ile $(120 + 1.5 * 19) = 148.5$'tir. Yani 72.5'ten az ve 148.5'ten fazla değerler aykırı değerler olarak değerlendirilir.

5. GRAFİKLER

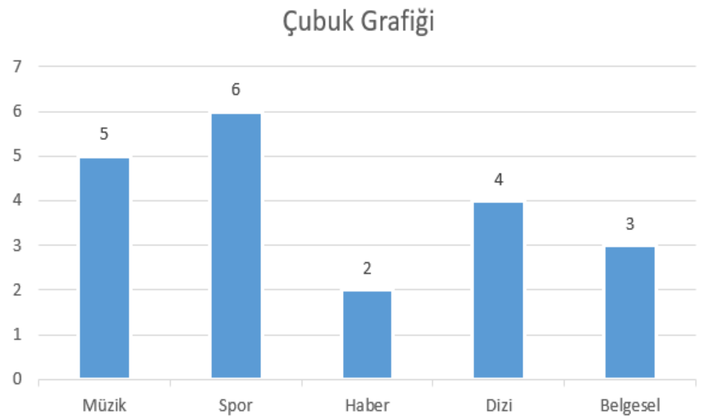
Veriyi özetlemenin görsel biçimidir. Genellikle kategorik değişkenler için daire ve çubuk grafiği, sayısal değişkenler için histogram, poligon ve kutu grafiği kullanılır.

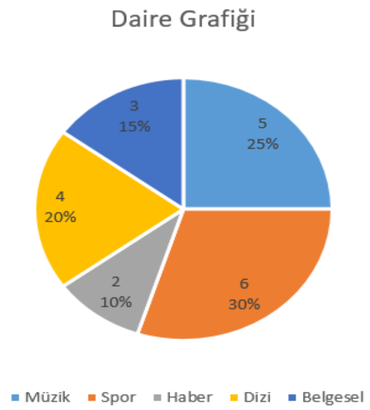
5.1. Daire (Pasta) Grafiği: Kategorik değişkenlerin frekans dağılımlarının grafiksel gösterimi için sıklıkla kullanılır. n gözleme sahip veriler için f_j: j inci sınıf frekansı, $\frac{f_j}{n} * 100$: j inci sınıfın dairedeki yüzdelik alan, $\frac{f_j}{n} * 360$: j inci sınıfın dairedeki açı değeri olarak alınır.

5.2. Çubuk (Sütun) grafiği: Bu grafikte yatay eksende değişkenin kategorileri, dikey eksende ise bunların frekanslarının gösterildiği eşit genişliklere sahip çubuklardan oluşan grafiklerdir.

Örnek: Bir sınıftaki 20 öğrencinin en çok izlediği TV programına ait frekans tablosuna göre daire ve çubuk grafiğini oluşturunuz.

TV programı	Frekans	Yüzdesi	Açısı
Müzik	5	$\frac{5}{20} * 100 = 25$	$\frac{5}{20} * 360 = 90$
Spor	6	$\frac{6}{20} * 100 = 30$	$\frac{6}{20} * 360 = 108$
Haber	2	$\frac{1}{20} * 100 = 10$	$\frac{2}{20} * 360 = 36$
Dizi	4	$\frac{4}{20} * 100 = 20$	$\frac{4}{20} * 360 = 72$
Belgesel	3	$\frac{3}{20} * 100 = 15$	$\frac{3}{20} * 360 = 54$
Toplam	20	100	360

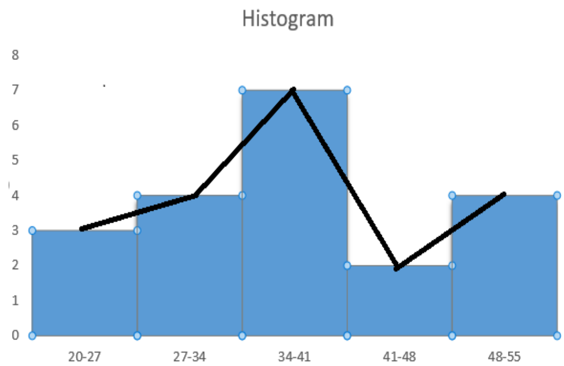




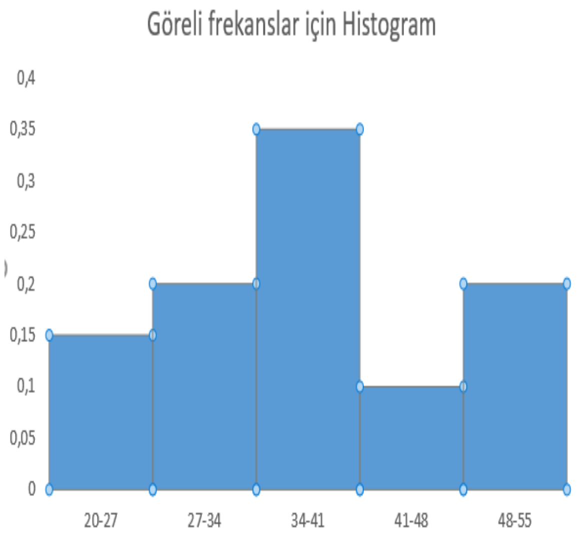
5.3. Histogram ve Poligon: Histogram bir dikdörtgenler dizisidir. Bu dikdörtgenlerin tabanları sınıflandırılmış frekans tablosundaki her bir sınıfın genişliğini, yükseklikleri ise sınıfın frekansını gösterir. Bu dikdörtgenlerin üst kenarlarının orta noktaları birleştirilmek suretiyle elde edilen grafiğe *frekans poligonu (diyagram)* denir. Göreli ya da birikimli frekans dağılımlarına ait histogramları elde etmek için bu histogramdaki dikey eksene göreli ya da veya birikimli frekanslar yerleştirilmelidir. Birikimli frekans poligonlarına *ojiv* eğrileri de denir.

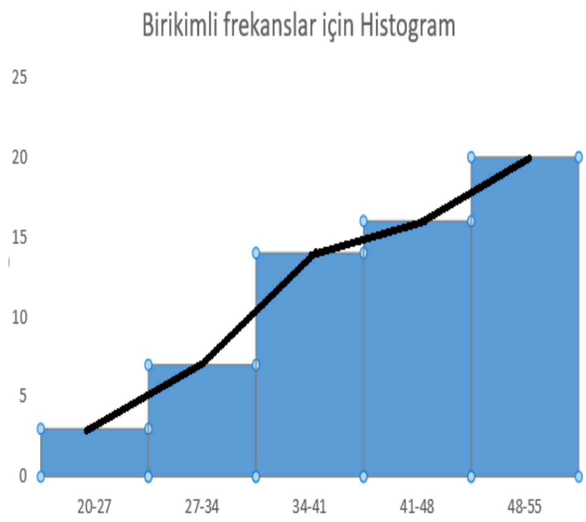
Örnek: 20 kişinin yaşlarına ilişkin sınıflandırılmış frekans tablosuna göre histogram ve poligon grafiklerini çiziniz.

Yaş	Frekans	Birikimli frekans	Göreli frekans
20-27	3	3	$^3/_{20} = 0.15$
27-34	4	7	$^4/_{20} = 0.20$
34-41	7	14	$^7/_{20} = 0.35$
41-48	2	16	$^2/_{20} = 0.10$

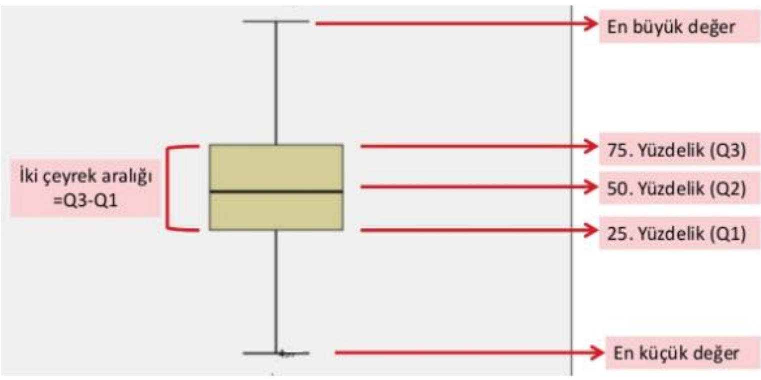


48-55	4	20	$^4/_{20} = 0.20$





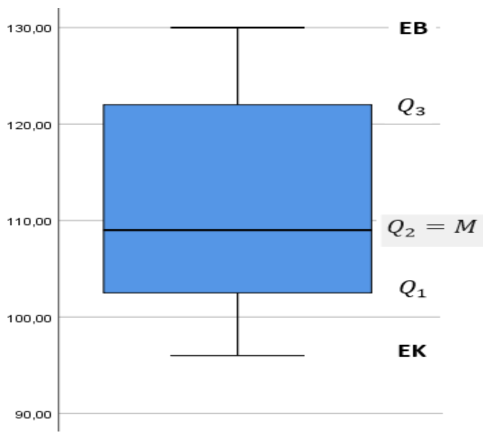
5.4. Kutu Grafiği: Bir veri setinin merkez, yayılım, simetri ve aykırı değer gibi birçok özelliği hakkında bilgi veren kutu şeklindeki gösterimlerdir. Kutunun içerisinde yer alan medyan çizgisi verinin konumu hakkında bilgi verirken Kartil (Çeyreklik) aralığı verinin yayılımı hakkında bilgi verir.



Örnek: 12 gözleme sahip veri setinin kutu grafiğini çiziniz.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
96	97	101	104	105	108	110	115	120	124	125	130

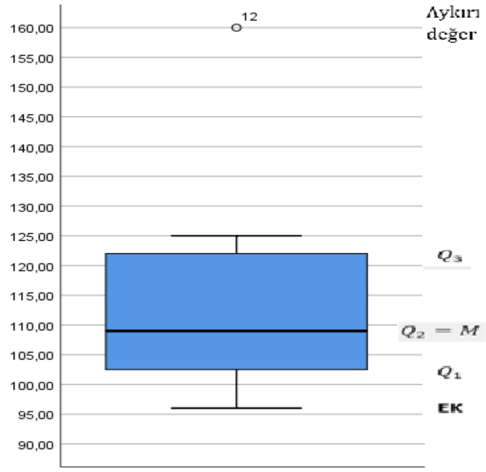
$Q_1 = 101.75$
$Q_2 = 109$
$Q_3 = 123$
$KA = 123 - 101.75 = 21.25$
$Q_1 - 1.5KA = 69.86$
$Q_3 + 1.5KA = 154.86$
$69.86\ ile\ 154.86$ aralığının dışında kalan değerler aykırı değerlerdir.
Veri setimizde aykırı değerler bulunmamaktadır.



Örnek: 12 gözleme sahip veri setinin kutu grafiğini çiziniz.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
96	97	101	104	105	108	110	115	120	124	125	160

$Q_1 = 101.75$
$Q_2 = 109$
$Q_3 = 123$
$KA = 123 - 101.75 = 21.25$
$Q_1 - 1.5KA = 69.86 \; ile \; Q_3 + 1.5KA = 154.86$
aralığının dışında kalan değerler aykırı değerlerdir. Veri setimizde aykırı değer 12 inci gözlem olup 160 değeridir.



6. OLASILIĞA GİRİŞ

6.1. Küme Kuramı

Küme: İyi tanımlı nesneler topluluğuna küme denir. Büyük harflerle gösterilir. Kümedeki her bir nesneye de kümenin elemanı denir. Genel olarak uygulamalarda bir kümeyi belirlemek için iki yol vardır:

- Küme sonlu elemanlı ise onun elemanları parantez içinde yazarak gösterilebilir.
- Kümenim öğesi olabilecek herhangi bir nesnenin sağlamak zorunda olduğu özellik parantez içinde tanımlanarak küme belirtilebilir.

Örnek: Elemanları 2, 4, 6 ve 8 tamsayıları olan bir A kümesi:
$A = \{2, 4, 6{,}8\}$ şeklinde yazılır.

Örnek: B kümesi A'nın elemanlarının küplerinden meydana gelmektedir. O halde B kümesi;
$B = \{8, 64, 216{,}512\}$ ya da;
$B = \{x^3 | x = 2, 4, 6, 8\}$ şeklinde yazılır.

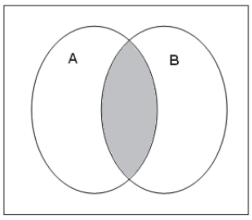
Boş Küme: Elemanı olamayan kümeye denir ve ∅ ya da { } ile gösterilir.

Alt Küme: A kümesinin her elemanı B kümesinin de elemanı ise A ya B kümesinin alt kümesi denir ve $A \subset B$ ile gösterilir. Boş küme her kümenin alt kümesidir.

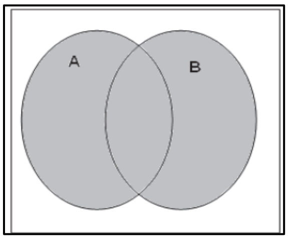
Evrensel Küme: Üzerinde çalışılan kümelerin hepsini kapsayan kümeye denir ve E ile gösterilir. Üzerinde çalışılan kümelerin her birini alt küme kabul eden n elemanlı sonlu bir küme olan E evrensel kümesinin E ve ∅' yi de kapsayan 2^n farklı alt kümesi vardır.

Örnek: $E = \{1, 3, 5\}$ ile evrensel kümeyi gösterelim.
Üç elemanlı bir evrensel kümenin alt kümelerinin sayısı $2^3 = 8$ olmak üzere;
$\emptyset, \{1\}, \{3\}, \{5\}, \{1, 3\}, \{1, 5\}, \{3, 5\}, \{1, 3, 5\}$ kümelerinin tümü E'nin alt kümeleridir.

Kesişim Kümesi: Hem A'nın hem de B'nin elemanlarının oluşturduğu kümeye denir ve $A \cap B = \{x \mid x \in A\ ve\ x \in B\ \}$ ile gösterilir.



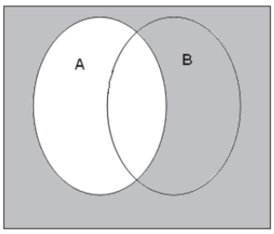
Birleşim Kümesi: A ya da B den an az birine ait elemanlarının oluşturduğu kümeye denir ve $A \cup B = \{x \mid x \in A\ veya\ \ x \in B\ \}$ ile gösterilir.



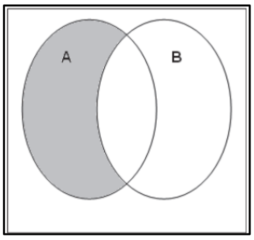
Ayrık Küme: Ortak elemanı olmayan yani $A \cap B = \emptyset$ ise A ve B kümeleri ayrıktır denir.

Küme Tümleyeni: A kümesinin tümleyeni A^t ya da A' ile gösterilir ve A kümesinde bulunmayıp diğer tüm kümelerde bulunun elemanlardan oluşur.

$$A' = \{x \mid x \in E\ ve\ x \notin A\ \}$$



Fark Kümesi: A kümesinde olup B kümesinde olmayan elemanlardan oluşan kümeye denir. $A - B = A \setminus B = \{x \mid x \in A\ veya\ \ x \in B\ \}$ ile gösterilir.



Küme Özellikleri: A, B ve C kümeleri E evrensel kümenin alt kümeleri olmak üzere

- *<u>Özdeşlik:</u>* $A \cup \emptyset = A$, $A \cup E = E$, $A \cap \emptyset = \emptyset$, $A \cap E = A$
- *<u>Tümleme:</u>* $A \cup A' = E$, $A \cap A' = \emptyset$, $(A')' = A$
- *<u>Değişme:</u>* $A \cup B = B \cup A$, $A \cap B = B \cap A$
- *<u>Birleşme:</u>* $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$
- *<u>Dağılma:</u>* $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- *<u>De Morgan:</u>* $(A \cup B)' = A' \cap B'$, $(A \cap B)' = A' \cup B'$

Örnek: Bir zar atılsın. $A = \{1, 2, 3\}$ ve $B = \{4, 5\}$ olayları tanımlansın. De Morgan Kuralının sağlandığını gösteriniz.

$(A \cup B)' = \{6\}$
$A' = \{4, 5, 6\}$
$B' = \{1, 2, 3, 6\}$
$A' \cap B' = \{6\}$

$(A \cup B)' = A' \cap B'$

$(A \cap B)' = \{1, 2, 3,4,5, 6\}$
$A' = \{4, 5, 6\}$
$B' = \{1, 2, 3, 6\}$

$(A \cap B)' = A' \cup B'$
$A' \cup B' = \{1, 2, 3,4,5, 6\}$