

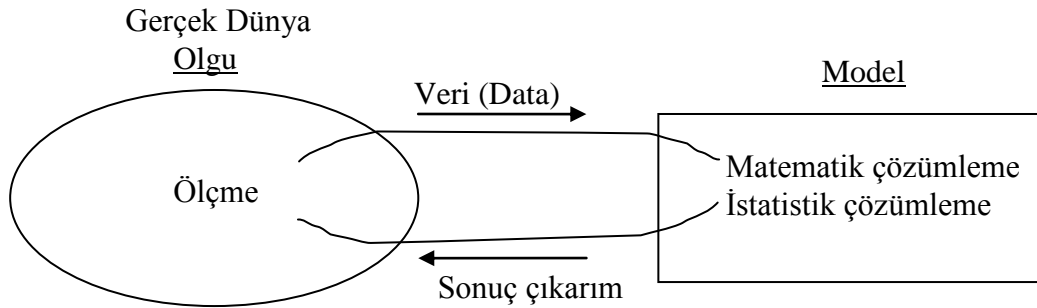
1. Giriş

İstatistik, rasgelelik içeren olaylar, süreçler, sistemler hakkında modeller kurmada, gözlemlere dayanarak bu modellerin geçerliliğini sınamada ve bu modellerden sonuç çıkarmada gerekli bazı bilgi ve yöntemleri sağlayan bir bilim dalıdır.

Tüm bilim dallarında amaç, gerçek dünyadaki olguları anlamak ve anlatmaktır. Bu anlama-anlatma işine **modelleme**, anlatımın kendisine de **model** denir. Modellemede, dilden sonra, aklın kullandığı ifade araçlarından en önde gelenleri matematik ve istatistiktir.

Model, gerçek dünyadaki bir olgunun ilgili olduğu bilim sahasının (fizik, kimya, biyoloji, jeoloji, astronomi, ekonomi, sosyoloji,...) kavram ve kanunlarına bağlı olarak ifade edilmesidir. Model gerçek dünyadaki bir olgunun belli bir anlatımıdır. Gerçek dünyanın çok karmaşık olması sebebiyle modeller, anlatmak istedikleri olgu ve sistemleri basitleştirerek belli varsayımlar altında ele alır. Modeller gerçeğin kendileri değildir ve ne kadar karmaşık görünseler de gerçeğin bir eksik anlatımıdır. Model denilen şey model kurucunun gerçeği anlayışının bir ürünüdür. Bu nedenle bazı durumlarda, gerçek dünyadaki bir olgu ile ilgili farklı modeller kurulabilir. Işık için tanecik ve dalga modellerinde olduğu gibi.

Olguları modellemede düşünce tarzı aşağıdaki gibidir.



Bir modelin yararlı olması için, verilerden sonuçların nasıl çıkarılacağına dair bir çözüm yönteminin bilinmesi gerekir. Eğer model stokastik ise çözümleme istatistiğin bir sorunudur. Verilerin nasıl toplanacağı da istatistiğin bir sorunudur. İstatistik yukarıdaki döngünün her safhasında yer almaktadır. Olguya temas ölçme ile olmaktadır. **Ölçme**, içinde istatistik de barındıran başlı başına bir konudur.

Fizik, kimya, biyoloji, jeoloji, astronomi, ekonomi, sosyoloji ve başka birçok bilim dalının gerçek dünyada ilgilendiği kendi konuları (sahaları) vardır ve çoğunun arakesiti boş değildir. Matematik ve İstatistiğin gerçek dünyada bir konusu olmamasına rağmen, gerçek dünyadaki

olguları anlama ve anlatmada, yani modellemede insan aklının en güçlü iki aracıdır. İstatistik, rasgelelik içeren olguların modellenmesinde öne çıkmaktadır.

Veriden (gözlemden) öğrenme sanatı olarak bahsedilen istatistik,

a) **Betimsel İstatistik:** Verinin tanımlanması ve özetlenmesi

b) **Çıkarımsal İstatistik:** Kitle ile ilgili sonuç çıkarımı yapabilmek için verinin kullanılması

olarak iki kısma ayrılır.

1.2. Bazı Temel Kavramlar

Her bilim dalının olduğu gibi istatistik biliminin de kendine ait temel kavramları vardır.

Belirli bir özelliğe sahip bireylerin veya birimlerin tümünün oluşturduğu topluluğa **kitle** denir.

Örnekleme yöntemlerinden yararlanılarak bir kitleden seçilen, aynı özellikleri taşıyan ve kitleyi temsil edebilecek nitelikteki ve nicelikteki bireylerin oluşturduğu topluluğa **örneklem** denir. Üzerinde araştırma yapılan kitle bazen sayılamayacak kadar çok birim içerebilir. Bu durumda bu birimlerin tamsayımı mümkün veya gerekli olmayabilir. Mümkün olsa bile zaman ve maliyet gibi kısıtlayıcılar tam sayımı imkansız kılabilir. Bu gibi durumlarda kitleden örneklem alınarak, elde edilen sonuçlar ile kitle için tahmin yapılabilir.

Kitleyi tanımlayan sayısal değerlere **parametre** denir. Örneğin kitle ortalaması (μ), kitle standart sapması (σ) birer parametredir. Genelde parametre değerleri bilinmez. Bu değerleri hesaplamak maliyet ve zaman açısından zordur. Bu nedenle, bilinmeyen parametrelerin uygun örneklem istatistikleri ile tahmin edilmesi yoluna gidilir.

Rasgelelik içeren olgulardan elde edilen ölçüm (gözlem) değerlerine **istatistiksel veri** veya kısaca **veri** (data) denir. Veriler deneyler sonucu veya doğal şartlarda olguları gözlemekle elde edilir.

Kitlede yer alan birimlerin belli karakteristiklerini belirten ve farklı değerler alabilen nitelik ya da niceliklerine **değişken** denir. Örneğin, bir ilköğretim okulundaki öğrencilerin kitlesi göz önüne alındığında bu okuldaki öğrencilere ait yaş, ağırlık, not ortalaması, göz rengi, saç rengi

öğrenciden öğrenciye değişecektir. Buna göre, yaş, ağırlık, not ortalaması, saç rengi, göz rengi birer değişkendir.

Sayısal değerler alabilen değişkenlere **nicel değişken** denir. Örneğin yaş, ağırlık, not ortalaması gibi. Nicel değişkenler, **kesikli** ve **sürekli** değişkenler olarak ikiye ayrılır. Belli bir aralıkta her değeri alan değişkenler sürekli değişkenler, her değeri alamayan değişkenler kesikli değişkenler olarak adlandırılır. Örneğin okuldaki öğrencilerin sayısı kesikli, öğrencilerin ağırlıkları sürekli değişkenlerdir.

Bir değişken sayısal değerlerle ölçülemiyorsa bunlara **nitel** ya da **kategorik** değişken adı verilir. Örneğin göz rengi, saç rengi, cinsiyet gibi.

Birimlerden bilgi toplanmasında değişkenlerin ölçülmesi önemli bir konudur. Değişkenlerin ölçülmesi, genel olarak dört başlık altında açıklanabilir;

a) Sınıflama Ölçme Düzeyi: Bu ölçme düzeyinde birimlere özelliklerine göre isimler verilir. Ölçme, isimlendirilerek yani sınıflandırılarak yapılır, herhangi bir sıralama yapılmaz. Örneğin cinsiyet değişkeni, kadın ve erkek gibi iki sınıfta ifade edilir. Bu çerçevede medeni durum, meslek, doğum yeri gibi değişkenlere ait ölçümler için sınıflama ölçme düzeyi kullanılır.

b) Sıralama Ölçme Düzeyi: Sıralama ölçme düzeyinde değişkenlerin aldığı değerler önem derecesi ya da üstünlüklerine göre sıralanır. Katılım düzeyi (Kesinlikle Katılıyorum, Katılıyorum, Kararsızım, Katılmıyorum, Kesinlikle Katılmıyorum), sıklık düzeyi (Hiç, Nadiren, Genellikle, Her Zaman), öğrenim durumu (İlköğretim, Lise, Lisans, Yüksek Lisans), vb. değişkenler için sıralama ölçme düzeyi kullanılır.

c) Eşit Aralıklı Ölçme Düzeyi: Sıcaklık, başarı, performans gibi nicel değişkenleri ölçmek için kullanılır. Bu ölçekte bir başlangıç noktası bulunmaz. Yani, “0” değeri eşit aralıklı ölçme düzeyinde yokluk ifade etmez. Örneğin termometrede görülen “0 °C” belirli bir anlam taşır.

d) Oranlama Ölçme Düzeyi: Aylık gelir, ağırlık, uzunluk, hız gibi değişkenleri ölçmek için kullanılır. Bu ölçme düzeyinde başlangıç “0” noktasıdır. Oranlama ölçme düzeyinde yer alan “0 Kg” bir yokluk ifadesidir.

Sınıflama ve sıralama ölçme düzeyi nitel değişkenler için, eşit aralıklı ve oranlama ölçme düzeyi nicel değişkenler için kullanılır.

1.3. Verilerin Düzenlenmesi ve Özetlenmesi

1.3.1. Frekans Tabloları

Uygulamada karşılaşılan birçok veri kümesi çok sayıda gözlem içerir. Gözlem sayısı çok olduğunda, verileri yorumlamak zorlaşır. Verilerin düzenlenmeden önceki biçimine **ham veri** denir. Gözlemlerin sınıflandırılması, özümsemesi gereken bilgi miktarını azaltarak verilerin yorumlanmasını kolaylaştırır. Gözlemleri sınıflandırmanın en doğru yolu frekans tablolarıdır. Verilerin bölüdüğü alt aralıklara sınıf, her bir sınıftaki gözlem sayısına da frekans (ya da sıklık) denir. Belli bir sınıfın birikimli frekansı ise o sınıf ile ondan önceki sınıflardaki toplam gözlem sayısıdır.

1.3.1.1. Nitel Verilerde Frekans Tabloları

Nitel veriler sınıflama ya da sıralama ölçme düzeyinde ölçülmüş verilerdir. Bu nedenle nitel veriler için frekans tablosu oluşturmak basittir. Frekans tablosunda sınıf, frekans ve göreceli frekans sütunları yer alır.

Örnek 3.1. 2016-2017 eğitim-öğretim yılı bahar döneminde Ankara Üniversitesi Fen Fakültesi İstatistik bölümünde açılan İST-206 (Bilgisayar Programlama IV) dersini alan 95 öğrenciden 47'si erkek, 48 tanesi kızdır. Bu dersi alan öğrencilerin cinsiyetlerine göre frekans tablosu aşağıdaki gibidir.

Sınıf(Cinsiyet)	Frekans(f_i)	Göreceli Frekans(p_i)
Kız	48	48/95
Erkek	47	47/95

Tablo 1. 2016-2017 eğitim öğretim yılında İst-206 dersini alan öğrencilerin cinsiyetlerine göre frekans tablosu

1.3.1.2. Nicel Verilerde Frekans Tabloları

Nicel verilerde frekans tablosu oluşturabilmek için bilinmesi gereken bazı tanımlar vardır. İlk olarak gözlemlerin genişliğinin bilinmesi gerekir. Yani en küçük ve en büyük değer bilinmesi gerekir.

Değişim Genişliği (R): Bir örneklemede en büyük değer ile en küçük değer arasındaki farka değişim genişliği denir.

$$R = \text{En büyük gözlem değeri} - \text{En küçük gözlem değeri}$$

n birimlik bir örneklemin frekans tablosunu oluşturabilmek için ilk olarak değişim genişliği R hesaplanmalıdır. Daha sonra bu R , eşit uzunluktaki aralıklara bölünerek sınıflar elde edilir.

Sınıf Aralığı: n birimlik bir rasgele örneklemin değişim genişliği R 'nin bölüdüğü aralıklara sınıf aralıkları denir. Sınıf aralığı c ile gösterilmek üzere;

$$c = \frac{\text{Değişim genişliği}(R)}{\text{İstenilen sınıf sayısı}(k)}$$

ile hesaplanır.

Frekans tablosu oluştururken, sınıf sayısını belirlemek için genel geçer bir kural olmamasına rağmen kullanılabilir çeşitli yöntemler vardır. Bir frekans tablosunda sınıf sayısı k ile gösterilmek üzere, bunlardan bir tanesi $2^k > n$ özelliğini kullanmaktır. Bir diğeri, $k \geq \sqrt{n}$ özelliğini kullanarak sınıf sayısını belirlemektir.

Alt Sınır (A_s): Bir sınıf aralığının en küçük değeridir.

Üst Sınır ($Ü_s$): Bir sınıf aralığının en büyük değeridir.

Sınıf Ortası (S_i): Bir sınıf aralığının merkezine ya da orta noktasına sınıf ortası denir. Bir frekans tablosunda k tane sınıf aralığı varsa sınıf ortaları S_1, S_2, \dots, S_k ile gösterilir.

Frekans (f_i): Bir sınıf aralığına düşen veri sayısına frekans denir. Bir frekans tablosunda k tane sınıf varsa sınıf frekansları f_1, f_2, \dots, f_k ile gösterilir. Frekansların toplamı toplam veri sayısını verir. Yani, $\sum_{i=1}^k f_i = n$ dir.

Görel Frekans (Frekans Yüzdesi- p_i): Her sınıfa düşen sınıf frekansının toplam frekansa oranıdır. Bir frekans tablosunda k tane sınıf varsa sınıflar için görel frekanslar p_1, p_2, \dots, p_k ile gösterilir ve $p_i = \frac{f_i}{n}, i = 1, 2, \dots, k$ olmak üzere $\sum_{i=1}^k p_i = 1$ 'dir.

Örnek 3.2: Bir bölgede bulunan bir sağlık ocağına kayıtlı 1 yaşındaki 100 kız çocuğuna ait ağırlıklar kilogram cinsinden aşağıda verilmiştir. Verilere ait frekans tablosunu oluşturunuz.

9.70	11.47	8.11	10.42	11.40	8.78	7.38	10.50	9.03	9.10
9.38	8.81	8.69	8.78	9.80	9.77	5.67	10.50	9.56	10.29
8.22	7.16	9.24	13.16	9.46	9.11	11.67	8.88	10.81	8.80
7.26	7.84	6.22	8.54	8.82	8.05	8.09	9.77	10.38	9.39
8.66	9.33	6.60	10.23	9.31	11.06	10.05	7.48	7.91	6.11
8.79	7.20	7.67	8.92	11.36	7.91	9.02	9.71	7.86	8.17
6.28	8.41	9.10	6.84	8.42	8.59	7.76	8.84	7.86	7.67
10.63	8.87	9.25	11.35	8.72	9.10	8.95	11.01	10.34	8.20
9.54	8.47	8.09	7.67	8.43	6.90	9.51	7.80	8.83	8.05
6.81	9.25	10.64	9.09	8.48	7.74	7.57	11.51	8.44	8.47

Tablo 2. 100 kız çocuğuna ait ağırlıklar (kg)

Sınıf sayısı olarak $k = 10$ alınarak frekans tablosu oluşturulsun.

Gözlemler içinde en küçük değer=5.67; en büyük değer=13.16 dir. Buna göre, gözlemlerin değişim genişliği;

$$R = 13.16 - 5.67 = 7.49$$

dir.

Sınıf aralığı,

$$c = \frac{\text{değişim genişliği}}{\text{sınıf sayısı}(k)} = \frac{7.49}{10} = 0.749 \cong 0.75$$

olarak alınabilir. Buna göre frekans tablosu; ilk sınıf aralığının alt sınırı gözlemlerin en küçük değeri alınarak oluşturulmaya başlanır. Bir sonraki sınıfın alt sınırı önceki sınıfın alt sınırı ile sınıf aralığı değerinin eklenmesi ile bulunur. Sınıf aralıklarındaki frekanslar, alt sınır değeri sınıfa dahil, üst sınır değeri sınıfa dahil değil şeklinde alınarak oluşturulur. x ile bebeklerin ağırlıkları gösterilmek üzere, gözlemlere göre oluşturulacak frekans tablosu aşağıdaki gibi olacaktır.

Bebeklerin ağırlıkları	S_i	f_i	p_i
5.6700 $\leq x < 6.42$	6.0450	4	0.04
6.4200 $\leq x < 7.17$	6.7950	5	0.05
7.1700 $\leq x < 7.92$	7.5450	16	0.16
7.9200 $\leq x < 8.67$	8.2950	18	0.18
8.6700 $\leq x < 9.42$	9.0450	28	0.28
9.4200 $\leq x < 10.17$	9.7950	10	0.10
10.1700 $\leq x < 10.92$	10.5450	10	0.10
10.9200 $\leq x < 11.67$	11.2950	7	0.07
11.6700 $\leq x < 12.42$	12.0450	1	0.01
12.4200 $\leq x < 13.17$	12.7950	1	0.01

Tablo 3. Bir yaşındaki kız bebeklerin ağırlıklarına ilişkin frekans tablosu

1.3.1.3 Birikimli (Eklemeli) Frekans Tablosu

Her sınıfın üst sınırından daha az değer gösteren verilerin toplamı o sınıf için birikimli sıklık değerinin vermiş olur.

Bebeklerin ağırlıkları	f_i	p_i	Eklemeli f_i	Eklemeli p_i
6.42 den az	4	0.04	4	0.04
7.17 den az	5	0.05	9	0.09
7.92 den az	16	0.16	25	0.25
8.67 den az	18	0.18	43	0.43
9.42 den az	28	0.28	71	0.71
10.17 den az	10	0.10	81	0.81
10.92 den az	10	0.10	91	0.91
11.67 den az	7	0.07	98	0.98
12.42 den az	1	0.01	99	0.99
13.17 den az	1	0.01	100	1

Tablo 4. Bir yaşındaki kız bebeklerin ağırlıklarına ilişkin Birikimli Frekans Tablosu

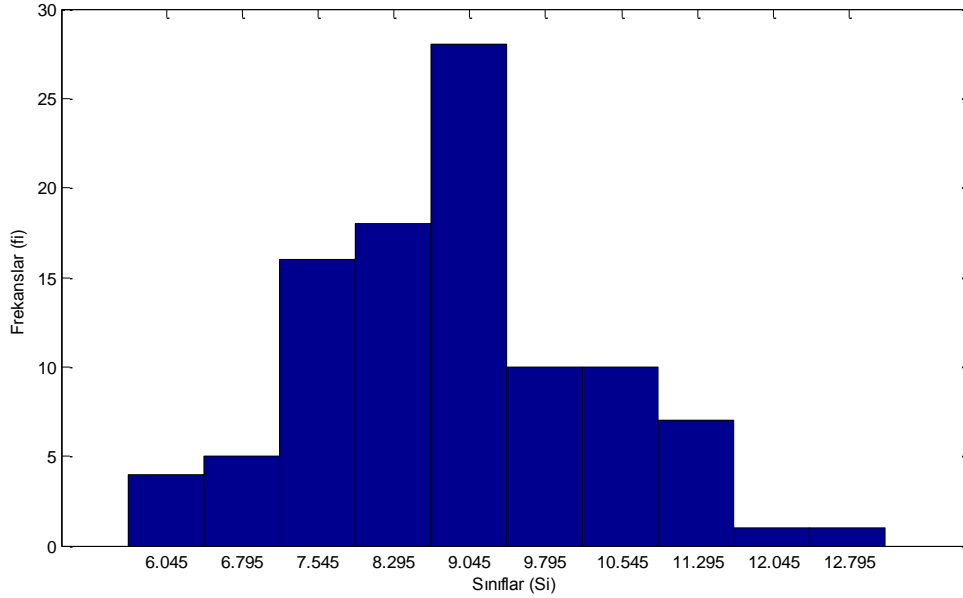
Birikimli frekans tablosuna bakılarak bir yaşındaki kız bebeklerin ağırlıklarına ilişkin yorum yapılabilir. Verilen birikimli frekans tablosuna göre bu sağlık ocağına getirilen bir yaşındaki kız bebeklerin %81'inin ağırlıklarının 10.17'den az olduğu söylenebilir. Yine aynı şekilde getirilen bir yaşındaki kız bebeklerin % 9'unun ağırlıklarının 10.92'den çok olduğu söylenebilir (Çünkü, %91'i 10.92 den daha düşük ağırlıkta.).

1.3.2. Grafikler

Grafikler toplanan verilerin kullanıma sunulmasında kullanılan, göze hitap eden geometrik şekillerdir. İstatistikte kullanılan çok çeşitli grafik türleri vardır. Burada bunların içinden histogram (dağılım dikdörtgenleri), birikimli sıklık grafiği, gövde yaprak grafiği, kutu grafiği, saçılım grafiği, çubuk grafiği, daire (pasta) grafiği ve çizgi grafiği verilecektir.

1.3.2.1. Histogram

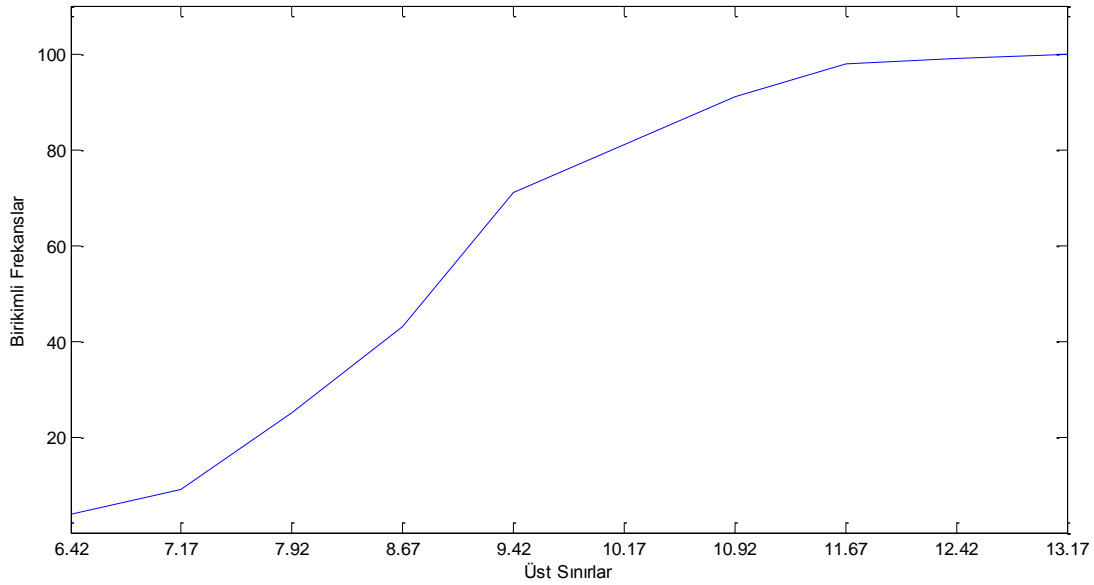
Histogram koordinat sisteminde tabanları x ekseninde frekans tablosundaki her bir sınıfın sınıf aralığı büyüklüğünde, yükseklikleri bulunduğu sınıfın frekanslarıyla orantılı olarak yan yana çizilen dikdörtgenlerden oluşur. Yukarıda frekans tablosu verilen örneğe ilişkin histogram aşağıda verilmiştir.



Grafik 1. Bir yaşındaki kız bebeklerin ağırlıklarına ilişkin Histogram Grafiği

1.3.2.2. Birikimli Frekans Grafiği

Birikimli frekans grafiğinin oluşturulabilmesi için ilk olarak birikimli frekans tablosunun hazırlanması gerekir. Birikimli frekans tablosundaki üst sınıf değerleri x eksenine birikimli frekans değerleri ise y eksenine işaretlenir ve elde edilen bu noktalar birleştirilir ise birikimli frekans grafiği çizilmiş olur.



Grafik 2. Bir yaşındaki kız bebeklerin ağırlıklarına ilişkin Birikimli Sıklık Grafiği

1.3.2.3. Gövde Yaprak Grafiği

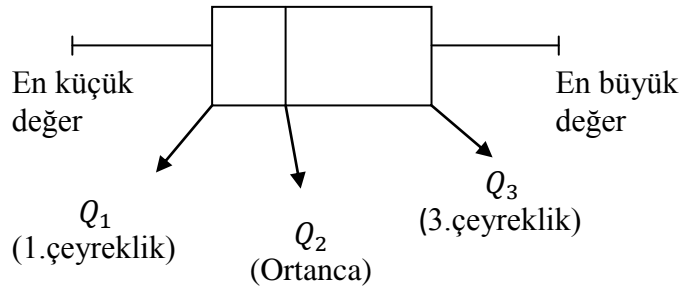
Verileri küçükten büyüye doğru sıralayıp düzenli hale getiren yöntemlerden bir tanesidir. Bu grafikte verilerin nasıl bir dağılım gösterdiğini gözlemlemek mümkündür. Bu grafik türünde sayılar basamaklarına ayrılır ve bir kısmı gövde bir kısmı yaprak şeklinde gösterilir. Örneğin, veriler iki basamaklı ise onlar basamağı gövde iken birler basamağı yaprakları oluşturur. Verilerimiz; 19, 48, 27, 23, 36, 49, 38, 16, 47, 44, 25 biçiminde olsun. Buna göre gövde yaprak grafiği,

1		6,9
2		3,5,7
3		6,8
4		4,7,8,9

biçiminde olacaktır.

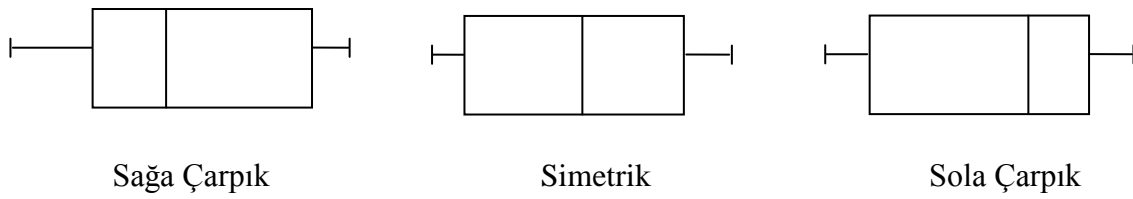
1.3.2.4 Kutu grafiği

Bir veri grubundan elde edilen en küçük değer, en büyük değer, birinci çeyreklik (verilerin %25' i bu değer altında değer alır), üçüncü çeyreklik (verilerin %25' i bu değer üstünde değer alır) ve ortanca (verilerin %50' si bu değer altında %50' si bu değer üstünde değer alır) değerlerini içeren bir grafikdir.



Grafik 3. Kutu Grafiği

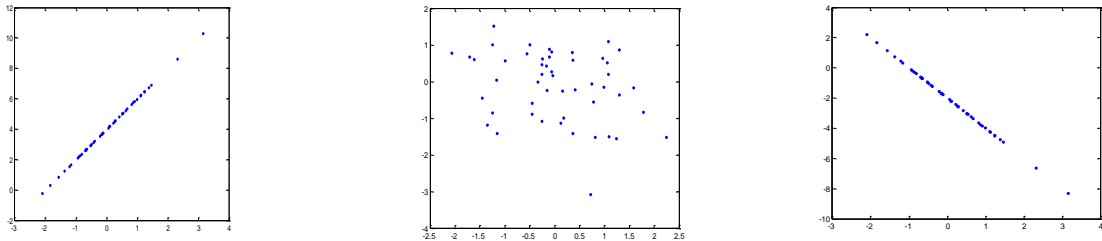
Kutu grafiğine bakılarak dağılımın şekli hakkında yorum yapılabilir.



Grafik 4. Kutu grafiğine göre dağılımın şekli

1.3.2.5. Serpme (Saçılım) Grafiği

Serpme ya da saçılım grafiği iki farklı değişkenin arasındaki ilişkiyi görmek için çizilen bir grafik türüdür. Değişkenlere karşılık gelen değerler (x,y) gibi ikililer ile ifade edilir ve iki eksen den birine bir değişken diğerine diğer değişken yerleştirilir. x değişkenine karşılık gelen y değişkeninin değeri bulunur ve noktalarla ifade edilir. Bu noktaların şekli iki değişken arasındaki ilişki hakkında bize bilgi verir.



İki değişken arasında pozitif bir ilişki vardır. Biri artarken diğeri de artıyor. biri azalırken diğeri de azalıyor şeklinde yorum yapılabilir.

İyi değişken arasında pozitif ya da negatif doğrusal bir ilişkiden bahsedilemez.

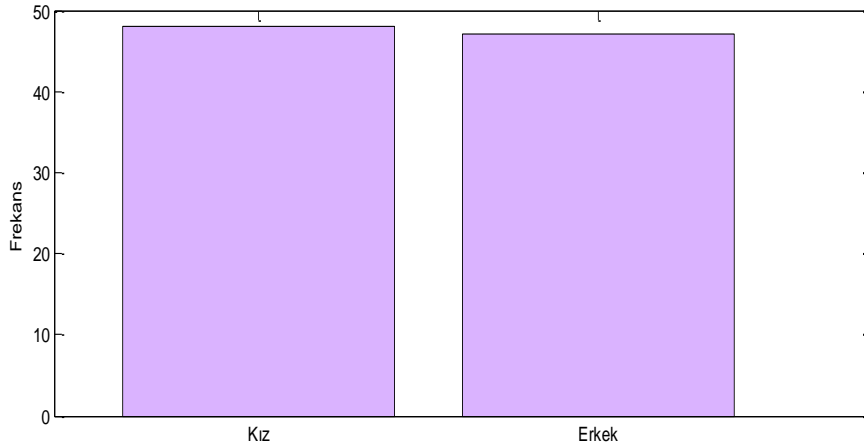
İki değişken arasında negatif bir ilişki vardır. Biri artarken diğeri azalıyor, biri azalırken diğeri artıyor şeklinde yorum yapılabilir.

Grafik 5. Saçılım Grafiğine göre iki değişken arasındaki ilişki

1.3.2.6. Çubuk Grafiği

Çubuk grafiğinde sınıflar tabanları eşit ve birbirine bitişik olmayan dikdörtgenlerle temsil edilir.

Örnek 3.3: İst-206 dersini alan öğrencilerin cinsiyetlerine göre frekans tablosu Tablo 1 ile verilmişti. Bu frekans tablosuna ilişkin çizilecek çubuk grafiği aşağıdaki gibi olacaktır.

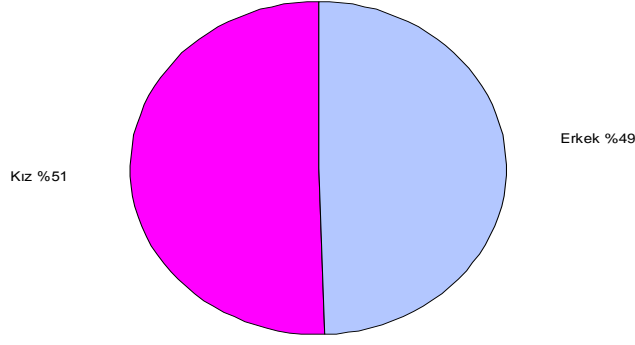


Grafik 6. İst-206 dersini alan öğrencilerin cinsiyetlerine göre çubuk grafikleri

1.3.2.7. Pasta Grafiği

Veriler nitel olduğunda göreceli frekansları göstermek için kullanılan grafik türüdür. Bir daire oluşturulur ve daha sonra bu daire her bir farklı tür veri değeri için dilimlere bölünür. Bir veri değerinin göreceli frekansı bu veri diliminin alanıyla gösterilir, bu alan dairenin toplam alanının veri değerinin göreceli frekansı ile çarpılmasıyla elde edilen değere eşittir.

Örnek 3.4. İst-206 dersini alan öğrencilerin cinsiyetlerine göre Frekans tablosu Tablo 1 ile verilmişti. Bu frekans tablosuna ilişkin çizilecek pasta grafiği aşağıdaki gibi olacaktır.

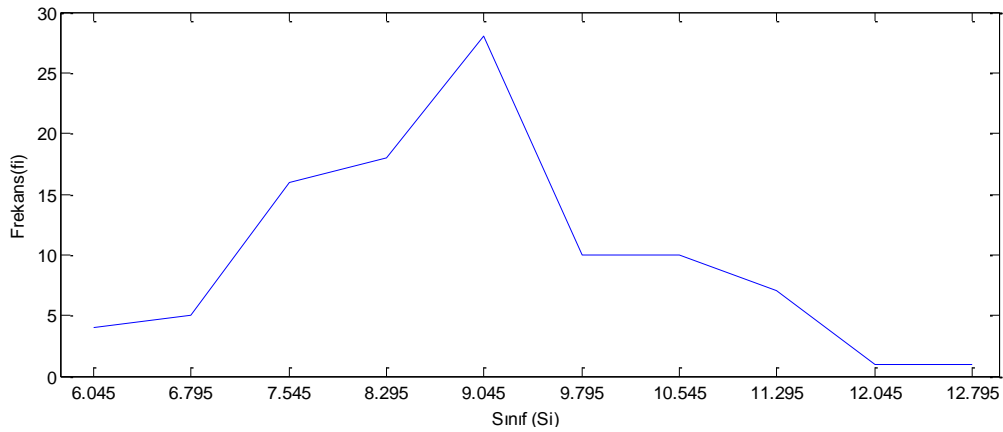


Grafik 7. İst-206 dersini alan öğrencilerin cinsiyetlerine göre pasta grafiği

1.3.2.8. Çizgi Grafiği

Frekans dağılımında her sınıfın düzlemde bir nokta ile temsil edilip, sonra bu noktaların birleştirilmesiyle elde edilen şekle çizgi grafiği denir. Sürekli veriler için kullanılan bir grafik türüdür.

Örnek 3.5. Tablo 3 ile bir yaşındaki kız çocuklarının ağırlıklarına ilişkin frekans tablosu verilmişti. Bu frekans tablosuna ilişkin çizgi grafiğini çiziniz.



Grafik 8. Bir yaşındaki kız çocuklarının ağırlıklarına ilişkin Çizgi Grafiği