İSTATİSTİK DERS NOTU

İçindekiler

İSTATİSTİK DERS NOTU

1. TEMEL KAVRAMLAR

1.1. İstatistiğin Tanımı ve Kullanım Alanları

1.1.1. İstatistik Nedir?

Genel anlamıyla istatistik , yaşama ilişkin düşüncelerin sistematize edilmiş biçimidir. İstatistik iş yaşamında da, işletmedeki her türlü faaliyetin kantitatif (nicel) temellerini hazırlar. Bu bakış açısıyla istatistik; “daha etkin kararlar vermek amacıyla verileri toplama, sınıflama, sunma, çözümleme ve yorumlama biçimi” olarak tanımlanabilir (Tütek ve Gümüşoğlu, 2008, s.1).

Diğer bir tanıma göre istatistik, verilerin toplanması, işlenmesi, analiz ve yorumunda kullanılan metodlar bütünüdür (Güler, 2007, s.1).

1.1.2. İstatistiğin Kullanım Alanları

İstatistik günlük hayatımızın hemen her alanında kullanılmaktadır. Aşağıda istatistiğin kullanım alanlarından bazılarına yer verilmiştir.

- IMKB ’de hisse senetlerinin analizi
- Siyasi partilerin seçimlere ilişkin kamuoyu yoklamaları
- Kalite Kontrol
- Şirketin muhasebe kayıtlarının denetlenmesi
- Pazarlama araştırmaları
- Ekonomik göstergelerin takibi
- Tıbbi araştırmalar
- Mühendislik araştırmaları
- Bilimsel çalışmalar
- Uzay araştırmaları
- Demografik (Nüfus özellikleri) araştırmalar

1.2. Temel Kavramlar

1.2.1. Araştırma

İlgilenilen konuya ilişkin sorunların saptanması, çözüm yollarının planlanması, uygulamaya konulması ve sonuçlarının değerlendirilmesine yönelik yapılan çalışmalardır. Araştırmanın konusu ve amacı belirgin olmalıdır. Zaman, personel ve maliyet dikkate alınarak, araştırmanın sınırı iyi belirlenmelidir. Araştırmada görev alan araştırıcılar ya da uzmanlar yeterli bilgi düzeyine sahip olmalıdır. Araştırmanın sonucunda elde edilen bilgiler doğru biçinde değerlendirilmelidir (Balce ve Demir, 2007, s.3).

1.2.2. Popülasyon (Kitle, Ana Kitle, Ana Kütle, Yığın)

Araştırma kapsamına giren, aynı özellikleri taşıyan birimlerin ya da bireylerin oluşturduğu topluluğa KİTLE denir. Kitlenin büyüklüğü araştırmanın özelliğine göre değişir. Nüfus sayımı için kitle Türkiye’dir. Denizli’deki üniversite öğrencilerinin giderleri için kitle Pamukkale Üniversitesi öğrencileridir (Balce ve Demir, 2007, s.3).

1.2.3. Örneklem ve Örnekleme

Kitle büyüklüğüne bağlı olarak her zaman tüm birimler (bireyler) hakkında bilgi sahibi olmak mümkün değildir. Bundan dolayı geniş kitlelerde araştırmalar; zaman, maliyet, personel, ulaşım, vb. nedenlerden dolayı tüm birimler yerine daha az sayıdaki birimler seçilerek yürütülür (Balce ve Demir, 2007, s.4).

Bir kitleden, belirli yöntemler kullanılarak seçilen aynı özellikleri taşıyan bir kısım bireyin oluşturduğu topluluğa ÖRNEKLEM denir (Balce ve Demir, 2007, s.4).

Örnekleme; ana kitleyi nitelik ve nicelik yönünden temsil edebilecek bir kümenin çekilmesi işlemidir (İslamoğlu, 2009, s.159). Örnekleme konusu 1.6'da detaylı olarak ele alınmıştır.

1.2.4. Tamsayım

Bir araştırma kapsamında, kitledeki tüm birimlerine ulaşılarak istenen bilginin elde edilmesi işlemidir. Bunun yapılabilmesi için incelenecek kitlenin büyüklüğünün, belirlenen maliyet ve zaman gibi kısıtlara uygun olması gerekir. Bazı durumlarda (nüfus sayımları gibi) kitle büyük olsa bile tam sayım yapılması zorunlu olmaktadır. Gelişen teknoloji ile birlikte bu tür tam sayımlar daha kolay yapılabilir hale gelmiştir (Balce ve Demir, 2007, s.4).

1.2.5. Gözlem (Denek)

Kitle ya da örneklemde yer alan her birime gözlem ya da denek denir. Gözlem (ya da denek) sayısı aşağıdaki biçimde simgeleştirilmektedir.

Kitledeki Gözlem Sayısı : **N**

Örneklemdeki Gözlem Sayısı: **n**

1.2.6. Parametre ve İstatistik

Kitle özelliklerinin sayısal değerlerine PARAMETRE denir. Araştırma kitle yerine örneklem üzerinde uygulanıyorsa, parametre değerleri tahmin edilir. Bu durumda, örneklemden elde edilen sayısal değerlere İSTATİSTİK denir. Örnek olarak, sıkça kullanılan bazı parametreler ve istatistikler aşağıda verilmektedir:

Parametre	***İstatistik***
Kitle Ortalaması : μ	Örneklem Ortalaması : $\bar{x}$
Kitle Varyansı : σ	Örneklem Varyansı : s^2

1.2.7. Değişkenler

Değişkenin kelime anlamı; Değişme özelliği gösteren, çok değişen, değişebilir, kararsız, değişici şeklindedir. Matematiksel tanımı ise; Gözlemden gözleme değişik değerler alabilen objelere, özelliklere ya da durumlara "Değişken" denir.

İstatistikte **değişken***(variable)* terimi, deneklere ait özellikler anlamında kullanılır. Örneğin araştırmaya katılanların cinsiyeti, medeni hali, gelir düzeyi, eğitim düzeyi, coğrafik konumu gibi demografik özellikleri ve boyu, kilosu, vizeden aldığı not, sorulara verdikleri cevaplar gibi özellikler birer değişkeni ifade ederler.

1.2.7.1. Nicel (Kantitatif) Değişkenler

Birimlerin ölçüm ve tartım sonucu değerleri saptanan sayısal özelliklerini belirten değişkenlerdir. Bu değişkenler değerleri, mekanik ve elektronik araçlara sayısal olarak aralıklı ölçekli yada orantılı ölçekli verileridir. Nicel değişkenlerin verilerine nicel veri adı verilir.Örneğin birimlerin, boy uzunluğu, vücut ağırlığı, kilosu, kan basıncı gibi özellikler nicel değişkenlerdir.

1.2.7.1.1. Kesikli (Süreksiz) Değişken

Bu değişkenler miktar yönünden değişiklik yerine tür yönünden değişiklik gösterir. Dolayısıyla bir obje ya da birey bir özelliğe sahiptir ya da değildir. Yani kesin değerler alırlar. Nitel değişkenlerin hemen hepsi süreksiz değişkendir. Cinsiyet, medeni durum, göz rengi gibi. Birinin diğerine göre daha çok veya az olması mümkün değildir.

1.2.7.1.2. Sürekli Değişken

İki ayrı ölçüm arası kuramsal olarak sonsuz parçaya bölünebilir. Ölçüm söz konusu olduğu için sürekli değişken değerleri her zaman tam değeri vermezler. Rasyonel sayılar kümesinin elemanları ile belirtilirler.Yaş, uzunluk ve ağırlık gibi

Ölçme tartma yoluyla elde edilen dolayısıyla nokta içermesi mümkün olan verilerdir. Örneğin; balıkların ağırlıkları, tohum ağırlıkları, tohum çapı, bitki boyu, ortamdan bakterilerin tüketmiş oldukları şeker miktarı, ineklerin yıllık süt verimleri gibi.

1.2.7.2. Nitel (Kalitatif) Değişkenler

Nitel değişkenler; birimlerin kalite, kategorik, yada isimsel olarak belirtilebilen karakteristik özelliklerini, durumlarını ve pozisyonlarını belirten değişkenlerdir. Bu değişkenlerin verileri isimsel ya da sıralı ölçekle elde edilmişlerdir ve iki yada daha fazla kategoriye (alt seçenek, sınıf, grup) ayrılarak sayımla elde edilir. Bu değişkenlerin verilerine nitel veriler adı verilir. Örneğin birimlerin, cinsiyeti, kan grubu, medeni durum, göz rengi, mesleği, yerleşim yeri, tuttuğu futbol takımı (fanatikler için) gibi nitelik bildiren drumları açıklayan değişkenlerdir.

1.2.7.3. Değişkenlerin Ölçümü

İstatistik analize başlamadan önce ilk yapılacak şey, değişkenlerin nasıl ölçüldüğünün belirlenmesidir. Bu sorunun yanıtı "sfigmomanometre, Sahli yöntemi, termometre, otoanalizer, ..." gibi cihaz ya da yöntemler ya da "santimetre, yıl, IU/mL" gibi birimler olabilir; ancak istatistik biliminde **"ölçüm"**den kastedilen, bu değildir. İstatistikte ölçüm denildiği zaman anlaşılan, değişkenin alabileceği değerlerle ilgili kısıtlamalardır. Örneğin bir insanın cinsiyeti “bayan” yada “bay” değerini, yaşı 1, 2, 3 … değerlerini, boyu 1.75 yada 1.85 değerlerini alabilir.

Değişkenlerin alabileceği değerlerin neler olabileceği, neler olamayacağı, yani nasıl ölçüldüğünü belirlemek, yapılacak istatistik analizin seçimi için çok önemlidir. Ölçüm özelliklerine göre değişkenleri başlıca üç gruba ayırabiliriz: **nominal, ordinal ve sayısal değişkenler**.

1.2.7.3.1. Kullanılan Ölçek Türleri

1.2.7.3.1.1. Nominal Ölçek

Nominal ölçekte veriler sadece niteliklerine göre isimlendirilip gruplandırılabilir. Bu ölçekle ölçülen veriler ile matematiksel işlem yapılamaz. Cinsiyet, medeni durum, din , parça numarası, çalışma durumu, meslek vb. kalitatif veriler nominal ölçekle ölçülürler. Bu verilerin frekansları ve modları belirlenebilir (Tütek ve Gümüşoğlu, 2008, s.3).

Nominal bir değişkende ölçüm düzeyleri arasında bir sıralama ya da uzaklık-yakınlık gibi belirli bir mesafe yoktur.

1.2.7.3.1.2. Ordinal Ölçek

Bu ölçekte veriler nominal özellik taşımakla birlikte aynı zamanda sıraya konulabilir (Tütek ve Gümüşoğlu, 2008, s.4). Ordinal bir değişkende ölçüm düzeyleri arasında bir sıralama vardır, ama düzeyler arasındaki mesafeler belirli değildir.

Bu ölçeğin amacı, bir konu hakkındaki düşünceleri belirli bir öncelik sırasına koymaya hizmet etmektir. Bu ölçek, belli bir özelliğe göre nesnelerin yada yargıların konumlarını, aralarındaki uzaklığı belirtmeksizin dile getiren bir ölçek türüdür (İslamoğlu, 2009, s.142).

Örneğin restaurant, otel, hastane gibi hizmet işletmeleri yiyecek kalitesi, servis, tesisler, konukseverlik gibi özellikler açısından “mükemmel”, “iyi”, “orta”, “kötü”, “çok kötü" olarak sınıflandırılabilirler. Bu sıralamada "mükemmel" için 1, "iyi" için 2, "orta" için 3, "kötü" için 4, "çok kötü" için 5 değeri kullanıldığında; "mükemmel" olarak değerlendirilen bir A otelinin "iyi" olarak değerlendirilen bir B otelinden iki kat iyi olduğu söylenemez. Bu ölçekle ölçülen verilerin frekansı, modu, medyanı, yüzdelik değeri ve korelasyonu hesaplanabilir (Tütek ve Gümüşoğlu, 2008, s.4).

1.2.7.3.1.3. Aralık Ölçeği

Bu ölçekte veriler ordinal özellik taşımakla birlikte, aynı zamanda bu veriler için ölçekte tanımlanan birimlerin birbirinden eşit uzaklıkta bulunduğu bir ölçü birimi ve rastgele bir başlangıç noktası tanımlanabilir (Tütek ve Gümüşoğlu, 2008, s.5). Örneğin su buz haline 0, kaynama haline 100 değeri verilerek aradaki ısı farkı 100 eşit parçaya bölünüp her aralığa sırası ile bir derece verilir. Ancak burada dikkat edilmesi gereken nokta şudur: Bu ölçekle ölçülmüş iki nesneden biri 10, öteki 20 puan almış ise, biri ötekinin yarısı yada iki katı değildir (İslamoğlu, 2009, s.144).

Bu ölçekle ölçülen verilerin aritmetik ortalama, standart sapma, yatıklık ve basıklık ölçüleri hesaplanabilir (İslamoğlu, 2009, s.144).

1.2.7.3.1.4. Oran Ölçeği

Bu ölçek, objektif bir başlangıç noktasını temel alan ve üzerinde ölçülecek nokta veya objeleri bir birinin katı olarak ifade eden bir ölçektir (İslamoğlu, 2009, s.144). Bu ölçek, aralıklı ölçeğin özelliklerini taşımasının yanı sıra bu verilerden iki ölçümlemenin oranlanmasıyla anlamlı kıyaslamalar elde edilir. Uzunluk, parasal değerler, zaman, güç değerleri oran ölçeğiyle ölçülürler (Tütek ve Gümüşoğlu, 2008, s.4).

1.2.7.3.2. Ölçmede Dikkat Edilmesi Gereken Hususlar, Ölçek Güvenilirliği ve Ölçek Geçerliliği

Toplumbilimde bağımlı ve bağımsız değişkenleri ölçmek her zaman kolay olmadığı için, kullanılan ölçeklerin istenen bilgiyi toplamaya elverişli olup olmadığı büyük önem taşır. Bu nedenle, ölçekleri hem seçerken hem de kullanırken iki önemli noktaya dikkat edilmelidir (İslamoğlu, 2009, s.152):

- Kullanılan ölçek, istenen bilgiyi tam, doğru ve objektif olarak ölçebilecek mi?
- Kullanılan ölçek, araştırmanın hipotezlerini test etmeye hizmet edecek mi? Ve uygulanması düşünülen test tekniklerine elverişli mi?

Bir ölçeğin güvenilirliği, onun sağlamlığını ve tutarlılığını ifade eder. Bir göstergenin ölçüm aracının kendi yüzünden çeşitlilik göstermemesi, kullanılan ölçeğin güvenilir olduğunu gösterir (İslamoğlu, 2009, s.153).

Güvenilirlik zaman, temsil ve eşdeğerlik açısından üç ayrı özellik taşır. Zaman özelliği, aynı ölçeğin farklı zamanlarda uygulanması sonucunda aynı sonuçları vermesidir. Temsil, aynı ölçeğin farklı gruplara uygulanması halinde aynı sonuçları vermesi demektir. Eşdeğerlilik, ölçümler arasında tutarlılığın sağlanıp sağlanmadığını ifade eder (İslamoğlu, 2009, s.153).

Geçerlilik, kullanılan ölçeğin, ölçülmek isteneni tam olarak ölçüp ölçmemesine bağlıdır (İslamoğlu, 2009, s.153).

1.3. İstatistik Türleri

İstatistik teknik ve yöntemlerini kullanma amaçlarına göre iki gruba ayırabiliriz.

- Tanımlayıcı (Betimleyici) İstatistik
- Öngörüleyici (Tahminleyici) İstatistik

1.3.1. Tanımlayıcı (Betimleyici) İstatistik

Tanımlayıcı istatistik sayısal verileri sınıflama ve özetlemede kullanılan yordamlardır. Verileri tablo, grafik veya sayısal olarak anlamlı bir biçimde özetler. Bazı veriler frekans dağılımı olarak düzenlenebilir (Tütek ve Gümüşoğlu, 2008, s.6).

Verilerden ortalama değer ve bazı özel orta değerler hesaplanabilir. Örneğin, medyan bir grup sayısal veriyi ikiye bölen (%50 - %50) orta noktadaki değerdir (Tütek ve Gümüşoğlu, 2008, s.6).

1.3.2. Öngörüleyici (Tahminleyici) İstatistik

Öngörüleyici istatistik, gözlemlenmiş durumlardan elde edilen verilerle, gözlemlenmemiş durumlar için sonuç çıkarır ve popülasyon hakkında öngörüleme yapar. Popülasyon bireylerden oluşabilir; Celal Bayar Üniversitesi'nde okuyan öğrenciler, istatistik 1 dersini alan öğrenciler, Sağmalcılar ceza evindeki tutuklular gibi. Popülasyon objelerden de oluşabilir: Ford fabrikasında geçen hafta üretilen Ford Ka arabaları gibi. Popülasyon bir grup ölçümden oluşabilir: Efes Pilsen Basketbol takımındaki oyuncuların boyları vb (Tütek ve Gümüşoğlu, 2008, s.6).

1.4. Veri Kaynakları

İstatistiksel çalışmalarda, ihtiyaç duyulan veriler, iç ve dış kaynaklardan elde edilirler.

1.4.1. İç Kaynaklar

Genelde örgütlerde, özelde de işletmelerde gerçekleştirilen istatistiksel araştırmalar için gerekli veriler çoğunlukla örgütün yada işletmenin kayıtlarından elde edilir. İşletmenin muhasebe, üretim, stok, bordro, satış ve müşteri hesapları kayıtlarından alınan verilerine iç veriler denir. İç verilerin toplanması genelde sorun yaratmaz (Tütek ve Gümüşoğlu, 2008, s.7).

1.4.2. Dış Kaynaklar

İç veriler genellikle önemli olsalar da birçok durumda dış kaynaklardan bazı bilgilerle desteklenmelidirler. Devletten, belediyelerden, bankalardan, ticaret ve sanayi odalarından ve diğer özel kurumlardan elde edilen verilere dış veriler denir. Dış kaynaklardan veri toplanırken bazı sorunlarla karşılaşılır. İki ayrı dış kaynaktan elde edilen veriler birbirine uymayabilir yada istenilen maliyet ve süre içinde uygun dış veri bulunmayabilir (Tütek ve Gümüşoğlu, 2008, s.7).

1.4.2.1. Birincil Veriler

Araştırmaya konu olan ana kitleden doğrudan elde edilen bilgiler yada verileri toplayan kuruluşlardan elde edilen veriler birincil veri olarak tanımlanırlar (Tütek ve Gümüşoğlu, 2008, s.7). Bazı bilgiler ikincil kaynaklardan ve başka örgütlerden bulunamadığından, bunları doğrudan birincil kaynaklardan toplama zorunluluğu vardır. "Tüketicilerin otomobil satın alma kararlarında aile reisi dışındaki aile bireylerinin bu karar üzerinde etkili olup olmadıkları" araştırılacaksa birincil kaynaklara yönelmek gerekir. Yani, bu bilgiler ilgili bireylerden elde edilecektir (İslamoğlu, 2009, s.103). Genellikle birincil kaynaktan sağlanan veriler tercih edilmektedir. İkincil kaynakta yayın hataları olabileceği gibi birincil kaynakta veriler düzgün ve tam tanımlamalarla verilir. DPT, DİE bültenleri ve yayınları TÜFE ve TEFE' yi kendileri hesaplayarak yayınladıkları için birincil verileri içerirler.

1.4.2.2. İkincil Veriler

Verileri toplayan örgütten başka bir örgüt tarafından yayınlanmış veriler ise ikincil veri olarak tanımlanır. Eğer TÜFE, gazetelerin finans sayfalarından elde ediliyorsa ikincil veri sayılır (Tütek ve Gümüşoğlu, 2008, s.7).

1.5. Veri Toplama Yöntemleri

Birincil ve ikincil veri kaynaklarının araştırılması ve verilerin toplanması için kullanılan bazı yöntemler mevcuttur. Bu yöntemler; literatür araştırması, gözlem, deney ve anket yöntemleridir.

1.5.1. Literatür Araştırması

Araştırma konusu ile ilgili benzer bilimsel araştırmalar ve raporlar, indeksler vb. kaynaklar bilimsel literatürde, ilgili kuruluşların sitelerinde detaylı olarak araştırılmalıdır. Bu şekilde bulunabilecek kaynaklardan veriler hazır olarak elde edilebileceği gibi, veri toplama araçlarının tasarımı için faydalı bilgiler de elde edilebilir.

1.5.2. Gözlem

Gözlem; kendiliğinden oluşan ya da bilinçli olarak hazırlanan olayları, belirdikleri sırada sistemli ve amaçlı bir biçimde inceleyerek bilgi toplama yöntemidir. Biz mezhebe mensup insanların dini törenlerini incelemek için bilgi toplayan bir araştırmacı, o mezhebe mensup insanların dini ayinlerine katılarak ya da bu ayinleri izleyerek bilgi toplayabilir (İslamoğlu, 2009, s.104).

Gözlem Yönteminin Avantajları:

- Anket yöntemine göre daha objektiftir,
- Hızlı bilgi toplamaya olanak sağlar,
- Uygulanması kolaydır.

Gözlem Yönteminin Dezavantajları:

- Her tür bilginin toplanmasına uygun değildir,
- Mekanik ve elektronik araçların kullanılmaması halinde, olayın ayrıntılarını yakalamak veya hatırlamak zordur,
- Gözlem altında olduklarını anlayanlar, olağan davranışlarını değiştirebilirler,
- Bazı durumlarda maliyetleri yüksektir.

1.5.3. Deney

Deney, bir hipotezin sınanması amacı ile koşulları deneyi yapan tarafından hazırlanan ve bağımsız değişkenin bağımlı değişken üzerindeki etkinin şiddetini ya da yönünü ortaya koymayı amaçlayan bir gözlem türüdür (İslamoğlu, 2009, s.107).

1.5.4. Anket

Anket, birincil kaynaklardan bilgi toplamak için hazırlanan sistematik bir soru formudur. Amacı, araştırmanın problemini çözecek ve ele alınan hipotezleri test edecek bilgileri sistematik bir biçimde toplamak ve saklamaktır (İslamoğlu, 2009, s.113).

1.5.4.1. Anketlerin Uygulanma Türleri

1.5.4.1.1. Yüz Yüze Görüşme

Anketörle deneklerin aynı fiziki ortamda bulunduğu, anket sorularının doğrudan sözlü olarak yöneltilmesi ve yine sözlü olarak alınan cevapların anket formuna not edilmesi şeklindeki bir veri toplama yöntemidir. Yöntemin avantajları (İslamoğlu, 2009, s.114):

- Bu yöntem hem soruları açıklığa kavuşturma hem de denetim sağlama bakımından öteki yöntemlere göre daha esnektir,
- Cevaplayıcı ile işbirliği olanağı vardır,
- Gözlem yapmaya olanak tanır,
- Cevaplayıcıdan derinlemesine bilgi edinmek için, ek soru sorulmasına imkan tanır,
- En üst düzeyde cevaplama oranı sağlar.

Yöntemin dezavantajları ise (İslamoğlu, 2009, s.114):

- Görüşmecinin hatalı davranışı görüşmeyi etkiler,
- Ücretli anketör çalıştırılıyorsa, anketör anketi, hayali bilgilerle doldurabilir,
- Maliyeti yüksek olabilir,
- Cevaplayıcıya ulaşmak çok zaman alabilir,
- Büyük çaplı araştırmalar için yeterli nitelik ve sayıda anketör bulmak zor olabilir,
- Görüşmeyi yapanın yanlı tutumu denetlenemez.

1.5.4.1.2. Telefonla anket

Anket sorularının telefonla cevaplayıcıya sorulduğu ve alınan cevapların anket formuna işlendiği bir veri toplama yöntemidir. Yöntemin avantajları (İslamoğlu, 2009, s.115) :

- Ucuzdur, kısa zamanda bilgi sağlar,
- Denetimi, yüz yüze görüşmeye göre daha kolaydır,
- Posta ile ankete göre daha esnektir.

Yöntemin dezavantajları ise (İslamoğlu, 2009, s.115):

- Herkesin telefonu olmadığı ya da herkesin telefon rehberinde ismi olmadığı için, örnekleme hatalarına düşülebilir,
- Ayrıntılı bilgi edinme olanağı sınırlıdır,
- Cevaplayıcı ile kurulacak ilişki yüz yüze görüşmede olduğu kadar sağlıklı olmayabilir,
- Uzun görüşmelere, şekil ve resim gibi yardımcı araçların kullanımına uygun değildir.

1.5.4.1.3. Posta İle Anket

Anketlerin cevaplayıcılara posta yolu ile iletildiği bir veri toplama yöntemidir. Cevaplayıcı anket formunu doldurup araştırmacıya geri gönderir. Yöntemin avantajları (İslamoğlu, 2009, s.115):

- Maliyeti, öteki yöntemler göre daha düşüktür,
- Cevaplayıcı daha bol zamana sahip olduğundan, anketi daha dikkatli cevaplayabilir,

- Anketör cevaplayıcıyı etkilemez,
- Sır sayılabilecek bilgiler daha kolay toplanır.

Yöntemin dezavantajları ise (İslamoğlu, 2009, s.116):

- Anketin geri dönüş oranı düşüktür,
- Cevaplayıcıların cevapları denetlenemez,
- Soruların tam anlaşılıp anlaşılmadıkları denetlenemez,
- Sorulara kimlerin cevap verdikleri bilinemez.

1.5.4.1.4. Internet'te Anket

Anketin bir web sitesinde online doldurulması yada e-posta yoluyla cevaplayıcılara anket formunun gönderilmesi şeklindeki veri toplama yöntemidir. Yöntemin avantajları (İslamoğlu, 2009, s.116):

- Maliyeti son derece düşüktür,
- Gerektiğinde resim ve şekil göstermeye elverişlidir,
- Hızlı bilgi toplar,
- Anketör kullanılmadığı için anketörün hatalı tutumundan etkilenmez.

Yöntemin dezavantajları (İslamoğlu, 2009, s.116):

- Herkesin internet bağlantısı olmadığından örnekleme hatalarına düşülebilir,
- Geri dönüş oranı düşüktür,
- Kimlerin cevap verdiği denetlenemez,
- Soruların doğru anlaşılıp anlaşılmadığı belli olmaz,
- Internet kullanımını bilmeyenler cevap veremez,
- Kimlerin cevapladığı belli olacağından, sır sayılabilecek sorulara cevap verilmek istenmeyebilir.

1.6. Örnekleme

Bir araştırmada ele alınan problemi çözmek ya da gerekli hipotezleri test etmek için, ihtiyaç duyulan bilgiler ikincil kaynaklardan derlenemiyorsa, bu bilgiler birincil kaynaklardan toplanmak zorundadır. Birincil kaynaklardan bilgi toplamak ise; zaman, maliyet ve diğer nedenlerden ötürü zor bir iş olduğundan, ana kitlenin tümünü gözlem altına alma ya da ana kitleyi tam olarak sayma yerine, ana kitleyi nitelik ve nicelik yönünden temsil eden bir örneğin çekilmesi yoluna gidilir. Böylece örnekten elde edilen bilgilerin, belirli olasılık kademelerinde ana kitle için de geçerli olduğu kabul edilebilir. Bu anlamda örnekleme; ana kitleyi nitelik ve nicelik yönünden temsil edebilecek bir kümenin çekilmesi işlemidir (İslamoğlu, 2009, s. 159).

Örneklemenin amacı tahmin yapmaktır. Tahminde bulunma, örnekleme yoluyla alınacak kısıtlı ve dar bir bilgiye dayanarak daha geniş çaptaki verilerin karakteristikleri, özellikleri hakkında genelleme yapmaktır. Örneğin, bir çuval dolusu pirinç içinden alınacak bir avuç pirincin incelenmesiyle tüm çuval dolusu pirinç hakkında çeşitli özellikleri itibariyle genelleme yapılabilir. Burada, çuvaldaki toplam pirinç miktarı ana kitle (popülasyon)'dir. İncelemek amacıyla aldığımız bir avuç dolusu pirinç ise örnektir (Çakıcı ve diğ., 2000, s. 148).

Pozivist yaklaşımın öngördüğü örnekleme, örnekleme kuramına uygun olan örneklemedir. Dolayısıyla, ana kitleyi her bakımdan temsil edecek en doğru örneği elde etmeyi amaçlar. Bu da örnekleme kuramındaki örnekleme süreçlerine tam olarak uymakla sağlanır (İslamoğlu, 2009, s. 160).

Bu anlayışa göre örnekleme süreci şu aşamalardan oluşmaktadır (İslamoğlu, 2009, s. 160):

- Ana kitlenin tanımlanması,
- Örnekleme çerçevesinin belirlenmesi,
- Örnek bireylerinin belirlenmesi,
- Örnekleme yönteminin tanımlanması,
- Örnek büyüklüğünün saptanması,
- Örnekleme planının belirlenmesi,
- Örneklerin seçimi.

1.6.1. Ana Kitlenin Tanımı

Bir araştırmada, çözüm aranan problemle ya da test edilecek hipotezlerle ilgili sağlıklı ve doğru bilgilerin kimlerden toplanabileceğinin belirlenmesi işlemine ana kitlenin tanımlanması denir (İslamoğlu, 2009, s. 160).

1.6.2. Örnekleme Çerçevesi

Ana kitleden öngörülen miktarda tesadüfi olarak örnek çekilebilmesi için, ana kitleyi kapsayan bir listenin elde bulunması gerekir. Sözgelimi, doktorlar ya da avukatlardan bilgi toplanacaksa, tabipler odasından ya da barodan, mesleğe kayıtlı olanların listesi alınabilir. Bir mahalleye ilişkin liste ise, muhtarlıklardan elde edilebilir (İslamoğlu, 2009, s. 162).

1.6.3. Örnek Bireylerinin Belirlenmesi

Örnek bireylerinin belirlenmesinde en kritik nokta, ana kitleyi oluşturan bireylerin örneğe girme ihtimalini eşit hale getirememe durumunda, sistematik hatanın kaçınılmaz olarak ortaya çıkmasıdır. Örnekleme kuramı, ana kitleyi oluşturan her bireyin örneğe girme şansının eşit olmasını öngörür (İslamoğlu, 2009, s. 162).

1.6.4. Örnekleme Yöntemi

Araştırmanın güvenilirliği bir de, aranan bilginin özelliği gereği örnekleme yönteminin doğru seçilmesine bağlıdır. Bu, aynı zamanda araştırmanın maliyeti ile de ilgilidir. Örnekleme yöntemleri önce tesadüfi ve tesadüfi olmayan olmak üzere iki gruba ayrılır (İslamoğlu, 2009, s. 162).

Örnekleme kuramına göre, örneklerden elde edilen bilgilerin matematik ve istatistik tekniklerle test edilip genelleme yapılabilmesi için örneklemenin tesadüfi örneklemeye uygun olarak yapılması gerekir. Ancak, bu tesadüfi olmayan örneklemenin işe yaramaz olduğu anlamına gelmemelidir. Araştırmanın amacına, türüne, maliyetine, kapsamına ve genelleme iddiasına göre tesadüfi olmayan örnekleme yöntemleri de kullanılabilir (İslamoğlu, 2009, s. 162).

1.6.4.1. Tesadüfi Örnekleme Yöntemleri

Ana kitle hakkında tahminde bulunurken yapılan tahminin geçerli olabilmesi için örneklemenin tesadüfi (rastgele) olması gerekir. Tesadüfi örnekleme kavramı, örnekte yer alacak elemanların belirlenmesinde hiçbir dış etkinin rolünün olmamasını ifade eder. Diğer bir ifadeyle, bir ana kitleden herhangi bir örnek alan bir kişi, örnekte bulunsun veya bulunmasın hiçbir eleman veya birim üzerinde etkiye sahip olmayacaktır (Çakıcı ve diğ, 2000, s. 149).

1.6.4.1.1. Basit Tesadüfi Örnekleme

Bu yöntemin kullanılabilmesi için, ele alınan sorun ya da hipotezlerle ilgili bilgilerin ana kitleye göre homojen (türdeş) olması gerekir (İslamoğlu, 2009, s. 162).

Basit tesadüfi örneklemede, ana kitledeki her birimin örneğe girme olasılığını eşit hale getirmek esastır. Bunu sağlamak için değişik yollara başvurulabilir. Ana kitle listesi bilgisayara yüklenerek buradan rastgele çekiliş yapılabilir. Bireylere birer numara verilerek bu numaralar bir torbaya konarak rastgele bu numaralar çekilebilir (İslamoğlu, 2009, s. 164). Rastgele sayılar tablosu kullanılarak da örneğe alınacak birimler seçilebilir.

Yöntemin Yararlı Yönleri

- ✓ Evrendeki her elemanın eşit seçilme şansı vardır,
- ✓ Evren çok büyük ve karmaşık değilse seçme işlemi kolaydır,
- ✓ Bu yöntemle yapılan örneklemede istatistiksel işlemler ağırlıksız olarak yapıldığı için değerlendirme işleminde kolay olur.

Yöntemin Sakıncalı Yönleri

- ✓ Evren çok büyükse evreni listelemek ve seçmek güçtür,
- ✓ İncelenen özellik evrendeki elemanların bazı özelliklerine göre değişiklik gösterebilir,
- ✓ Örnekleme seçilecek bireyler çok geniş bir bölgede dağınık bir şekilde yerleşmiş olabilirler,
- ✓ Evrende azınlıkta olan grupların yeterince temsil edilmemesi söz konusu olabilir.

1.6.4.1.2. Zümrelere Göre Örnekleme (Tabakalı Örnekleme)

Elde edilecek bilgi, ana kitle itibari ile türdeş değilse yani, bilgi ana kitleyi oluşturan değişik özellikteki gruplara göre farklılık gösteriyorsa, zümrelere göre örnekleme yöntemi kullanılır. Ana kitle zümrelere (tabakalara) ayrılırken dikkat edilmesi gereken husus, elde edilen bilgi ya da ele alınan sorun hangi kriterlere göre farklılık gösteriyorsa, zümrelerin de bu kriterlere göre oluşturulmasıdır. Sözgelimi, ele alınan sorun ya da sorunu çözecek bilgiler sosyo-ekonomik ya da sosyo-kültürel faktörlere göre farklı ise, zümreler de bu faktörlere göre oluşturulmalıdır (İslamoğlu, 2009, s. 164). Zümreler belirlendikten sonra, her zümreden belirli sayıda birim rastgele örnekleme ile çekilmelidir.

Her tabakaya eşit sayıda birey düşmesi olanaksız olacağından, her tabakadan kaç bireyin örnekleme alınacağı sorunu çıkar. Bu durumda iki yol izlenebilir. Birincisinde, tabakalardaki birey sayısı göz önüne alınmadan her tabakadan eşit sayıda birey örnekleme alınır. Buna *orantısız seçim* denir. *Orantısız seçimde istatistiksel değerlendirmenin kesinlikle ağırlıklı olarak yapılması gerekir.* İkincisinde ise, örnekleme alınacak bireyleri tabakalardaki birey sayısına orantılı olarak seçmektir. Başka bir deyişle, çok kişi içeren tabakadan çok, az kişi içeren tabakadan az kişiyi örnekleme almaktır.

Örneklem seçimi orantılı yapıldığında aritmetik ortalama ağırlıksız, standart sapma ise ağırlıklı olarak hesaplanır. Orantılı seçim, işlemleri kolaylaştırdığı için tercih edilen bir yoldur.

Tabakalı rastgele örnekleme yöntemine tabakalar arasında gerçek bir farklılık olduğunda başvurulmalıdır. Bu yöntemin sakıncalı yanları çok azdır. Bunlar;

- ✓ Tabakalardaki birey sayısının bilinmediği durumlarda seçim işlemlerinin güçleşmesi,
- ✓ Örnekleme seçilecek birimlerin çok büyük bir bölgede dağınık olarak oturması durumunda araştırmanın uygulama aşamasının güçleşmesidir.

1.6.4.1.3. Kümelere Göre Örnekleme

Bu örnekleme yönteminde, en güvenilir örneği elde etme yerine, en düşük maliyetle en doğru örneği elde etmek amaçlanır. Ana kitle önce kümelere ayrılır, sonra kümelerden bireylere geçilir (İslamoğlu, 2009, s. 164).

Bu tür örnekleme, sağlıklı bir ana kitle çerçevesinin elde bulunmaması ya da çok büyük ana kitleden çekim yapmanın çok zor ve maliyetinin yüksek olması durumunda uygulanır. Bu tür örnekleme tesadüfi örnekleme olmasına karşın, ana kitleyi ne ölçüde temsil ettiği tartışılabilir. Bu nedenle de çok dikkatli yapılması gerekir (İslamoğlu, 2009, s. 164).

Küme örnekleme yönteminde genel kural kümedeki birim sayısının az olması yani kümelerin küçük olmasıdır. Kümelerin küçük olması küme sayısını artıracak, bu da değişik özellikteki kümelerin örnekleme girme şansını arttıracaktır. *Örneğin* 5 000 aile içeren bir bölgeyi 1000'er ailelik 5 kümeye ayırıp buradan 1 kümeyi örnekleme alma yerine, 250'şer ailelik 20 kümeye ayırıp 4 küme seçmek daha uygundur.

1.6.4.1.4. Sistematik Örnekleme

Örneklem seçim işlemlerinin kolay olması nedeniyle özellikle evren büyük olduğunda kullanılan bir örnekleme yöntemidir. Bu yöntemin en çok kullanıldığı durumlar:

- ✓ Çok sayıda birim içeren kayıt sistemlerinin incelenmesinde. *Örneğin,* hasta dosyaları, hasta ya da işçi kayıtları, kayıt defterleri, fişler , listeler gibi.
- ✓ Birim sayısı çok fazla olduğu için listelenmesi güç ya da olanaksız olan durumlarda. *Örneğin*, büyük bir kentte ev seçimi, sokak seçimi, işyeri seçimi otomobil seçimi gibi.

Seçim işlemlerinde evren büyüklüğü(**<u>N</u>**) örneklem büyüklüğüne (**<u>n</u>**) bölünerek kaç birimde bir birimin örnekleme alınacağı saptanır. *Örneğin,* 15 000 hasta dosyası bulunan bir arşivden 500 dosya örnekleme seçilecekse (15 000 / 500 = 30) her 30 dosyada bir dosya örnekleme alınacaktır. Başlangıç sayısı rastgele sayılar tablosundan 1 – 30 arasında bir sayı seçilerek bulunur. Seçilen sayı 8 ise önce 8'inci dosya örnekleme alınır, sonra her 30 dosya 1 dosya örnekleme alınır. Böylece örnekleme çıkan dosya numaraları 8, 38, 68, 98, ……14 978 olacaktır.

<u>Bu yöntemi kullanacak araştırıcılar şu noktaları göz önünde bulundurmalıdırlar</u>

- ✓ Başlangıç sayısı dağılımı büyük oranda etkiler.Örneğin dosyalar küçük yaştan büyük yaşa doğru sıralanmışsa ve araştırıcı yaş ortalamasını öğrenmek istiyorsa 3.33.63.93…. sırasında elde edilecek ortalama ile 28,58,88,118…. Sırasından elde edilecek ortalama farklı sonuçlar vermektedir.
- ✓ Birden çok kurumda dosyalar incelenecekse ve her kurumda diyelim 30 dosya varsa ve her kurum dosyaları küçük yaştan büyük yaşa doğru sıralanmış ise başlangıç sayısı dağılımı yine etkiler.
- ✓ Birden çok kurumda dosyalar incelenecekse bir kurum dosyaları büyükten küçüğe bir kurum dosyaları küçükten büyüğe doğru sıralanmışsa bir diğeri de sırasız olarak dizilmişse araştırıcı bunların sırasını belirli bir düzene soktuktan sonra seçim işlemine geçmelidir.

1.6.4.2. *Tesadüfi Olmayan Örnekleme Yöntemleri*

Tesadüfi örneklemenin uygulanamadığı durumlarda uygulanan ölçekleme yöntemlerine verilen addır. Nicel araştırmalarda kullanılmaları, örnekleme kuramının genelleyici özelliğine uymadığından, kullanılmaları eleştirilir. Buna karşılık nitel araştırmacıların ve ön araştırma yapanların sıkça bu yöntemleri kullandıkları görülmektedir (İslamoğlu, 2009, s. 166).

1.6.4.2.1. Kolay Yoldan Örnekleme (Kolayda Örnekleme)

Bu tür örnekleme en düşük maliyetli ve uygulanması en kolay örneklemedir. Süpermarket çıkışındaki tüketicilere, sokakta dolaşan insanlara uygulanan anketler buna örnek gösterilebilir (İslamoğlu, 2009, s. 167).

1.6.4.2.2. Yargısal Örnekleme

Ana kitleyi temsil edeceği varsayımı ile yapılan örneklemedir. Yeterince deneyimlilerin ve uzman kişilerin görüşleriyle hangi örneklerin ana kitleyi temsil edebileceği belirlenir.

1.6.4.2.3. Kota Örneklemesi

Kota örneklemesinin uygulanabilmesi için, toplanacak bilginin hangi seçim kriterine göre değerlendirileceğine karar vermek gerekir. Örneğin, toplanacak bilgiler yaş, cinsiyet ya da öğrenim düzeylerine göre farklılık gösteriyorsa, bunların ana kitle dağılımına uygun örnek dağılımı elde edilmelidir. Sözgelimi, yüksek öğretimli olanların ana kitle içindeki oranı % 10 ise, örnek içindeki dağılımı da %10 olmalıdır (İslamoğlu, 2009, s. 168).

1.6.4.2.4. Kartopu Örneklemesi

Bu örneklemede, araştırmacı ilk adımda tamamen rastlantısal olarak seçtiği çekirdek bir örnekle yola çıkar. İkinci adımda, çekirdekte yer almış örneklerin önerileri ile yeni örneklere ulaşır ve işleme böylece devam edilir. Sözgelimi, matematik öğretimi ile ilgili bir araştırma yapacaksa, ilk adımda bulabildiği üç beş matematik öğretmeni ile görüşür ve daha sonra onlardan başkaca önerdikleri matematik öğretmenlerinin isim ve adreslerini alarak onlarla görüşür (İslamoğlu, 2009, s. 168).

1.6.5. Örnek Büyüklüğünün Belirlenmesi

Bazı durumlarda, yapılan tahminlerde belli bir doğruluk veya isabet derecesine ulaşmak hedef olarak alınabilir. Bu gibi durumlarda istenen hedefe ulaşabilmek için ne büyüklükte örnek alınması gerektiği araştırma konusudur (Çakıcı ve diğ., 2003, s.23).

Örnek büyüklüğünün hesaplanmasında aşağıdaki formüller kullanılabilir.

$$n = \left.\frac{Z^2\,\sigma^2}{(\bar{x}-\mu)^2}\right\} \textit{Ana kitle varyansı biliniyorsa}$$

$$n = \left.\frac{Z^2\,s^2}{(\bar{x}-\mu)^2}\right\} \textit{Ana kitle varyansı bilinmiyorsa}$$

Z : belirlenen güven düzeyi için Z tablosundan bakılan değer,
σ^2: ana kitle varyansı,
$\bar{x}$: örnek ortalaması,
μ : ana kitle ortalaması,
$(\bar{x}-\mu)$: göze alınan örnekleme hatası.

Örnek : Bir iş kolunda çalışan işçilerin ortalama saat ücretleri gerçek ana kitle ortalama saat ücretinden en fazla 100 TL sapma gösterecek şekilde ve %90 güven aralığında tahmin edilmek isteniyor. Geçmiş kayıtlara dayanarak bu iş kolu için hesaplanan standart sapma $\sigma = 500\ TL$ olduğuna göre örnek büyüklüğü ne olmalıdır.

$$n = \frac{z^2\sigma^2}{(\bar{x}-\mu)^2} = \frac{(1,64)^2(500)^2}{(100)^2} = \frac{(2,69)(250.000)}{10.000} = 67,25 \cong 68$$

Bu sonuca göre, gerçek ortalamanın $\mp$100 TL sınırları içinde ve % 90 güven aralığında istenen amacı gerçekleştirmek için örnek büyüklüğü 68 olmalıdır.

1.6.6. Örnekleme Planının Belirlenmesi

Örnek büyüklüğü de belirlendikten sonra örnekleme planı oluşturulur. Bu planda hangi bilgilerin kimler tarafından, ne zaman kimlerden toplanacağı ve örnekleme için ayrılan bütçe belirlenir. Araştırmada görev yapan herkes toplayacağı bilgileri, bilgi toplama yerlerini, hangi bilgileri toplayacaklarını, bilgi toplayacağı kişileri ve kullanabilecekleri kaynakları bu planda görebilmelidir. Araştırma bu plan çerçevesinde gerçekleştirilir.

1.6.7. Örnekleme Hataları

Örnekleme yöntemlerine dayalı olarak yapılan tahminlerde iki tip hata ile karşılaşılır. Birinci tip hata, tesadüfi hata olarak adlandırılır. Bu tür hatalar örnek sayısı arttırılarak giderilebilir. Sistematik hata olarak adlandırılan ikinci tip hatalar ise; örnekleme sürecindeki hatalardan kaynaklanır ve sonradan giderilmeleri mümkün değildir. Bu hatalar (İslamoğlu, 2009, s. 168):

- Örnekleme yönteminin doğru seçilmeyişinden,
- Ana kitlenin yanlış tanımlanmasından,
- Örnekleme çerçevesinin yanlış belirlenmesinden,
- Örneklerin doğru çekilmeyişinden,
- Örnek büyüklüğünün doğru hesaplanmayışından kaynaklanırlar.

2. VERİLERİN ANALİZİ VE SUNUMU

2.1. Kalitatif ve Kantitatif Verilerin Özetlenmesi ve Sunumu

Toplanan ham veriler sınıflandırılmadan hiçbir anlam taşımazlar. Veriler kalitatif ya da kantitatif olabilir. Kalitatif veriler için nümerik ya da nümerik olmayan kategoriler elde edilebilir. Kategoriler kodlarla tanımlanabilir. Kalitatif veriler, nominal ve ordinal ölçekte; kantitatif veriler, aralık yada oran ölçeğinde ölçülebilirler (Tütek ve Gümüşoğlu, 2008, s.9).

Kalitatif ve kantitatif veriler tablo, grafik biçiminde ya da nümerik olarak özetlenebilirler (Tütek ve Gümüşoğlu, 2008, s.9)

2.1.1. Sıklık (Frekans) Dağılımı

Çeşitli yollarla toplanan veriler, özellikleri hakkında bilgi edinmek amacıyla, düzenlemeler yapılarak özet halinde sunulurlar. Uygulamalarda veri kümeleri çok sayıda gözlem içerebilmektedir. Bu nedenle verilerin özetlenmesi, ilgilenilen olay ya da problem açısından ilk yapılacak iş ve son derece önemli bir işlemdir. Verilerin düzenlendiği çizelgelere ***sıklık (frekans) çizelgeleri***, verilerin gösterdiği dağılıma ***sıklık (frekans) dağılımı*** denir. Verilerin yapısına (nitel, nicel, vb...) göre sıklık çizelgeleri düzenlenir (Balce ve Demir, 2007, s. 7).

2.1.1.1. Nicel Verilerde Sıklık (Frekans) Dağılımı

Bir araştırma sonunda elde edilen sürekli nicel veriler, düzenlenmemiş ham ya da sınıflandırılmamış verilerdir. Aşağıda örnek olarak sunulan veriler düzenlenmemiş ham verilerdir. Konunun daha iyi anlaşılması için bu örnek üzerinden uygulamalar yapılacaktır. **ÖRNEK:** Bir finans analisti, bilgisayar donanım ve yazılım şirketlerinin Araştırma-Geliştirme(AR-GE) faaliyetlerine ayırdıkları kaynak miktarıyla ilgilenmektedir. Bu analist yüksek teknolojiye sahip 50 firmayı örneklem olarak belirlemiş ve bir önceki yıl gelirlerinden AR-GE'ye ayırdıkları kaynak miktarlarını (1000 TL) elde etmiştir. Analistin amacı bu veri kümesini özetleyerek bir takım bilgilere ulaşmaktır. Veriler Tablo 1'de verilmiştir (Balce ve Demir, 2007, s. 7).

Tablo1. Firmaların AR-GE faaliyetlerine ayırdıkları kaynak miktarları

Firma No	Miktar	Firma No	Miktar	Firma No	Miktar	Firma No	Miktar	Firma No	Miktar
1	13.5	**11**	8.0	**21**	8.2	**31**	9.6	**41**	7.1
2	8.4	**12**	7.9	**22**	8.0	**32**	7.2	**42**	13.2
3	10.5	**13**	6.8	**23**	7.7	**33**	8.8	**43**	7.7
4	9.0	**14**	9.5	**24**	7.4	**34**	11.3	**44**	5.9
5	9.2	**15**	8.1	**25**	6.5	**35**	8.5	**45**	5.2
6	9.7	**16**	13.5	**26**	9.5	**36**	9.4	**46**	5.6
7	6.6	**17**	9.9	**27**	8.2	**37**	10.5	**47**	11.7
8	10.6	**18**	6.9	**28**	6.9	**38**	6.9	**48**	6.0
9	10.1	**19**	7.5	**29**	7.2	**39**	6.5	**49**	7.8
10	7.1	**20**	11.1	**30**	8.2	**40**	7.5	**50**	6.5

Kaynak: (Balce ve Demir, 2007, s. 7)

Bu verilerin kaynak miktarına göre küçükten büyüğe doğru sıralanmış şekli, Tablo2'de verilmektedir.

Tablo2. Sıralanmış Kaynak Miktarı Verileri

Firma No	Miktar	Firma No	Miktar	Firma No	Miktar	Firma No	Miktar	Firma No	Miktar
45	5.2	**28**	6.9	**43**	7.7	**35**	8.5	**9**	10.1
46	5.6	**38**	6.9	**49**	7.8	**33**	8.8	**3**	10.5
44	5.9	**10**	7.1	**12**	7.9	**4**	9.0	**37**	10.5
48	6.0	**41**	7.1	**11**	8.0	**5**	9.2	**8**	10.6
25	6.5	**29**	7.2	**22**	8.0	**36**	9.4	**20**	11.1
39	6.5	**32**	7.2	**15**	8.1	**14**	9.5	**34**	11.3
50	6.5	**24**	7.4	**21**	8.2	**26**	9.5	**47**	11.7
7	6.6	**19**	7.5	**27**	8.2	**31**	9.6	**42**	13.2
13	6.8	**40**	7.5	**30**	8.2	**6**	9.7	**1**	13.5
18	6.9	**23**	7.7	**2**	8.4	**17**	9.9	**16**	13.5

Kaynak: (Balce ve Demir, 2007, s. 7)

Sıklık Çizelgesinin Elde Edilişi

Sıklık çizelgesinin elde edilişinde kullanılan tanımlar adımsal olarak verilerek yukarıdaki örnekle bu tanımlar pekiştirilmeye çalışılacaktır (Balce ve Demir, 2007, s.7).

Dağılım Sınırları: Bir dağılımda (veri kümesinde) yer alan en küçük ve en büyük denek değerleridir (Balce ve Demir, 2007, s. 7).

En büyük değer (Maksimum): 13.5 (dağılımın üst sınırı)
En küçük değer (Minimum) : 5.2 (dağılımın alt sınırı)

Dağılım Genişliği (DG): Dağılım sınırları arasındaki farktır (Balce ve Demir, 2007, s.7).

DG=En büyük değer - En küçük değer =13.5 - 5.2 = 8.3

Sınıf: Eşit ya da birbirine yakın değerli deneklerin oluşturduğu her bir gruba SINIF denir. Sınıf sayısı, **k** ile gösterilir. Sınıf sayısının genellikle 7 ile 20 arasında olması istenir. Araştırmacı tarafından belirtilen sınıf sayısı, çok sayıda veri olduğunda aşağıda verilen Sturges'in formülü ile de bulunabilir (Balce ve Demir, 2007, s. 7).

k=1+3.3log(n)

Sınıfın Alt Sınırı: Bir sınıfta yer alan en küçük değerdir (Balce ve Demir, 2007, s. 7).
Sınıfın Üst Sınırı: Bir sınıfta yer alan en büyük değerdir (Balce ve Demir, 2007, s. 7).
Sınıf Aralığı: Ard Arda gelen iki sınıfın alt sınırları ya da üst sınırları arasındaki farktır. Sınıf aralığı, c ile gösterilir. Örneğimizde sınıf sayısı 8 olarak alınsın. Bu durumda Sınıf Aralığı (Balce ve Demir, 2007, s. 7):

$$c = \frac{DG + a}{SINIFSAYISI}$$

a: veri kümesindeki verilerin ondalık kısmındaki hane sayısı ile ilgilidir. Örneğimizde tam kısımdan sonra 1 hane olduğu için a=0.1 alınır (Balce ve Demir, 2007, s. 7).

$$c = \frac{8.3 + 0.1}{8} = \frac{8.4}{8} = 1.05$$

Sınıf sayısı ve aralığı belirlendikten sonra, **ilk sınıfın alt sınırı saptanır**. Bu değer genellikle dağılımın en küçük değeridir. Sınıf aralığı (c=1.1) ard arda eklenerek **diğer sınıfların alt sınırları** bulunur. **İlk sınıfın üst sınırı** ise ikinci sınıfın alt sınırının **son hanesinden** 1 çıkarılarak bulunur. Diğer sınıfların üst sınırları da ardı ardına sınıf aralığı eklenerek bulunur (Balce ve Demir, 2007, s. 7).

Sıklık: Bir sınıfta yer alan denek sayısı o sınıfın sıklığıdır. f ile gösterilir. Sıklıklar toplamı denek sayısına eşittir (Balce ve Demir, 2007, s. 7).

$$\sum_{i=1}^{n} f_i = n$$

Örneğimizde ise k=8 tane sınıf olduğundan $\sum_{i=1}^{8} f_i = 50$ 'dir.

Sınıflar oluşturulduktan sonra sınıfların sıklıkları bulunur. Bunun için önce her sınıfın sıklığı işaretleme (çeteleme) ile bulunur. İşaretlerin sayısı **sınıf sıklıklarını** verir. Seçilen **sınıf sayısının uygun olup olmadığını** tespit etmek için son sınıfta en büyük denek değerinin yer alıp almadığın bakılmalıdır. Eğer son sınıfta en büyük denek değeri yer almıyorsa sınıf sayısı daha büyük olmalıdır. En büyük denek değeri son sınıftan bir önceki sınıfta da yer alabilir. Bu durumda seçilen sınıf sayısı veriler için büyüktür. Bu iki durumda sınıf sayısı artırılır ya da azaltılır (Balce ve Demir, 2007, s. 7).

Sınıf (Orta) Değeri (m): Bir sınıfın alt ve üst sınırlarının ortalaması o sınıfın sınıf değeri ya da sınıf orta değeridir. Sınıf değeri bir sınıfı tek bir değerle temsil eder ve m ile gösterilir (Balce ve Demir, 2007, s. 7).

$$m_i = \frac{AS_i + ÜS_i}{2} \quad , \ i = 1,2,\ldots,k$$

AS_i: i.sınıfın alt sınırı
$ÜS_i$: i.sınıfın üst sınırı

Göreli Sıklık(Sıklık Yüzdesi): Her sınıfa düşen denek sayısının toplam denek sayısına göre yüzdesidir. Göreli Sıklıklar p_i ile gösterilir. Toplamları 1 olmalıdır (Balce ve Demir, 2007, s. 7).

$$p_i = \frac{f_i}{n} \ , \ i = 1,2,\ldots,k \qquad \sum_{i=1}^{n} p_i = 1$$

Çizelge 1. AR-GE faaliyetleri için ayrılan kaynak miktarı verileri için sıklık çizelgesi

Sınıf	Alt Sınır (AS)	Üst Sınır (ÜS)	Sınıf Orta Değeri (m_i)	Çeteleme	Sıklık (f_i)	Göreli Sıklık ($p_i=f_i/n$)
1	5,2	6,24	(5,20+6,24)/2=5,72	////	4	4/50=0,08
2	6,25	7,29	6,77	///// ///// //	12	0,24
3	7,3	8,34	7,82	///// ///// ///	13	0,26
4	8,35	9,39	8,87	/////	5	0,10
5	9,4	10,44	9,92	///// //	7	0,14
6	10,45	11,49	10,97	/////	5	0,10
7	11,5	12,54	12,02	/	1	0,02
8	12,55	13,59	13,07	///	3	0,06
				Toplam	**50**	**1,00**

Kaynak: (Balce ve Demir, 2007, s. 7)

Çizelge 1 için örnek yorumlar:

İkinci sınıfta AR-GE'ye 6.25 ile 7.29 bin YTL arasında yatırım yapan firmalar yer alır ve bu sınıftaki firmalar ortalama 6.77 bin YTL AR-GE'ye yatırım yapar.

AR-GE'ye yatırım yapan 50 firmanın 12 tanesi ya da %24'ü AR-GE'ye ortalama 6.77 bin TL yatırım yapmışlardır.

AR-GE’ye yatırım yapan firmaların %24’ü AR-GE’ye ***tahminen*** ortalama 6.77 bin TL yatırım yapmaktadırlar. (kitle için yorum)

Sınıf Ara Değerleri (SA): Sınıflar arasındaki değerlerdir. Birinci sınıfın üst sınırı ile ikinci sınıfın alt sınırının ortalaması, birinci sınıf ile ikinci sınıf arasındaki sınıf ara değerini verir. İlk sınıfın ara değeri birinci sınıftan önce bir sınıf varmış diye kabul edilerek, hesaplanın birinci ve ikinci sınıflar arasındaki değerinden sınıf aralığı (c) çıkarılarak bulunur. Son sınıf ara değeri ise, sanki son sınıftan sonra bir sınıf daha varmış gibi kabul edilerek, hesaplanan son sınıf ara değerine sınıf aralığı (c) eklenerek bulunur. Sınıf ara değerlerinin sayısı (k+1)’dir (Balce ve Demir, 2007, s. 7).

Birikimli Sıklık: Sınıf sıklıklarının üst üste eklenmesi ile oluşan sıklıklar birikimli sıklıklardır (Balce ve Demir, 2007, s. 7).

Den Daha Az Birikimli Sıklık: *Sınıf ara değerinden* daha az değeri olan sınıf sıklıklarının birinci sınıftan başlayarak eklenmesi ile elde edilir (Balce ve Demir, 2007, s. 7).

Den Daha Çok Birikimli Sıklık: *Sınıf ara değerinden* daha çok değeri olan sınıf sıklıklarının birinci sınıftan başlayarak eklenmesi ile elde edilir (Balce ve Demir, 2007, s. 7).

Bir sınıf ara değerine karşı gelen den daha az ve den daha çok birikimli sıklıklar toplamı denek sayısına eşittir. Birikimli sıklıklar denek sayısına oranlanırsa **den daha çok birikimli sıklık yüzdeleri** ve **den daha çok birikimli sıklık yüzdeleri** elde edilir. *Birikimli sıklık yüzdeleri sınıf ara değerinden daha az ya da daha çok büyük değerli denek değerlerinin yüzdesini verir* (Balce ve Demir, 2007, s. 7).

Alt sınır, üst sınır, sınıf değeri, sıklık ve göreli sıklık bilgilerinin oluşturduğu çizelgeye Sıklık Çizelgesi (Çizelge 1) ve Alt sınır, üst sınır; sınıf ara değeri, birikimli den daha az ve den daha çok sıklıklar ve birikimli den daha az ve den daha çok sıklık yüzdelerinden oluşan çizelgeye Birikimli Sıklık Çizelgesi denir (Çizelge 2) (Balce ve Demir, 2007, s. 7).

Çizelge2. AR-GE faaliyetleri için ayrılan kaynak mikatarı verileri için Birikimli Sıklık Çizelgesi

Sınıf	AS	ÜS	Sınıf Ara Değeri	f_i	Birikimli Sıklıklar Den Daha Az	Birikimli Sıklıklar Den Daha Çok	Birikimli Sıklık Yüzdeleri Den Daha Az	Birikimli Sıklık Yüzdeleri Den Daha Çok
			5.145		0	50	0.00	1.00
1	5.20	6.24		4				
			6.245		4	46	0.08	0.92
2	6.25	7.29		12				
			7.295		16	34	0.32	0.68
3	7.30	8.34		13				
			8.345		29	21	0.58	0.42
4	8.35	9.39		5				
			9.345		34	16	0.68	0.32
5	9.40	10.44		7				
			10.445		41	9	0.82	0.18
6	10.45	11.49		5				
			11.495		46	4	0.92	0.08
7	11.50	12.54		1				
			12.545		47	3	0.94	0.06
8	12.55	13.59		3				
			13.595		50	0	1.00	0.00

Kaynak: (Balce ve Demir, 2007, s. 7)

Çizelge 2 için örnek yorumlar:

50 firmadan sadece 4 tanesi ya da %8'i AR-GE'ye 6.245 bin YTL'**den daha az** yatırım yapmıştır.

AR-GE'ye yatırım yapan firmalardan tahmini olarak %8'i AR-GE'ye 6.245 bin YTL'**den daha az** yatırım yapmaktadır. (kitle için yorum)

50 firmadan sadece 46 tanesi ya da %92'si AR-GE'ye 6.245 bin YTL'**den daha çok** yatırım yapmıştır.

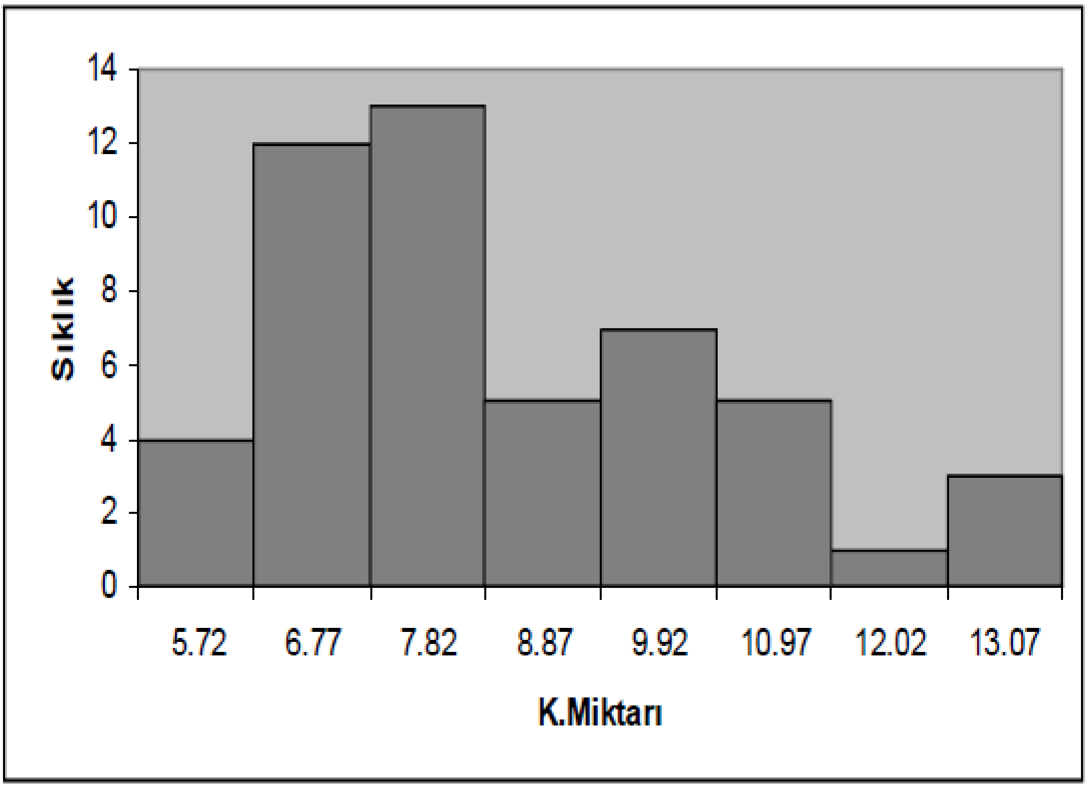
AR-GE'ye yatırım yapan firmalardan tahmini olarak %92'i AR-GE'ye 6.245 bin YTL'**den daha çok** yatırım yapmaktadır. (kitle için yorum)

Dikkat: Rakamların geldiği yerlere dikkat ederek benzer yorumlar diğer sınıflar içinde yapılabilir (Balce ve Demir, 2007, s. 7).

Veriler ondalıklı değil ise sınıf ara değerinin anlamı yoktur. Çizelge2'deki sınıf ara değeri kolonu olmaz ve birikimli sıklık ve birikimli sıklık yüzdeleri alt sınır ve üst sınır değerlerine göre yorumlanır (Balce ve Demir, 2007, s. 7).

2.1.2. Histogram

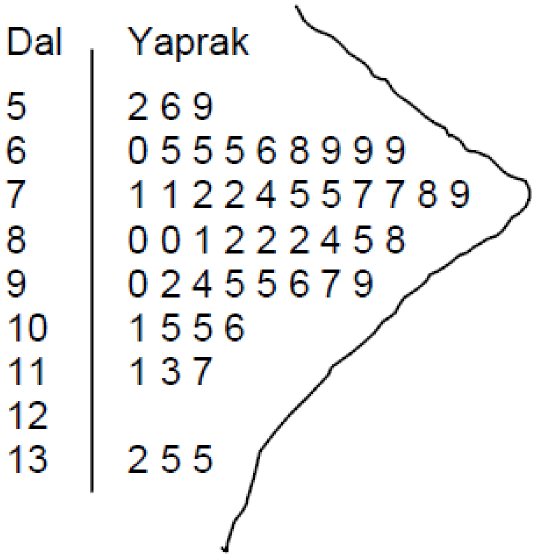
Kantitatif verilerin en yaygın gösterim biçimi histogram ya da çubuk grafiğidir. İlgilenilen değişken yatay eksende, frekans yada göreli frekans düşey eksende gösterilir. Her sınıfın frekans yada göreli frekansı; tabanı sınıf genişliği, yüksekliği ise sınıfın frekans yada göreli frekansı olan dikdörtgen ile gösterilir (Tütek ve Gümüşoğlu, 2008, s.13).



Çizelge1'de verilen göreli sıklık ve sınıf değerleri kullanılarak histogram çizilmiş ve yukarıda verilmiştir. Tablodaki bilgiler görsel olarak bu histogramdan da yorumlanabilir. Histogramdan veri dağılımının sağa çarpık olduğu gözlenmektedir. Büyük olasılık ile verilerin ortalaması ortancadan büyüktür. Eğer ortalama ve ortanca hesaplanırsa bu kolaylıkla görülebilir (Balce ve Demir, 2007, s. 12).

DAL YAPRAK ÇİZİMİ:

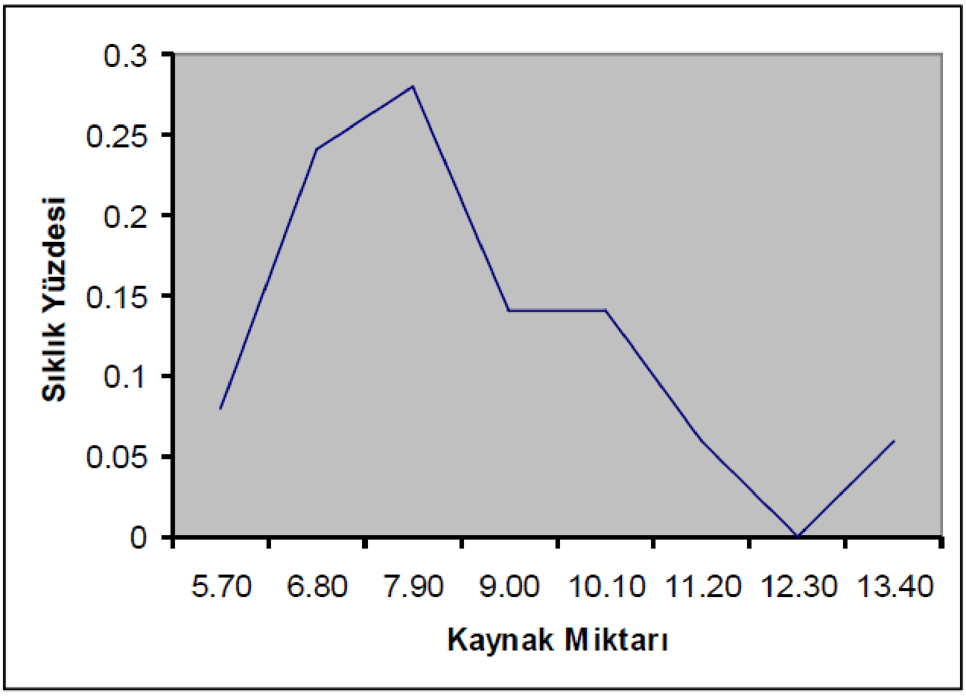
Veri kümesinin görsel olarak bir diğer gösterimi Dal-yaprak çizimidir.



Dal kısmında ondalıktan önceki verinin tam kısmı, yaprak kısmına ise ondalıktan sonraki kısım yazılarak dal-yaprak çizimi gerçekleştirilir. Benzer yorumlar bu çizimden de yapılır.Veri kümesinin sağa çarpıklığı buradan da gözlenebilir. 50 firmanın çoğunluğunun %9.9'dan daha az AR-GE'ye yatırım yaptığı söylenebilir. Bu durumdaki firma sayısı ise 40 tanedir. Grafikten yorumlar subjektif'tir. Diğer benzer ya da farklı yorumlar üretmek mümkündür (Balce ve Demir, 2007, s. 12).

2.1.3. Frekans (Dağılım) Poligonu

Frekans poligonunun oluşturulması için sınıfların orta noktaları kullanılır.



2.1.4. Pasta Grafikler

Kantitatif verilerin oran ya da yüzde şeklinde sunulmasında kullanılan bir grafik türüdür.

NİTEL VERİLERDE SIKLIK ÇİZELGESİ

SINIFLANABİLEN VERİLERDE SIKLIK ÇİZELGESİ

Sınıflanan verilerde, sınıflar bağımsız olarak elde edildiği için her sınıfa düşen denek sayıları sıklık çizelgesini oluşturur. Sınıflar bağımsız olduğu için sıklık çizelgesinde sadece sınıf ve sıklık, göreli sıklık kolonları yer alır (Balce ve Demir, 2007, s. 14).

ÖRNEK: 2003 turizm sezonunda ülkemize gelen 800 turistin ülkeye gelişte yararlandıkları taşıt türlerine göre dağılımının sıklık çizelgesi aşağıdaki çizelgede verilmiştir.

Çizelge 3. 800 turistin yararlandıkları taşıt türlerine göre sıklık çizelgesi

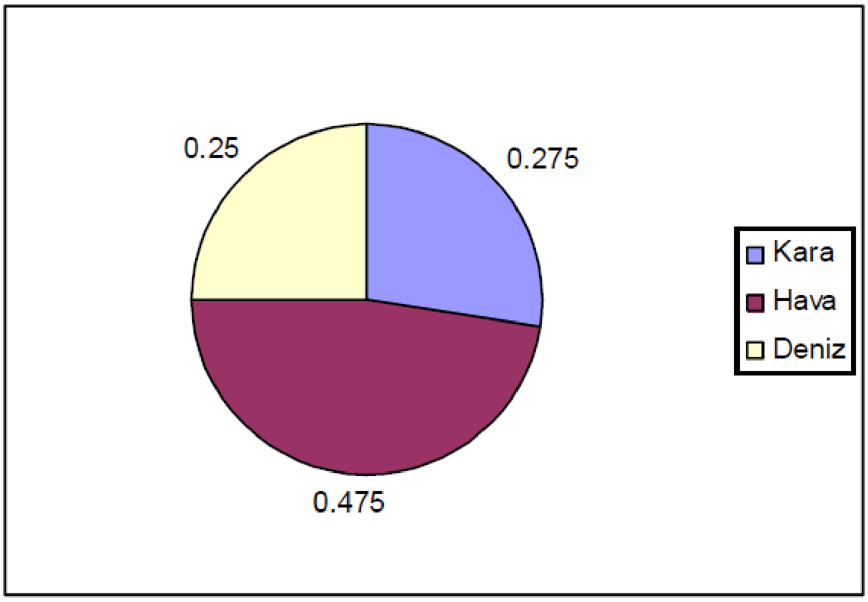
Taşıt Türü	Sıklık (f_i)	Göreli Sıklık ($p_i=f_i/n$)
Kara Taşıtı	220	0.275
Hava Taşıtı	380	0.475
Deniz Taşıtı	200	0.250
Toplam	800	1.000

Yorumlar:

2003 turizm sezonunda ülkemize gelen 800 turistin %47.5'i (380 tanesi) hava yolu ile seyahat etmişlerdir.

2003 turizm sezonunda ülkemize gelen turistlerin yaklaşık %25'inin deniz yolu ile seyahat ettikleri tahmin edilmektedir (Balce ve Demir, 2007, s. 14).

Pasta Grafik / Daire Dilimleri Grafiği



2.2. Sıralanabilen Verilerde Sıklık Çizelgesi

Veriler belli bir sıralama ölçütüne göre sıralanabilen sınıflara ayrılır ve her sınıfa düşen deneklerin sayısı saptanırsa sıralanabilen verilerde sıklık çizelgesi elde edilir (Balce ve Demir, 2007, s. 15).

ÖRNEK: Ekonomi dersinin genel sınav sonuçlarının sıklık çizelgesi aşağıdaki çizelgede verilmiştir.

Çizelge 4. 77 öğrencinin Ekonomi sınav sonuçlarının sıklık çizelgesi

Notlar	Sıklık (f_i)	Göreli Sıklık ($p_i=f_i/n$)	Ve Daha Az Birikimli Sıklığı	Ve Daha Çok Birikimli Sıklığı
F	29	0.377	29 (%37.7)	77 (%100)
C	31	0.402	60 (%77.9)	48 (%62.3)
B_2	7	0.091	67 (%87.0)	17 (%22.0)
B_1	6	0.071	73 (%94.8)	10 (%12.9)
A_2	3	0.039	76 (%98.7)	4 (%5.2)
A_1	1	0.013	77 (%100)	1 (%1.3)
	77	**1.000**		

Yukarıdaki sıklık çizelgesinden yararlanarak bazı yorumlar yapılabilir: Ekonomi dersini alan öğrencilerin %3.9'u dersi A2 notu ile başarmıştır. Ekonomi dersini alan öğrencilerin %62.3'ü C **ve daha yüksek** not almıştır.

EŞİT OLMAYAN ARALIKLI VE AÇIK UÇLU SIKLIK ÇİZELGESİ

Eşit olmayan aralıklı sıklık çizelgeleri, uçlara doğru sıklık yoğun olduğunda ya da dağılım aşırı derecede çarpıklaştığında, uçların birinde ayrıntıların yok olduğu diğerinde gereksiz olduğu durumlarda düzenlemelidir. Örneğin, gelir dağılımı eşit olmayan aralıklı çizelgesi oluşturulur. Düşük gelirlerde sıklık yayılması olmasına karşın yüksek gelirlerde sıklıklar azalır (Balce ve Demir, 2007, s. 15).

Eşit olmayan aralıklı sıklık çizelgelerinde, sınıflandırmada yorum ve grafik çiziminde kolaylık sağlanması için aralıklar en küçük aralığın katları olarak alınmalı ve değişik aralık sayısı da az olmalıdır (Balce ve Demir, 2007, s. 15).

Dağılım sınırları belli olmayan verilerin sıklık çizelgeleri açık uçlu düzenlenir. Açık uçlu sıklık çizelgeleri, sınıf değerlerine bağlı olan hesaplamalarda güçlük yaratabilir. Böyle durumlarda kesin olmayan sınıf değerleri tahmin edilir. Açık uçlu sınıf sıklıkları küçük olduğu için, tahmin edilen değerlerden ötürü hatalı sonuç bulma olasılığı küçük olabilir (Balce ve Demir, 2007, s. 15).

ÖRNEK: Bir ildeki 30 ilkokulun derslik sayılarının dağılımı incelensin.

Derslik Sayısı:	5	6	7	8	9	10	11	12	13	14	15
İlkokul Sayısı:	6	8	0	1	0	9	0	3	0	2	1

Verilerin, derslik sayısı 7, 9 11 ve 13 olan ilkokul sayısı sıfır olduğu için eşit olmayan aralıklı sıklık çizelgesi düzenlenebilir:

Derslik Sayısı	İlkokul Sayısı
5	6
6	8
7-9	1
10-12	12
13-15	3

ÖRNEK: 108 kişinin yetenek test puanlarına göre dağılımı, açık uçlu sıklık çizelgesine örnektir:

Puan	Sıklık(f_i)
100'den az	7
100-119	11
120-139	24
140-159	36
160-179	19
180-199	8
200 ve çok	3

2.3. Merkezi Eğilim Ölçüleri

Konum ölçüleri, verilerin dağılımdaki yerlerini, birbirlerine olan uzaklıklarını kısacası konumlarını belirlemek için kullanılan ölçülerdir. Bu ölçüler Tablo1'de verilen veriler üzerinde uygulanarak aşağıda tek tek ele alınmaktadır.

2.3.1. Aritmetik Ortalama

Aritmetik ortalama istatistikte ve günlük hayatta çok kullanılan ve bilinen bir merkezi eğilim ölçüsüdür. Bir seri verinin aritmetik ortalaması ve sadece ortalaması denildiği zaman gözlemlerin merkezi yerinde bulunan değer anlaşılmaktadır (Çakıcı ve diğ., 2000, s.31).

2.3.1.1. Basit Serilerde Aritmetik Ortalama

Bu değeri elde etmek için gözlem değerleri toplanarak gözlem sayısına bölünür (Çakıcı ve diğ., 2000, s.31).

$$\overline{x} = \sum_{i=1}^{n} x_i / n$$

Örnek 1: 15, 12, 14, 18, 16 sayılarının aritmetik ortalaması:

$$\bar{x} = \frac{15+12+14+18+16}{5} = \frac{75}{5} = 15 \; olarak \; bulunur.$$

2.3.1.2. Sınıflandırılmış Serilerde Aritmetik Ortalama

Sınıflandırılmış verilerde aritmetik ortalama:

$\bar{x} = \frac{\sum_{i=1}^{n} f_i . m_i}{\sum_{i=1}^{n} f_i}$ formülüne göre hesaplanır :

Örnek 2 :

Sınıflar	**Frekanslar (f_i)**	**Sınıf Orta Değerleri (m_i)**	**$f_i.m_i$**
10-20	3	15	3.15=45
21-31	5	26	5.26=130
32-42	8	37	8.37=296
43-53	4	48	4.48=192
54-64	2	59	2.59=118
Toplam	**22**		**781**

Bu verilere göre aritmetik ortalama :

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i . m_i}{\sum_{i=1}^{n} f_i} = \frac{781}{22} = 35{,}5 \text{ bulunur.}$$

2.3.2. Tepe Değeri (Mod)

2.3.2.1. Sınıflandırılmamış verilerde tepe değeri

Sınıflandırılmamış verilerde tepe değeri en sık tekrar eden değerdir.

Örnek 3 : 15, 22, 18, 19, 15, 18, 20, 15 serisi için tepe değerini bulmak için her değerin frekansı (sıklığı) bulunur:

Değer	**Frekans**
15	**3**
18	2
19	1
20	1
22	1

En fazla tekrarlanan değer 15'dir (3 kez) bu nedenle bu dağılımın tepe değeri (mod) 15'dir.

2.3.2.2. Sınıflandırılmış verilerde ise tepe değeri

Sınıflandırılmış verilerde ise tepe değeri aşağıdaki gibi hesaplanır.

Mod= $A_s + c.\frac{d_1}{d_1+d_2}$

A_s : $En\ büyük\ sıklığın\ bulunduğu\ sınıfın\ alt\ sınırı$

d_1: $En\ büyük\ sıklık - bir\ önceki\ sıklık$

d_2: $En\ büyük\ sıklık - bir\ sonraki\ sıklık$

c : Sınıf aralığı

Örnek 4 :

Veriler (x_i)	**Frekanslar (f_i)**
10-20	3
21-31	5
32-42	8
43-53	4
54-64	2
Toplam	**22**

$A_s = 32, d_1 = 8 - 5 = 3, d_2 = 8 - 4 = 4$ olduğuna göre:

$Mod = A_s + c.\frac{d_1}{d_1+d_2} = 32 + 10.\frac{3}{3+4} = 32 + 4{,}28 = 36{,}28$

2.3.3. Ortanca (Medyan)

2.3.3.1. Sınıflandırılmamış serilerde ortanca

Sınıflandırılmamış serilerde ortanca hesaplanırken, veriler öncelikle küçükten büyüğe doğru sıralanırlar. n çift ise n/2. değer ile n/2+1. değerlerin ortalaması, n tek ise $\frac{(n+1)}{2}$. değer ortancadır.

Örnek 5 : 25, 12, 32, 18, 28, 25,42, 26 serisinin ortanca değerini bulunuz.

Öncelikle sayılar sıralanır:

12, 18, 25, 25, 26, 28, 32, 42

n=8 olduğu için n/2=8/2=4. Ve n/2+1=8/2+1=4+1= 5. Değerlerin ortalaması alınır:

4.değer=25, 5.değer=26 olduğundan

Ortanca=(25+26)/2=25,5 bulunur.

Örnek 6 : 25, 12, 32, 18, 28, 25,42 serisinin ortanca değerini bulunuz.

Öncelikle sayılar sıralanır:

12, 18, 25, 25, 28, 32, 42

n=7 olduğu için (n+1)/2=(7+1)/2=8/2=4.değer ortancadır:

ortanca=25

2.3.3.2. Sınıflandırılmış Serilerde Ortanca

Sınıflandırılmış verilerde ortanca aşağıdaki gibi hesaplanır.

$$Medyan = L + C.\frac{\left(\frac{n}{2}\right)-F}{f_{med}}$$

$L = Medyan\ sınıfının\ alt\ sınırı$

$c = Sınıf\ aralığı$

$n = Dağılımda\ değişkene\ ait\ değerlerin\ sayısı$

$F = Medyan\ sınıfından\ önceki\ sınıfların\ birikimli\ frekansı$

$f_{med} = Medyan\ sınıfının\ frekansı$

Medyanı kapsayan sınıfa medyan sınıfı adı verilir.

Örnek 7 :

Veriler (x_i)	**Frekanslar (f_i)**	**Birikimli Frekanslar**
10-20	3	3
21-31	5	8
32-42	8	16
43-53	4	20
54-64	2	22
Toplam	**22**	

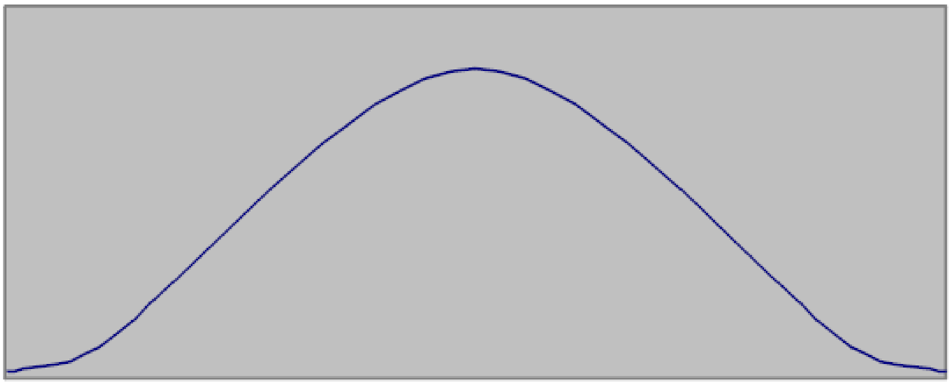
n=22 çift sayı olduğu için n/2=22/2=11.eleman medyandır. 11.elemanı içeren sınıf medyan sınıfıdır. Bu durumda medyan sınıfı 32-42 sınıfıdır.

L=32, c=10, F=8, f_{med}=8 olduğuna göre:

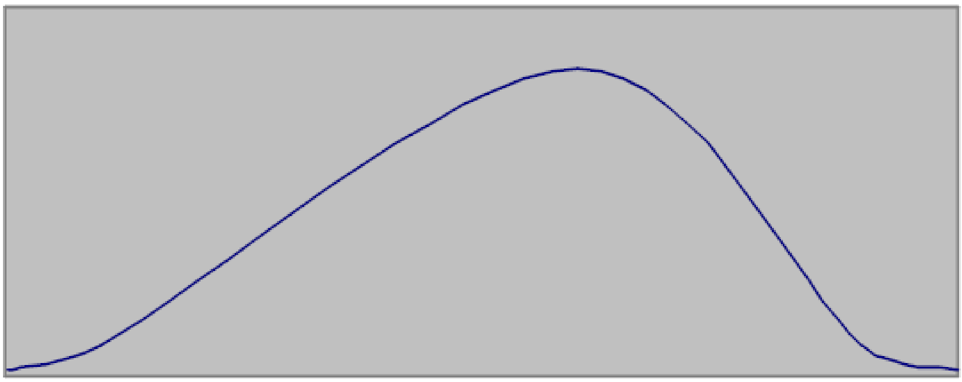
$$Medyan = L + C.\frac{\left(\frac{n}{2}\right) - F}{f_{med}} = 32 + 10.\frac{\left(\frac{22}{2}\right) - 8}{8} = 32 + 10.\frac{11 - 8}{8} = 32 + \frac{30}{8}$$
$$= 32 + 3{,}75 = 35{,}75\ bulunur.$$

2.3.4. Ortalama, ortanca ve tepe değeri arasındaki bağıntı

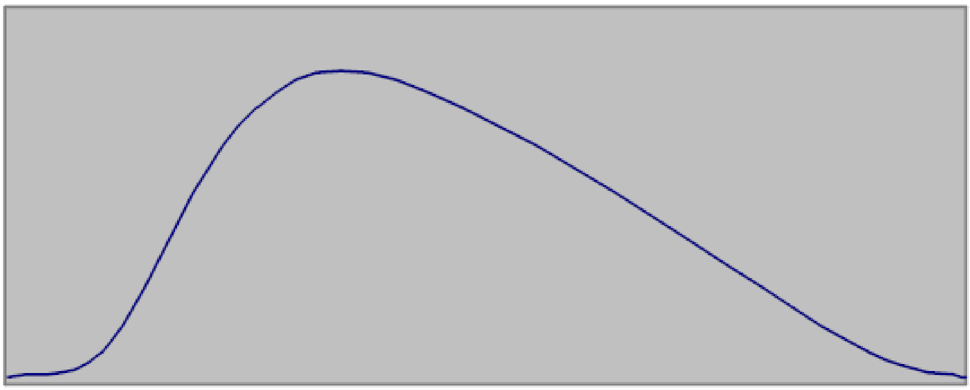
Ortalama=Mod=Medyan ise sıklık dağılımı simetrikdir.



Ortalama<Medyan<Mod ise dağılım sola çarpıktır.



Ortalama>Medyan>Mod ise dağılım sağa çarpıktır.



Örneğimiz için hesaplanan ortalama=35,5; mod=36,28; medyan=35,75 değerleri dikkate alındığında, ortalama<medyan<mod olduğu görülür. Yani dağılım sola çarpıktır.

2.4. Dağılım Ölçüleri

Konum ölçüleri veri kümesinin ya da dağılımının merkezi hakkında bilgi varmektedir. Yalnızca bu ölçülere bakılarak verinin dağılımı hakkında tam bir bilgi sahibi olmak mümkün değildir. Veriler ne derecede dağılmakta ya da yayılım göstermektedirler?; ortalamadan uzaklıkları ne kadardır? Gibi sorulara cevap vermek için değişim ölçüleri hesaplanmalıdır.

ÖRNEK 8 : 2 firmanın 7 ustabaşısının yıllık kazançları aşağıda veilmiştir.

A FİRMASI: 345 307 329 360 341 338 325
B FİRMASI: 349 275 316 397 353 338 317

İki firmaya ilişkin ortalama ve ortanca değerleri hesaplandığında eşit bulunmaktadır.

Ortalama= 335 ve Ortanca (Medyan) =338.

Konum ölçülerine baktığımızda 2 firma arasında bir farklılık görünmüyor. Gerçekte verilerin yayılımına dikkat ettiğimizde, A firmasında çalışanların birbirlerine daha yakın kazanç elde ettiklerini, buna karşın B firmasındakilerin ise kazançların birbirlerinden daha uzak olduğunu görmekteyiz. B firmasına çalışanların kazançları daha fazla farklılık göstermektedir. Az sayıdaki bu verilere bakarak bu sonuca varabiliyoruz. Ancak veri sayısının fazla olduğu durumda, bakarak bir sonuça ulaşmamız imkansızlaşacaktır.

2.4.1. DAĞILIM GENİŞLİĞİ (ARALIK)

Bir veri kümesindeki en büyük değer ile en küçük değer arasındaki farktır.

DG= Maksimum – Minimum

Örnek 8’deki A ve B Firmaları için Dağılım genişlikleri :

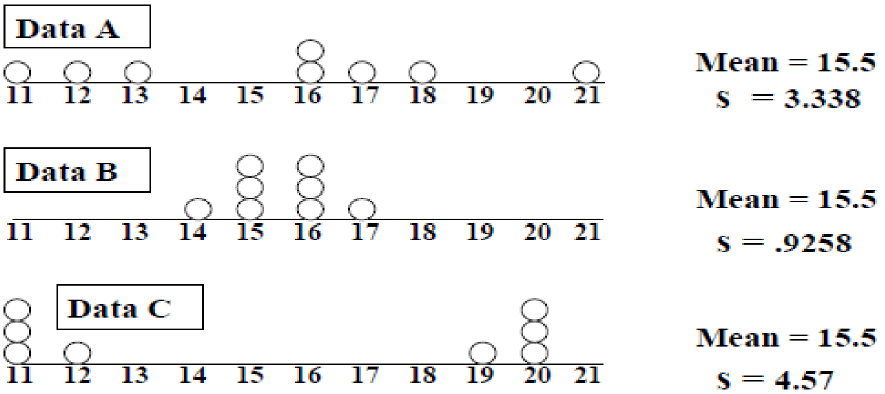
A Firması için DG=360 – 307=53

B Firması için DG= 397 – 275=122 bulunur.

2.4.2. ORTALAMA MUTLAK SAPMA:

Verilerin ortalamadan sapmalarını gösteren bir dağılım ölçüsüdür. Değişkenlik düzeyinin anlaşılması için kullanılır.

Aşağıda ortalamaları aynı olan üç ayrı dağılım görülmektedir:



Üç dağılımında ortalamasının 15,5 olduğu görülmektedir. Ancak ilk dağılım ile son dağılımda verilerin ortalamadan uzak olduğu, ikinci dağılımda ise verilerin ortalamaya çok yakın olduğu görülmektedir. Bu durum, ikinci dağılımın diğerlerine göre daha homojen olduğunu göstermektedir.

Burada, dağılımın yapısı hakkında sadece ortalamanın verdiği bilginin yetersiz olduğu anlaşılmaktadır. Bu nedenle dağılımın değişkenliğini gösteren bazı dağılım ölçüleri geliştirilmiştir. Bu değişkenlik ölçüleri: Ortalama mutlak sapma, varyans, standart sapma ve değişim katsayısı ölçüleridir.

2.4.2.1. Basit Serilerde Ortalama Mutlak Sapma

Verilerin ortalamadan farklarının (sapmalarının) mutlak değerlerinin ortalamasıdır.

$$OMS = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$$

Örnek 8'deki A Firması için Ortalama Mutlak Sapma :

Öncelikle aritmetik ortalama bulunur:

$$\bar{x} = \frac{345+307+329+360+341+338+325}{7} = \frac{2345}{7} = 335$$

Daha sonra her verinin ortalamadan farkları bulunur ve mutlak değerleri toplanır.

Yıllık Kazanç (x_i)	$x_i - \bar{x}$	$\lvert x_i - \bar{x}\rvert$
345	345 – 335 = 10	10
307	307 – 335 = - 28	28
329	329 – 335 = - 6	6
360	360 – 335 = 25	25
341	341 – 335 = 6	6
338	338 – 335 = 3	3
325	325 – 335 = -10	10
Toplam	**0**	**88**

Ortalama Mutlak Sapma :

$$OMS = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n} = \frac{88}{7} \cong 12{,}57 \text{ bulunur.}$$

Örnek 8'deki B Firması için Ortalama Mutlak Sapma :

Öncelikle aritmetik ortalama bulunur:

$$\bar{x} = \frac{349+275+316+397+353+338+317}{7} = \frac{2.345}{7} = 335$$

Daha sonra her verinin ortalamadan farkları bulunur ve mutlak değerleri toplanır.

Yıllık Kazanç (x_i)	$x_i - \bar{x}$	$\lvert x_i - \bar{x}\rvert$
349	349 – 335 = 14	14
275	275 – 335 = -60	60
316	316 – 335 = -19	19
397	397 – 335 = 62	62
353	353 – 335 = 18	18
338	338 – 335 = 3	3
317	317 – 335 =-18	18
Toplam	**0**	**194**

Ortalama Mutlak Sapma :

$$OMS = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n} = \frac{194}{7} \cong 27{,}71 \text{ bulunur.}$$

A ve B firmalarının verileri aynı ortalamaya sahip olmasına karşın, verilerin ortalama mutlak sapmalarının farklı olduğu görülmektedir. A firmasının ortalama mutlak sapmasının (12,57) B firmasının ortalama mutlak sapmasından (27,71) daha küçük olduğu görülüyor. Bunun anlamı A firmasının verilerinin ortalamadan daha az sapması, yani değişkenliğinin daha az olmasıdır.

2.4.2.2. Sınıflandırılmış Serilerde Ortalama Mutlak Sapma

$$OMS = \frac{\sum_{i=1}^{n} f_i|m_i - \bar{x}|}{\sum_{i=1}^{n} f_i}$$

Örnek 9 : Bir işletmenin son 30 yılına ilişkin satış gelirleri için düzenlenen frekans tablosu aşağıdaki gibidir. Bu tabloya göre ortalam mutlak sapma değerini hesaplayınız.

Satış Gelirleri (x_i)	Frekans (f_i)
30 – 32	1
33 – 35	3
36 – 38	5
39 – 41	2
42 – 44	1

Öncelikle dağılımın aritmetik ortalaması hesaplanmalıdır. Bunun içinde sınıf orta değerlerinin ve $f_i m_i$ değerlerinin hesaplanması gerekir:

Satış Gelirleri (x_i)	Frekans (f_i)	Sınıf Orta Değerleri (m_i)	$f_i m_i$
30 – 32	1	$\frac{30+32}{2}=\frac{62}{2}=31$	1x31=31
33 – 35	3	$\frac{33+35}{2}=\frac{68}{2}=34$	3x34=102
36 – 38	5	$\frac{36+38}{2}=\frac{74}{2}=37$	5x37=185
39 – 41	2	$\frac{39+41}{2}=\frac{80}{2}=40$	2x40=80
42 – 44	1	$\frac{42+44}{2}=\frac{86}{2}=43$	1x43=43
Toplam	**12**		**441**

$$\bar{x}=\frac{\sum_{i=1}^{n} f_i . m_i}{\sum_{i=1}^{n} f_i}=\frac{441}{12}=36{,}75$$

Daha Sonra sınıf orta noktaları ile ortalama arasındaki farklar ve mutlak değerleri hesaplanır:

Satış Gelirleri (x_i)	Frekans (f_i)	SOD (m_i)	$m_i-\bar{x}$	$\lvert m_i-\bar{x}\rvert$	$f_i\lvert m_i-\bar{x}\rvert$
30 – 32	1	31	31-36,75 = -5,75	5,75	1x5,75 = 5,75
33 – 35	3	34	34-36,75 = -2,75	2,75	3x2,75 = 8,25
36 – 38	5	37	37-36,75 = 0,25	0,25	5x0,25 = 1,25
39 – 41	2	40	40-36,75 = 3,25	3,25	2x3,25 = 6,50
42 – 44	1	43	43-36,75 = 6,25	6,25	1x6,25 = 6,25
Toplam	**12**				**28**

Ortalama Mutlak Sapma : $OMS=\frac{\sum_{i=1}^{k} f_i\lvert m_i-\bar{x}\rvert}{\sum_{i=1}^{k} f_i}=\frac{28}{12}\cong 2{,}33$

2.4.3. VARYANS ve STANDART SAPMA

Bir veri dağılımındaki değişimin önemli bir ölçüsü varyanstır. Varyansın karekökü alınarak standart sapma elde edilir.

2.4.3.1. Basit Serilerde Varyans Ve Standart Sapma

Varyans :

Örnek İçin : $S^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$

Ana Kitle İçin : $\sigma^2=\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{N}$

S^2: Örnek Varyansı,

σ^2: Ana Kitle Varyansı

Örnek 8'deki A Firması için Varyans :

Öncelikle aritmetik ortalama bulunur:

$$\bar{x} = \frac{345+307+329+360+341+338+325}{7} = \frac{2345}{7} = 335$$

Daha sonra her verinin ortalamadan farkları ve bu farkların karesi bulunur. Bu değerler toplanır.

Yıllık Kazanç (x_i)	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
345	345 – 335 = 10	100
307	307 – 335 = -28	784
329	329 – 335 = -6	36
360	360 – 335 = 25	625
341	341 – 335 = 6	36
338	338 – 335 = 3	9
325	325 – 335 =-10	100
Toplam	**0**	**1.690**

Varyans :

$$S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{1.690}{7-1} = \frac{1.690}{6} \cong 281,67$$ bulunur.

Daha önce A firması için hesaplanan ortalama mutlak sapma ile karşılaştırıldığında (*1.257,14*), varyansın daha büyük olduğu görülmektedir. Bunun nedeni ortalamadan sapmaların daha fazla cezalandırılmasıdır.

Örnek 8'deki B Firması için Ortalama Mutlak Sapma :

Öncelikle aritmetik ortalama bulunur:

$$\bar{x} = \frac{349+275+316+397+353+338+317}{7} = \frac{2.345}{7} = 335$$

Daha sonra her verinin ortalamadan farkları bulunur ve mutlak değerleri toplanır.

Yıllık Kazanç (x_i)	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
349	349 – 335 = 14	196
275	275 – 335 = -60	3.600
316	316 – 335 = -19	361
397	397 – 335 = 62	3.844
353	353 – 335 = 18	324
338	338 – 335 = 3	9
317	317 – 335 =-18	324
Toplam	**0**	**8.658**

Ortalama Mutlak Sapma :

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{8.658}{7-1} = \frac{8.658}{6} = 1.443$$ bulunur.

Görüldüğü gibi A firmasının varyansı (281,67), B firmasının varyansından (1.443) daha küçüktür. Bu durumda A firmasının değerlerinin daha homojen olduğu söylenebilir.

Standart Sapma :

Örnek İçin : $S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$

Ana Kitle İçin : $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{N}}$

Örnek 8'deki A Firması için Standart Sapma :

Standart Sapma:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{1.690}{7-1}} = \sqrt{\frac{1.690}{6}} \cong \sqrt{281,67} \cong 16,78$$ bulunur.

Örnek 8'deki B Firması için Standart Sapma :

Standart Sapma:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{8.658}{7-1}} = \sqrt{\frac{8.658}{6}} = \sqrt{1.443} \cong 38$$ bulunur.

Örneklem için neden (n-1)'e bölünüyor ?

20 koltuk bulunan bir salona öğrenciler sırasıyla girsinler. İlk öğrencinin 20 koltuktan birini seçme serbestliği var. İkinci öğrencinin kalan 19 koltuktan birini seçme serbestliği var. Bu şekilde devam edilirse, 19. öğrencinin kalan 2 koltuktan birini seçme serbestliği var.

Ancak 20. öğrencinin koltuk seçme serbestliği kalmamıştır. Demek ki 19 öğrencinin seçme serbestliği varken 1 öğrencinin yoktur.

2.4.3.2. Sınıflandırılmış verilerde Varyans ve Standart Sapma

Varyans :

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

Örnek 9'daki işletmenin verilerine göre varyans değerini hesaplayınız.

Satış Gelirleri (x_i)	Frekans (f_i)
30 – 32	1
33 – 35	3
36 – 38	5
39 – 41	2
42 – 44	1

Öncelikle dağılımın aritmetik ortalaması hesaplanmalıdır. Bunun içinde sınıf orta değerlerinin ve $f_i m_i$ değerlerinin hesaplanması gerekir:

Satış Gelirleri (x_i)	Frekans (f_i)	Sınıf Orta Değerleri (m_i)	$f_i m_i$
30 – 32	1	$\frac{30+32}{2} = \frac{62}{2} = 31$	1x31=31
33 – 35	3	$\frac{33+35}{2} = \frac{68}{2} = 34$	3x34=102
36 – 38	5	$\frac{36+38}{2} = \frac{74}{2} = 37$	5x37=185
39 – 41	2	$\frac{39+41}{2} = \frac{80}{2} = 40$	2x40=80
42 – 44	1	$\frac{42+44}{2} = \frac{86}{2} = 43$	1x43=43
Toplam	**12**		**441**

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i . m_i}{\sum_{i=1}^{n} f_i} = \frac{441}{12} = 36,75$$

Daha Sonra sınıf orta noktaları ile ortalama arasındaki farklar ve farkların karesi hesaplanır:

Satış Gelirleri (x_i)	Frekans (f_i)	SOD (m_i)	$m_i - \bar{x}$	$(m_i - \bar{x})^2$	$f_i(m_i - \bar{x})^2$
30 – 32	1	31	31-36,75 = -5,75	33,06	1x33,06=33,06
33 – 35	3	34	34-36,75 = -2,75	7,56	3x7,56=22,68
36 – 38	5	37	37-36,75 = 0,25	0,06	5x0,06= 0,30
39 – 41	2	40	40-36,75 = 3,25	10,56	2x10,56=21,12
42 – 44	1	43	43-36,75 = 6,25	39,06	1x39,06=39,06
Toplam	**12**				**116,22**

Varyans : $S^2 = \frac{\sum_{i=1}^{k} f_i (m_i - \bar{x})^2}{\sum_{i=1}^{n} f_i} = \frac{116,22}{12} = 9,685$ bulunur.

Standart Sapma :

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

Örnek 9'daki işletme için Standart Sapma :

Standart Sapma:

$$S = \sqrt{\frac{\sum_{i=1}^{k} f_i (m_i - \bar{x})^2}{\sum_{i=1}^{n} f_i}} = \sqrt{\frac{116,22}{12}} = \sqrt{9,685} \cong 3,112$$ bulunur.

2.4.4. Değişim Katsayısı

Standart sapma ölçü biriminin etkisindedir. Dolayısıyla hesaplanan standart sapmanın küçük olması birimlerin homojen (bir birine benzeyen) olması, büyük olması birimlerin heterojen (bir birine benzemeyen) olması anlamına gelmez. Bir ana kitledeki birimlerin homojen olup olmadığının belirlenmesinde serinin standart sapması kadar o gözlem kümesinin ölçü birimi ve aritmetik ortalaması da önem kazanır.

Aşağıda verilen iki seriyi karşılaştıralım:

Birinci Seri
1
2
3
4
$\overline{x_1} = 2,5$
$\sigma_1 = 1,118$

İkinci Seri
1000
2000
3000
4000
$\overline{x_2} = 2.500$
$\sigma_2 = 1.118$

Bu iki seri dağılım açısından birbirinin aynısıdır. İkinci serideki değerler birinci serideki değerlerin 1.000 katıdır. Ancak standart sapmalar karşılaştırıldığında ikinci serinin standart sapmasının birinci serinin standart sapmasından daha büyük olduğu görülmektedir. Standart sapmalara göre birinci serinin daha homojen olduğu söylenebilir. Ancak bu doğru değildir. Çünkü birinci seride değerler kg cinsinden, ikinci seride ise gr cinsinden verilebilir. Aslında her iki seri de aynı derecede homjendir.

Değişim katsayısı standart sapmanın bu dezavantajını ortadan kaldıran ve iki veya daha fazla serinin dağılımlarını karşılaştırarak hangi serideki birimlerin daha homojen olduklarını belirlemek amacıyla kullanılan bir ölçüdür.

Değişim Katsayısı (DK) :

$DK = \frac{\sigma}{\bar{x}} \times 100$ formülü ile hesaplanır.

Yukarıdaki iki seri için DK hesaplanırsa:

Birinici seri için $\mathrm{DK}_1 = \frac{\sigma}{\bar{x}} \times 100 = \frac{1{,}118}{2{,}5} \times 100 = 44{,}72$ ‘dir.

İkinci seri için $\mathrm{DK}_2 = \frac{\sigma}{\bar{x}} \times 100 = \frac{1.118}{2.500} \times 100 = 44{,}72$ ‘dir.

Görüldüğü gibi her iki serinin DK aynı olmaktadır. Bu her iki serinin aynı ölçüde homojen olduğunu gösterir.

Başka iki seriyi inceleyelim:

Seri 1
57.022
60.954
64.435
61.167
61.018
$\overline{x_1} = 60.919{,}2$
$\sigma_1 = 2.627{,}8$

Seri 2
3.700
3.600
3.600
3.250
3.500
$\overline{x_2} = 3.530$
$\sigma_2 = 171{,}7$

$DK_1 = \frac{\sigma}{\bar{x}} \times 100 = \frac{2.627{,}8}{60.919{,}2} \times 100 = 4{,}31$;

$DK_2 = \frac{\sigma}{\bar{x}} \times 100 = \frac{171{,}7}{3.530} \times 100 = 4{,}86$ bulunur.

Bu iki değer karşılaştırıldığında seri 1’in daha homojen olduğu söylenebilir.

Örnek 8’deki A firması için hesaplanan ortalama=335, standart sapma=16,78 bulunmuştu. Aynı şekilde B firması için ortalama=335, standart sapma=38 bulunmuştu. Bu iki seriyi homojenlik açısından karşılaştırmak için değişim katsayılarını hesaplayalım.

$DK_A = \frac{\sigma}{\bar{x}} \times 100 = \frac{16{,}78}{335} \times 100 \cong 5$;

$DK_B = \frac{\sigma}{\bar{x}} \times 100 = \frac{38}{335} \times 100 = 11{,}34$ bulunur.

Değişim katsayılarına göre A firmasına ait verilerin daha homojen olduğu anlaşılmaktadır. Her iki dağılımın ortalaması aynı olduğu için karşılaştırma standart sapma değerlerine göre de yapılabilirdi. Ancak verilerin ortalamaları ve standart sapmaları arasında büyük farklılıklar olduğunda, karşılaştırma için değişim katsayısının kullanılması daha uygun olacaktır.

2.4.5. Çarpıklık ve Basıklık Katsayısı

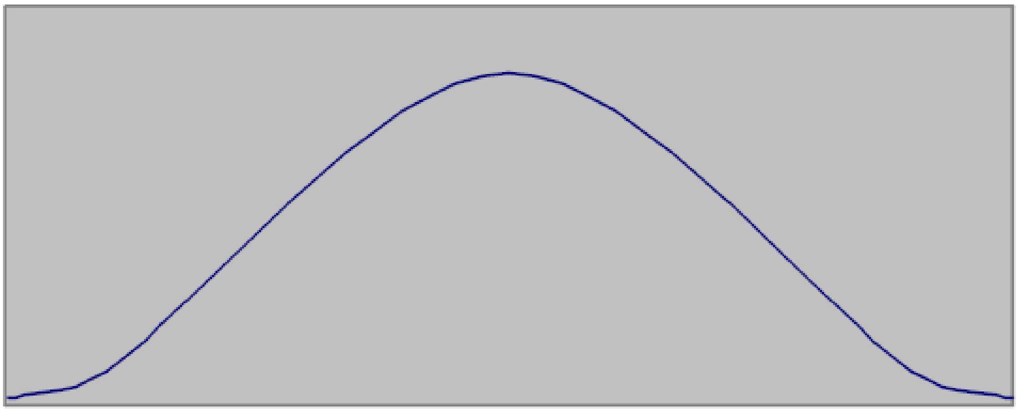
Dağılımın ortalamaya göre biçimine ilişkin bazı bilgileri çarpıklık ve basıklık katsayıları ile öğrenebiliriz.

2.4.5.1. Çarpıklık Katsayısı

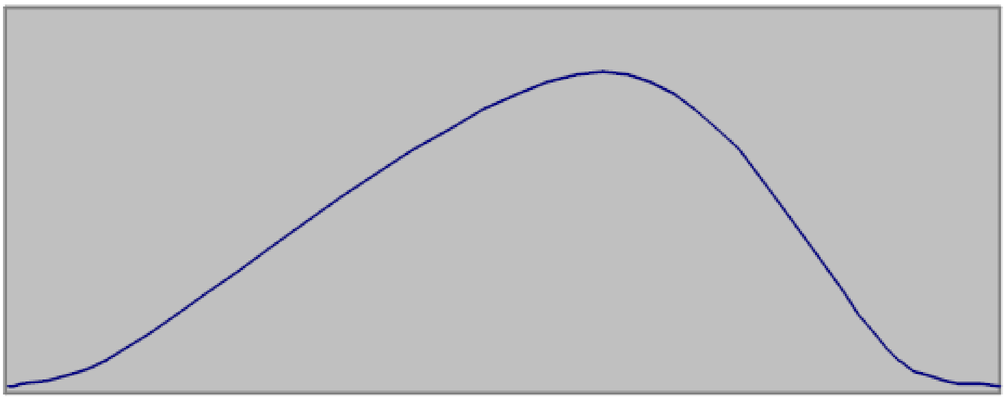
Çarpıklık katsayısı aşağıda verilen eşitlik ile elde edilebilir.

$$ÇK = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3 / n}{s^3}$$

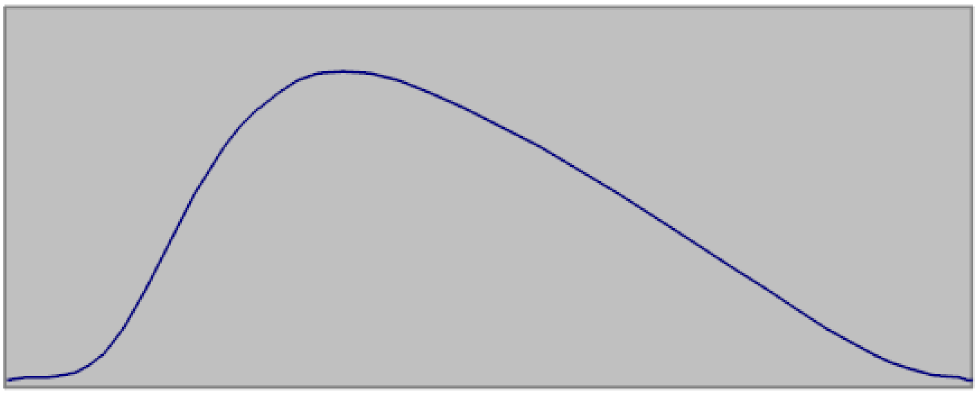
ÇK=0 ise dağılım simetriktir.



ÇK<0 ise dağılım sola doğru çarpık ya da - yöne eğilimlidir.



ÇK>0 ise dağılım sağa doğru çarpık ya da + yöne eğilimlidir.



Örnek 8'deki A Firması için Çapıklık Katsayısı :

Öncelikle aritmetik ortalama bulunur:

$$\bar{x} = \frac{345 + 307 + 329 + 360 + 341 + 338 + 325}{7} = \frac{2345}{7} = 335$$

Daha sonra her verinin ortalamadan farkları ve bu farkların karesi bulunur. Bu değerler toplanır.

Yıllık Kazanç (x_i)	$x_i - \bar{x}$	$(x_i - \bar{x})^3$
345	345 – 335 = 10	1.000
307	307 – 335 = -28	-21.952
329	329 – 335 = -6	-216
360	360 – 335 = 25	15.625
341	341 – 335 = 6	216
338	338 – 335 = 3	27
325	325 – 335 =-10	-1.000
Toplam	**0**	**-6.300**

Standart sapma (S) daha önce *16,78* olarak hesaplanmıştı. Bu durumda çarpıklık katsayısı:

$$ÇK = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3 / n}{S^3} = \frac{-6.300/7}{16,78^3} = \frac{-900}{4724,718} \cong -0,19$$

Bu çarpıklık katsayısı dağılımın sola çarpık olduğunu göstermektedir.

Örnek 8'deki B Firması için Çapıklık Katsayısı :

Öncelikle aritmetik ortalama bulunur:

$$\bar{x} = \frac{349 + 275 + 316 + 397 + 353 + 338 + 317}{7} = \frac{2.345}{7} = 335$$

Yıllık Kazanç (x_i)	$x_i - \bar{x}$	$(x_i - \bar{x})^3$
349	349 – 335 = 14	2.744
275	275 – 335 = -60	-216.000
316	316 – 335 = -19	-6.859
397	397 – 335 = 62	238.328
353	353 – 335 = 18	5.832
338	338 – 335 = 3	27
317	317 – 335 =-18	-5.832
Toplam		**18.240**

Standart sapma (S) daha önce *38* olarak hesaplanmıştı. Bu durumda çarpıklık katsayısı:

$$ÇK = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3 / n}{S^3} = \frac{18.240/7}{38^3} = \frac{2.605,714}{54.872} \cong 0,047$$

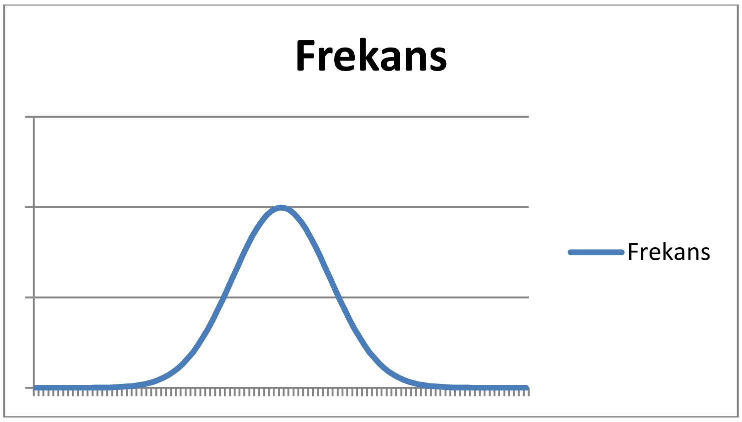
Bu çarpıklık katsayısı dağılımın sağa çarpık olduğunu göstermektedir.

2.4.5.2. Basıklık Katsayısı

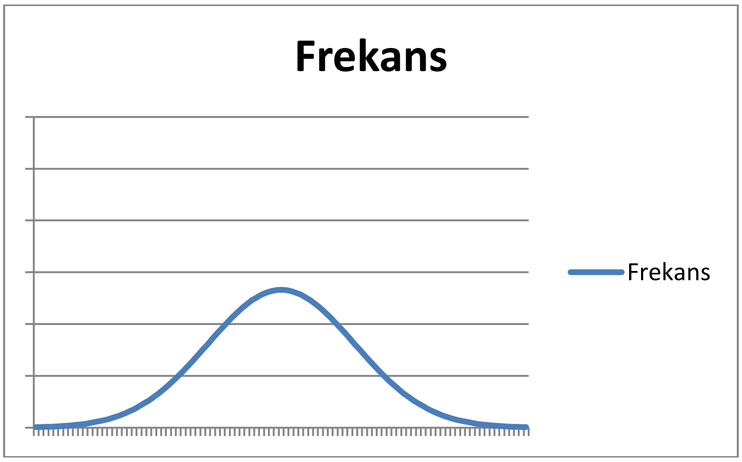
Basıklık Katsayısı ise aşağıdaki eşitlik ile hesaplanabilir:

$$BK = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4 / n}{s^4} - 3$$

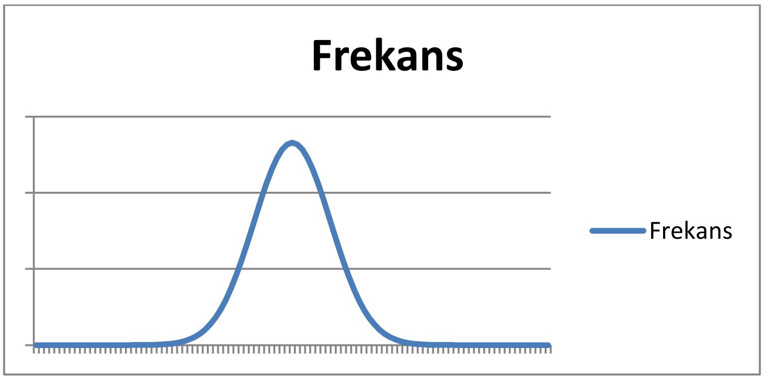
BK=0 ise dağılımın yüksekliği standart normal dağılıma uygundur.



BK<0 ise dağılım standart normal dağılımdan daha basıktır.



BK>0 ise dağılım standart normal dağılımdan daha sivridir.



Örnek 8'deki A Firması için Basıklık Katsayısı :

Öncelikle aritmetik ortalama bulunur:

$$\bar{x} = \frac{345 + 307 + 329 + 360 + 341 + 338 + 325}{7} = \frac{2345}{7} = 335$$

Daha sonra her verinin ortalamadan farkları ve bu farkların karesi bulunur. Bu değerler toplanır.

Yıllık Kazanç (x_i)	$x_i - \bar{x}$	$(x_i - \bar{x})^4$
345	345 – 335 = 10	10.000
307	307 – 335 = -28	614.656
329	329 – 335 = -6	1.296
360	360 – 335 = 25	390.625
341	341 – 335 = 6	1.296
338	338 – 335 = 3	81
325	325 – 335 =-10	10.000
Toplam	**0**	**1.027.954**

Standart sapma (S) daha önce *16,78* olarak hesaplanmıştı. Bu durumda çarpıklık katsayısı:

$$BK = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4 / n}{S^4} - 3 = \frac{1.027.954/7}{16{,}78^4} - 3 = \frac{146.850{,}57}{79.280{,}76} - 3$$

$$\cong 1{,}852 - 3 = -1{,}147$$

Bu basıklık katsayısı dağılımın standart normal dağılımdan daha basık olduğunu göstermektedir.

Örnek 8'deki B Firması için Çapıklık Katsayısı :

Öncelikle aritmetik ortalama bulunur:

$$\bar{x} = \frac{349+275+316+397+353+338+317}{7} = \frac{2.345}{7} = 335$$

Yıllık Kazanç (x_i)	$x_i - \bar{x}$	$(x_i - \bar{x})^3$
349	349 – 335 = 14	38.416
275	275 – 335 = -60	12.960.000
316	316 – 335 = -19	130.321
397	397 – 335 = 62	14.776.336
353	353 – 335 = 18	104.976
338	338 – 335 = 3	81
317	317 – 335 =-18	104.976
Toplam		**28.115.106**

Standart sapma (S) daha önce *38* olarak hesaplanmıştı. Bu durumda basıklık katsayısı:

$$BK = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4 / n}{S^4} - 3 = \frac{28.115.106 / 7}{38^4} - 3$$

$$= \frac{4.016.443{,}714}{2.085.136} - 3 \cong 1{,}93 - 3 \cong -1{,}07$$

Bu basıklık katsayısı dağılımın standart normal dağılımdan daha basık olduğunu göstermektedir. A ve B firması karşılaştırıldığında A'nın verilerinin daha basık olduğu söylenebilir.

2.4.6. Örnek 8'deki A ve B Firmalarının Verilerine İlişkin Toplu Değerlendirme

A Firması	
Ortalama	335
Medyan	338
Dağılım Genişliği	53
OMS	12,57
Varyans (S^2)	281,67
Standart Sapma (S)	16,78
Değişim Katsayısı (DK)	5
Çarpıklık Katsayısı (ÇK)	-0,19
Basıklık Katsayısı (BK)	-1,147

B Firması	
Ortalama	335
Medyan	338
Dağılım Genişliği	122
OMS	27,71
Varyans (S^2)	1.443
Standart Sapma (S)	38
Değişim Katsayısı (DK)	11,34
Çarpıklık Katsayısı (ÇK)	0,047
Basıklık Katsayısı (BK)	-1,07

A ve B firmalarının verileri aynı ortalama ve medyana sahiptir. Ancak Ortalama mutlak sapma, varyans, standart sapma ve değişim katsayıları A firmasına ait verilerin daha homojen (bir birine benzer) olduğunu göstermektedir. Çapıklık katsayıları A firmasına ait verilerin sola, B firmasına ait verilerin ise sağa çarpık dağıldığını göstermektedir. Basıklık katsayıları her iki firmaya ait verilerin standart normal dağılımdan daha basık olduğunu göstermektedir. Basıklık katsayılarına göre A firmasına ait verilerin, B firmasına ait verilerden daha basık olduğu anlaşılmaktadır.

3. OLASILIK TEORİSİ

Herhangi bir olayın meydana gelme şansını ölçmeyle ilgilenen olasılık, istatistiğin önemli bir bölümünü oluşturmaktadır. İstatistiğin çıkarsama (öngörü) temelini oluşturan olasılık, belirsizlik durumunda sağlıklı kararlar vermeyi sağladığı için, planlama çalışmalarında yoğun bir biçimde kullanılmaktadır. Örneğin bir firmanın gelecek yıldaki satış kestirimleri, bir kısmı gerçekleşecek bir kısmı gerçekleşmeyecek bir çok varsayıma dayalıdır. Bu nedenlerden dolayı olasılık kuramı, bizlere belirsizlik altında ya da mevcut bilgilerin tam ve sağlıklı olmaması gibi durumlarda doğru ve sağlıklı kararlar verebilmede yardımcı olacaktır (Balce ve Demir, 2007, s.33).

3.1. Olasılığın Tanımı

Olasılık, basit olarak, bir olayın ortaya çıkma şansıdır ve belirsizliğin ölçüsü olarak kabul edilir. Sıfır ile bir arasında farklı değerler alır; gerçekleşme şansı yüksek olan olaylar için olasılık bire yakındır; tersi durumda ise olasılık sıfıra yakındır (Tütek ve Gümüşoğlu, 2008, s.47).

Olasılık biçimi üç grupta incelenebilir (Tütek ve Gümüşoğlu, 2008, s.47):

- Klasik Olasılık
- Deney Olasılığı
- Subjektif Olasılık

3.2. Olasılık Türleri

3.2.1. Klasik Olasılık

Klasik olasılık, deney yapmadan, mantıksal bir temele dayanarak bulunur. Bir olay karşılıklı olarak birbirinin dışında ve eşit şanslı n farklı biçimde sonuçlanıyorsa ve n sonucun n_A sayıda olanı belirli bir A özelliği taşıyorsa A sonucunun ortaya çıkma olasılığı $P(A) = \frac{n_A}{n}$ ‘dır.

Örneğin, bir zarın atılışında altı gelme olasılığı 1/6’dır. Zarın atılışında olanaklı sonuç sayısı n=6’dır. Altı gelme sonucunun ortaya çıkma sayısı n_A=1 olduğundan altı gelme olasılığı$P(A) = \frac{n_A}{n} = \frac{1}{6}$ olur.

3.2.2. Deney Olasılığı (Objektif Olasılık)

İş hayatında karşılaşılan sorunlar, para yada zar atışı gibi, birbirinin dışında ve sonuçları eşit şanslı olaylar değildir. Bu nedenle olasılıkların saptanmasında deney yada gözlem yapılması zorunluluğu ortaya çıkmıştır.

Çok büyük sayıdaki deney yada gözlem sonucunda bir olayın göreli frekansına o olayın objektif (deneysel) olasılığı denir. Buna göre A olayının objektif olasılığı $P(A) = \frac{m}{n}$ olarak yazılır. Burada n, toplam yapılan deney sayısı; m, olayın gerçekleşme sayısıdır.

Örneğin, bir firmanın gelecek dönem pazarın %20’sinden fazlasını ele etme olasılığı firmanın geçmiş yıllardaki pazar payları incelenerek bulunabilir. Burada n, incelenen yıl sayısı; m, Pazar payının %20’den büyük olduğu yıl sayısıdır. n=15 ve m=5 ise firmanın gelecek yıl Pazar payının %20’den fazla olması olasılığı $P(A) = \frac{m}{n} = \frac{5}{15} = \frac{1}{3}$ ’tür.

3.2.3. Subjektif Olasılık

Subjektif olasılık, kişiden kişiye, bazı durumlarda aynı kişi için günden güne değişebilen bir olasılıktır. İşletme kararlarında çeşitli olayların gerçekleşme şansını belirlemek için geçmiş veriler elde edilemez. O zaman karar vericiler olaylara kendi bilgi, sezgi, yargı ve deneyimlerinden yola çıkarak olasılık atamak zorundadırlar.

Örneğin sigara üretimi yapan bir firmanın tesislerini büyütüp büyütmeme kararı, yakın bir gelecekte hükümetin sigara yasağına ilişkin politikasının değişip değişmeyeceği ile yakından ilgilidir. Sigara yasağının kalkmasına ilişkin olasılık subjektif bir olasılık olmak zorundadır. Çünkü sigara yasağı yakın bir geçmişte getirilmiştir ve bu yasağın kaldırılması yada sürdürülmesine ilişkin veriler bulunmamaktadır.

3.3. Bileşik Olaylar Ve Olasılık Teorileri

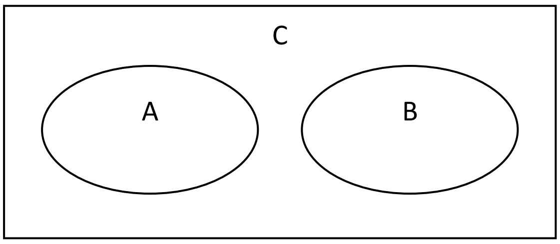
- Olasılık Sayılarının Özellikleri:
  - Herhangi bir olayla ilgili olasılık 0’dan küçük 1’den büyük olamaz. Yani $0 \leq P(A_i) \leq 1$’dir.
  - Bir deneyde olayların olasılıkları toplamı 1’e eşittir. Yani $\sum_{i=1}^{n} P(A_i) = 1$’dir.

- Bütünleme/Tamamlama Özelliği:

Herhangi bir A olayının bütünleyicisi, A olayının dışındaki tüm olayları ifade eder. Örneğin, bir madeni parada yazının bütünleyicisi turadır. Turanın bütünleyicisi ise yazıdır. A'nın dışındaki olayların gösteriliş şekli $\bar{A}$'dır. $\bar{A}$'ın olasılığı ise $P(\bar{A}) = 1 - P(A)$'ya eşittir.

- Birbirini Engelleyen/Dışlayan Olaylar:

A ve B şeklindeki iki olay bir arada gerçekleşemiyorsa bu olaylar karşılıklı olarak birbirini engelleyen olaylardır. Bu durum aşağıdaki gibi gösterilebilir.



Şekil 1: Birbirini Engelleyen Olaylar.

Şekil 1'de A ile gösterilen küme A olayının gerçekleşebileceği muhtemel yolların sayısını belirtmektedir. Aynı şekilde, B ile gösterilen küme de B olayının gerçekleşebileceği muhtemel yolların sayısını gösterir. A ve B'nin dışındaki C alanı A ve B'nin dışındaki olayların gerçekleşme sayısını gösterir. Dikkat edilirse A ve B'nin bir arada gerçekleşebileceği bir alan mevcut değildir. Şu halde, bu iki olay karşılıklı olarak birbirini engelleyen olaylardır.

- Bağımlı ve Bağımsız Olaylar:

Bağımsız olaylar birbirleri üzerinde etkili olmayan olaylardır. Bir olayın gerçekleşmesi, diğer bir olayın gerçekleşmesini etkilemiyorsa veya değiştirmiyorsa bunlara bağımsız olaylar denir. Buna karşılık, bir olayın gerçekleşmesi diğer bir olayın gerçekleşme olasılığını değiştiriyorsa bunlara bağımlı olaylar denir. Burada ikinci olayın gerçekleşmesi olasılığı birinci olayın gerçekleşmesine bağlı olarak şartlı (koşullu) bir olasılıktır.

Örnek : İçinde dört kırmızı, altı yeşil top bulunan bir kap düşünelim. Bu topların ayrı ayrı olasılıkları kırmızı top K ile, yeşil top Y ile ifade edilirse P(K)=4/10 ve P(Y)=6/10 olacaktır. Eğer çekilen kırmızı bir top tekrar kap içine konmazsa, yani, çekiliş iadesiz olursa, ikinci çekilişteki yeşil top çekme olasılığı artık 6/10 değil 6/9 olacaktır ve bu olasılık ilk çekilişte kırmızı top gelmesine bağlı şartlı bir olasılıktır.

Burada dikkati çeken husus şartlı (koşullu) olasılığın ana kitle sayısında azalma olması halinde ortaya çıkmasıdır.

3.4. Olasılık Kuralları

Olasılığın toplama ve çarpım kuralı olmak üzere iki temel kuralı vardır. Bu kurallar, olayların birleştirilerek incelenmesine olanak verir.

3.4.1. Toplama Kuralı

Toplama kuralının uygulanabilmesi için olayların birbirini dışlayan olaylar olması gerekir. Eğer A ve B birbirlerini dışlayan iki olay ise, toplama kuralı, A yada B'nin ortaya çıkma olasılıklarının her bir olayın olasılıklarının toplamı olduğunu belirtir:

$P(A \; yada \; B) = P(A) + P(B)$ olur.

A ve B birbirini dışlamayan olaylar olduğunda A yada B'nin ortaya çıkma olasılığı :

$P(A\ yada\ B) = P(A) + P(B) - P(A\ ve\ B)$ olur.

Örnek : İçinde üç kırmızı, dört yeşil ve beş mavi top bulunan bir torbadan çekilen bir topun kırmızı veya yeşil olma olasılığı nedir?

Torbadan çekilecek bir topun aynı anda hem kırmızı hem de yeşil olması mümkün değildir. Bu durumda:

P(K)=3/12; P(Y)=4/12; P(M)=5/12'dir.

torbadan çekilen bir topun kırmızı veya yeşil olma olasılığı:

$P(K\ yada\ Y) = P(K) + P(Y) = \frac{3}{12} + \frac{4}{12} = \frac{7}{12}$'dir.

Örnek : A olayının 52'lik oyun kağıdından bir kupa çekilmesi olayı olduğunu, B olayının ise aynı deste içinden bir papaz çekilmesi olayı olduğunu düşünelim. Çekilen bir kağıdın kupa veya papaz olması olasılığı nedir?

Bilindiği gibi, bir deste 52'lik oyun kağıdı içinde 13 adet kupa ve 4 adet de papaz kartı bulunmaktadır. Bu durumda, kupa papazı çekilmesi halinde her iki olayda gerçekleşmiş olacaktır. Bu tür olaylar karşılıklı olarak birbirini engellemeyen olaylardır. Sonuç olarak bir deste oyun kağıdından çekilecek bir kupa veya papaz kartının olasılıkları gösterilirken kupa papazının iki defa hesaplamaya girmesinden kaçınmak gerekir.

P(Kupa)=13/52; P(Papaz)=4/52; P(Kupa ve Papaz)=1/52

P(Kupa veya Papaz)=P(Kupa)+P(Papaz)-P(Kupa ve Papaz)

P(Kupa veya Papaz) $= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52}$'dir.

Örnek : Bir paketleme fabrikası otomatik makineyle yarım kiloluk plastik torbalara bezelye, havuç, patates doldurmaktadır. Paketlerin çoğu yarım kilo gelmekte, ancak sebzelerin boyutları nedeniyle bazı paketler daha ağır yada hafif gelmektedir. Geçmişte yapılan kalite kontrol çalışmalarında Tablo 1'deki veriler elde edilmiştir.

Tablo 1 : Örnek 1 için veriler

Ağırlık	Olay	Torba Sayısı	Olasılık
Eksik	E	100	0,025
Tam	T	3.600	0,900
Fazla	F	300	0,075
Toplam		4.000	1,000

Üretim sürecinden seçilen herhangi bir torbanın eksik yada fazla olma olasılığı nedir?

P(E)=0,025; P(T)=0,900; P(F)=0,075 olduğuna göre:

P(E yada F)=P(E)+P(F)=0,025+0,075=0,10 olacaktır.

Örnek : Küçük bir montaj atölyesinde 50 işçi çalışmaktadır. Her işçinin kendisine verilen işi zamanında ve son kontrole hazır biçimde bitirmesi beklenmektedir. Bazen işçiler performans ölçütlerine erişemeyerek ya geç kalır, ya da hatalı montaj gerçekleştirirler. Performans değerleme döneminin sonunda 50 işçiden 5'i geç bitirmiş, 6'sı hatalı montaj yapmış, 2'si hem gecikmiş hem de hatalı montaj yapmıştır. Bir işçinin geç yada hatalı montaj yapma olasılığı nedir?

P(G)=5/50=0,10; P(H)=6/50=0,12; P(G ve H)=2/50=0,04 olarak bulunur.

Bir işçinin geç yada hatalı montaj yapma olasılığı:

P(G yada H)= P(G) + P(H) - P(G ve H) = 0,10 + 0,12 – 0,04 = 0,18'dir.

3.4.2. Çarpım Kuralı

A ve B olayları bağımlı olduğunda, A ve B olaylarının ortaya çıkma olasılığı:

P(A ve B)=P(A/B) x P(B) yada P(B ve A)=P(B/A) x P(A)'dır.

A ve B olayları bağımsız olduğunda, A ve B olaylarının ortaya çıkma olasılığı:
P(A ve B)=P(A) x P(B)’dir.

Örnek : Hilesiz bir madeni para ile hilesiz bir zarın birlikte atıldığını düşünelim. Paranın tura ve zarın da 6 yazılı yüzünün gelmesi olasılığı nedir?

Bu olaylar bağımsız olduğundan bileşik olasılık:

$\mathrm{P(A\ ve\ B)} = \mathrm{P(A)x\ P(B)} = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$’dir.

Örnek : 4 kırmızı ve 6 yeşil top bulunan bir kutudan iki defa ardı ardına iadeli çekiliş yapılacaktır. Çekilen ilk topun kırmızı, ikinci topun yeşil olması olasılığı nedir?

P(K)=4/10; P(Y)=6/10

Çekiliş iadeli olduğu için bu iki olay bağımsızdır. Bu nedenle bileşik olasılık:

$P(K\ ve\ Y) = P(K) \times P(Y) = \frac{4}{10} \times \frac{6}{10} = \frac{24}{100}$ olacaktır.

Örnek : 4 kırmızı ve 6 yeşil top bulunan bir kutudan iki defa ardı ardına iadesiz çekiliş yapılacaktır. Çekilen ilk topun kırmızı, ikinci topun yeşil olması olasılığı nedir?

P(K)=4/10; P(Y/K)=6/9

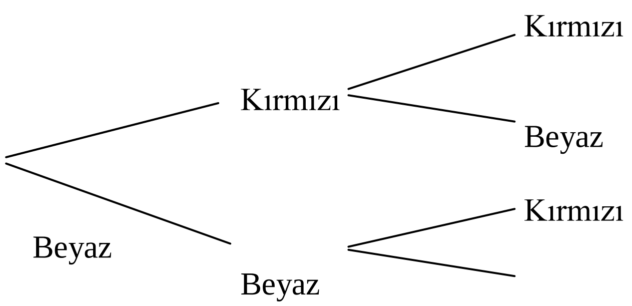
Çekiliş iadeli olduğu için bu iki olay bağımsızdır. Bu nedenle bileşik olasılık:

$P(K\ ve\ Y) = P(K) \times P(Y/K) = \frac{4}{10} \times \frac{6}{9} = \frac{24}{90} = \frac{4}{15}$ ‘dir.

4. OLASILIK DAĞILIMLARI

4.1. Rastlantısal (Tesadüfi) Değişken

Tesadüfi değişken, değeri bir deneydeki tesadüfi sonuçlarla ortaya çıkan bir değişken olarak tanımlanabilir. Örneğin, bir kutuda bulunan bir kırmızı bir beyaz topu göz önüne alalım. İadeli çekilişle yapacağımız iki çekilişin muhtemel sonuçları aşağıdaki şekilde gerçekleşecektir.



Eğer, her bir çekilişte beyaz topun kaç kere geleceğini bilmek istersek tesadüfi değişken her bir çekilişteki beyaz top sayısı olacaktır.

Tesadüfi değişken başlıca iki şekilde ortaya çıkar. Bunlardan biri kesikli tesadüfi değişken, diğeri de sürekli tesadüfi değişkendir. 0, 1, 2 gibi tam sayı değerler alabilir.

Kesikli tesadüfi değişken, değeri sayılabilen sınırlı nicel değerlerle ifade edilen bir değişkendir. Buna karşılık sürekli tesadüfi değişken ise belli bir aralıkta herhangi bir değeri aldığı kabul edilen bir değişkendir. Tesadüfi değişken sınırsız sayıda sayısal değer alabilir. Bu nedenle sürekli tesadüfi değişken adı verilir. 3,00; 3,01;3,85;3,98 gibi değerler alabilir.

4.2. Olasılık Dağılımı

Eğer, tesadüfi olarak yapılan bir denemenin mümkün olabilen tüm sonuçlarını biliyorsak ve eğer, tesadüfi değişkenin alabileceği tüm değerleri biliyorsak bu tesadüfi değişkenin gerçekleşme olasılığı gösterilebilir. Yani X tesadüfi değişkeninin alabileceği mümkün olabilen değerleri x_i ile gösterecek olursak $P(X=x_i)$ yazılabilir. Tesadüfi değişkenin alacağı bütün değerleri

$$0 \leq P(X = x_i) \leq 1$$

olmak üzere ve

$\sum_{i=1}^{n} P(X = x_i) = 1$ sağlayacak şekilde ele alınırsa X tesadüfi değişkenin olasılık dağılımı gösterilebilir. Değişkenin alabileceği her değer için bulunacak olasılıklar bir tablo haline getirilerek bu olasılık dağılımı elde edilebilir.

4.2.1. Kesikli Olasılık Dağılımları

4.2.1.1. Binom Dağılımı

Binom dağılımı, her denemenin sadece iki olanaklı sonucunun bulunduğu denemeler dizisini içeren bir model olarak düşünebiliriz. Binom dağılımında, başarı yada başarısızlık, red yada kabul, doğru yada yanlış, satın alma yada almama gibi iki sonuçlu durumlar söz konusudur. Teknik olarak binom tesadüfi değişkenin aşağıdaki üç özelliğe sahip olması gerekmektedir:

- Birbirlerinden istatistiksel olarak bağımsız belirli, n sayıda yinelenmiş denemeler olmalıdır.
- Her deneme başarı yada başarısızlıkla sonuçlanmalıdır.
- Tüm denemeler aynı başarı olasılığına (p) sahip olmalıdır ki böylece başarısızlık olasılığı, 1-p'de her deneme için sabit olsun.

Buna göre, X binom değişken olduğunda 0, 1, 2, … , n değerleri alabilir ve X'in dağılımı:

$P(X = x) = C_{x}^{n} \times p^{x} \times (1-p)^{n-x}$ şeklinde olur.

X'in ortalaması $\bar{X} = n \times p$, varyansı ise $var(X) = n \times p \times (1-p)$'dir.

Örnek: İçinde dört mavi, altı beyaz top bulunan bir torbadan iadeli çekilişle dört top çekilecektir.

a) Çekilen dört toptan üçünün mavi olma olasılığı nedir?
b) En az bir topun mavi olma olasılığı nedir?
c) En fazla iki topun mavi olma olasılığı nedir?

Problem binom dağılımı formülüyle kolaylıkla çözülebilir.

P=4/10=0,40; 1-p=1-0,40=0,60; n=4

a) $P(X = 3) = C_{x}^{n} \times p^{x} \times (1-p)^{n-x} = C_{3}^{4} \times (0{,}40)^{3} \times (0{,}60)^{4-3} = 4 \times 0{,}064 \times 0{,}6$ $= 0{,}1536$'dır.
b) En az bir topun mavi olma olasılığı sorulduğu için X değişkeni 1, 2, 3, 4 değerlerini alabilir. Bu nedenle bu dört olayın olasılıkları toplamı hesaplanmalıdır. Yada hiç mavi top gelmeme (X=0) olasılığının bu olayın tümleyeni olmasından faydalanarak aşağıdaki gibi bu olasılık hesaplanabilir.

P(X>=0)=P(X=1)+P(X=2)+P(X=3)+P(X=4) yada P(X>0)=1-P(X=0) ikinci formülü uygulamak daha kolaydır.

$P(X=0)=C_x^n \times p^x \times (1-p)^{n-x} = C_0^4 \times (0{,}40)^0 \times (0{,}60)^{4-0} = 1 \times 1 \times 0{,}1296 =$ 0,1296'dır.

$P(X>0)=1-P(X=0)=1-0{,}1296=0{,}8704$ bulunur.

a) En fazla iki topun mavi olma olasılığı sorulduğundan X değişkeni 0, 1, 2 değerlerini alabilir. Bu durumda:

$P(X \le 2)=P(X=0)+P(X=1)+P(X=2)$ olacaktır.

$P(X=0)=C_x^n \times p^x \times (1-p)^{n-x} = C_0^4 \times (0{,}40)^0 \times (0{,}60)^{4-0} = 1 \times 1 \times 0{,}1296 =$ 0,1296'dır.

$P(X=1)=C_x^n \times p^x \times (1-p)^{n-x} = C_1^4 \times (0{,}40)^1 \times (0{,}60)^{4-1} = 4 \times 0{,}40 \times 0{,}216 =$ 0,3456'dır.

$P(X=2)=C_x^n \times p^x \times (1-p)^{n-x} = C_2^4 \times (0{,}40)^2 \times (0{,}60)^{4-2} = 6 \times 0{,}16 \times 0{,}36 =$ 0,3456'dır.

$P(X \le 2)=P(X=0)+P(X=1)+P(X=2)=0{,}1296+0{,}3456+0{,}3456=0{,}8208$ bulunur.

Örnek: Bir üretim hattında bozuk parça üretilme olasılığı, geçmiş verilerden %10 olarak hesaplanmaktadır. Her gün hatta üretilen parçalardan beşi rastgele olarak seçilmekte ve kontrol edilmektedir.

a) Seçilen 5 parçadan bir tanesinin bozuk olma olasılığı nedir?

n=5, p=0,10 olduğundan X=1 için :

$$P(X=1)=C_x^n \times p^x \times (1-p)^{n-x} = C_1^5 \times (0{,}10)^1 \times (1-0{,}10)^{5-1} = 5 \times 0{,}10 \times (0{,}90)^4$$
$$= 5 \times 0{,}10 \times 0{,}6561 = 0{,}328 \approx 0{,}33 \approx \%33$$

b) Seçilen örneğin %20 yada daha azının (0,20x5=1 parça) bozuk olma olasılığı nedir?

$P(X=0)=C_x^n \times p^x \times (1-p)^{n-x} = C_0^5 \times (0{,}10)^0 \times (1-0{,}10)^{5-0} = 1 \times 1 \times 0{,}9^5 = 0{,}59,$

$P(X=1)=C_x^n \times p^x \times (1-p)^{n-x} = C_1^5 \times (0{,}10)^1 \times (1-0{,}10)^{5-1} = 5 \times 0{,}1 \times 0{,}9^4 =$ $5 \times 0{,}1 \times 0{,}6561 = 0{,}328 \cong 0{,}33$'dür.

$P(X \le 1)=P(X=0)+P(X=1)=0{,}59+0{,}33=$0,92'dir.

c) Günlük beklenen bozuk parça sayısı kaçtır?

$\bar{X}=n \times p = 5 \times 0{,}10 = 0{,}5\ bozuk\ parça.$

d) Örnek bozuk parçaların, beklenen bozuk parça sayısı etrafında varyansı kaçtır?

$VAR(X)=n \times p \times (1-p) = 5 \times 0{,}10 \times (1-0{,}10) = 5 \times 0{,}1 \times 0{,}9 = 0{,}45$'dir.

4.2.1.2. Hipergeometrik Dağılım

Binom dağılımında tesadüfi değişkenle ilgili olayların bağımsız olduklarını gördük. Ayrıca, bu dağılımla ilgili denemelerde çekilişler iadeli yapılmakta idi. Eğer, sınırlı sayıda bir grup eleman arasından iadesiz çekilişle tesadüfi bir değişken alınırsa bununla ilgili olasılıkların hesaplanmasında hipergeometrik dağılımı kullanmak uygun olur. Diğer bir ifadeyle olayların bağımsız olmadığı düşünülüyorsa tesadüfi değişken için hipergeometrik dağılım uygundur.

Hipergeometrik dağılımın olasılık fonksiyonu:

$P(X = x) = \frac{C_x^{N_1} \times C_{n-x}^{N_2}}{C_n^N}$ 'dir.

Burada,
x: X tesadüfi değişkeninin değeri,
N1: Ana kitle içinde başarılı sonuç verecek durumların toplam sayısı,
N2: Ana kitle içinde başarısız sonuç verecek durumların toplam sayısı,
n: Örnekteki eleman sayısı,
N: Ana kitledeki eleman sayısı. (N=N1+N2)

Örnek: Bir torbada dört yeşil ve altı kırmızı bilye olduğunu düşünelim. Dört bilyenin iadesiz olarak çekilmesi durumunda:

a) Çekilen dört bilyeden tam olarak üçünün kırmızı olması olasılığı nedir?
b) En az bir bilyenin kırmızı olması olasılığı nedir?

N1: Torbadaki kırmızı bilye sayısı=6
N2: Torbadaki kırmızı olmayan bilye sayısı=4
n: Çekilecel bilye sayısı=4
$N=N_1+N_2= 6 + 4 = 10$

a) $C_3^6 = \frac{6!}{3! \times (6-3)!} = 20;\ C_1^4 = \frac{4!}{1! \times (4-1)!} = 4;\ C_4^{10} = \frac{10!}{4! \times (10-4)!} = 210$

$P(X = 3) = \frac{C_x^{N_1} \times C_{n-x}^{N_2}}{C_n^N} = \frac{C_3^6 \times C_1^4}{C_4^{10}} = \frac{20 \times 4}{210} = 0{,}381$

b) Bu olasılığın hesaplanmasında 0 kırmızı bilye gelme olasılığı bulunarak 1'den çıkarılır. Elde edilen olasılık 1, 2, 3, 4 kırmızı bilye gelmesi olasılığıdır. X=0 için olasılığı hesaplayalım:

$C_0^6 = \frac{6!}{0! \times (6-0)!} = 1;\ C_4^4 = \frac{4!}{4! \times (4-4)!} = 1;\ C_4^{10} = \frac{10!}{4! \times (10-4)!} = 210$

$P(X = 0) = \frac{C_x^{N_1} \times C_{n-x}^{N_2}}{C_n^N} = \frac{C_0^6 \times C_4^4}{C_4^{10}} = \frac{1 \times 1}{210} = 0{,}0047$

$P(X \geq 1) = 1 - P(X = 0) = 1 - 0{,}0047 = 0{,}9952$ bulunur.

4.2.1.3. Poisson Dağılımı

Bundan önce incelediğimiz binom dağılımı ile bir denemede gerçekleşen olayların olasılıklarının bulunmasını göstermiştik. Bazı durumlarda belli bir zamanda veya mekanda gerçekleşen olayların olasılıklarının hesaplanması istenebilir. Bu gibi durumlarda istenen olasılıkların hesaplanmasında poisson dağılımından yararlanılır.

$P(X = x) = \frac{\lambda^x \times e^{-\lambda}}{x!}$ formülü ile ifade edilir.

Burada,

x= X tesadüfi değişkeninin değeri,
λ= zaman veya yer birimi cinsinden olayların ortalama sayısı,
e= doğal logaritma tabanı (2,71828)'dır.

$e^{-\lambda}$ değeri için Ek2'deki tablo kullanılabilir. Bu tablo ile hesaplamaya gerek kalmadan doğrudan değer okunur ve formülde yerine konabilir.

Poisson dağılımının uygulanabileceği özel problemler birim zamanda gerçekleşen bazı olayların olasılıklarının hesaplanması ile ilgilidir. Örneğin, bir telefon santraline birim zamanda gelen çağrı sinyallerinin sayısı, paralı geçişte bir yol yada köprü girişindeki kulübelere birim zamanda ulaşan araçların sayısı, gümrük yada banka veznelerinde birim zamanda işlem yaptırmak isteyen kişi sayısı vb. örnekler verilebilir. Zaman faktörü dışında, örneğin, bir top kumaşta metre kareye düşen kusur sayısı, bir sayfa yazıda bulunan hata sayısı, piyasaya yeni çıkan bir ilacı kullanan insanlar arasında alerji görülenlerin sayısı vb. örnekler verilebilir.

Bunlar ve benzeri olaylarda birim ölçü başına gerçekleşen sayıların olasılıklarını poisson dağılımı ile çözebilmek için bazı şartların bulunması gerekir. Bunlar:

- λ yani birim zaman veya alan başına gerçekleşen olayların ortalama sayısı sabittir.
- Bir zaman periyodunda gerçekleşen olayların sayısı, önceki olay sayılarından bağımsızdır.
- λ zaman veya alanın uzunluğu ile orantılıdır ve uzunluk kısaldıkça küçülür. Öyle ki çok kısa bir uzunlukta birden fazla olayın gerçekleşme olasılığı pratik olarak 0 olur.

Örnek: Bir turistik otelin resepsiyonuna 1 dakikada ortalama iki telefon gelmektedir. Buna göre:

a) Herhangi bir dakikada üç telefon gelmesi olasılığı nedir?
b) İki dakikada 5 telefon gelme olasılığı nedir?

a) $P(X=3)=\frac{\lambda^x \times e^{-\lambda}}{x!}=\frac{2^3 \times e^{-2}}{3!}=\frac{8 \times 0{,}13534}{6}=0{,}18045$'dir.
b) İki dakikada 5 telefon gelme olasılığı sorulmuş bu nedenle λ, 2 değil 4 olarak hesaba katılmalıdır. 1 dakikadaki ortalama telefon sayısı 2 ise 2 dakikadaki ortalama telefon sayısı 4 olacaktır.

$\mathrm{P}(\mathrm{X}=5)=\frac{\lambda^x \times e^{-\lambda}}{x!}=\frac{4^5 \times e^{-4}}{5!}=\frac{1024 \times 0{,}01832}{120}=0{,}1563$ bulunur.

Örnek: Trafiğin yoğun olduğu bir karayolunda günde ortalama üç kaza meydana gelmektedir. Buna göre:

a) Belli bir günde hiç kaza olmaması olasılığı kaçtır?
b) Belli bir günde üç veya dört kaza olması olasılığı nedir?

a) $P(X=0)=\frac{\lambda^x \times e^{-\lambda}}{x!}=\frac{3^0 \times e^{-3}}{0!}=\frac{1 \times 0{,}04979}{1}=0{,}0498$ bulunur.
b) Olaylar birbirini engelleyen olaylar olduğu için $P(X=3\ veya\ X=4)=P(X=3)+P(X=4)$ olacaktır.

$P(X=3)=\frac{\lambda^x \times e^{-\lambda}}{x!}=\frac{3^3 \times e^{-3}}{3!}=\frac{27 \times 0{,}04979}{6}=0{,}2240$'dır.

$P(X=4)=\frac{\lambda^x \times e^{-\lambda}}{x!}=\frac{3^4 \times e^{-3}}{4!}=\frac{81 \times 0{,}04979}{24}=0{,}1680$'dir.

$P(X=3\ veya\ X=4)=P(X=3)+P(X=4)=0{,}2240+0{,}1680=0{,}3920$ bulunur.

4.2.2. Sürekli Olasılık Dağılımları

Buraya kadar gördüğümüz olasılık dağılımları kesikli olasılık dağılımları idi. Şimdi de sürekli olasılık dağılımlarından istatistikte önemli yeri olan normal dağılımı görelim.

4.2.2.1. Normal Dağılım

Günlük hayatta karşılaşılan olaylar çeşitli olasılık dağılımları gösterirler. Bunlar arasında normal dağılım özel bir yere sahiptir. Çünkü, olayların çoğunda dağılım normal dağılıma uyar. Örneğin, insanların boyları, ağırlıkları çeşitli ölçü ve hacimlerle ifade edilen pek çok dağılımda veriler normal dağılımın şekline uyarlar.

Normal dağılımın başlıca özellikleri:

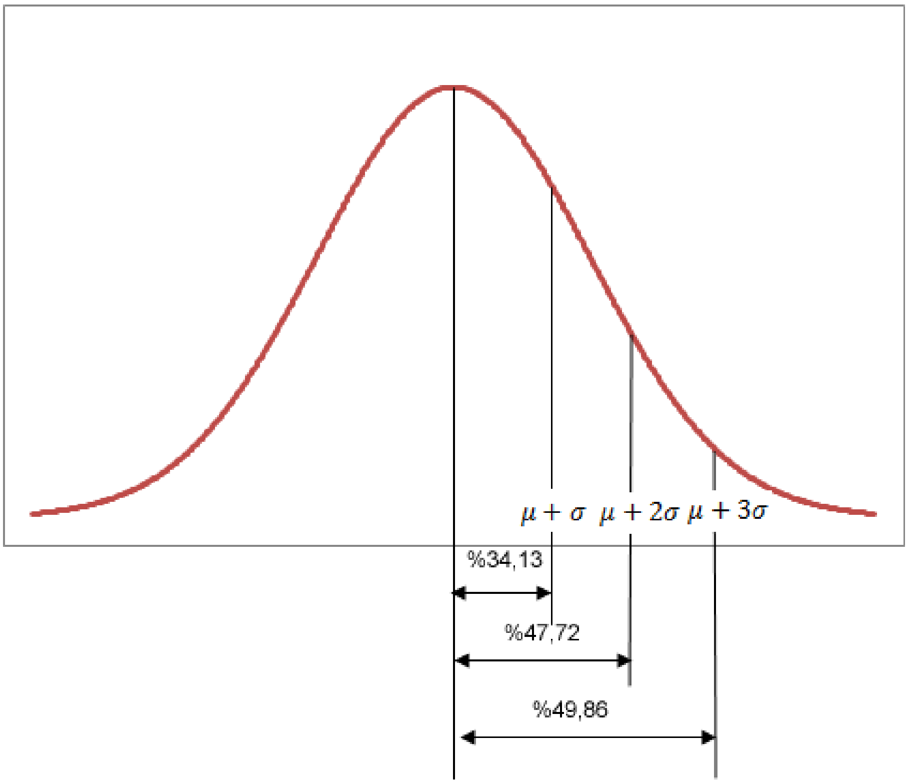
- Dağılım süreklidir. Değişken sonsuz sayıda bölünebilen değerler alır.
- Dağılım simetriktir. Aritmetik ortalama, mod ve medyan birbirine eşittir.
- Dağılımın iki ucu yatay eksene giderek yaklaşan değerler alır ve ancak sonsuzda yatay ekseni keser.

- Tek modlu bir dağılımdır.

Bu özellikleri taşıyan dağılımların normal dağılıma sahip olduğu söylenebilir.

Ortalama (μ) ve standart sapmanın (σ) değerine bağlı olarak sonsuz sayıda normal dağılım mevcuttur. Bu nedenle, verilen herhangi bir normal dağılım (μ *ve* σ *parametreleri ile*) standart forma çevrilebilir. Böyle bir dağılıma standart normal dağılım adı verilir. Standart normal dağılım ortalaması 0 ve standart sapması 1 olan bir dağılımdır.

Bu dağılıma ait alanlar tablo haline getirilmiş ve ek tablolar arasında gösterilmiştir. Z, ortalamadan uzaklaşan standart sapmaların sayısını ifade eder. Yani, örneğin, Z=1 olması halinde μ ile $\mu + \sigma$ arasındaki alan tablodan doğrudan okunabilir. Bu değer, 0,3413'tür.



Şekil : Normal Dağılım Eğrisi

Normal dağılımdaki tesadüfi değişkenin bir değerinin ortalamadan kaç standart sapma uzakta olduğunu bulmak için aşağıdaki formül kullanılır:

$$Z = \frac{X - \mu}{\sigma}$$

Bu eşitlikte sembollerin anlamları şöyledir:
Z: standart sapmaların sayısı,
X: tesadüfi değişkenin değeri,
σ: Normal dağılımın standart sapması,
μ: Normal dağılımın ortalaması.

Bu eşitlik yardımıyla yatay eksen üzerinde ortalamadan sapmaların değerleri Z birim cinsinden bulunmuş olur. Değişkenin sahip olduğu gerçek birimlerin (TL-Kg. vb.) kullanılmasına ihtiyaç kalmaz.

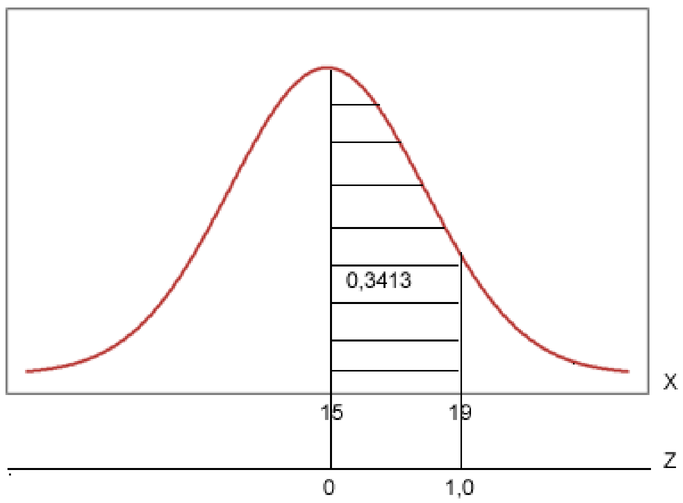
Örnek: Bir futbol maçına giden bir grup gencin yaş ortalamasının 15, standart sapmasının 4 yaş olduğunun yapılan bir anketle belirlendiğini varsayalım. Gençlerin yaşlarının normal dağılım gösterdiğini kabul ederek aşağıdaki soruları cevaplayınız.

a) Yaşları 15 ve 19 arasında olan gençlerin oranı nedir?
b) Yaşları 21'i geçen gençlerin oranı nedir?
c) Yaşları 12'den az olanların olasılığı nedir?
d) Yaşların 13 ile 14 arasında olma olasılığı nedir?

a) 15-19 arasındaki yaşların olasılıklarını bulabilmek için ortalama (15) ile X (19) arasındaki alanı veren Z değerini bulmak gerekir:

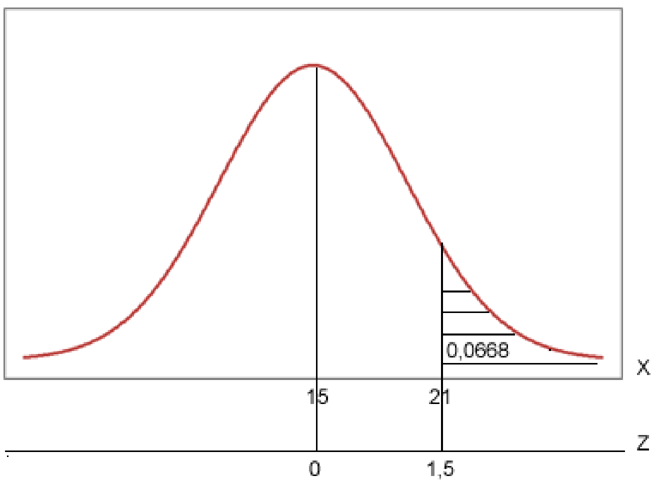
$Z = \frac{X-\mu}{\sigma} = \frac{19-15}{4} = \frac{4}{4} = 1.0$

Bu Z değerine karşılık gelen tablo değeri 0,3413 olduğundan yaşları 15 ile 19 arasında olan gençlerin oranı % 34,13'tür.



b) Normal eğri altında X'in belli bir değerinden daha büyük olan alanını bulabilmek için önce söz konusu X değeri ile ortalama arasındaki alan bulunur. Daha sonra bu değer 0,5000 değerinden çıkarılır.

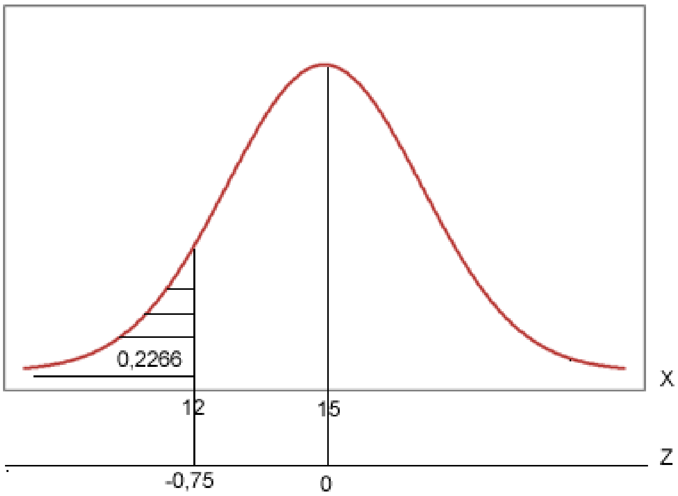
$Z = \frac{X-\mu}{\sigma} = \frac{21-15}{4} = \frac{6}{4} = 1{,}5$ bulunur. Bu Z değeri için alan 0,4332'dir. Bu durumda, 21 yaşın üzerindeki alan 0,5000-0,4332=0,0668'dir. Yani yaşı 21'i geçenlerin oranı % 6,68'dir.



c) Normal eğri altında X'in belli bir değerinden daha küçük olan alanını bulabilmek için önce ortalama ile söz konusu X değeri arasındaki alan bulunur. Daha sonra bu alan 0,5000 değerinden çıkarılır.

$Z = \frac{X-\mu}{\sigma} = \frac{12-15}{4} = \frac{-3}{4} = -0{,}75$ bulunur.

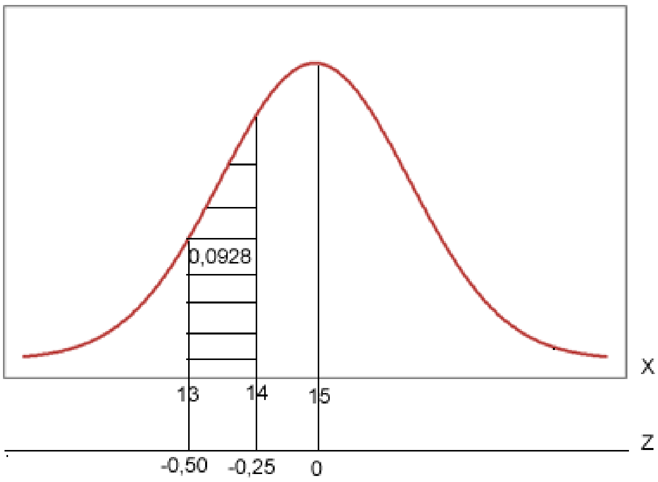
Z'nin -0,75 değeri için tablodan okunan alan 0,2734'tür. Z'nin negatif ya da pozitif değer alması sadece ortalamanın solunda veya sağında olan değerlerle ilgilendiğimizi gösterir. Dağılım simetrik olduğundan Z'nin pozitif ve negatif değerleri arasında fark yoktur. Bu durumda, yaşları 12'den daha az olan gençlerin olasılığı: 0,5000-0,2734=0,2266'dır.



d) Normal eğri altında kalan iki X değeri arasındaki alanın bulunmasından her iki değer birlikte ister ortalamanın altında ister üstünde olsun kademeli bir işlem takip edilir. Bu örnekte, önce X'in 13 ile 15 değerleri arasındaki alanı bulunur. Sonra, X'in 14 ile 15 değerleri arasındaki alan bulunarak birinci alandan çıkarılır. Böylece, geriye kalan taralı alan elde edilmiş olur.

$Z_1 = \frac{X-\mu}{\sigma} = \frac{13-15}{4} = \frac{-2}{4} = -0{,}5$; $Z_2 = \frac{X-\mu}{\sigma} = \frac{14-15}{4} = \frac{-1}{4} = -0{,}25$ bulunur.

Z'nin -0,50 değeri için tablodan 0,1915 ve Z'nin -0,25 değeri için tablodan 0,0987 değeri okunur. Yaşların 13 ile 14 arasında olma olasılığı, 0,1915-0,0987=0,0928 yani % 9,28 olarak bulunur.



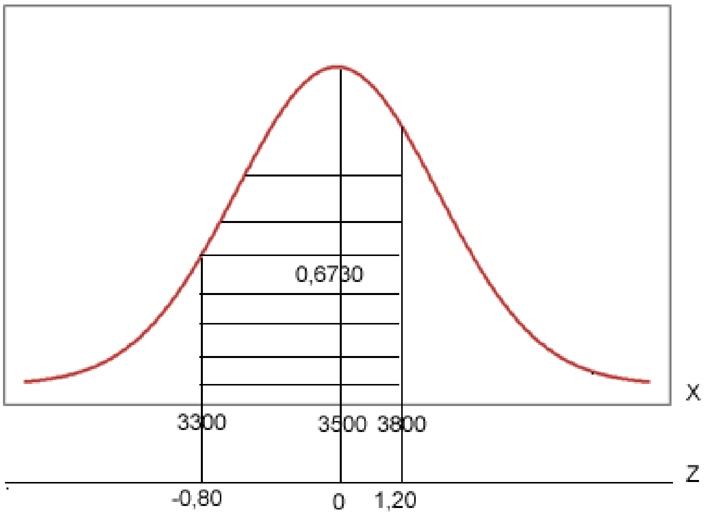
Örnek : Bir işletmenin A malından haftalık satış miktarı ortalama 3500 adet, standart sapması 250 adettir. Satışların normal dağılım gösterdiğini kabul ederek haftalık satışların 3300 adet ile 3800 adet arasında olma olasılığı nedir?

Bu örnekte istenen alan ortalamanın her iki tarafında bulunduğundan Z değerleri iki taraf için ayrı ayrı hesaplanarak alanlar bulunacaktır, sonra bulunan alanların toplamı hesaplanacaktır.

$Z_1 = \frac{X-\mu}{\sigma} = \frac{3300-3500}{250} = \frac{-200}{250} = -0{,}80;\ Z_2 = \frac{X-\mu}{\sigma} = \frac{3800-3500}{250} = \frac{300}{250} = 1{,}20$ bulunur.

Z_1'e karşılık gelen alan 0,2881, Z_2'ye karşılık gelen değer ise 0,3849'dur. Bu durumda, haftalık satışların 3300 ile 3800 arasında olma olasılığı:

$P(3300 < X < 3800) = 0{,}2881 + 0{,}3849 = 0{,}6730$ olacaktır. Yani olasılık % 67,30'dur.



PROBLEMLER

1. Bir kentte araçların % 40'ının kasko sigortası yaptırdığı bilinmektedir. Tesadüfi olarak alınan 10 araçtan,
   a) Dört tanesinin sigortalı olması,
   b) En az 2 tanesinin sigortalı olması,
   c) Dokuz ve daha fazlasının sigortalı olması,
   d) Bir taneden fazlasının sigortalı olmaması olasılıklarını bulunuz.

Çözüm: Bu örnekte binom dağılımını kullanmak gerekir: Hesaplamalarda başarı olasılığı (p) için 0,4 değeri kullanılır.

a) $P(X = 4) = C\binom{10}{4} \times 0{,}4^4 \times (1-0{,}4)^{10-4} = \frac{10!}{4! \times (10-4)!} \times 0{,}0256 \times 0{,}6^6$
$= 21 \times 0{,}0256 \times 0{,}046656 = 0{,}025082 \cong 0{,}025$'dir.

b) $P(X \geq 2) = 1 - P(< 2)$ şeklinde hesaplanabilir.
$P(X < 2) = P(X = 0) + P(X = 1)$
$P(X = 0) = C\binom{10}{0} \times 0{,}4^0 \times (1-0{,}4)^{10-0} = 1 \times 1 \times 0{,}6^{10} \cong 0{,}006$'dır.
$P(X = 1) = C\binom{10}{1} \times 0{,}4^1 \times (1-0{,}4)^{10-1} = 10 \times 0{,}4 \times 0{,}6^9 = 10 \times 0{,}6 \times 0{,}010 =$ 0,06'dır.
$P(X < 2) = P(X = 0) + P(X = 1) = 0{,}006 + 0{,}06 = 0{,}066$ olur.
$P(X \geq 2) = 1 - P(< 2) = 1 - 0{,}066 = 0{,}934$ olarak hesaplanır.

c) $P(X \geq 9) = P(X = 9) + P(X = 10)$ şeklinde hesaplanabilir.
$P(X = 9) = C_{9}^{10} \times 0{,}4^{9} \times (1 - 0{,}4)^{10-9} = 10 \times 0{,}00026 \times 0{,}6 = 0{,}0016$'dır.
$P(X = 10) = C_{10}^{10} \times 0{,}4^{10} \times (1 - 0{,}4)^{10-10} = 1 \times 0{,}0001 \times 1 = 0{,}0001$'dir.
$P(X \geq 9) = P(X = 9) + P(X = 10) = 0{,}0016 + 0{,}0001 = 0{,}0017$ olarak hesaplanır.

d) Sigortalı olmama olasılığı sorulduğu için başarıl olasılığı (p) için 0,4 değil 1-0,4=0,6 değeri kullanılmalıdır.
$P(X > 1) = 1 - P(X \leq 1)$ şeklinde hesaplanabilir.
$P(X \leq 1) = P(X = 0) + P(X = 1)$ olacaktır.
$P(X = 0) = C_{0}^{10} \times 0{,}4^{0} \times (1 - 0{,}4)^{10-0} = 1 \times 1 \times 0{,}6^{10} \cong 0{,}006$'dır.
$P(X = 1) = C_{1}^{10} \times 0{,}4^{1} \times (1 - 0{,}4)^{10-1} = 10 \times 0{,}4 \times 0{,}6^{9} = 10 \times 0{,}4 \times 0{,}10 \cong$ 0,04'tür.
$P(X \leq 1) = P(X = 0) + P(X = 1) = 0{,}006 + 0{,}04 = 0{,}046$ bulunur.
$P(X > 1) = 1 - P(X \leq 1) = 1 - 0{,}045 = 0{,}954$ olarak hesaplanır.

2. Şehirlerarası telefon santralinde çalışan bir operatöre dakikada ortalama üç konuşma talebi gelmektedir.
a) Bir dakika içinde hiç konuşma talebi gelmemesi olasılığını,
b) Bir dakika içinde iki konuşma talebi gelmesi olasılığını,
c) Üç dakikada dört konuşma talebi gelmesi olasılığını bulunuz.

Çözüm : Bu problemde poisson dağılımı kullanılmalıdır.
a) $P(X = 0) = \frac{\lambda^{x} \times e^{-\lambda}}{x!} = \frac{3^{0} \times e^{-3}}{0!} = \frac{1 \times 0{,}04979}{1} = 0{,}04979$ olarak hesaplanır.
b) $P(X = 2) = \frac{\lambda^{x} \times e^{-\lambda}}{x!} = \frac{3^{2} \times e^{-3}}{2!} = \frac{9 \times 0{,}04979}{2} \cong 0{,}224$ olarak hesaplanır.
c) Üç dakikada 4 konuşma talebi gelme olasılığı sorulduğuna göre λ için $3 \times 3 = 9$ değeri kullanılmalıdır.
$P(X = 4) = \frac{\lambda^{x} \times e^{-\lambda}}{x!} = \frac{9^{4} \times e^{-9}}{4!} = \frac{6561 \times 0{,}00012}{24} \cong 0{,}033$ olarak hesaplanır.

3. Bir kutuda 100 bilye bulunmaktadır. Bunlardan beş tanesi mavi, gerisi yeşildir. Eğer iadesiz olarak sekiz bilye çekilirse bunlardan iki tanesinin mavi olması olasılığı nedir?

Çözüm : Bu problemde iadesiz çekiliş yapıldığı için hipergeometrik dağılım kullanılır.
$N_1=5$, $N_2=95$, $N=100$, $n=8$, $x=2$'dir. Buna göre:
$$P(X = 2) = \frac{C_{x}^{N_1} \times C_{n-x}^{N_2}}{C_{n}^{N}} = \frac{C_{2}^{5} \times C_{8-2}^{95}}{C_{8}^{100}} = \frac{10 \times 869.107.785}{186.087.894.300} \cong 0{,}047 \text{ olarak hesaplanır.}$$

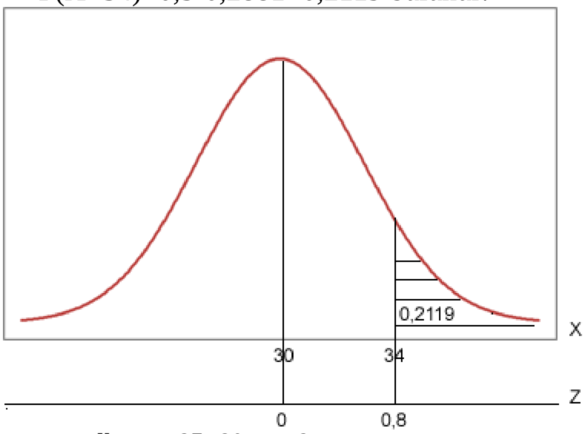
4. Bir kutuda 100 bilye bulunmaktadır. Bunlardan beş tanesi mavi, gerisi yeşildir. Eğer iadesiz olarak sekiz bilye çekilirse bunlardan iki tanesinin mavi olması olasılığı nedir?

Çözüm : Bu problemde iadesiz çekiliş yapıldığı için hipergeometrik dağılım kullanılır.
$N_1=5$, $N_2=95$, $N=100$, $n=8$, $x=2$'dir. Buna göre:
$$P(X = 2) = \frac{C_{x}^{N_1} \times C_{n-x}^{N_2}}{C_{n}^{N}} = \frac{C_{2}^{5} \times C_{8-2}^{95}}{C_{8}^{100}} = \frac{10 \times 869.107.785}{186.087.894.300} \cong 0{,}047 \text{ olarak hesaplanır.}$$

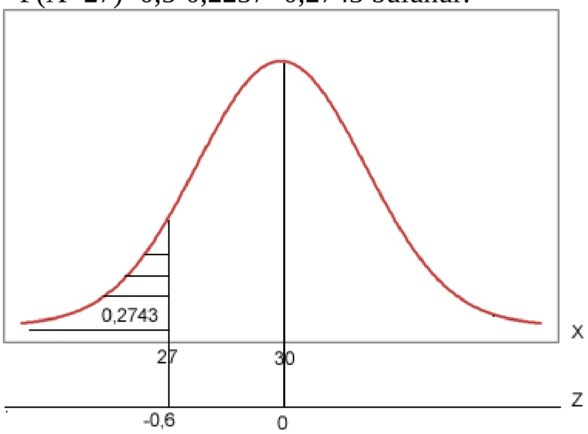
5. Ortalaması 30 ve standart sapması 5 olan bir normal dağılım için aşağıdaki olasılıkları hesaplayınız.
a) P(X>34),
b) P(X<27),
c) P(31<X<33),
d) P(27<X<32),
e) P(25<X<29).

Çözüm :

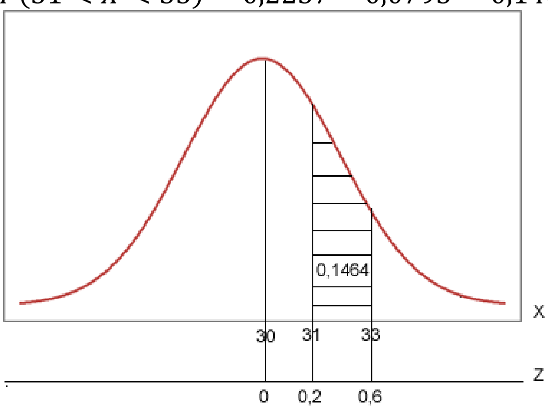
a) $Z = \frac{X-\mu}{\sigma} = \frac{34-30}{5} = \frac{4}{5} = 0{,}8$'dir. Z=0,8 için tablo değeri 0,2881'dir. P(X>34)=0,5-0,2881=0,2119 bulunur.



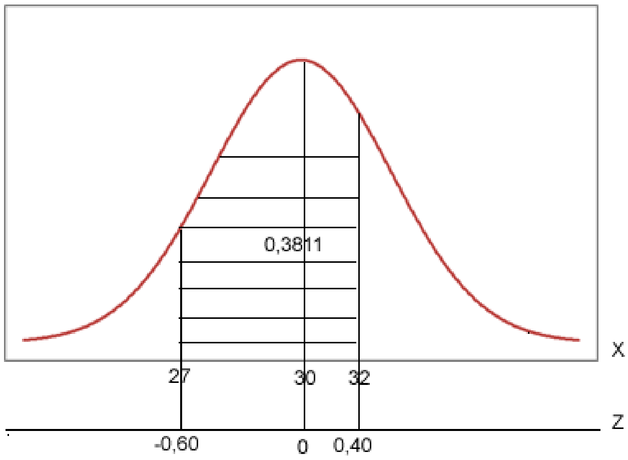
b) $Z = \frac{X-\mu}{\sigma} = \frac{27-30}{5} = \frac{-3}{5} = -0{,}6$'dır. Z=0,6 için tablo değeri 0,2257'dir. P(X<27)=0,5-0,2257=0,2743 bulunur.



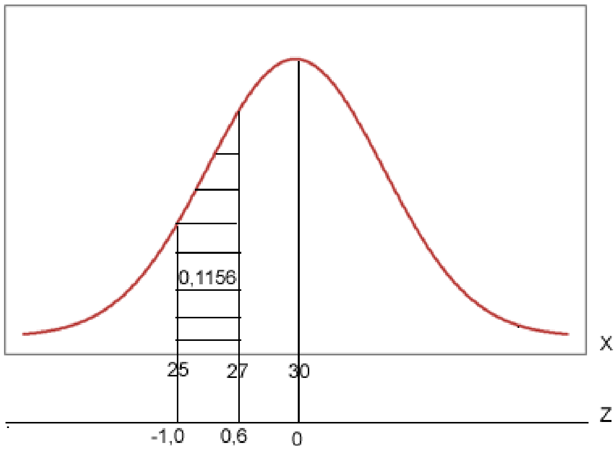
c) $Z_1 = \frac{X_1-\mu}{\sigma} = \frac{31-30}{5} = \frac{1}{5} = 0{,}2$'dir. $Z_2 = \frac{X_2-\mu}{\sigma} = \frac{33-30}{5} = \frac{3}{5} = 0{,}6$'dır. Z=0,2 için tablo değeri 0,0793, Z=0,6 için tablo değeri 0,2257'dir.
$P(31 < X < 33) = 0{,}2257 - 0{,}0793 = 0{,}1464$ olarak hesaplanır.



d) $Z_1 = \frac{X_1-\mu}{\sigma} = \frac{27-30}{5} = \frac{-3}{5} = -0{,}6$’dır. $Z_2 = \frac{X_2-\mu}{\sigma} = \frac{32-30}{5} = \frac{2}{5} = 0{,}4$’tür. Z=0,4 için tablo değeri 0,1554, Z=0,6 için tablo değeri 0,2257’dir.
$P(27 < X < 32) = 0{,}1554 + 0{,}2257 = 0{,}3811$ olarak hesaplanır.



e) $Z_1 = \frac{X_1-\mu}{\sigma} = \frac{25-30}{5} = \frac{-5}{5} = -1{,}0$’dir. $Z_2 = \frac{X_2-\mu}{\sigma} = \frac{27-30}{5} = \frac{-3}{5} = -0{,}6$’dır. Z=1,0 için tablo değeri 0,3413, Z=0,6 için tablo değeri 0,2257’dir.
$P(25 < X < 27) = 0{,}3413 - 0{,}2257 = 0{,}1156$ olarak hesaplanır.



6. Bilgisayar monitörü üreten bir fabrikada her 100 monitörden 30’u kusurlu çıkmaktadır. Tesadüfi olarak alınan 5 adetlik bir grup monitör içinde:
a) Tam olarak 2 monitörün,
b) En az dört monitörün kusurlu olması olasılığını hesaplayınız.

Çözüm : Bu problemde binom dağılımı kullanılır. n=5 ve $p = \frac{30}{100} = 0{,}3$’tür.

a) $P(X = 2) = C_2^5 \times 0{,}3^2 \times (1 - 0{,}3)^{5-2} = 10 \times 0{,}09 \times 0{,}7^3 = 10 \times 0{,}09 \times 0{,}343 = 0{,}3087$’dir.

b) $P(X \geq 4) = P(X = 4) + P(X = 5)$ şeklinde hesaplanabilir.
$P(X = 4) = C_4^5 \times 0{,}3^4 \times (1 - 0{,}3)^{5-4} = 5 \times 0{,}0081 \times 0{,}7 = 0{,}02835$’dir.
$P(X = 5) = C_5^5 \times 0{,}3^5 \times (1 - 0{,}3)^{5-5} = 5 \times 0{,}0024 \times 1 = 0{,}01215$’dir.
$P(X \geq 4) = P(X = 4) + P(X = 5) = 0{,}02835 + 0{,}01215 = 0{,}0405$ olarak hesaplanır.

7. ABC sigorta şirketi şikayetlerin % 8'inin hırsızlıklardan kaynaklanan zararlardan meydana geldiğini bilmektedir. 5 şikayeti kapsayan tesadüfi bir örnekte:
a) En çok 2 hırsızlıktan kaynaklı şikayet olması olasılığını,
b) Üçten fazla hırsızlıktan kaynaklanan şikayet olması olasılığını bulunuz.

Çözüm : Bu örnekte binom dağılımı kullanılmalıdır.

a) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$ şeklinde hesaplanabilir.
$P(X = 0) = C_0^5 \times 0{,}08^0 \times (1 - 0{,}08)^{5-0} = 1 \times 1 \times 0{,}92^5 = 0{,}659$'dur.
$P(X = 1) = C_1^5 \times 0{,}08^1 \times (1 - 0{,}08)^{5-1} = 5 \times 0{,}08 \times 0{,}92^4 = 5 \times 0{,}08 \times 0{,}716 = 0{,}286$'dır.
$P(X = 2) = C_2^5 \times 0{,}08^2 \times (1 - 0{,}08)^{5-2} = 10 \times 0{,}08^2 \times 0{,}92^3 = 10 \times 0{,}0064 \times 0{,}778 = 0{,}049$'dur.
$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0{,}659 + 0{,}286 + 0{,}049 = 0{,}994$ olarak hesaplanır.

b) P(X>3)=P(X=4)+P(X=5) şeklinde hesaplanabilir.

$P(X = 4) = C_4^5 \times 0{,}08^4 \times (1 - 0{,}08)^{5-4} = 5 \times 0{,}08^4 \times 0{,}92 = 5 \times 0{,}00004 \times 0{,}92 = 0{,}0002$'dir.
$P(X = 5) = C_5^5 \times 0{,}08^5 \times (1 - 0{,}08)^{5-5} = 1 \times 0{,}08^5 \times 0{,}92^0 = 1 \times 0{,}000003 \times 1 = 0{,}000003$'tür.

P(X>3)=P(X=4)+P(X=5)=0,0002+0,000003=0,000203 olarak bulunur.

8. Belli bir otoban giriş rampasına gelen arabaların miktarının poisson dağılımlı olduğunu varsayalım. Varsayım doğru ise ve bir saat içinde gelen ortalama araç sayısı beş ise 1 saat içinde
a) 5 arabanın gelme olasılığını,
b) 2'den az arabanın gelme olasılığını,
c) Hiç araba gelmeme olasılığını,
d) 2 ile 5 arasında araba gelme olasılığını bulunuz.

Çözüm : Bu problemde poisson dağılımı kullanılır. $\lambda = 5$:

a) $P(X = 5) = \frac{\lambda^x \times e^{-\lambda}}{x!} = \frac{5^5 \times e^{-5}}{5!} = \frac{3125 \times 0{,}00674}{120} = 0{,}175$'dir.
b) $P(X < 2) = P(X = 0) + P(X = 1)$ şeklinde hesaplanabilir.

$P(X = 0) = \frac{\lambda^x \times e^{-\lambda}}{x!} = \frac{5^0 \times e^{-5}}{0!} = \frac{1 \times 0{,}00674}{1} = 0{,}00674$'tür.
$P(X = 1) = \frac{\lambda^x \times e^{-\lambda}}{x!} = \frac{5^1 \times e^{-5}}{1!} = \frac{5 \times 0{,}00674}{1} = 0{,}0337$'dir.

$P(X < 2) = P(X = 0) + P(X = 1) = 0{,}00674 + 0{,}0337 = 0{,}04044$ olarak hesaplanır.

c) $P(X = 0) = \frac{\lambda^x \times e^{-\lambda}}{x!} = \frac{5^0 \times e^{-5}}{0!} = \frac{1 \times 0{,}00674}{1} = 0{,}00674$'tür.
d) P(2<X<5)=P(X=3)+P(X=4) şeklinde hesaplanabilir.

$P(X = 3) = \frac{\lambda^x \times e^{-\lambda}}{x!} = \frac{5^3 \times e^{-5}}{3!} = \frac{125 \times 0{,}00674}{6} = \frac{0{,}8425}{6} = 0{,}1404$'tür.

$P(X=4)=\frac{\lambda^x \times e^{-\lambda}}{x!}=\frac{5^4 \times e^{-5}}{4!}=\frac{625 \times 0{,}00674}{24}=\frac{4{,}2125}{24}=0{,}1755$’dir.

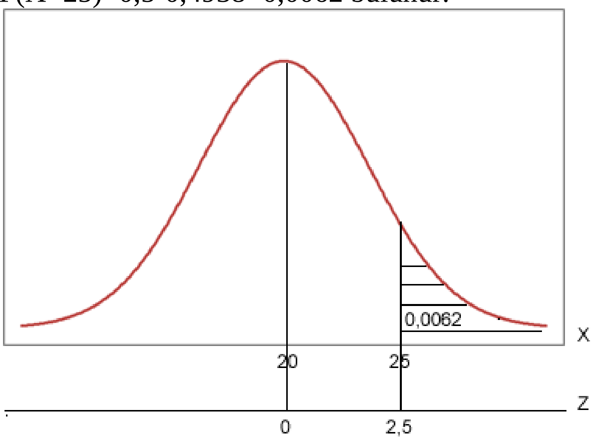
P(2<X<5)=P(X=3)+P(X=4)=0,1404+0,1755=0,3159 bulunur.

9. Ortalaması 20 ve standart sapması 2 olan bir normal dağılım için aşağıdaki olasılıkları hesaplayınız.

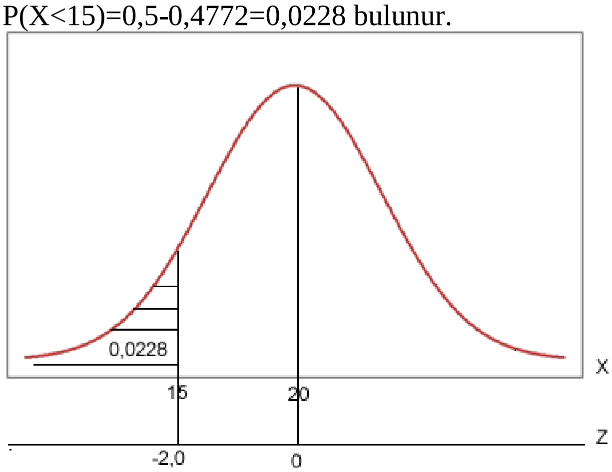
a) P(X>25),
b) P(X<16),
c) P(14<X<15),
d) P(15<X<25),
e) P(24<X<26).

Çözüm :

a) $Z=\frac{X-\mu}{\sigma}=\frac{25-20}{2}=\frac{5}{2}=2{,}5$’dur. Z=2,5 için tablo değeri 0,4938’dir.
P(X>25)=0,5-0,4938=0,0062 bulunur.

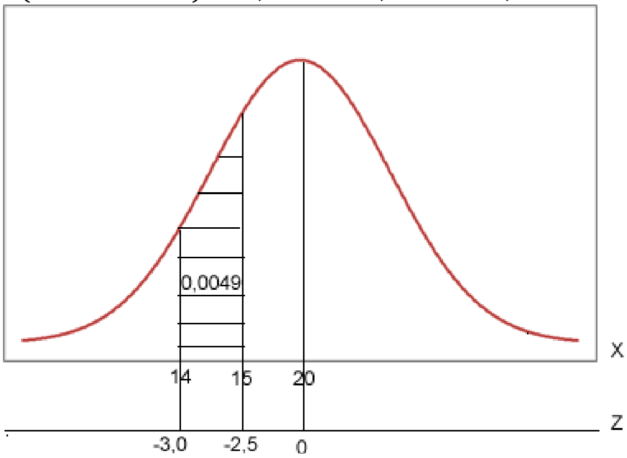


b) $Z=\frac{X-\mu}{\sigma}=\frac{16-20}{2}=\frac{-4}{2}=-2{,}0$’dir. Z=2,0 için tablo değeri 0,4772’dir.
P(X<15)=0,5-0,4772=0,0228 bulunur.



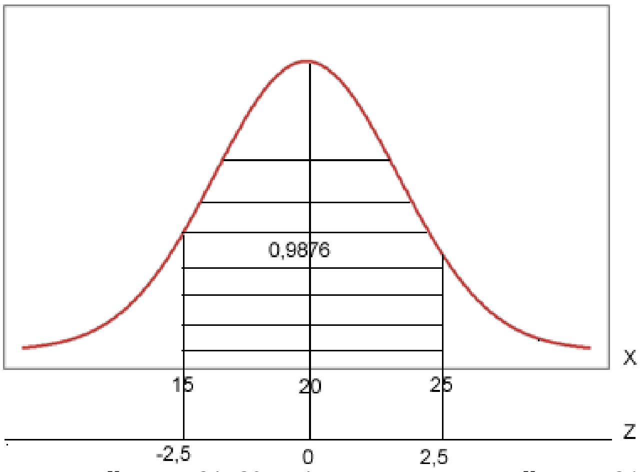
c) $Z_1 = \frac{X_1-\mu}{\sigma} = \frac{14-20}{2} = \frac{-6}{2} = -3,0$’tür. $Z_2 = \frac{X_2-\mu}{\sigma} = \frac{15-20}{2} = \frac{-5}{2} = -2,5$’dur. Z=3,0 için tablo değeri 0,4987, Z=2,5 için tablo değeri 0,4938’dir.

$P(14 < X < 15) = 0,4987 - 0,4938 = 0,0049$ olarak hesaplanır.



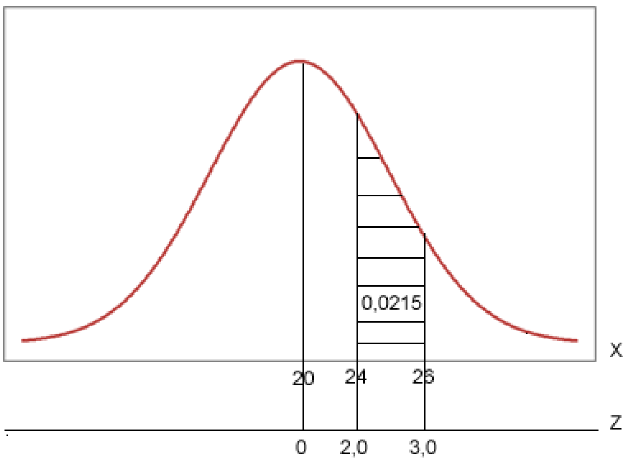
d) $Z_1 = \frac{X_1-\mu}{\sigma} = \frac{15-20}{2} = \frac{-5}{2} = -2,5$’dur. $Z_2 = \frac{X_2-\mu}{\sigma} = \frac{25-20}{2} = \frac{5}{2} = 2,5$’dur. Z=2,5 için tablo değeri 0,4938’dir.

$P(15 < X < 25) = 0,4938 + 0,4938 = 0,9876$ olarak hesaplanır.



e) $Z_1 = \frac{X_1-\mu}{\sigma} = \frac{24-20}{2} = \frac{4}{2} = 2,0$’dır. $Z_2 = \frac{X_2-\mu}{\sigma} = \frac{26-20}{2} = \frac{6}{2} = 3,0$’dır. Z=2,0 için tablo değeri 0,4772, Z=3,0 için tablo değeri 0,4987’dir.

$P(24 < X < 26) = 0,4987 - 0,4772 = 0,0215$ olarak hesaplanır.



5. KAYNAKLAR

I. Tütek, H., Gümüşoğlu, Ş., 2008, *İşletme İstatistiği,* Beta Basım Yayım Dağıtım A.Ş., İstanbul.

II. Güler, F., 2007, *Temel İstatistik,* Beta Basım Yayım Dağıtım A.Ş., İstanbul.

III. İslamoğlu, A.H., 2009, *Sosyal Bilimlerde Araştırma Yöntemleri (SPSS Uygulamalı),* Beta Basım Yayım Dağıtım A.Ş., İstanbul.

IV. Balce, A.O., Demir, S., 2007, *İstatistik Ders Notları*, Pamukkale Üniversitesi İktisadi Ve İdari Bilimler Fakültesi, Ekonometri Bölümü, Denizli

V. Çakıcı, M., Oğuzhan, A., Özdil, T., 2000, *Temel İstatistik I*, Özal Matbaası, İstanbul.

VI. Çakıcı, M., Oğuzhan, A., Özdil, T., 2000, *Temel İstatistik II*, Özal Matbaası, İstanbul.