TEMEL İSTATİSTİK I

DERS NOTLARI

Prof.Dr.YÜKSEL TERZİ

ONDOKUZ MAYIS ÜNİVERSİTESİ

FEN-EDEBİYAT FAKÜLTESİ

İSTATİSTİK BÖLÜMÜ

SAMSUN

2018

İÇİNDEKİLER

I.BÖLÜM

1. İSTATİSTİK KAVRAMLAR

1.1. Giriş

Günümüzde artan rekabet koşulları ve teknolojik gelişmeler bireyleri, işletmeleri ve ülkeleri çeşitli konularda verecekleri kararlarla ilgili olarak verileri toplamaya ve bunları analiz etmeye zorunlu kılmaktadır. İşte gerek verilerin toplanması gerekse toplanan verilerin analiz edilmesi ve verilecek kararlarda kullanılması istatistiğin önemini ortaya koymaktadır. Dar anlamda bir konuda toplanan verilerden o konu ile ilgili karar verinceye kadar yapılan tüm işlemleri kapsayan geniş anlamda ele alınan istatistik (***statistics***) günümüzde bu yelpazede çok farklı anlamlarda kullanılmaktadır.

Günlük yaşamımızda pek çok olayla ilgilenir ve birçok sorular sorup cevaplar ararız. Örneğin; "Yaklaşmakta olan bayram tatilinde havalar nasıl olacak? Üniversite öğrencilerinin üniversiteye giriş sınavlarında aldıkları puanlarla üniversitedeki başarıları arasında ne-kadar ilişki vardır?" gibi sorular. Cevaplarından tam emin olmadığımız ve cevapları hakkında bilgimizin çok sınırlı olduğu soruları güvenilir biçimde cevaplandırmanın en kolay yolu, bunlarla ilgili bilgi toplamak ve bilgileri çözümleyerek sonuçlar çıkarmaktır. Topladığımız bilgiler genelde sayısaldır ya da anlam kolaylığı için biz sayılarla gösteririz. Mesela bir okula devam eden öğrencilerin boy uzunlukları, ağırlıkları, yaşları, zeka seviyeleri gibi bilgileri toplamaktır.

İstatistik sözcüğünün kökeni konusunda kesin bir görüş birliği yoktur. Bazı bilimciler sözcüğün Latince'de "devlet ve durum'' anlamına gelen *status* kelimesinden, ya da Yunanca "gözlem" anlamına gelen "statizien" kelimesinden türemiştir.

1.2. İstatistiğin Tanımı

İstatistikle ilgili çeşitli tanımlar yapılmıştır. Bunlardan bazıları da bilim dışı, ciddiyetsiz tanımlardır.

"Üç çeşit yalan vardır: yalan, kuyruklu yalan ve istatistik"

Disraeli

"İyi ambalajlanmış bir istatistik, Hitler'in büyük yalanından daha etkilidir"
Darell Huff

Bu tanımlar istatistiğin işe yaramadığını değil, her şeyin istatistikle kanıtlanamayacağını vurgulamaktadır. Örneğin Üsküdar-Eminönü arasında seyahat eden vapur seferleri ile İstanbul'daki boşanma oranları arasında bulunan yüksek bir korelasyon (ilişki), bu iki olayın birbiri ile ilişkili olduğunu düşündürülebilir. Ancak bu reddedilmesi gereken bir durumdur. ***Rakamlar yalan söylemez, fakat yalancılar rakam söyler.***

Ortalama sıcaklığın 40C° olduğu bir çöle üstündeki ince bir tişörtle giden bir kişi aldanacaktır. Çünkü çölde sıcaklık gündüz 50C°'ye çıkmakta, gece ise -10C°'ye düşmektedir.

İstatistik sözcüğü farklı yaklaşımlara göre değişik anlamlar taşır. Günlük dilde istatistik denildiği zaman, belirli bir olaya ilişkin derlenmiş sayısal bilgiler akla gelir. Örneğin dışalım, turizm, inşaat istatistikleri gibi.

İstatistiğin birçok tanımı mevcuttur. İstatistiğin değişik anlamlarda kullanılması doğaldır. Çünkü günümüzde bütün çalışma alanlarında kullanılmaktadır. Günlük yaşamımızda pek çok olayla ilgilenir ve birçok sorular sorup cevaplar ararız. Örneğin; "Yaklaşmakta olan bayram tatilinde havalar nasıl olacak? Üniversite öğrencilerinin üniversiteye giriş sınavlarında aldıkları puanlarla üniversitedeki başarıları arasında ne-kadar ilişki vardır?" gibi sorular. Cevaplarından tam emin olmadığımız ve cevapları hakkında bilgimizin çok sınırlı olduğu soruları güvenilir biçimde cevaplandırmanın en kolay yolu, bunlarla ilgili bilgi toplamak ve bilgileri çözümleyerek sonuçlar çıkarmaktır. Topladığımız bilgiler genelde sayısaldır ya da anlam kolaylığı için biz sayılarla gösteririz. Mesela bir okula devam eden öğrencilerin boy uzunlukları, ağırlıkları, yaşları, zeka seviyeleri gibi bilgileri toplamaktır.

Belirli amaçlarla toplanan sayısal bilgilere **veri** denir. Verileri inceleme işi ile uğraşan bilime **istatistik** denir. Başka bir deyişle istatistik; belirli amaçlar için veri toplama, toplanan verileri düzenleme, çözümleme yorumlama teknik ve yöntemleridir. İstatistik, bilimsel yöntemin en güçlü temel araçlarından biridir. Bir başka açıdan istatistik; hem teknik ve yöntemler geliştiren bir bilim, hem de yöntemler topluluğu sayılabilir.

İstatistikle ilgili çeşitli tanımlar yapılmıştır.

- ✓ Yığın olayları incelemek, olaylarla ilgili toplanan verileri analiz etmek, olayların sebep ve sonuçlarını açıklamak, aralarındaki ilişkileri ortaya koymak için kendine özgü yöntemleri olan bir bilim dalıdır.
- ✓ Olayların nicel ve nitel yönlerinin tablolar, grafikler veya sayısal değerler şeklindeki özet ifadelerdir. Ölüm istatistikleri, doğum istatistikleri, göç istatistikleri, kaza istatistikleri gibi özet bilgiler istatistikler olarak adlandırılır.
- ✓ Örneği oluşturan birimlerden hesaplanmış, ana kütleyi tanımlayan değerlere karşılık gelen değerlerdir. Örnek ortalaması, örnek varyansı gibi.
- ✓ ***İstatistik, geçmişi ve şimdiki durumu çeşitli sayısal tekniklerle analiz ederek gelecek hakkında karar vermeyi sağlayan bir bilim dalıdır.***
- ✓ İstatistik, çeşitli olaylarla ilgili toplanmış veriler demektir. Nüfus, istihdam, kaza istatistikleri, yıllık imalat sanayi anketleri veya Merkez Bankası'nın periyodik

olarak yayınladığı bültenler örnek olarak verilebilir. Belirli bir konu ile ilgili sayısal verilerin en uygun şekilde derlenerek açık ve anlaşılır bir şekilde ifade edilmesidir.

- ✓ İstatistik verilerin toplanması, organize edilmesi, özetlenmesi, sunulması, analiz edilmesi ve bu verilerden bir sonuca varılabilmesi ile ilgili olarak kullanılan bilimsel yöntemler topluluğudur.
- ✓ Belirli amaçlarla toplanan sayısal bilgilere veri denir. Verileri inceleme işi ile uğraşan bilime **istatistik** denir. Başka bir deyişle istatistik; belirli amaçlar için veri toplama, toplanan verileri düzenleme, çözümleme yorumlama teknik ve yöntemleridir. İstatistik, bilimsel yöntemin en güçlü temel araçlarından biridir. Bir başka açıdan istatistik; hem teknik ve yöntemler geliştiren bir bilim, hem de yöntemler topluluğu sayılabilir.

Sosyal Bilimler açısından bakıldığında istatistik; çeşitli anlamlarda kullanılan bir sözcüktür. Günlük yaşamda istatistik deyimi ile çeşitli olaylara ilişkin olarak toplanan rakamlar yani verileri örneğin tarım istatistikleri, turizm istatistikleri gibi rakamlarla istatistik anlatılmak istenir. İstatistik, bu anlamda daima çoğul olarak kullanılır. Ancak istatistik, herhangi bir rakam değildir. Belirli bir olayın gözlenmesiyle onun hacmi, bölünüşü, büyüklüğü vs. hakkında elde edilen rakamların ifade eder. Bu yüzden sözgelimi milli piyangodan ikramiye kazananların listesi ve logaritma cetvellerine istatistik denilemez.

Daha geniş anlamda, istatistik sözcüğü belirli olaylar hakkında nicel bilgilerin toplanmasında, işlenmesinde, analiz ve yorumunda kullanılan bütün yöntemi ifade eder. Ancak bu anlamda istatistik yöntemi şeklinde belirtilmelidir.

İstatistik için yapılabilecek bir başka tanımda örneklem değeri olması ile ilgilidir. İstatistik hakkında bilgi edinilmek istenen bütüne anakütle, ondan alınan parçalara da örneklem adı verilir. Dolayısı ile bir özellik hem anakütlede hem de örneklem de söz konusu olabilmektedir. İşte örnekleme ait olan ve örneklemi karakterize eden değerlere istatistik adı verilir. İstatistikte genellikle örneklem incelenerek anakütle hakkında bilgi sahibi olmaya çalışıldığından, yani kısaca tümevarım yöntemi kullanıldığından istatistik için buna uygun bir tanımda yapılabilmektedir. Buna göre istatistik, örneklem istatistiğinden hareketle anakütle parametresini tahmin etmeye çalışan bir bilim dalıdır şeklinde tanımlanabilir.

İstatistik teknik ve yöntemlerini kullanılış amaçlarına göre, iki genel grupta toplayabiliriz:

Tanımlayıcı İstatistik (descriptive statistics):

Tüm istatistikleri elde etmek için kullanılan ve sonuçları yorum yapmadan veren istatistiklere verilen addır. Tanımlayıcı istatistik verilerin kapsamlı açıklaması şeklinde de tanımlanabilir.

Elde edilen verilerin sınıflandırılması, frekans dağılımlarının yapılması, bu dağılımların ortalamalar, yüzdelikler ve standart sapma gibi ölçülerle tanımlanması ve bunların tablo ya da grafiklerle okuyuculara sunulması tanımlayıcı istatistiğin konularıdır.

Çıkarımsal İstatistik (inferential statistics):

Kitleye ilişkin genellemelerin yapılmasını sağlayan yöntemlerdir. Örneklemden elde edilen bulgularla, örneklemin seçildiği anakütle hakkında tahminlerde bulunma, karşılaştırmalar yapma ve kararlara varma işlemleri çıkarımsal istatistiğin konularıdır.

Bir bilim olarak istatistik, uygulamalı matematiğin bir dalı olup bütün alanlarla ilişkilidir. İstatistiğin genel bir kuramlar, teknikler ve yöntemler bütünü vardır. Bunlar gözlem yapılabilen her alana uygulanır. Ancak, bazı teknik ve yöntemler bazı alanlarda karşılaşılan özel durumlara daha uygun düşer ve daha fazla kullanılır.

Dar alanda istatistik, çeşitli olaylarla ilgili toplanmış veriler demektir. Nüfus, istihdam, kaza istatistikleri, yıllık imalat sanayii anketleri veya Merkez Bankası'nın periyodik olarak yayınladığı bültenler örnek olarak verilebilir. Belirli bir konu ile ilgili sayısal verilerin en uygun şekilde derlenerek açık ve anlaşılır bir şekilde ifade edilmesi büyük önem taşısa da, bu konu genel anlamdaki istatistiğin sadece ilk safhasını meydana getirmektedir.

Günümüzde istatistik, deney yada gözlemlere dayalı tüm bilim dallarında geniş bir uygulama alanına sahiptir. Fen bilimleri (biyoloji, fizik, kimya gibi), sağlık bilimleri (tıp, diş, eczacılık gibi), eğitim bilimleri ve sosyal bilimleri (sosyoloji, psikoloji, coğrafya gibi) çok geniş bir alanda istatistik kullanılmaktadır.

1.3. İSTATİSTİĞİN TARİHSEL GELİŞİMİ

İstatistiğin uygulamada kullanılışı çok uzun bir gelişime sahip olmasına rağmen bilim olarak 19.yy. ve 20.yy. da ortaya çıkmıştır. İnsanlar topluluklar halinde yaşamaya başlayıp devletler kurunca, yönetenler işleri daha düzenli yürütülebilmek için bir takım bilgilere ihtiyaç duydular. Bunlar başlangıçta toplumdaki birey sayısı, asker sayısı hayvan sayısı vb. hususları kapsıyordu.

Zamanla bu bilgiler yenilendi ve gelişti. Eski Mısır'da bazı devlet görevlileri bütün aile reislerinin listesini tutuyorlardı. Yine Mısır'da M.Ö. 3000 yıllarında piramit inşaatına gerekli iş gücü talebini garanti etmek için ilk nüfus sayımı yapılmıştır.

Osmanlı devletinde de ilk dönemlerden itibaren istatistiksel bilgilerin toplanmasına önem verilerek, Orhan Bey zamanında çeşitli sayımlar yapılmıştır.

İstatistiğin gerçek ilerlemesi Yeni Çağ' dan itibaren başlamıştır. 17.yy.'da Fransa'da ilk defa maliye ve dış ticaret istatistikleri düzenlenmesine başlanmıştır. İlk bilimsel nüfus sayıma A.B.D.'de 1790 yılında yapılmıştır.

İstatistik yöntem bilimi, istatistiksel işlemlerin uygulamasından çok sonra ortaya çıkmış ve çeşitli aşamalardan geçerek bugünkü durumunu almıştır. 17.yy.'ın ilk yarısından itibaren bazı Alman Üniversitelerinde okutulan " Devletlerin Özellikleri "adlı yeni bir derste çeşitli ülkelerin tarihi, mali. askeri ve idari özellikleri hakkında bilgi veriliyordu. Bir müddet sonra bu konuya, statüs (devlet) 'den gelme statistik (istatistik) denilmeye başlandı.

Tipik olaylar, birbirinin tam benzeri olan ve tek nedenle bağlı olarak meydana gelen olaylardır. Olayları meydana getiren gelen ve tesadüfi nedenler vardır. Genel yada temel nedenler her olaya eşit derecede ve tek yönde etki eden nedenlerdir. Tesadüfi nedenler ise her olaya farklı derecede ve çift yönde etki eden nedenlerdir. İşte tipik olaylar sadece genel nedenlerin etkisi ile ortaya çıkan ve birbirleri ile tamamen aynı olan dolayısı ile birbirinin incelenmesi ile hepsi hakkında kesin bilgi elde edilen ve tekrarı gerektirmeyen olaylardır. Örneğin; bir kişinin kan grubunu belirlemek için o kişiden defalarca ve vücudunun farklı bölgelerden kan alarak tahlil etmeye gerek yoktur. Aynı şekilde deniz seviyesinde suyun kaynama derecesini belirlemek için olayı defalarca tekrarlamaya gerek yoktur. Cisimlerin yer çekimi nedeni ile yere düşmelerini belirlemek için olayı defalarca tekrarlamaya veya farklı cisimler için bunu denemeye gerek yoktur. Bu tür olaylarda değişkenlik olmadığından benzerlikleri bulup bunları genelemeye çalışmak söz konusu olmadığı için bunları istatistiğin konusu içine girmezler.

1.4. İstatistiğin Konusu

Yığın Olay:İstatistik diğer bilim dalları gibi olayları konu alır. Olay varsa istatistik vardır. Ancak her olay istatistiğe konu oluşturmaz.

Bir olaylar kümesindeki tek bir olay kümedeki diğer olayları temsil edemiyorsa, bu tür olaylara yığın olay denir. İstatistik yığın olaylarla ilgilenir. Örneğin firmaların yıllık ciroları, trafik kazaları, evlenmeler, boşanmalar, doğumlar, ölümler gibi her gün karşılaşılan olaylar yığın olaydır.

Tipik Olay:Eğer bir olaylar kümesinde tek bir olay, tüm olaylar kümesini temsil edebiliyorsa, bu tür olaylara tipik olay denir. Örneğin ideal koşullar altında ve uygun bir laboratuar ortamında iki hidrojen ve bir oksijen atomu bir araya gelirse su elde edilir. Deneyin her tekrarında aynı sonuç elde edileceğinden tek bir deney ilgili olaylar kümesini temsil eder. İstatistik tipik olaylarla ilgilenmez.

Neden İstatistik?

İçinde yaşadığımız Dünya hızla değişmektedir. Günlük gazetelere bir göz atar ve radyo Tv yayınlarını izlerse ele alınan konu ne olursa olsun, sayısal ifadelerin sık sık kullanıldığını görürüz. Ayrıca, günlük yaşantımız süresince karşılaştığımız "ne kadar", " ne zaman", "nerede" 7 "nasıl" ve "kaç tane" gibi soruları çoğu kez sayısal ifadelerle cevaplayabiliriz. İstatistik sayısal bilgileri inceleyen bilim olduğuna göre, istatistik bilgisi en azından anakütlede olup bitenleri anlama ve bunları başkalarına anlatmada yardımcı olur.

Öte yandan, çeşitli alanlarda karşılaşılan sorunlara çözüm yolları bulma ihtiyacı gün geçtikçe artmaktadır. Bunun yanı sıra, sorunları çözümlemede "pratik muhakeme" ya da geleneklere dayanma yerine gözlemlerde bulunarak sonuçları bilimsel yollarla inceleme ihtiyaç ve eğilimini de gün geçtikçe güçlendirmektedir. Böyle bir ortamda farklı alanlarda günlük meselelerin ötesinde işlem ve sorunlarla uğraşan kişilerin, doğal bir parçası olan istatistiğin yöntemlerini bilmesi bir zorunluluktur.

Bir araştırmanın düzeni teknik yönden hatalıysa, hiçbir istatistik teknik ve yöntemi böyle bir araştırmadan geçer ve güvenilir sonuçlar çıkarma olanağı sağlayamaz; başı ve sonu belirsiz verileri anlamlı hale getiremez. Bu nedenle istatistik bilgi ve anlayışı yalnız veriler toplandıktan sonra değil, araştırmanın düzenlenmesi ve yürütülmesi aşamalarında da gereklidir. Bunu sağlamanın en uygun yolu da istatistik teknik ve yöntemlerini gereğince öğrenmektir.

Bir bölgede veya ülkede sağlık durumunun saptanması, başka bölgeler veya ülkelerle karşılaştırılması, buna paralel olarak sağlık tedbirlerinin alınması ve uygulamaların kontrolü, toplumun değişik özelliklerinin sağlık sorunları üzerine etkilerinin saptanması hep istatistik yöntemler kullanılarak yapılır. Koruyucu hekimlikte istatistik yöntemlerin bilinmesi ve uygulanması bu açıdan oldukça önemlidir.

Tedavi edici hekimlikte hastaların yaş, cinsiyet, ırk, ekonomik durumu, mesleği, yaşadığı çevre gibi değişik özelliklerinin hastalığın gelişimi üzerine etkileri, tedavi yöntemlerinin etkinliği v.b durumlar için istatistik analizlerden yararlanmak, sonuçların yorumuna objektiflik kazandırmak, çalışmanın inandırıcılığı yönünden çok önemlidir.

Karşımıza çıkan bir istatistik sonucu aşağıdaki gibi sorgulanmalıdır.

- Önce bilinçli sapma olup olmadığına bakılır. İşe yarayan veriler seçilmiş, işe yaramayan veriler örtbas edilmiş olabilir.
- Uygun olmayan ölçü kullanılmış olabilir. Medyan yerine aritmetik ortalama gibi.
- Örneklem güvenilir bir sonuç verecek kadar büyük mü? Korelasyon bir anlam verecek kadar büyük mü?

- Güvenilirlik ölçüsü (olası hata, standart hata) verilmeden önünüze konan bir korelasyon ciddiye alınmaz.
- Ortalamanın türü muhakkak belirtilmelidir.
- Bazen oranlar verilir, ancak sayılar ortada görünmez. Bu yanıltıcı olabilir. Mesela bir büroda çalışan bayanların %33'ü (1/3) bankacılar ile evlenmiştir. Ancak burada sayı belirtilmemiştir. Bu işyerinde 3 bayan vardır ve biri bankacı ile evlenmiştir.
- Bazen olgulardaki değişikliğe neden olan faktör görünmez. Mesela Ekim ayı perakende satışları geçen yılın Ekim ayına göre artış göstermiştir. Ancak burada özel olan durum, bu seneki ekim ayının ramazan ayına denk gelmesidir. Bu ayrıntı açıklamada yer almamıştır.
- İstatistiği değerlendirirken elde edilmiş sayılardan sonuca giderken bir saptırma yapılıp yapılmadığına bakılır. Mesela Çin'de bir bölgenin nüfusu 28 milyon bulunmuştur. Aynı bölgenin 5 yıl sonraki nüfus sayımı ise 105 milyon çıkmıştır. Ancak burada hangi sayımın yapıldığı belirtilmemiştir. Birinci sayım vergi ve askerlik için, ikinci sayım ise gıda yardımı üzerine yapılmıştır.

1.5. İstatistik Metodunun Aşamaları

İstatistik metodu dört aşamada uygulanır. Bu aşamalar:

i). Bilgilerin Toplanması (Röleveler)

Bu aşama araştırmasının konusunun ve birimlerinin kesin tarifi ile başlar. Rölevenin ne zaman yapılacağına ve kapsamının ne olacağına bu aşamada karar verilir.

ii). Bilgilerin Organize Edilmesi

Bu aşamada toplanmış olan ham veriler matematik ve istatistik analizlere elverişli, düzenli bir hale getirilir. Verilerin tasnif edilmesi ve gruplandırılması bu aşamada yapılması gereken işlerdir.

iii). Verilerin Sunulması

Düzenli ve gruplanmış verilerin tablo ve grafik halinde sunulması ve bu işlemlerle ilgili metotlar bu aşamada uygulanır.

iv). İstatistik Tahlil

Çeşitli metotlar kullanarak düzenli verilerin derinlemesine analizini yapmak, olaylarla ilgili eğilimleri ortaya çıkarmak, istatistik testler yardımıyla sonuca varmak ve karar vermek bu aşamanın incelediği konulardır. Bu metotlar, istatistik metodolojisinin önemli bir kısmını meydana getirir.

1.6. TEMEL KAVRAMLAR

İstatistiğin iyi anlaşılması için istatistikte çok sık kullanılan bazı kavramların anlamlarının, birbirleri ile olan ilişkilerinin ve farklılıklarının iyi bilinmesi gerekir. İstatistik **yığın olaylarla** ilgilenir. Yığın olay, bir olaylar kümesinde tek bir olayın diğerlerini bağlı olarak da ait olduğu kümeyi temsil edemeyen olaylardır.

1.6.1. Veri ve Bilgi

Gözlem: Birimlerde incelenen özelliğin gözlenmesi veya ölçülmesi suretiyle elde edilen değerlerdir.

Veri: İki veya daha fazla denek üzerinden elde edilen bir veya daha fazla değişkene ait sayısal değerler kümesi veridir. Yani bir gözlem veya deney sonucunda ölçümlerle elde edilmiş olan bilgilerdir. Bir klinikte muayene edilen şahıslara ait tansiyon değerleri veridir.

İstatistikte en çok eşanlamlı kullanılan kavramlar veri ve bilgi kavramıdır. **Veri** (data-gözlem) incelenen birimlerin çeşitli özelliklerine ait sembolik değerlerdir. Semboller yerine çoğunlukla rakamlar kullanılır. İstatistiksel olarak veri, analiz ve yorumlama için kullanılan bilgidir.**Bilgi** (information) ise birimlerden elde edilen verilerin işlenerek anlamlı hale getirilmiş halidir. Bir başka ifade ile ham verilerin işlenmiş halidir. Elde edilen veriler birtakım işlemlere tabi tutulduktan sonra yani süzgeçten geçirildikten sonra bilgiye dönüştüklerine göre hacimce küçülürken değerce büyümektedirler. Dolayısı ile veriler hacim olarak büyük değer olarak küçük iken, bilgi aksine hacim olarak küçük değer olarak büyüktür.

Faktör: Birimlerin incelemeye alınan özellikleri üzerinde etkileri olduğu kabul edilen dış etmenlerdir. Birimin incelenen özelliği (diyelimki tansiyon) dışında birimin yaşı, cinsiyeti, sosyo-ekonomik durumu, diğer etmenler faktör olarak alınır.

Risk Faktörü: Bir olayın ortaya çıkmasında kesin etkisi olup olmadığı bilinmeyen, ancak varlığında olayın ortaya çıkmasını etkilediğinden şüphelenilen faktörlere risk faktörü denir. Örneğin sigara akciğer kanseri için bir risk faktörüdür.

1.6.2. Birim (olgu=denek=case)

Üzerinde gözlem ve ölçüm yapılan ve anakütleyi oluşturan en küçük öğeye birim adı verilir. Yığın olay niteliğindeki her bir olaya birim denir. Kütleyi oluşturan ve sayısal olarak incelebilen kollektif olaylardan her birine birim denir.

Birimler canlı yada cansız varlıklar olabileceği gibi, kurum, kuruluş da olabilir. Her ilaç(veya her kutu) bir birimi oluşturabilir. Bir olayın birim olabilmesi için kesinlikle ölçülmeye ve sayılmaya elverişli olması gerekir. İnsan, hayvan gibi canlı bir yaratık, bina, ağaç, araba gibi her hangi bir şey, aile, banka, şirket gibi sosyal bir kuruluş veya doğum, ölüm, evlenme, boşanma, suç işleme gibi bir olay birime örnek olarak gösterilebilir. Renkler, kokular ve rüya birim olmazlar.

i) Maddesel Bir Varlığa Sahip Olan ve Olmayan Birimler:

Eğer birimler insan, araba, ve benzeri gibi canlı yada cansız maddesel bir varlığa sahipse, bu tür birimlere maddesel varlığa sahip birimler denir. Eğer birimler doğum, ölüm, trafik kazası gibi olay niteliğindeyse bu tür birimlere maddesel varlığa sahip olmayan birimler adı verilir.

ii) Doğal-Doğal Olmayan Birimler:

Nitelikleri açısından bir bütün oluşturan, parçalanmaları yada birleştirilmeleri halinde niteliklerini kaybeden birimlere doğal birim adı verilir. Örneğin bir canlı parçalandığında canlı olma niteliğini kaybeder.

Nitelikleri açısından bir bütün oluşturmayan, parçalanmaları yada birleştirilmeleri halinde niteliklerini kaybetmeyen birimlere doğal olmayan birim adı verilir. Bir arsa kaç parçaya bölünürse bölünsün, o parçaları yine arsadır.

iii) Sürekli-Ani Birimler:

Belirli bir zaman aralığı içinde herhangi bir anda gözlenebilen birimlere sürekli birimler adı verilir. Örneğin insan, bina, firma gibi birimler sürekli birimlerdir. Bu tür birimler varlıklarını sürdükleri sürece gözlenebilirler. Maddesel bir varlığa sahip olan birimler aynı zamanda sürekli birimlerdir.

Evlenme, boşanma, trafik kazası gibi ömürleri kısa olan ve aniden ortaya çıkan birimlere de ani birimler denir. Ani birimler maddesel bir varlığa sahip olmayan birimlerdir.

1.6.3. Vasıf

Birimlerin birbirlerinden ayırt edilmesini sağlayan özelliklerine (*characteristic*) **vasıf** (*quality attribute*) denir. Örneğin; bir sınıftaki öğrencilerin her biri bir birimdir. Öğrencilerin boy uzunlukları, yaşları cinsiyetleri, babalarının meslekleri, üniversiteye girişte aldıkları puanlar, bitirdikleri liseler, bir konudaki düşünceleri, kan grupları, IQ puanları, tuttukları futbol takımı, sevdikleri renkler, okudukları gazete, beğendikleri film türü v.s. birer vasıftır. Birimlerin tanımında da olduğu gibi vasıfların tanımında da çok dikkatli davranmak gerekir. İncelenecek vasfın tanımında tereddütlerin ortaya çıkması yapılan çalışmanın güvenilirliğini zedeleyeceğinden bu konuda gerekli hassasiyetin gösterilmesi gerekmektedir.

Vasıfları farklı şekillerde gruplandırmak olanaklıdır. Ancak bu konudaki en önemli ayrım nitel ve nicel vasıf ayrımıdır. Buradaki esas nokta da vasıfların sayılarla ifade edilip edilememesidir. Sayılarla ifade edilebilen nicel, edilemeyenler niteldir.

Ancak vasıfları aldıkları değerler açısından cinsiyet vasfında olduğu gibi bekar, evli, dul, boşanmış şeklinde çok sonuçlu vasıflar diye de sınıflandırmak olanaklıdır.

1.6.4. Gözlem Sonucu ve Şık

Gözlem sonucu yada görünüm ve şık kavramları da istatistikte çok karşılaşılan ve eş anlamlı olarak kullanılan iki kavramdır. Görünüm bir bireyin bir özelliğine ait veri iken, şık çeşitli vasıfların bireylerden bağımsız olarak ortaya çıkış şekilleri yada dereceleridir. Yani görünüm bireye ait bir değer iken, şık özelliğe aittir. Örneğin bir sınıftaki bir öğrencinin boy uzunluğu o öğrenciye ait bir görünümdür. Boy uzunluğu vasfına ait değerler de şık tır. Ali'nin boyu 170 cm dir dendiğinde bu sonuç Ali'nin boyunun görünümüdür. Genel olarak boy uzunluğu vasfına ait 170 cm. ise şıktır.

Bir başka örnek medeni durumu şıkları bekar, evli, dul ve boşanmış olup herkes için geçerli olan değerlerdir. Ali'nin medeni durumu bekardır dendiğinde bu medeni durumun Ali'deki görünümüdür.

1.6.5. Anakütle (Evren-Toplum-Popülasyon)

Üzerinde inceleme veya araştırma yapılacak olayın gözlenebileceği tüm birimlerin yer aldığı topluluktur. Anakütle yığın olay niteliğinde ve aynı cins birimlerin oluşturduğu topluluktur. Bir fabrikanın ürettiği aynı türden ilaçlar anakütleyi oluşturur.

Herhangi bir gözlem yada inceleme kapsamına giren obje yada bireylerin tümüne anakütle ya da kütle denir. Gözlemin amacına bağlı olarak, anakütle küçülebilir yada büyüyebilir. Örneğin bir araştırmacı Türkiye'de 5 yaşında çocukların boy uzunlukları üzerinde bir inceleme yapmak isterse, araştırmacının anakütleyi Türkiye'de 5 yaşındaki çocukların tümünün oluşturduğu gruptur. Öte yandan başka bir araştırmacı Ankara şehrinde yaşamakta olan 60 yaşından büyük kişilerde bir inceleme yapmak ister ve elde edeceği sonuçları Ankara şehrinin dışında kalan 60 yaşındaki kişilere genelleme amacı taşımazsa, bu araştırmacının anakütleyi Ankara şehrinde yaşayan ve 60 yaşından büyük olanların oluşturduğu grup olur.

Belirli bir amaç için anakütle kabul edilen grup başka bir amaç için anakütle olmayabilir. Anakütlenin sınırlarını anakütleyi kimlerin ya da nelerin oluşturduğunu gözlemin amacı ve gözlem sonuçlarının kimlere genelleneceğini belirler. Anakütlenin sınırlarını belirlemek ve anakütlenin kimlerden yada nelerden oluştuğunu ve sayısını saptamak bazen kolay bazen de çok zor hatta olanaksız olabilir. Bu zorluklar özellikle anakütledeki obje ya da deney sayısını saptamada ortaya çıkar. Çünkü çoğu kez belirli bir anakütleye girmesi gereken obje ya da bireyleri teker teker bulup ortaya karmak ve saymak olanaksızdır. Örneğin Türkiye'deki kanserli hastalar Üzerinde inceleme yapmak isteyen bir araştırmacıyı düşünelim. Bu araştırmacı için anakütlenin sınırlarını çizmek, kimlerin anakütleye girip kimlerin anakütleye girmeyeceğini saptamak oldukça kolaydır. Çünkü belirli bir zamanda Türkiye'de yaşayan ve kanserli olan herkes anakütleye dahildir, bunun dışında kalanlar dahil değildir. Fakat Türkiye'deki kanserli hepsini teker teker saptamak ve toplam sayılarını bulmak olanaksızdır. Çünkü kanserli olduğu bilinenlerin yanında kanserli olup ta bilinmeyen daha birçok kimsenin de bulunduğu bir gerçektir. Türkiye'deki bütün insanları kısa bir zaman içinde muayene edip kanserli olanları doğru bir şekilde saptama olanağı olmadığına göre bu anakütledeki birey sayısını doğru olarak bulmak olanaksızdır.

Anakütledeki obje ya da birey sayısını tam bir doğrulukla saptamak, anakütledeki sınırlarını belirlemek ve anakütlenin kimlerden yada nelerden oluştuğunu belirtmek o kadar önemli olmayabilir, pek çok durumda anakütledeki obje ya da birey sayısını saptama yerine onu belirli yollarla tahmin etmeye çalışırız, incelememelerde de tahmin edilen bu sayıyı kullanırız. Böyle yapmak bir bakıma zorunluluktur. Ayrıca istatistikte bazı nitelikleri bilinen bir anakütledeki obje yada birey sayısını oldukça güvenilir ve geçerli bir şekilde tahminde yararlı olabilecek bazı teknikler geliştirilmiştir.

Pek çok araştırma amaçları için anakütleyi oluşturan obje yada bireylerin tümünü ayrı ayrı gözlemlemek olanaksız olduğu gibi zorunlu da değildir. Geliştirilmiş olan bazı teknik yöntemlerden yararlanarak anakütleden seçilecek daha küçük sayıda bir grubu gözleyip elde ettiğimiz sonuçlan anakütleye genelleme olanağı vardır. İstatistik teknik ve yöntemlerinin birçoğu da bu amaçla geliştirilmiştir. Bu tür istatistik teknik ve yöntemlerinin oluşturduğu bu kısma vardamdı istatistik denildiği bundan Önceki bölümde belirtilmişti.

Nüfus, yığın, anakütle (*population universe*) gibi adlarla da ifade edilen anakütle (*population*) incelenen konudaki olası tüm birimlerin oluşturduğu topluluktur. Anakütle birimler topluluğu yerine gözlem sonuçlarının oluşturduğu topluluk diye de tanımlanabilir. Anakütle N sembolü ile gösterilir ve anakütle hacmi yada büyüklüğü (*population size*) şeklinde ifade edilir.

Anakütleleri farklı şekillerde gruplamak olanaklı ise de en önemli ayrım sonlu ve sonsuz anakütle ayrımıdır. Sonlu anakütlede ilk ve son gözlem sonucunda bilinirken, sonsuz anakütlede ilk gözlem sonucu bilinirken son gözlem sonucu bilinmez. Bu ayrım sayılabilir sayıda birim içeren ve sayılamayan sayıda birim içeren anakütleler diye de yapılabilir. Örneğin bir sınıftaki öğrenciler bir sonlu anakütle iken, bir fabrikada üretilen ampuller yada Marmara denizindeki balıklar sonsuz anakütledir.

1.6.6. Örneklem (Sample)

Örneklem, anakütleden seçilen ve anakütleye göre daha az sayıda birimden oluşan topluluktur. Örneklem, gözlem sonuçları açısından da anakütlede ulaşılabilen yada elde edilebilen gözlem sonuçlarını oluşturduğu topluluk şeklinde tanımlanabilir. Örneklem istatistikte n sembolü ile gösterilir ve örneklem hacmi (*sample size*) veya örneklem büyüklüğü diye de ifade edilir. Örneklem ve anakütle hacimleri arasında $n<N$ durumu geçerli olup $n=N$ durumunda örneklem kavramı önemini kaybeder.

Dikkat edilecek olursa anakütle bir konudaki tüm birimleri yada olanaklı gözlem sonuçlarını kapsarken, örneklem ona göre daha az sayıda birim yada gözlem sonucundan meydana gelmektedir. Çünkü bir konudaki olanaklı tüm gözlem sonuçlarına ulaşmak her zaman söz konusu olamaz. Bir başka ifade ile sonsuz bir anakütlede tüm birimlere ait gözlem sonuçlarına ulaşılamazken, ulaşılabilenlerle yetinip örneklem elde edilir. Örneklem konusunda dikkat edilmesi gereken en önemli nokta örneklemin tarafsız olması ve anakütleyi iyi temsil etmesidir.

Herhangi bir anakütleden belirli bir yolla seçilmiş daha küçük sayıdaki obje ve bireylerin oluşturduğu gruba örneklem denir. Örneklemden edindiğimiz bilgilere dayanarak anakütle hakkında tahminde bulunuruz. Çünkü pek çok durumda asıl amacımız örneklem grubunu tanımlamak değil, anakütleyi tanımak, onunla ilgili sonuçlar çıkararak karar vermektir.

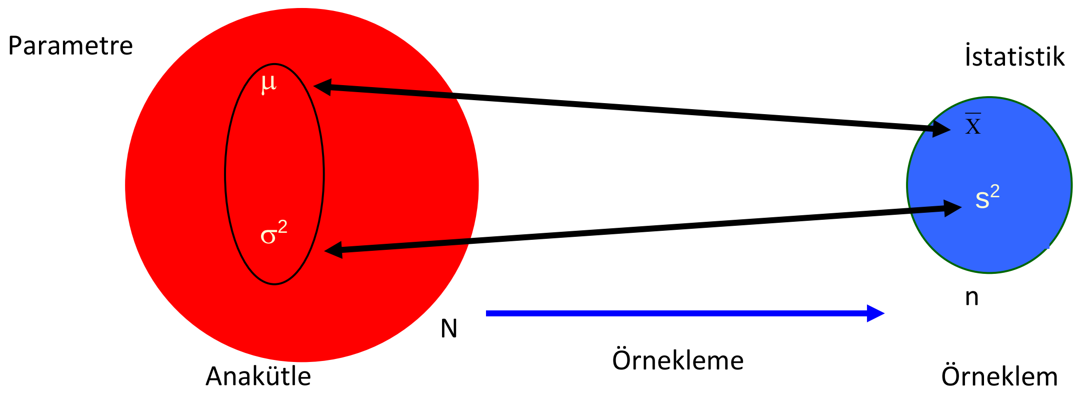
Örneklem grubu üzerinde gözlem sonuçlarını genellerden en az hata ile tahminde bulunmak için örneklemin anakütleyi temsil etmesi temel nitelikleri yansıtması gerekir. Örneklemin anakütleyi temsil etmesi içinde en başta yansız olması gerekir. Herhangi bir örneklem grubu seçildiği anakütleyi belirli bir alt gruba (alt anakütleye) ya da bazı niteliklere sahip olanlara gerçekte olduğundan daha çok ya da daha az ağırlık vermeden temel nitelikleriyle yansıtıyorsa, ya da temsil ediyorsa bu gibi örneklemelere yansız örneklemler denir. Öte yandan seçildiği anakütleyi temel nitelikleriyle tam yansıtmaya, bazı alt gruplara ya da belirli niteliklere taşımaya gerçekte olduğundan daha çok ya da daha az ağırlık veren örneklemelere de yanlı örneklemler denir. Yanlı örneklemelerden elde edilecek bilgiler anakütledeki durumu tam yansıtmayacağından yanıltıcı olur.

Bu anakütleden amaca uygun örneklem seçme işine örnekleme denir. Anakütlelerden yansız örneklemler seçebilmek için geliştirilmiş çeşitli örnekleme yöntemleri vardır. Bu yöntemler ve uygulaması istatistiğin çok ilginç bir o kadar da karmaşık çalışma alanlarından biridir. Bir örnekleme işleminde araştırmacının amacına anakütlenin yapısına ve olanaklara bağlı olarak bu yöntemlerden bir ya da birkaçı birlikte kullanılabilir.

Parametre: Anakütlenin özelliklerini belirleyen sayısal karakteristiklere parametre adı verilir. Anakütleyi tanımlamada kullanılabilen tipik değerlerdir. Anakütle aritmetik ortalaması, anakütle varyansı gibi değerlerdir.

İstatistik (statistics): Örneği oluşturan birimlerden hesaplanmış, anakütleyi tanımlayan değerlere karşılık gelen değerlerdir. Örnek ortalaması, örnek varyansı gibi.

Örnekleme (sampling): Bir örneklem yardımıyla ilgilenilen anakütleye ilişkin genelleme yapma sürecine örnekleme denir.



Şekil 1.1. Anakütle-örneklem ilişkisi

Şekil 1.1.'de görüldüğü gibi anakütlenin aritmetik ortalaması (μ) ve anakütlenin varyansı (σ^2) parametredir. Anakütledeki toplam gözlem sayısı N ile gösterilir. Örneklemin aritmetik ortalaması ($\overline{X}$) ve örneklemin varyansı (S^2) ise istatistiktir. Anakütleden seçilen örneklem sayısı ise **n** kadardır.

Tamsayım : Sonlu bir anakütlenin bütün birimlerinin incelenmesi yada sayılması işlemidir.

Örnek 1.1. Yeni bir ücret sisteminin uygulandığı 30 işçisi olan bir işletmede, işçilerin yeni ücret sisteminden memnuniyetleri araştırılmak istenmektedir. Burada tamsayım yapılabilir mi?

Çözüm : Burada anakütle N=30 işçiden oluşmaktadır ve küçük hacimli bir anakütledir. İşçilerin her birine ulaşmak ve bunlardan veri elde etmek kolaydır. Bu yüzden tamsayım yapılabilir.

Örnek 1.2. 25000 öğrencisi bulunan bir üniversitede, öğrencilerin kendilerine sunulan hizmetleri yeterli bulup bulmadıklarını belirlemek amacıyla, bir araştırma planlanmış ve rasgele seçilen 400 öğrenciden görüşleri alışmıştır. Bu araştırmada anakütle hacmi ve örneklem kaçtır? Neden tamsayım yapılmamıştır?

Çözüm : Anakütle büyük hacimli sonlu bir anakütledir ve N=25000, örneklem ise n=400 işçiden oluşmaktadır. 25000 öğrencinin görüşüne başvurmak, onlara ulaşmak zordur. Bu yüzden tamsayım oldukça zordur.

Örnek 1.3. Bir fabrikada üretilen bisküvi paketlerinin, planlanan ağırlıkta üretilip üretilmediğinin araştırılması amacıyla, üretilen paketler arasından 200 paket seçilmiştir. Bu araştırmada anakütle ve örneklem nedir? Tamsayımın yapılıp yapılmayacağını açıklayınız?

Çözüm : Anakütle sonsuzdur. Açıktır ki bu tür anakütleler üzerinde tam sayım yapılamaz., örnekleme zorunludur. Örneklem 250 bisküvi paketinden oluşan topluluktur.

Örnekleme Yapmayı Gerekli Kılan Sebepler:

Maliyet: Popülasyonun hepsini incelemek çok masraflı olabilir. Popülasyondan alınacak küçük örnekler yardımı ile gerçeğe yakın bilgiler elde edilebilir.

Zaman: Popülasyonla yapılacak bir çalışma çok uzun zamana ihtiyaç gösterebilir. Halbuki örnekle çalışılırsa kısa zamanda gerçeğe yakın bilgiler kısa zamanda elde edilebilir.

Örneğe giren birimlerin fiziksel zarara uğramaması: Birçok durumda gözlemlerin elde edilmesi deneklerin yok edilmesini gerektirebilir. Örneğin bir ilaç üzerinde deneme yapılıyorsa fabrikanın ürettiği tüm ilaçları denemeye almak ve yok etmek mümkün değildir.

Doğru veri etme: Küçük sayıda örneklerle çalışılırken daha hassas çalışma yapmak ve daha dikkatli ölçüm almak, daha hassas alet ve yöntemler kullanmak mümkündür. Yapılan işin denetlenmesi de daha kolay olur.

1.7. Değişken (Variable)

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. Değişken kitleyi oluşturan birimlerin ölçülebilen yada sayılabilen özelliklerine denir. Değişken birimden birime değişen bir sayı ile ifade edilebilen bir bilgidir.

İstatistik birimlerinin sahip oldukları özelliklere **değişken**, değişkenlerin aldıkları değerler ise **şık** adı verilir.

Gözlemden gözleme farklı değerler alabilen objelere niteliklere ya da durumlara değişken denir. Gözlemden gözleme farklı değerler alabilme iki durumda ortaya çıkar birinci şekilde bir obje ya da bireyin belirli bir niteliği üzerinde iki ayrı zamanda gözlemde bulunup ayrı ayrı değerler gözleyebiliriz. Bir şehirde sıcaklığın sabah ve öğleyin ayrı ayrı ölçülüp farklı sonuçlar bulunması ya da bir çocuğun ağırlığının 6 ay arayla iki kez ölçülüp farklı sonuçlar alınabilmesi gibi. Birinci örnekte değişkenimiz havanın sıcaklığı, ikincisinde ise ağırlıktır. Her iki durumda da gözlemler arasındaki farklar gözlemlenebileceğinden değişken tanımımıza uymaktadır.

İkinci durumda ise tek bir nitelikle ilgili gözlem işini yaklaşık olarak aynı zamanda ve başka obje ya da bireyler üzerinde ya da ortamlarda yapı farklı sonuçlar alabiliriz. Örneğin günün belirli bir zamanında Türkiye'de 25 şehirde hava sıcaklığını ölçüp farklı sonuçlar gözleyebiliriz ya da aynı yaştaki 15 çocuğu tarttığımız zaman ağırlıklarının az çok değiştiğini görürüz. İnceleme konusu yaptığımız sıcaklık ve ağırlık bu durumda da değişken tanımımıza uymaktadır.

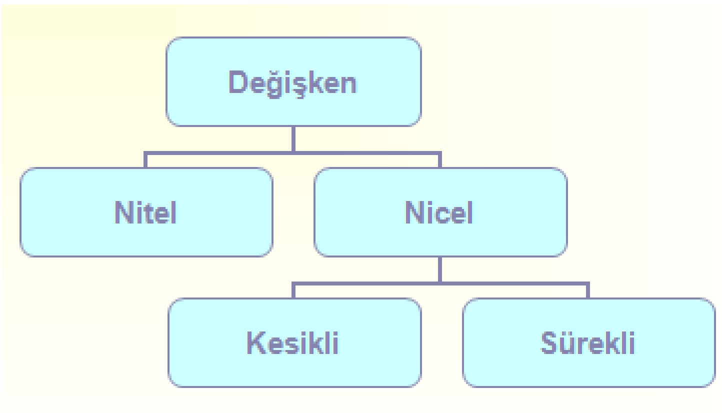
Yukarıda verdiğimiz tanım ve açıklamadan değişkenlerin türleri olabileceği sonucu çıkarılabilir. Aslında değişkenlerin türleri yerine sınıflamasından bahsetmek daha doğrudur. Değişkenler farklı şekillerde sınıflanabilir. Ancak burada her durum için geçerli sayılabilecek bir sınıflama yapma yerine uygulamada sıkça kullanılan bazı deyimler örneklerle açıklanacaktır. Değişkenler X, Y, Z, W gibi harflerle genelde gösterilirler.

İstatistikte **değişken***(variable)* terimi, deneklere ait özellikler anlamında kullanılır. Örneğin iki gruptan birine A ilacı, diğerine B ilacının verildiği ve deneklerin ilaca verdikleri yanıtın karşılaştırıldığı bir klinik çalışma düşünelim. Bu çalışmada hastaların yaşı, bir değişkendir. Aynı şekilde cinsiyet, boy, ağırlık, kan basıncı, hangi ilaç grubuna girdikleri, tedavi öncesi kan basıncı vb. hepsi birer değişkendir. Pratik anlamda düşünürsek, her hastaya ait veri giriş formundaki her bilgi bir değişkendir.

Eğer bir değişkenin şıkları mekana göre oluşuyorsa bu tür değişkenlere **mekan değişkeni** adı verilir. Örneğin doğum yeri ve üniversitelerin bulundukları şehirler mekan değişkenidir. Değişkenin şıkları zamana göre oluşuyorsa bu tür değişkenlere **zaman değişkeni** denir. Örneğin doğum yılları, üniversitelerin kuruluş yılları, aylara göre enflasyon rakamları zaman değişkenidir. Zaman ve mekan değişkenleri dışındaki tüm değişkenlere **maddesel değişken** adı verilir. Örneğin insanların medeni durumu, öğrenci notları, işletmenin maliyetleri maddesel değişkendirler.

Değişkenlerin Özellikleri

Değişkenler gözlenme biçimlerine Nitel ve Nicel değişken olarak ikiye ayrılır.



Şekil 1.2. Değişkenlerin özellikleri

Nicel değişkenler aldıkları değerlerin türüne göre sürekli veya kesikli değişkenler olarak isimlendirilirler**. Kesikli değişkenler sayılarak elde edilen veriler alırken, sürekli değişkenler ise bazı ölçümler sonunda elde edilen verileri alırlar.**

1.7.1.Nitel Değişkenler (Kalitatif-Qualitative Variables)

Bu değişkenler gözlemden gözleme farklılık gösterirler, ancak bu farklılık derece yönünden değil kalite ve çeşit yönündendir. Cinsiyet, medeni durum, göz rengi, din vb.

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 durumları açıklayan değişkenlerdir.

Sosyal bilimlerde incelenen ölçümlerin büyük bir kısmı nitel türden değişkenler üzerinde yapılan gözlemlerden elde edilir. Nitel veriler sınıflandırırken değişkenin kaç kategorisi (sınıf, alt seçenek) varsa belirlenir ve her bir kategoride kaç birim bulunduğu sayılarak belirlenir.

Nitel verileri el ile sınıflandırırken iki sütunlu bir tablo düzenlenir. Bu tabloda birinci sütuna değişkenin kategorileri yazılır. İkinci sütuna ise veri setindeki her değe ele alınarak bu değerlerin hangi kategoriye girdiği çeteleme yoluyla ait olduğu sınıfa yazılır. Tüm değerler tarandıktan sonra çeteleme değerleri sayılır ve her bir kategoride gözlenen birim sayısını gösteren frekanslar belirlenir.

1.7.2. Nicel Değişkenler (Kantitatif-Quantitative Variables)

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. Değişik derecelerde az ya da çok değerler alabilen değişkendir. Örneğin birimlerin, boy uzunluğu, vücut ağırlığı, kilosu, kan basıncı gibi özellikler nicel değişkenlerdir.

Bu tür değişkenler farklı derecelerde az yada çok değerler alan değişkenlerdir. Yaş, ağırlık, boy uzunluğu, yıllık yada aylık gelir, zeka düzeyi, matematik yada tarih bilgisi, havanın sıcaklığı, hava basıncı, hız, nüfus yoğunluğu vb. nicel değişkenlerdir. Bütün insanlar boy uzunluğu ve ağırlığa sahiptir ancak bunun miktarı kişiden kişiye ya da bir kişi için zamandan zamana değişebilir. Bütün şehirlerde insan yaşar sayısı şehirden şehre değişir. Bu gibi değişkenleri sayabildiğimiz gibi ölçerek derece sırasına koymak ve bir ölçek üzerinde işaretleme olanağı vardır. Bu tür değişkenlerin çoğu genellikle normal adını verdiğimiz türden bir dağılım gösterir. Bazen de bazı gerekçelerle bunların normal bir dağılım gösterdiği ya da göstereceği kabul edilir.

1.7.3. Kesikli Değişkenler

Kesikli değişkenler sayılarak elde edilen verileri alırlar.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.

1.7.4. Sürekli Değişkenler

Sürekli değişkenler bazı ölçümler sonunda elde edilen verilerin değerlerini alırlar. İki ayrı ölçüm arası kuramsal olarak sonsuz parçaya bölünebilir. Ölçüm söz konusu olduğu için sürekli değişken değerleri her zaman tam değeri vermez, 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. Noktadan sonraki hane sürekli veri olduğunu gösterir. Örneğin pamuk tohumu: 1.3; 1.32; 1.32076 gr. Bir ineğin bir senelik süt miktarı 3762 kg tartmaya dayandığı için süreklidir.

Kalitatif özellikler için sayma yoluyla veri elde edilir. Dolayısıyla bu veriler kesiklidir. Örneğin; doğan yavruların erkek yada dişi oldukları (cinsiyet özellikleri) incelendiğinde erkek ve dişilerin sayısı ancak elde edilebilir. Kantitatif özellik için ise hem sayma hem ölçme, hem de tartma yoluyla veri elde edilebilir. Örneğin kök sayısı, balık ağırlığı, tohum çapı sürekli değişkenlerdir.

Sürekli değişkenleri, sınıflayarak kesikli değişkenlere dönüştürebiliriz. Örneğin kalsiyum düzeyi sürekli değişkendir. Eğer kalsiyum düzeyinin rakam olarak değerinden çok, "normalden düşük", "normal" ya da "normalden yüksek" olması önemliyse, "8.9’dan düşük", "8.9-10.1 arasında" ve "10.1’den yüksek" olarak yalnız üç değer alabilen bir kesikli değişkene dönüştürebiliriz. Hastaların yaşlarını da "20’den küçük", "20-34", "35-50", "50’den büyük" gibi sınıflandırarak, kesikli değişkene dönüştürebiliriz.

Sürekli değişkenleri çok gerekmedikçe, kesikli değişkenlere dönüştürmek uygun değildir. İstatistik analiz sırasında hataya yol açmaz, ama daha az bilgi veren yöntemler kullanılmasını gerekli kılabilir. Bu nedenle çalışmalar sırasında verileri toplarken, deneklere ait sayısal değişkenleri sınıflandırmadan, gerçek değerleri ile kaydediniz.

Çizelge 1.1. Sürekli ve Kesikli Veriler

Özellik (birim)	**Değişken**	**Açıklama**
Doğum ağırlığı (kg)	X	Kantitatif özelliktir, sürekli veridir, tartma yoluyla elde edilebilir. 4.72, 5.18,5.02
100 tohum ağırlığı (g)	Y	Kantitatif özelliktir, sürekli veridir, tartma yoluyla elde edilebilir. 10.8, 11.2, 11.3
Laktasyon (yıllık) süt verimi (kg)	Z	Kantitatif özelliktir, sürekli veridir, tartma yoluyla elde edilebilir. 3572, 4080, ..kg
Başaktaki dane veya tohum sayısı (adet)	T	Kantitatif özelliktir, kesikli veridir, sayma yoluyla elde edilir. 38, 62, 78
Cinsiyet	W	Kalitatif özelliktir, kesikli veridir. 26 erkek, 18 dişi
Puldaki halka sayısı (adet)	Q	Kantitatif özelliktir, kesikli veridir, sayma yoluyla elde edilir. 8, 6, 10, 12

1.8. Ölçme Düzeyleri

Ölçme objelere ve ya bireylere, belirli bir özelliğe sahip oluş derecelerini belirtmek için, belirli kurallara uyarak sembolik değerler verme işlemidir. Ölçme gözlem sonuçlarının sayısal sembollerle ifade edilmesidir. Cinsiyet, kilo, sıcaklı birer ölçmedir. Ölçmede ölçme konusu olan şey özelliktir. Bu özellik nesneden nesneye, zamandan zamana değişebilir. Yani bir farklılık mevcuttur. Farklılık ölçme için temeldir.

Değişken için elde edilen değerlerin nasıl yorumlanacağına karar vermede yardımcı olur. Bir ölçüm adlandırma ölçeğinde yapılmışsa buna atfen verilen rakamsal değerler, sadece kısa gösterim içindir. Yanlış seçilen ölçme düzeyleri istatistik analizi etkiler.

İstatistik analize başlamadan önce ilk yapılacak şey, değişkenlerin nasıl ölçüldüğünün belirlenmesidir. İstatistikte ölçüm denildiği zaman anlaşılan, değişkenin alabileceği değerlerle ilgili kısıtlamalardır. Örneğin bir kadının gebelik sayısı 5.5

olamaz, ama yaşı 25.64 yıl olabilir. Ya da yeni doğan bir bebeğin APGAR skoru 0 ile 10 arasındaki tüm tam sayılar olabilir, ama 8.4 ya da 15 olamaz. Değişkenlerin alabileceği değerlerin neler olabileceği, neler olamayacağı, yani nasıl ölçüldüğünü belirlemek, yapılacak istatistik analizin seçimi için çok önemlidir. Dört farklı ölçme düzeyi vardır.

1.8.1. Adlandırma Ölçme Düzeyi (nominal level)

Bir sıralaması olmayan kategorileri temsil eden değişkenler sınıflama ölçeğinde tarif edilir. Rakamlar ise farklı bir sınıfı temsil eder. Verile kodlanabilir fakat sıralanamaz. Yani ölçüm düzeyleri arasında bir sıralama ya da uzaklık-yakınlık gibi belirli bir mesafe yoktur. Aritmetik işlem yapılamaz. Sadece = (eşittir) vardır. Aynı ismi taşıyan fertlerde belirli özelliği taşıma bakımından eşitlik vardır denir.

Örneğin, psikiyatride hastalar şizofrenik, paranoid, manyak depresif, psikonörotik gibi isimlendirilir. Bu isimler hastalığın tipine ait sembollerdir. Bu isimlendirme A,B,C,D, veya 1,2,3,4,5 diye de yapılabilirdi. Bu tip verilerde özellikle parametrik olmayan istatistikler kullanılır.Adlandırma ölçeğine örnek olarak nitel değişkenlerden cinsiyet, bölüm, fakülte, kan grubu, saç rengi verilebilir.

1.8.2. Sıralama Ölçme Düzeyi (ordinal level)

Belli bir sıralamaya göre gözlem sonuçlarının sıralanmasıdır. Sınıflar belli bir özelliğe sahip olma bakımından sıralanabiliyorsa, ölçek sıralama ölçeğindedir. Kategorilerin birinin diğerinden ne kadar büyük, ne kadar önemli olduğu bilinemez.

Ölçme sonucunda verilen sayısal değerler büyükten küçüğe sıralanabilir. Bir özelliğe sahip oluş derecesidir. Örneğin, yarışma 1.'si 2.'si 3.'sü, birinci tercih, ikinci tercih vb. Bu ölçekte bir büyüklük veya önemlilik söz konusudur. Kişilerin eğitim durumunu gösterirken, Eğitimsiz=1, İlkokul=2, Ortaokul=3, Lise=4, Üniversite=5, Yüksek lisans=6 kodları verildiğinde sayı büyüdükçe eğitim düzeyinin arttığı anlaşılır. Ancak 1 ile 2 arasındaki mesafe ile 5 ile 6 arsındaki mesafe aynı değildir. Yani sıralama ölçeğinde kod olarak kullanılan sayılar arasındaki mesafe önemli değildir. Sıralamada en iyiye büyük sayı verilebileceği gibi küçük sayıda verilebilir, bu kullanılacak analiz yöntemini değiştirmez.

Örneğin tümörlü hastaların evresinin girildiği "EVRE" adını verdiğimiz bir değişken olsun. Bu değişkene evre 1’den 4’e karşılık gelmek üzere, 1, 2, 3 ve 4 değerleri girilebilir. Bu değerler belirli bir sıra ifade etmektedir. Örneğin "evre 3, evre 2’den daha ileri evredir", "evre 1 en iyi, evre 4 en kötü evredir" vb. Ama değişkenin düzeyleri arasındaki mesafeler belirli değildir. Örneğin matematik işlem yapıldığında

2 - 1 = 4 - 3 = 1 doğrudur, ama "evre 2, evre1’den ne kadar ileriyse, evre 4 de evre 3’den o kadar ileridir" denemez.

Sıralama ölçme düzeyinde değişken değerleri yalnızca ">" ve "<" işlemleri için sayı gibi değerlendirilir; bunlar dışındaki matematik işlemler uygulanamaz. Sıralama ölçekle elde edilmiş verilere genellikle paramatrik olamayan teknikler uygulanır. Ayrıca parametrik test varsayımları yerine getirilemiyorsa, hangi ölçekle toplanmış olursa olsun parametrik olmayan teknikler tercih edilmelidir. Nominal ve ordinal ölçümle belirtilen değişkenlere **kategorik değişkenler** adı da verilmektedir.

1.8.3. Aralık Ölçme Düzeyi (interval level)

Nesnelerin belli bir başlangıç noktasına göre ve belli bir özelliğe sahip oluş derecesine göre eşit aralıklarla sıralandığı ölçektir. Aralık ölçeğinde başlangıç noktası keyfi seçilir ve bu noktadan itibaren belli bir ölçü birimiyle bölümlenerek genişletilir.

Isıyı ölçen santigard ölçeği bir aralık ölçeğidir. 0 ^{0}C başlangıç noktasıdır. Ölçümler arasındaki farklar birbirinin katı olarak ifade edilebilir. 0 ^{0}C-18 ^{0}C arasındaki fark 9 ^{0}C-18 ^{0}C arasındaki farkın iki katıdır. Ancak 18 ^{0}C ısı 9 ^{0}C ısının iki katıdır denilemez.

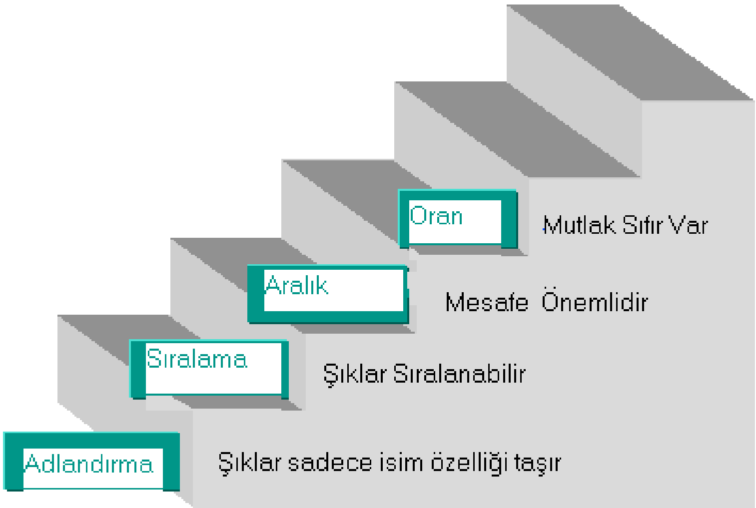
Aralık ölçeğinde sıfır noktası mutlak sıfır noktası değildir. İtibari bir sıfır noktasıdır. Sıcaklığın 0 ^{0}C olması sıcaklığın hiç bulunmaması anlamına gelmez. Bu ölçme düzeyinde toplama, çıkarma yapılabilir. Sayılar arası mesafelerin anlamı önemlidir. ***Bu tip sayılar toplanarak ortalama alınabilir. Ancak çarpma ve bölme işlemleri yapılamaz, oranlar bir anlam taşımaz.***

Üzerinde sıfır noktası 10. noktaya kaydırılan bir metre ile ölçüm yapılırsa, 2m uzunluk 210 cm, 1m uzunluk 110cm olur. Normal bir metrede 2m uzunluk 1m uzunluğun 2 katı olmasına rağmen, sıfır noktası kaydırılarak yapılan ölçümde 210 cm, 110cm’nin 2 katı değildir. İki aralık ölçeğinden birisiyle elde edilen ölçüm diğerine çevrilebilir. ^{0}C derece Fahranayta, hicri takvim miladi takvime çevrilebilir.

1.8.4. Oran Ölçme Düzeyi (ratio level)

Mutlak sıfır noktası olan bir ölçektir. Ölçek üzerindeki noktalar birbirinin katı olarak ifade edilebilir. Tüm matematiksel işlemler kullanılabilir. Metre, kg oran ölçeği için uygundur. Bir şeyin uzunluğu sıfır metre demek ölçülecek uzunluk yok demektir.

Sayılarak elde edilen değişkenlerin çoğu oran ölçeğindedir. Örneğin; geçen 6 aydaki hasta sayısı nedir dendiğinde, bu sıfır olabilir. ***Nicel değişkenlerin ölçme düzeyleri oran ölçeklidir.*** Örneğin boy uzunluğu, kilo, sınavdan alınan notlar, hava sıcaklığı gibi.



Şekil 1.3. Ölçeklerin Güç Hiyerarşisi

1.9. Oran (Ratio)

İki değişken arasındaki ilişkiyi, birinin diğerine bölünmesiyle ifade eden değişkendir. İncelenen birimler (N) içerisinde istenen özelliğe sahip olanların (n) payını gösterir. Oran aşağıdaki gibi bulunur:

$$Oran = \frac{n}{N}$$

Örnek 1.4. Bir sınıfta bulunan 80 kişiden 20 tanesi mavi gözlü ise bu sınıftaki mavi gözlü öğrencilerin oranı;

Oran=20/80=0,25 olur.

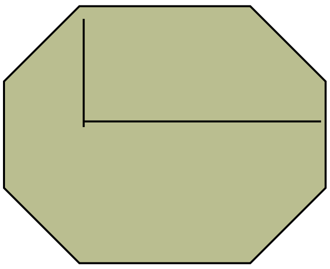
Yüzde : Oranın özel bir halidir. Oran olarak elde edilen değerin 100 ile çarpımının ifadesidir.

$$\text{Yüzde} = \frac{\text{n}}{\text{N}} \times 100$$

Örnek 1.5. Yukarıda ki örnekte sınıfta mavi gözlü öğrencileri yüzdesi;

$$\text{Yüzde} = \frac{20}{80} \times 100 = \%25$$

<u>Hız (Rate):</u> Olay sayısının maruz kalınan insan-yıla (toplam maruziyet süresi) oranıdır. Hız bir olayın oluş olasılığı olarak tanımlanır. Toplumda hız hesabı yapılırken kişi başına risk altındaki zaman izlenemez. Yıl ortası nüfus, yaklaşık olarak insan-yıl olarak toplam maruziyet süresini verir. Orandan farkı, zaman boyutunu kullanmasıdır. Kesrin payında belirli bir zaman içinde belirli bir toplumda saptanan toplam olgu sayısı, paydada ise bu vakaların içinden geldiği toplumda aynı zaman dilimi içinde hastalık ile karşılaşma ve yakalanma riski olan kişilerin (risk altındaki popülasyon) toplam sayısı verilir. Hız ifadesi geleceğe dair bir kestirim imkanı sunar. Saatte belirli bir kilometre mesafe kateden arabanın bir saat sonra nereye varabileceğini tahmin edebilmemiz gibi, hızı bilinen bir hastalığın gelecek yıl kaç kişiyi etkileyebileceği kestirilebilir.



$$\text{HIZ} = \frac{a}{a+b} \times k$$

Hız formülünde *k* 100, 1000, 100.000 gibi alınabilir.

Örnek 1.6 150 kişilik bir toplumda 70 erkek ve 80 bayan varsa;

Erkeklerin bayanlara oranını

Bu toplumda erkek bulunma hızını bulunuz?

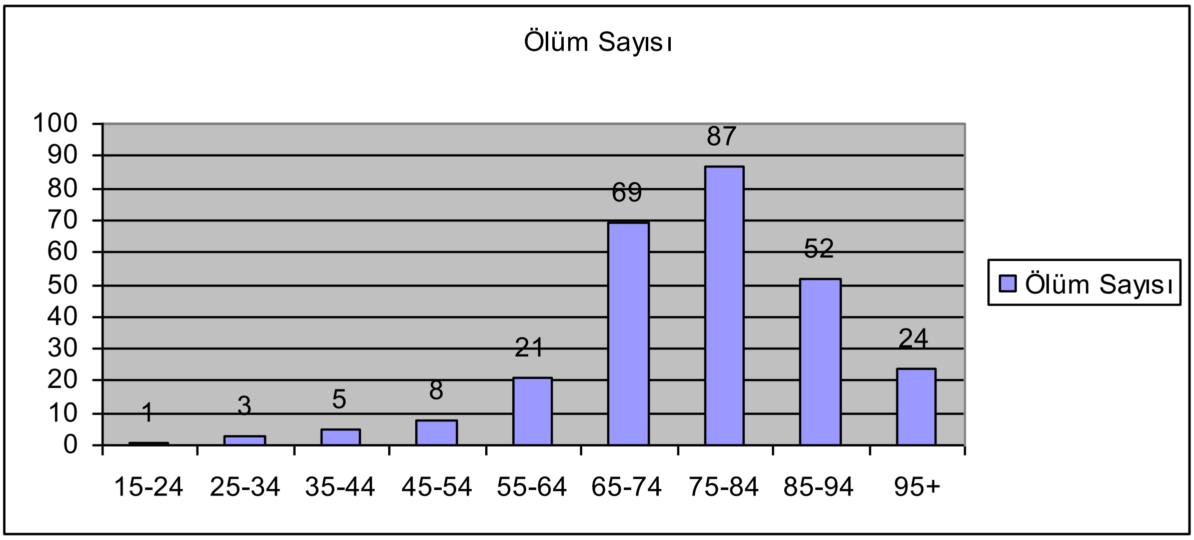
a- Oran=a/b=70/80=0,875

b- Hız=a/(a+b)=70/(70+80)=0,4667 %46,67 binde 466,7

Örnek 2.7. Bir toplumda 15 yaş üzerindeki gruplarda görülen ölüm sayıları aşağıdaki gibidir.

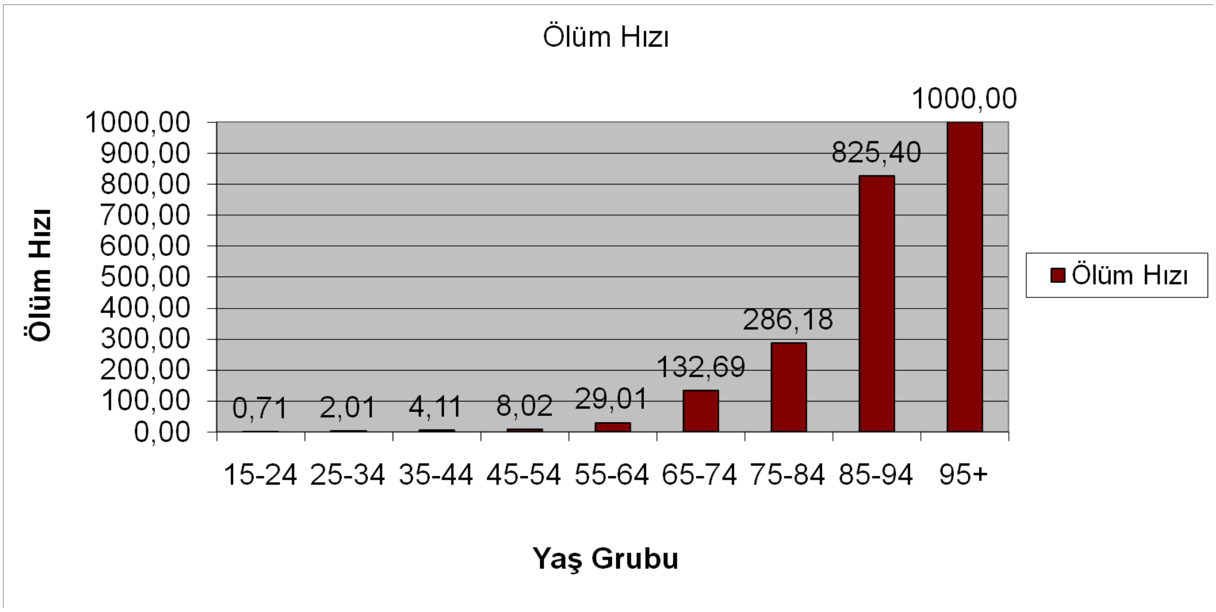
Yaş Grubu	15-24	25-34	35-44	45-54	55-64	65-74	75-84	85-94	95+
Ölüm Sayısı	1	3	5	8	21	69	87	52	24

Tabloda genç yaş gruplarında düşük olan ölüm sayıları yaş ilerledikçe artmakta, belirli bir yaştan sonra ise düşüşe geçmektedir. Veriye göre ölümün genç yaşta az görüldüğü, ileri yaşta fazla görüldüğü ve çok ileri yaşta ise azaldığı söylenebilir.



Yaş Grubu	15-24	25-34	35-44	45-54	55-64	65-74	75-84	85-94	95+
Ölüm Sayısı	1	3	5	8	21	69	87	52	24
Nüfus	1.417	1.496	1.218	998	724	520	304	63	24
Ölüm Hızı (binde)	0,7	2,0	4,1	8,0	29,0	132,7	286,2	825,4	1000,0

Yaş gruplarındaki kişi sayıları farklı verinin hız olarak gösterilmesi gerekir. Yani yaş gruplarındaki ölüm sayıları toplam kişi sayısına bölünerek her yaş grubu için ölüm hızları bulunur. Bulunan değerlere bakıldığında görüldüğü gibi yaş ilerledikçe ölümün arttığı görülmektedir. Yani ölümün yaşla birlikte düzenli olarak arttığı söylenebilir.



1.10. Örnek Problemler

1. Aşağıdakilerden hangisi doğal olmayan birimdir?

a)Araba b) Öğrenci c) Arsa d) Doğum e)Banknot

2. Aşağıdakilerden hangisi ani birimdir?

a) Banknot b) Otomobil c) Kitap d) Boykot e) Öğrenci

3. Aşağıdakilerden hangisi kesikli rassal değişkendir?

a)Sınavda bir sorunun çözülme süresi b) Bir evin fiyatı c)Bir şişe sütün ağırlığı
d) Bir kişinin boy uzunluğu e)Bir ailenin çocuk sayısı

4. Ana kütlenin özeliklerini belirleyen sayısal karakteristiklere ne ad verilir?

a)Parametre b) İstatistik c) Örnekler d)Tamsayım e)Gözlem birimi

5. Aşağıdaki tabloda verilen seri hangi istatistik serisine örnektir?

İller	Yüzölçümleri(km^2)
Afyon	14.718,63
Ankara	25.401,94
Bursa	10.886,38

a) Zaman serisi b) Birikimli seri c)Bileşik seri
d) Dağılma serisi e)Mekan serisi

6. Eşit aralıklı zaman noktalarında bir değişken ile ilgili elde edilen gözlem değerlerini zamana göre sıralanmış olarak gösteren serilere ne ad verilir?

a) Çapraz tablo b) Mekan serisi c) Bölünme serisi
d) Bileşik bölünme serisi e) Zaman serisi

7. Ağıdakilerden hangisi ani birimdir?
a) Renkler b)okul c) Deprem d)Evlenmelere) Kan grubu

8. Sonlu bir kütlenin tüm birimlerinin sayılması işlemine ne ad verilir?
a) Parametreb) Kontrol grubu c) Örnekleme d) İstatistik e) Tam sayım

9. Belirlenen amaçlar doğrultusunda hakkında bilgi edinilmek istenen yığının tümüne ne ad verilir?

a)Topluluk b)Örneklem c)Anakütle d) Örnekleme e)Örnek

10. Aşağıdakilerden hangisi <u>sürekli rassal</u> değişkendir?

a) Medeni durum b) Bir maddenin ağırlığı c) Fakülteler
d) Bir evdeki çocuk sayısı e) Kan grubu

11. Aşağıdakilerden hangisi <u>kesikli rassal</u> değişkendir?

a) Sınavda bir sorunun çözülme süresi b) Bir otomobilin fiyatı
c) Bir madenin ağırlığı d) Bir kişinin boy uzunluğu e) Bir galerideki araç sayısı

12. Anakütleden uygun tekniklerle seçilen alt birimlerin topluluğuna ne denir?

a)Değişken b) Kütle c) Örneklem d) İstatistik e) Birim

13. Cinsiyet (Erkek, Bayan) değişkeninin ölçme düzeyi nedir?

a) Adlandırma b) Sıralama c) Aralık d) Oran e) Hiçbiri

14. Eğitim düzeyi (ilkokul, Orta, Lise, Üniversite) değişkeninin ölçme düzeyi nedir?

a) Adlandırma b) Sıralama c) Aralık d) Oran e) Hiçbiri

15. Bir sınava giren 50 öğrenciye ait notların ölçme düzeyi ne olur?
a) Adlandırma b) Sıralama c) Aralık d) Oran e) Hiçbiri

II. BÖLÜM

2. İSTATİSTİK SERİLERİ VE FREKANSLAR

2.1. İstatistik Serileri

İstatistik serileri gözlem değerlerinin büyüklüklerine göre sıralanmasıyla oluşturulur. İstatistik serileri ilgili yığın olayın kavranması açısından etkin bir araç olup, istatistik analizlere temel oluşturur. Değişik ölçütler temel alınarak istatistik seriyle ilgili farklı sınıflandırmalar yapılabilir.

2.1.1. Zaman ve Mekan Serileri

Eğer gözlem sonuçları yıl, ay, hafta, gün yada saat gibi bir zaman değişkeninin şıklarına göre sıralanırsa oluşturulan seriye **zaman serisi** adı verilir. Yıllara göre ülke nüfusları, günün belli saatlerindeki trafik yoğunluğu, aylara göre ortalama sıcaklık değerleri zaman serisine örnek verilebilir.

Örnek 2.1. Eskişehir ili aylara göre 2000 yılı 6 aylık ortalama sıcaklık değerleri (C°) aşağıdaki gibidir.

Ocak	Şubat	Mart	Nisan	Mayıs	Haziran
-1,5	1,3	4,9	10,4	15,1	18,8

Eğer gözlem sonuçları ülke, bölge, şehir yada köy gibi mekan değişkeninin şıklarına göre sıralanırsa, elde edilen seriye mekan serisi adı verilir. Şehirlere göre elektrik tüketimi, bölgelere göre tahıl üretimi bu tür serilere örnek verilebilir.

Örnek 2.2. Bazı illerin denizden yükseklikleri (m)

Ankara	İzmir	Kars	Muğla	Rize	Sivas	Trabzon	Van	Gaziantep
891	29	1755	646	9	1285	30	1661	855

2.1.2. Dağılma Serileri

Gözlem sonuçlarının maddesel bir değişkenin (zaman ve mekan değişkenleri dışındaki değişkenler) şıklarına göre sıralanmasıyla oluşturulan serilere dağılma serileri adı verilir. Dağılma serileri Basit Seri, Frekans Serisi ve Gruplandırılmış (Sınıflandırılmış) Seriler olarak üç farklı biçimde gösterilir.

i) Basit Seri

İlgilenilen değişkenin değerlerinin basit (sade) biçimde gösterilmesidir. Eğer derlenen veriler ilgilenilen konunun dışında başka bir yönde örneğin gözlem sırasına göre sıralanmış ise, bu sıralamaya **liste** adı verilir.

Örnek 2.3. 20 öğrencinin istatistik notları aşağıdaki gibidir.

70,60,45,85,72,61,38,90,75,78,38,45,45,72,72,90,72,90,45,70

Eğer liste belirtilen amaçlar doğrultusunda düzenlenirse sonuçlara daha kısa sürede ulaşılır. Böyle bir sıralama (küçükten büyüğe doğru sılama gibi) sonucu elde edilen seriye **Basit Seri** adı verilir. Verilen notları daha kolay anlaşılması için liste halinde küçükten büyüğe doğru yazalım.

38	38	45	45	45	45	60	61	70	70	72	72	72	72	75	78	85	90	90	90

Soru. Bir bölgedeki 5 binanın değerleri (TL) olarak aşağıdaki gibi verilmiş olsun.

Bina değeri (TL): 125000 130000 135000 145000 160000

Bu seri basit seridir, çünkü değişkenimiz olan bina değerinin sadece fiyatı verilmiştir.

ii) Frekans Seri

Verilerin daha kolay anlaşılması için gözlem değerlerinin yanına kaç kez tekrarlandığı kaydedilerek oluşturulan seriye **frekans serisi**, tekrarlara da frekans denir.

Örnek 2.4. Örnek 2.3'deki basit seriyi frekans serisi biçiminde gösterelim.

Notlar	38	45	60	61	70	72	75	78	85	90	Toplam
Frekans	2	4	1	1	2	4	1	1	1	3	20

Frekans tablosu ile kolay yorum yapma sağlanır. Mesela sınavda 45 alan 4 kişi vardır.

Soru. Bir bölgedeki 20 binanın değerleri (TL) olarak aşağıdaki gibi verilmiş olsun.

Bina değeri (TL)	125000	130000	135000	145000	160000	Toplam
f	3	4	6	5	2	20

Bu seri frekans serisidir. 125000 TL fiyata sahip 3 bina, 130000 TL fiyata sahip 4 bina vardır. Bir frekans serisi benzer biçimde okunabilir.

Frekans serisi basit seriye dönüştürülebilir.

125000	125000	125000	130000	130000	130000	130000	135000	135000	135000
135000	135000	135000	145000	145000	145000	145000	145000	160000	160000

iii) Gruplanmış Seri

Deney yada gözlem sonuçları çok iken, deney yada gözlem sonuçlarının belirli aralıklar (sınıflar) içinde kalan şıklara göre düzenlenmesiyle oluşturulan istatistik serilerine **gruplandırılmış (sınıflandırılmış) seri** denir.

Eğer bir değişkene ait çok sayıda veri varsa bu veriler gruplanarak, verilerin daha kolay anlaşılması yoluna gidilir.

Not Sınıfı	Frekans
30-45	6
45-60	1
60-75	8
75-90	5
Toplam	20

Bir sınıfın alt ve üst sınırları arasındaki farka **sınıf aralığı** yada **sınıf büyüklüğü** adı verilir ve **h** ile gösterilir. Yukarıdaki tabloda 30 alt sınıf, 45 üst sınıf, 30-45 aralığı 15 olup bu değer sınıf aralığıdır.

Not: Gruplanmış bir seride bir sınıfın üst değeri diğer sınıfın alt değerine eşit verilmişse, alt sınır değeri ilgili sınıfa dahil, üst sınır değeri ise hariç tutulur.

Soru. Bir bölgeye ait arsa fiyatları (TL) aşağıdaki gibi olsun.

Arsa Fiyatları (TL)	f
20.000-30.000 den az	4
30.000-40.000 den az	6
40.000-50.000 den az	10
50.000-60.000 den az	3
60.000-70.000 den az	7
Toplam	30

Bu tabloda 20.000 TL ile 30.000 TL arasında 4 arsa, 30.000 TL ile 40.000 TL arasında 6 arsa, 40.000 TL ile 50.000 TL arasında 10 arsa, 50.000 TL ile 60.000 TL arasında 3 arsa ve 60.000 TL ile 70.000 TL arasında 7 olmak üzere toplam 30 arsa vardır.

Örnek 2.5. Aşağıda verilen frekans dağılımında ilgili sınıflara karşı gelen sınıf orta değerlerini bulunuz?

Sınıflar	f	Sınıf Orta Noktası
0-4	4	(0+4)/2=2
4-8	10	(4+8)/2=6
8-12	17	(8+12)/2=10
12-16	25	(12+16)/2=14
16-20	14	(16+20)/2=18
20-24	6	(20+24)/2=22
24-28	4	(24+28)/2=26
Toplam	80	

2.1.3. Birikimli Seriler

Bir frekans dağılımında, her sınıfın frekansına bir önceki sınıfın frekansı eklenerek oluşturulan seriye birikimli seri, bu tür oluşturulan frekanslara da birikimli frekanslar adı verilir. Eğer birikimli seriler küçükten büyüğe doğru oluşturulmuşsa

–den az, büyükten küçüğe doğru oluşturulmuşsa **–den çok** olarak isimlendirilir.

Örnek 2.6. Bir doğum evinde doğan 100 bebeğe ilişkin sınıflandırılmış seriyi ele alarak –den az ve –den çok birikimli serilerini bulunuz?

Ağırlık (kg)	Frekans	**-den az**	**-den çok**
1,50-1,75	5	5	95+5=100
1,75-2,00	6	6+5=11	89+6=95
2,00-2,25	10	10+11=21	79+10=89
2,25-2,50	10	10+21=31	69+10=79
2,50-2,75	35	35+31=66	34+35=69
2,75-3,00	15	15+66=81	19+15=34
3,00-3,25	13	13+81=94	6+13=19
3,25-3,50	2	2+94=96	4+2=6
3,50-3,75	4	4+96=100	4
Toplam	100		

Yukarıdaki tablodan –den az serisi yardımıyla 66 bebeğin ağırlıklarının 2,75kg az olduğu, -den çok serisi yardımıyla da 34 bebeğin 2,75 kg dan fazla olduğu görülebilir.

Örnek 2.7. Aşağıda verilen frekans serisi için –den az ve –den çok serilerini oluşturunuz?

X	5	10	15	20	25	30	Toplam
f	3	5	8	6	3	5	30

X	5	10	15	20	25	30
f	3	5	8	6	3	5
-den az	3	8	16	22	25	30
-den çok	30	27	22	14	8	5

2.1.4. Bileşik Seriler

Birimlerin birden fazla değişkene göre dağılımlarını bir arada gösteren serilerdir.

Örnek 2.8. Bir sınıftan rasgele seçilen 5 öğrencinin boy uzunlukları ve ağırlıkları aşağıdaki gibidir.

Öğrenci No	Uzunluk (m) X	Ağırlık (kg) Y
1	1,72	68
2	1,68	70
3	1,80	76
4	1,74	73
5	1,76	71

Verilen örnekte birim öğrencidir. Boy uzunluğu ve ağırlık ise aynı birim üzerinde tanımlanmış iki farklı değişkendir. Eğer uzunluk bağımsız, ağırlık bağımlı değişken olarak alınırsa frekans dağılımı aşağıdaki gibi olur.

Uzunluk (m) X	Ağırlık (kg) Y
1,68	70
1,72	68
1,74	73
1,76	71
1,80	76

2.2. Frekans Tabloları

İstatistiksel analizlerden önce istatistiksel varsayımların değerlendirilmesi gerekir. Bunun için aşağıdaki adımlar izlenir:

- ✓ Her değişken ile başlangıç analizi
- ✓ Verilerin temizlenmesi
- ✓ Veri kalitesinin gözden geçirilmesi
- ✓ Verilerin değişiminin belirlenmesi
- ✓ Örneklemenin belirlenmesi

Verilerin büyük değerler etrafında mı toplandığını yoksa dengeli mi dağıldığı hususu, frekans dağılımı ile ortaya konabilir. Hatalı veri girişi, hatalı şık kodlamaları ve eksik veri girişi frekans tablolarında rahatlıkla görülebilir.

Veri seti değerler veya kategoriler halinde organize edilir ve bunlar başlıklarla açıklandığında elde edilen veri gösterimine Tablo denir.

Ham veri : Çeşitli yöntemlerle toplanan ve sayısal olarak düzenlenmemiş verilere ham veriler denir.

Sıralanış : Ham sayısal verilerin artan veya azalan büyüklük sırasına konulmuş durumudur. En büyük ve en küçük sayılar arasındaki farka verilerin açıklığı denir. Örneğin bir sınıfta ki erkek öğrencilerin en uzun boyu olan 190 cm ve en kısa olan 165 cm ise, açıklık 190-165=25 cm dir.

Ham veriler yığını özetlenirken, sıklıkla yararlı bir işlem olan bunların sınıflara veya kategorilere ayrılması ve her sınıfa ait birey sayılarının belirlenmesi sınıf frekansı olarak adlandırılır. Verilerin sınıflara göre ve ilgili sınıfın frekansları ile birlikte bir tablo halinde düzenlenmesine frekans dağılımı veya frekans tablosu adı verilir.

Frekans tabloları ele alınan bir değişkene ilişkin ölçüm değerlerinden istenen sınıflara düşen değerlerin sayısını veya frekansını, oran veya % cinsinden belirlenmiş dağılımlarını gösteren tablolardır. Frekans tablolarının hazırlanmasında aşağıdaki esaslar dikkate alınmalıdır.

Herhangi bir gözlem değeri ancak bir sınıfta yer almalı, sınıflar çakışmamalıdır.

Hiçbir gözlem değeri belirlenen sınıfların dışında kalmamalı, tüm gözlemler sınıflarda yer almalıdır.

Frekans dağılımında kullanılan sınıf aralıkları mümkün olduğunca eşit olmalıdır. Eşit olmayan sınıf aralıkları dağılımların grafiksel olarak gösterilmesinde sorunlar yaratır. Ancak bazen çok sayıda boş sınıfın ayıklanması için eşit olmayan aralıkların kullanılması gerekebilir.

Basit bir serinin gruplanmış seriye dönüştürülmesi:

Frekans dağılımlarının yer aldığı frekans tablolarının oluşturulmasında izlenecek adımlar ve her adımda yapılması gereken işlemler aşağıdaki şekilde özetlenebilir:

1. Sınıf sayısı belirlenir (k).

Kaç sınıf yapılacağını yaklaşık hesaplamak için aşağıdaki formül kullanılır (Lind ve Mason, 1997):

$$Sınıf\ sayısı = 1 + 3{,}322 \times \log(N)$$

2. Her sınıfın sınıf aralığı veya sınıf genişliği belirlenir.

$$C = \frac{X_{max} - X_{min}}{1 + 3{,}322 \times \log(N)}$$

X_{max} : En büyük gözlem değeri

X_{min} : En küçük gözlem değeri

k: Sınıf sayısı

3. Frekans dağılımlarının kolay yorumlanması için bütün sınıfların aynı genişlikte olması önerilir.

4. Her sınıfın tanımında kullanılacak **sınıf orta değeri (SOD)** bulunur.

$$SODi = \frac{ASDi - ÜSDi}{2}$$

ASD_i : i. sınıfa ait alt sınır değeri

$ÜSD_i$: i. sınıfa ait alt sınır değeri

5. Gözlem değerleri tek tek sınıflara birer çizgi çizilerek dağıtılır ve her sınıfa ait frekanslar bu çizgiler sayılarak bulunur.

6. Her sınıftaki frekansların oransal frekansları bulunur.

$$\frac{f_i}{\sum_{i=1}^{k} f_i}$$

7. Oransal frekans değerleri 100 ile çarpılarak her sınıfa ait yüzde frekanslar bulunur.

8. Kümülatif frekanslar ve yüzdeleri bulunur.

Örnek 2.9. Bir araştırmada rasgele seçilen 35 otomobilin yaşları aşağıdaki gibidir. Otomobil dağılımlarını gösteren frekans tablosunu oluşturunuz?

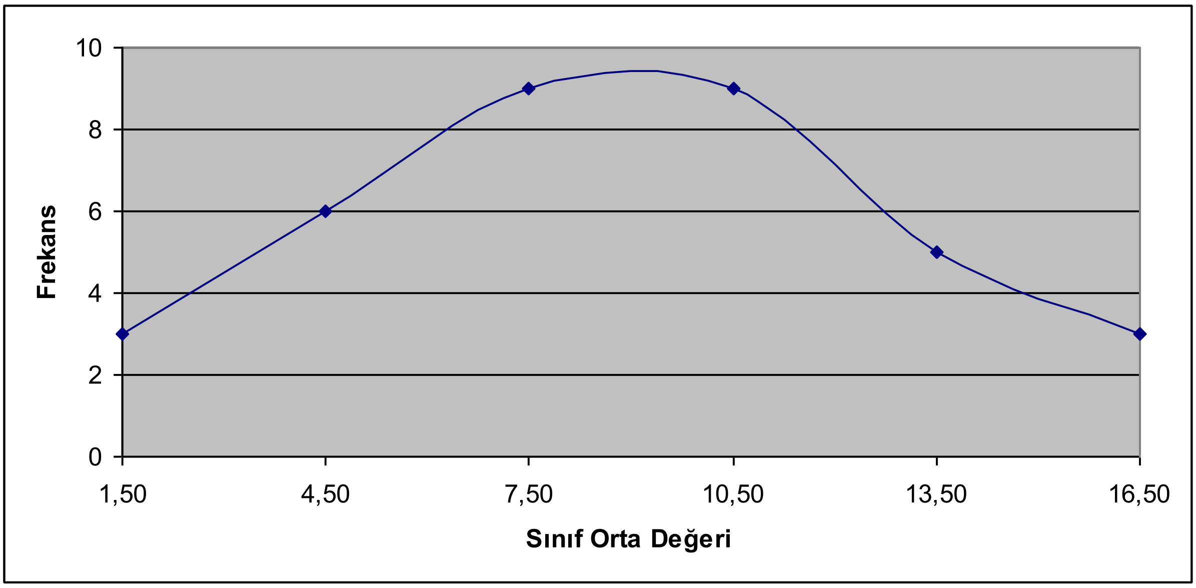
Yaşlar: 17, 14, 9, 8, 4, 9, 5, 6, 7, 12, 17, 10, 9, 9, 1, 11, 3, 2, 6, 7, 12, 13, 15, 5, 8, 7, 9, 12, 11, 6, 1, 3, 7, 4, 10

Sınıf Sayısı (k)=1 + 3.322*log$_{10}$(35)=6.13 ~6

$$C = \frac{17-1}{6} = 3$$

Sınıf No	Sınıf Sınırları *	Sınıf Değeri	Frekans			Kümülatif Frekans		
			Sayısal	Oran	%	Sayısal	Oran	%
1	0-3	1.5	///=3	0.09	9	3	0.09	9
2	3-6	4.5	//////=6	0.18	18	9	0.27	27
3	6-9	7.5	/////////=9	0.25	25	18	0.52	52
4	9-12	10.5	/////////=9	0.25	25	27	0.77	77
5	12-15	13.5	/////=5	0.14	14	32	0.91	91
6	15-18	16.5	///=3	0.09	9	35	1.00	100
Toplam			35	1.0	100			

* Alt sınır değeri ilgili sınıfa dahil, üst sınıf değeri ise dahil değildir.



Şekil 2.1. Çizgi grafiği

Örnek 2.10. Yüz yeni doğan çocuğun doğum ağırlıkları (kg) aşağıdaki gibi bulunmuştur.

2.6	2.6	3,0	2,7	3,8	2,7	3,1	2,7	4,8	4,9
3,6	4,0	4,8	2,7	3,9	2,4	2,0	3,4	2,0	3,9
3,6	4,7	2,3	2,4	3,7	2,7	1,7	2,9	2,4	2,1
5,1	3,2	3,1	4,1	3,0	4,3	3,8	3,3	2,4	3,2
1,7	5,0	4,4	2,6	3,0	3,6	4,7	2,6	2,8	2,7
3,1	2,0	2,4	3,5	2,2	3,9	2,9	3,6	2,6	4,0
2,4	3,3	4,1	2,4	2,2	3,1	4,2	3,9	2,4	3,3
3,8	3,5	**5,2**	2,5	3,4	3,2	3,8	4,1	2,7	3,7
3,3	4,3	3,4	2,1	2,9	3,8	2,2	3,2	2,8	2,9
2,7	5,0	3,1	2,9	2,7	3,5	2,0	2,5	4,7	2,5

Sınıf sayısı= 1+3.322.(log100)= 1+6,644=7.644 = 8 sınıf yapılabilir.

$$C = \frac{X_{max} - X_{min}}{1 + 3{,}322 \times \log(N)} = \frac{5{,}2 - 1{,}7}{8} = 0{,}5$$

Sınıflar oluşturulurken belirli kurallara uymak gerekir. Çok fazla sayıda sınıf yapılırsa veri iyi özetlenmemiş olur. Az sınıf yapılırsa bilgi kaybı çok fazla olur. Bu nedenle optimum sayıda sınıf oluşturmalıdır.

Sınıf genişliği 0.5 kg olacak şekilde doğum ağırlıklarını sıklık(frekans) tablosu halinde özetlemek için 8 sınıf yapılabilir.

Sınıflar	Sıklık(frekans)
1.5 - 1.9	2
2.0 - 2.4	18
2.5 - 2.9	24
3.0 - 3.4	19
3.5 - 3.9	18
4.0 - 4.4	9
4.5 - 4.9	6
5.0 - 5.4	4
Toplam	100

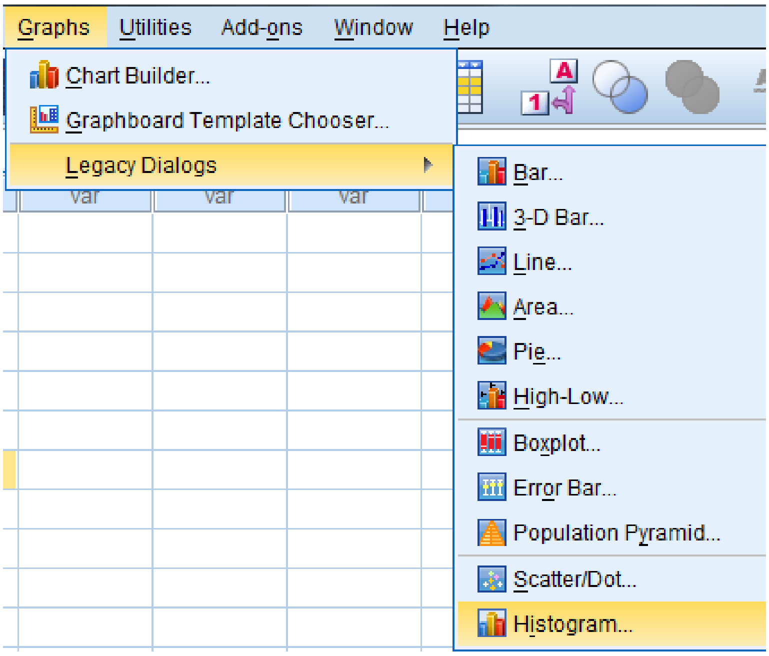
2.3. Grafikler

2.3.1.Basit Serilerinin Grafikle Gösterilmesi

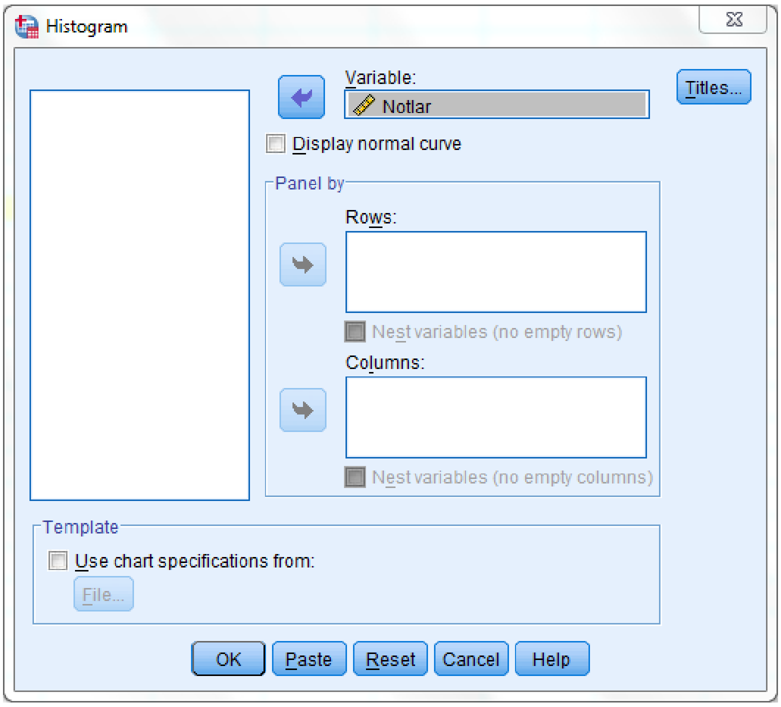
Histogram Grafik:

Örnek 2.11.24 öğrenciye ait bir sınavdan alınan notlar aşağıdaki gibidir. Verilerin **histogram** grafiğini SPSS'de çiziniz.

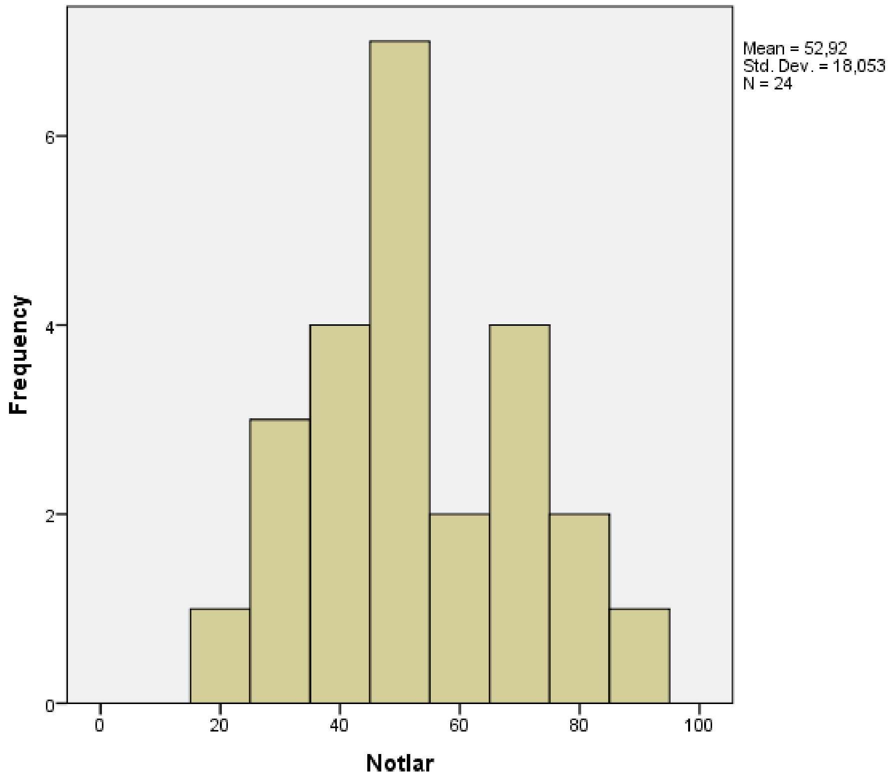
	Notlar
1	20
2	30
3	30
4	30
5	40
6	40
7	40
8	40
9	50
10	50
11	50
12	50
13	50
14	50
15	50
16	60
17	60
18	70
19	70
20	70
21	70
22	80
23	80
24	90



Graphs
Utilities
Add-ons
Window
Help
Chart Builder...
Graphboard Template Chooser...
Legacy Dialogs
Bar...
3-D Bar...
Line...
Area...
Pie...
High-Low...
Boxplot...
Error Bar...
Population Pyramid...
Scatter/Dot...
Histogram...



Histogram
Variable:
Notlar
Titles...
Display normal curve
Panel by
Rows:
Nest variables (no empty rows)
Columns:
Nest variables (no empty columns)
Template
Use chart specifications from:
File...
OK
Paste
Reset
Cancel
Help

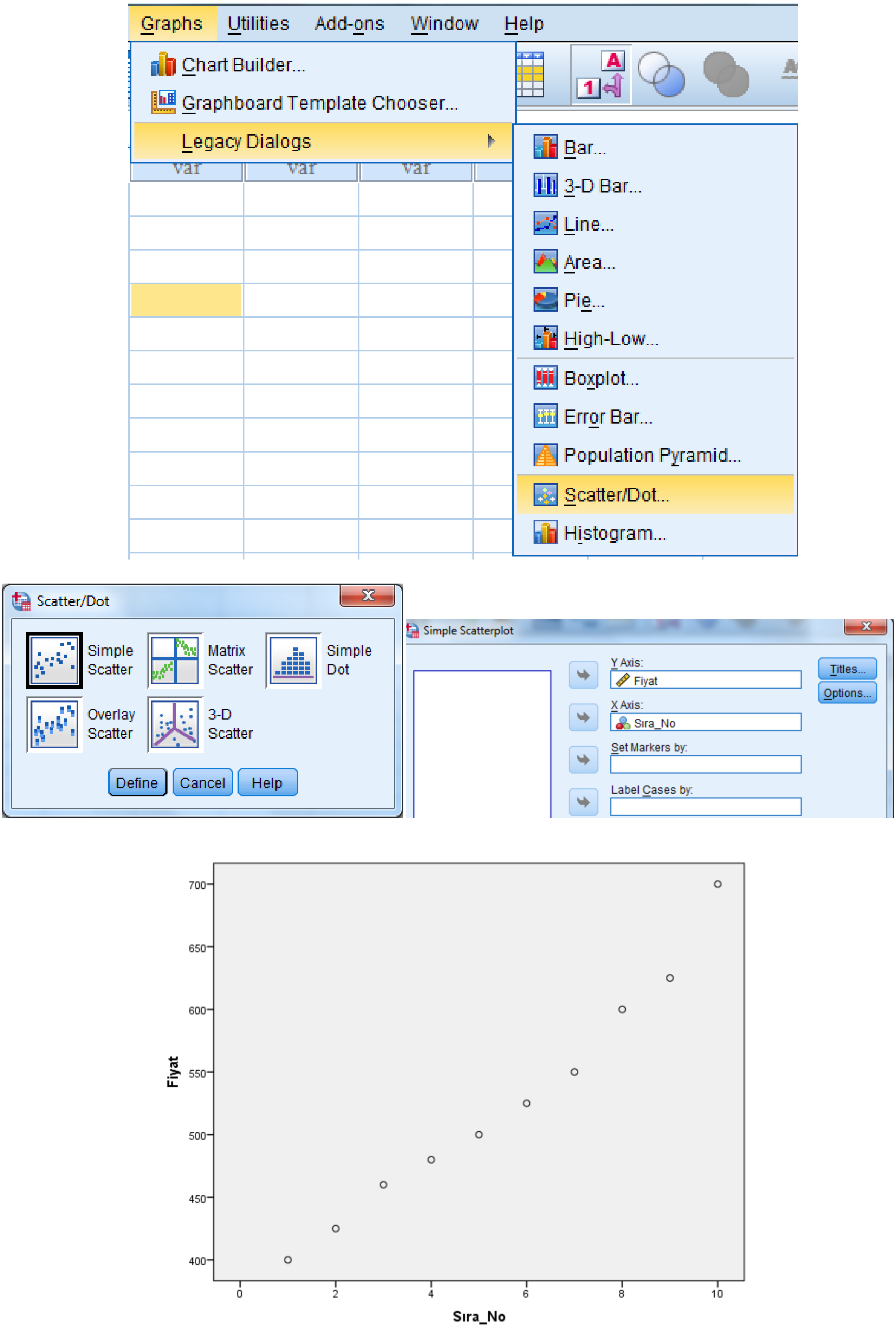


Şekil 2.2.Histogram grafiği

Serpme Diyagramı (Scatter Plot) Grafiği:

Örnek 2.12. Bir bölgedeki kiralık ev fiyatlarının listesi aşağıdaki gibidir. Bu verilerin **serpme diyagramını (Scatter plot)** SPSS'de çiziniz.

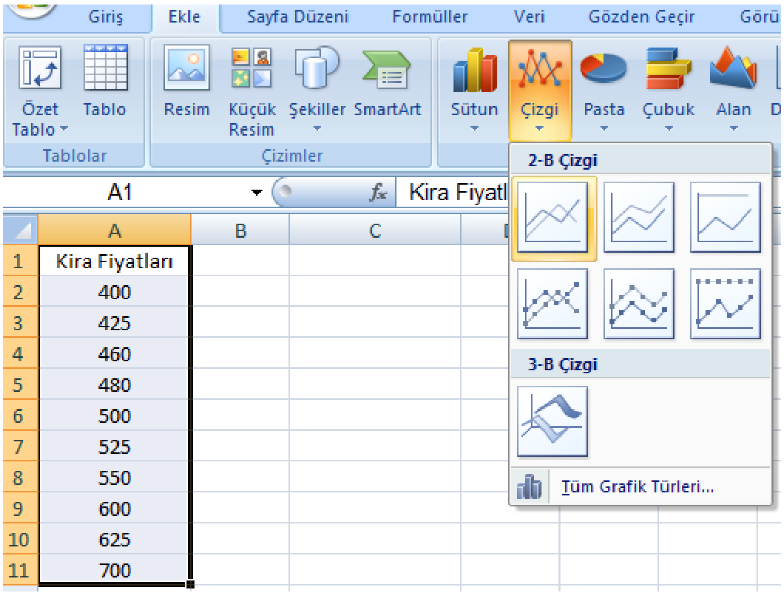
	Sıra_No	Fiyat
1	1	400
2	2	425
3	3	460
4	4	480
5	5	500
6	6	525
7	7	550
8	8	600
9	9	625
10	10	700

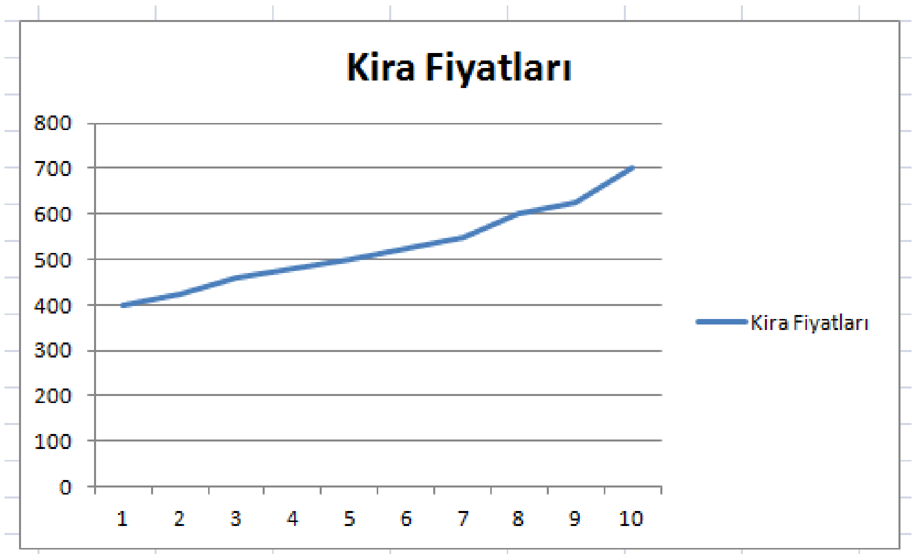


Şekil 2.3. Serpme Diyagramı Grafiği

Çizgi Grafiği:

Örnek 2.13.Örnek 2.12’daki verileri kullanarak Excel’de Çizgi grafiği çiziniz.



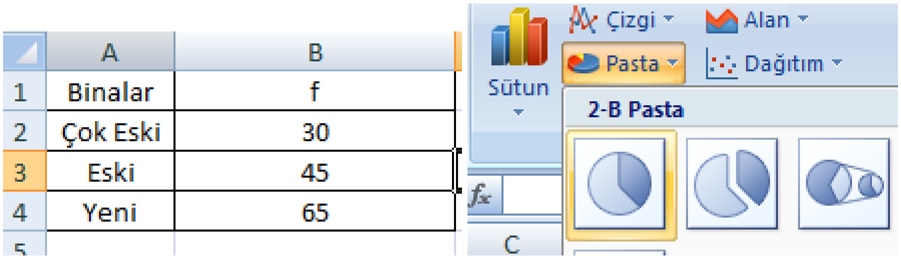


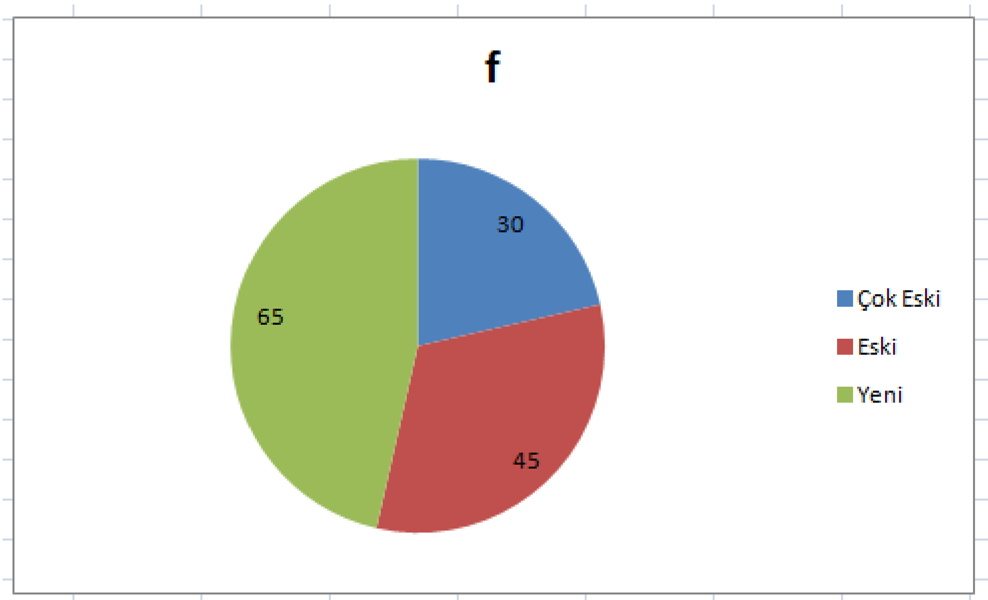
Şekil 2.4. Excel’de Çizgi Grafiği

<u>Pasta (Pie) Grafiği:</u>

Nitel değişkenlerin grafikle gösterildiği en önemli grafiklerden biridir. Grafikte genelde kategorilerin yüzdelik dilimleri gösterilir.

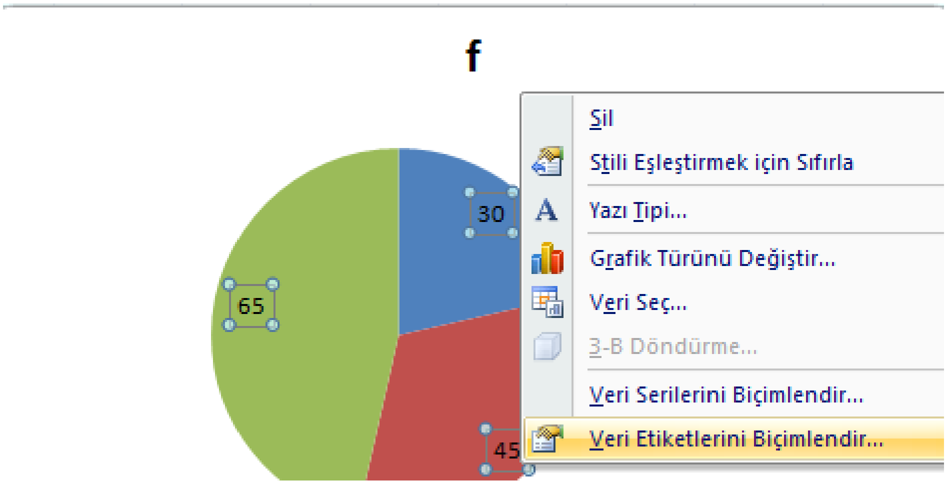
Örnek 2.14. Bir bölgedeki binaların durumuna ilişkin aşağıdaki sonuçlar elde edilmiştir. Bu verilere ait Excel'de pasta grafiğini çiziniz.

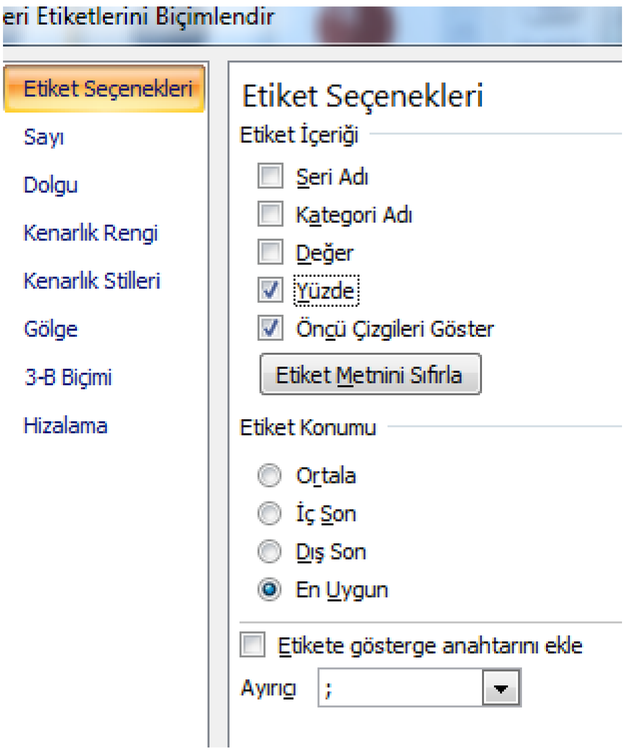


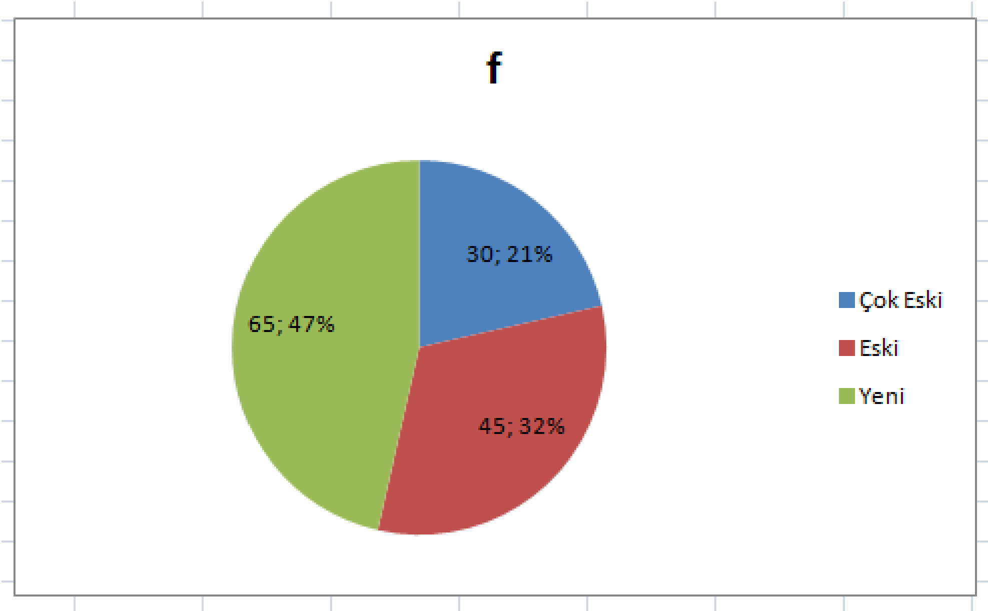


Grafikte frekanslar değil de yüzdeleri gösterilmek isteniyorsa aşağıdaki işlem yapılır.

Grafikteki frekanslar seçilip, farenin sağ tuşu tıklandığında veri etiketlerinin biçimlendir seçilir ve yüzde işaretlenir.


f
30
65
45
Sil
Stili Eşleştirmek için Sıfırla
Yazı Tipi...
Grafik Türünü Değiştir...
Veri Seç...
3-B Döndürme...
Veri Serilerini Biçimlendir...
Veri Etiketlerini Biçimlendir...


eri Etiketlerini Biçimlendir
Etiket Seçenekleri
Sayı
Dolgu
Kenarlık Rengi
Kenarlık Stilleri
Gölge
3-B Biçimi
Hizalama
Etiket Seçenekleri
Etiket İçeriği
Seri Adı
Kategori Adı
Değer
Yüzde
Öncü Çizgileri Göster
Etiket Metnini Sıfırla
Etiket Konumu
Ortala
İç Son
Dış Son
En Uygun
Etikete gösterge anahtarını ekle
Ayırıcı
;



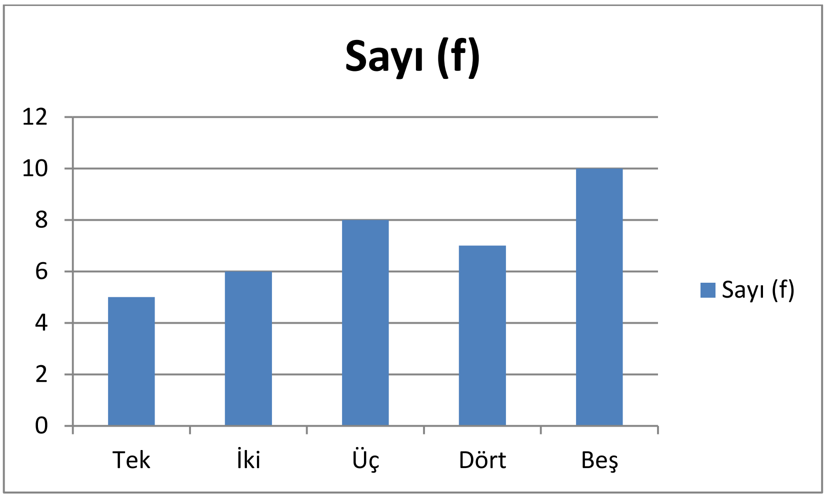
Şekil 2.5. Excel'de Pasta Grafiği

2.3.2. Frekans Serilerinin Grafikle Gösterimi

Çubuk Grafik: Bu tür grafiklerde frekans serilerinde frekanslar gözlem değerlerine göre değiştiğinden gözlem değerleri yatay eksende, frekanslar ise dikey eksende gösterilir.

Örnek 2.15. Bir sokakta yer alan binaların katlara göre dağılımı aşağıdaki gibidir. Çubuk (sütun) grafiğini Excel'de çiziniz.

Bina Katı	Sayı (f)
Tek	5
İki	6
Üç	8
Dört	7
Beş	10

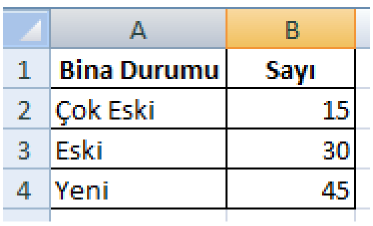


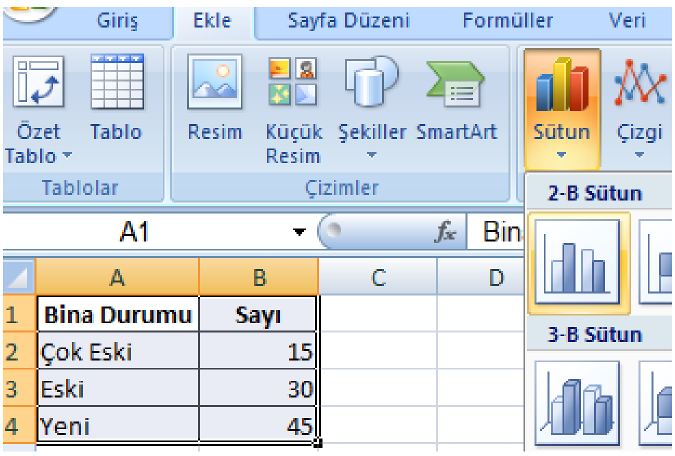
Şekil 2.6. Çubuk grafiği

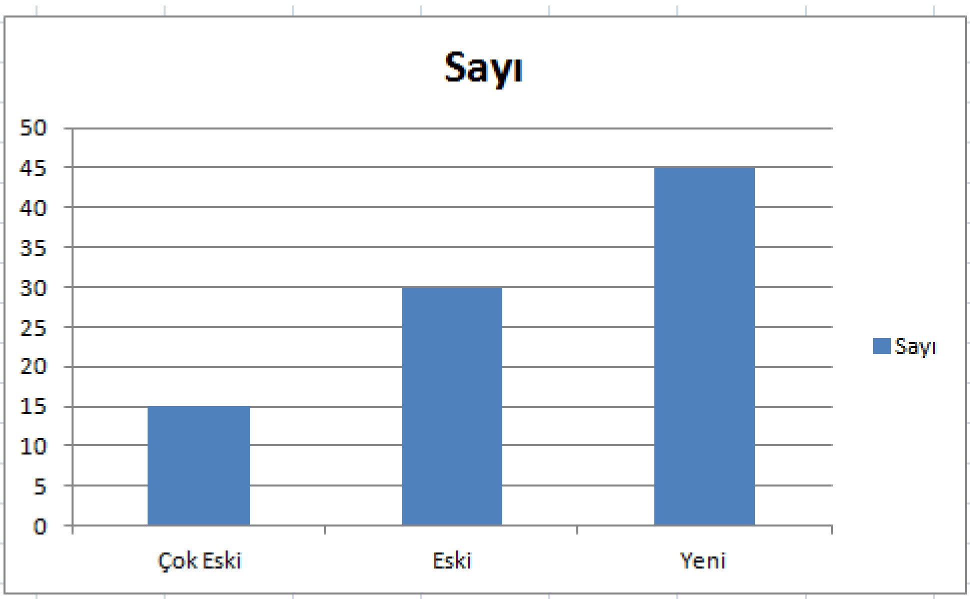
Sütun Grafiği:

Örnek 2.16. Bir sokakta yer alan binaların durumuna ilişkin aşağıdaki veriler elde edilmiştir. Bu verileri kullanarak, 47Excel'de sütun grafiği çiziniz.

Bina Durumu	Sayı
Çok Eski	15
Eski	30
Yeni	45





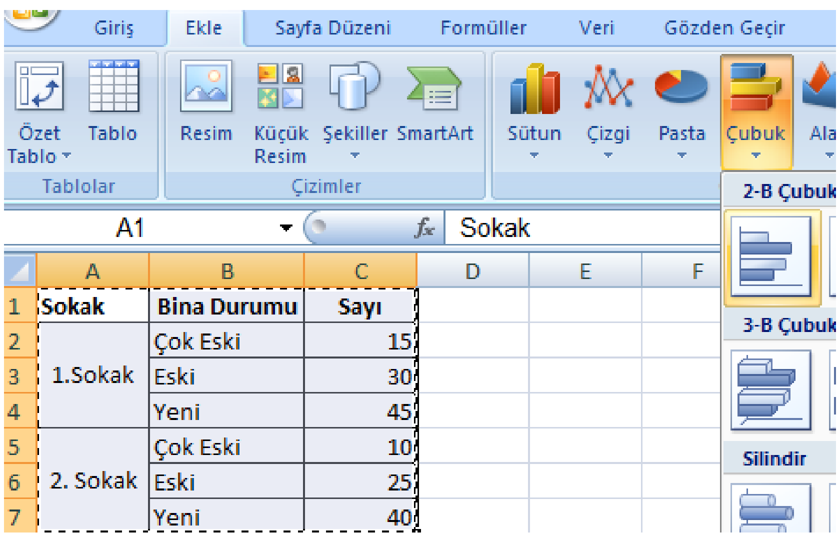


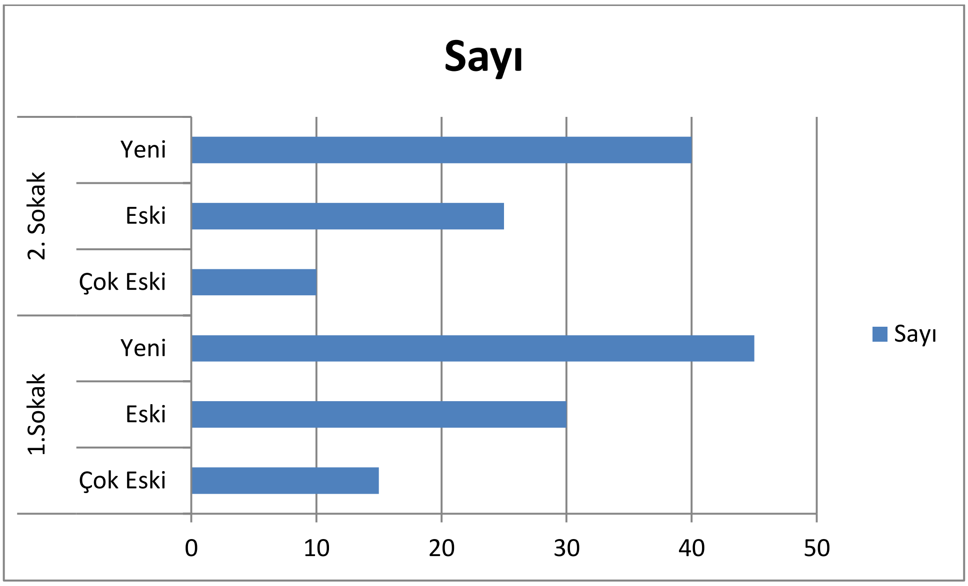
Şekil 2.7. Excel’de Sütun Grafiği

<u>Çubuk Grafiği:</u>

Örnek 2.17. Üç farklı sokaktaki binaların durumuna ilişkin aşağıdaki veriler elde edilmiştir. Bu verileri kullanarak, Excel’de çubuk grafiğini çiziniz.

Sokak	Bina Durumu	Sayı
1.Sokak	Çok Eski	15
	Eski	30
	Yeni	45
2. Sokak	Çok Eski	10
	Eski	25
	Yeni	40



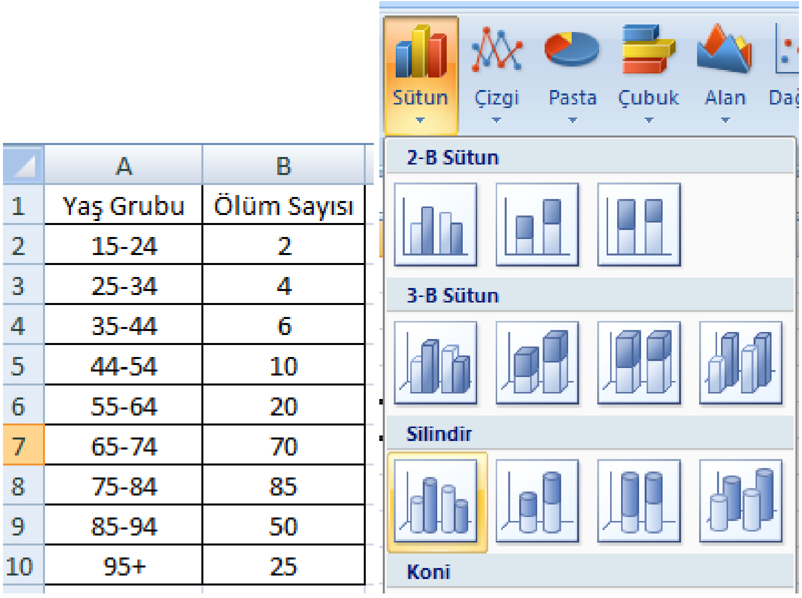


Şekil 2.8. Excel’de Çubuk Grafiği

2.3.3. Gruplandırılmış Serilerinin Grafikle Gösterilmesi

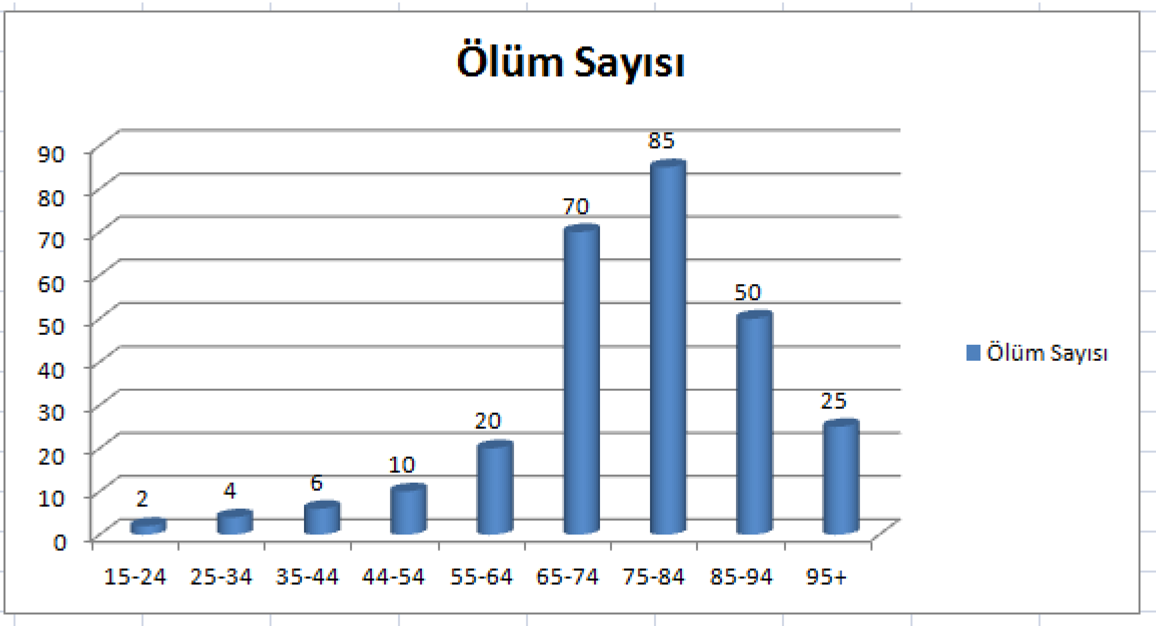
<u>Koni Grafiği:</u>

Örnek 2.18. Bir toplumda 15 yaş üzerindeki gruplarda görülen ölüm sayıları aşağıdaki gibidir. BU verilerin Excel'de koni grafiğini çiziniz.



	A	B
1	Yaş Grubu	Ölüm Sayısı
2	15-24	2
3	25-34	4
4	35-44	6
5	44-54	10
6	55-64	20
7	65-74	70
8	75-84	85
9	85-94	50
10	95+	25

Tabloda genç yaş gruplarında düşük olan ölüm sayıları yaş ilerledikçe artmakta, belirli bir yaştan sonra ise düşüşe geçmektedir. Veriye göre ölümün genç yaşta az görüldüğü, ileri yaşta fazla görüldüğü ve çok ileri yaşta ise azaldığı söylenebilir.



Şekil 2.9.Yaş gruplarına göre ölüm sayılarının koni grafiği

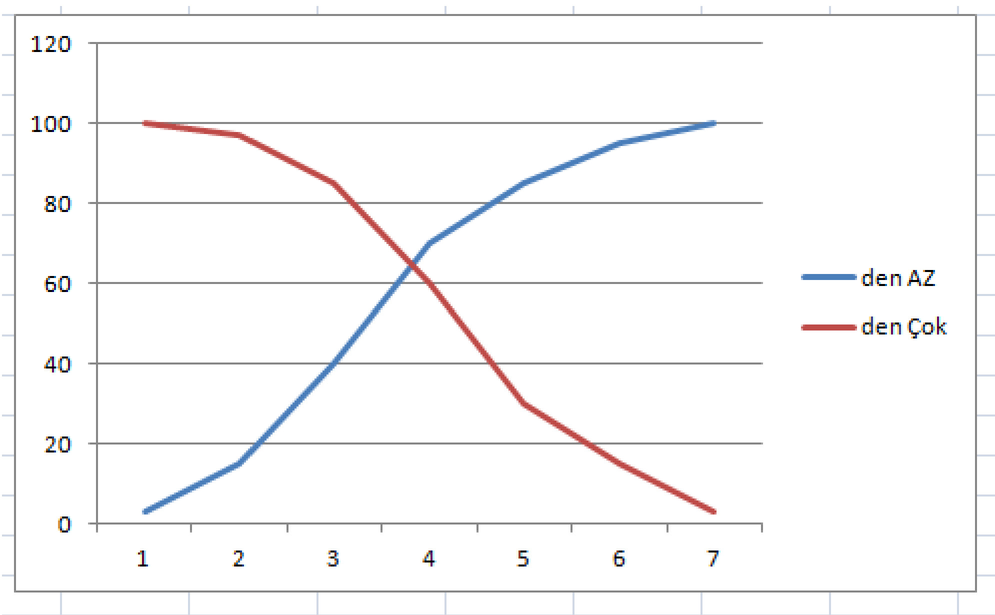
2.3.4. Birikimli Serilerinin Grafikle Gösterilmesi

Birikimli serilerin grafikleri çizilirken sınıflar yatay, birikimli frekanslar ise dik eksende gösterilir.

Örnek 2.19. Aşağıda verilen seri için –den az ve –den çok serilerini oluşturunuz ve grafiklerini çiziniz?

Sınıflar	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	3	12	25	30	15	10	5

Sınıflar	f	-den az	-den çok
0-10	3	3	100
10-20	12	15	97
20-30	25	40	85
30-40	30	70	60
40-50	15	85	30
50-60	10	95	15
60-70	5	100	5



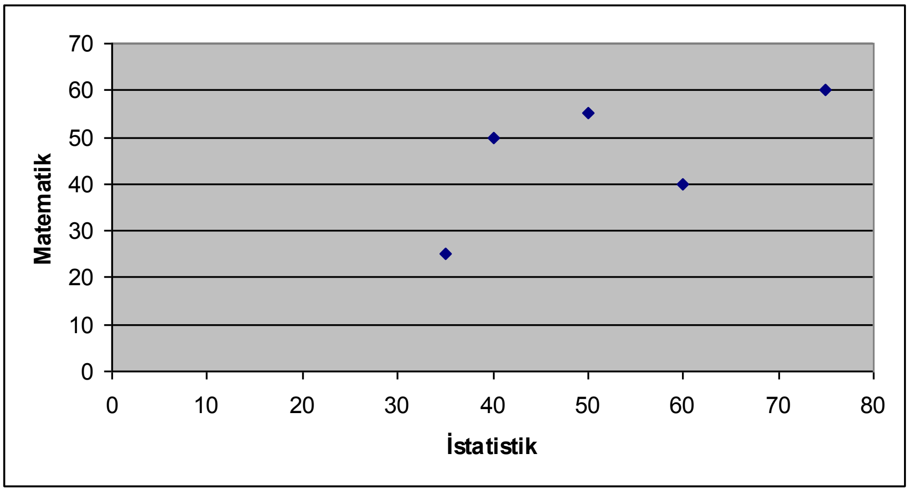
Şekil 2.10.Birikimli Serilerin Excel'de Çizgi Grafiği

2.3.5. Bileşik Serilerinin Grafikle Gösterilmesi

Serpilme (Dağıtım) Diyagramı: Bileşik serilerin grafikleri oluşturulurken ilk değişkenin değerleri yatay, diğer değişkenin değerleri ise dikey eksende yer alır. Bu değerlere koordinat sisteminde karşı gelen noktalar belirlenerek grafik çizilir.

Örnek 2.20. 5 öğrencinin istatistik ve matematik derslerinden aldıkları notlar aşağıdaki gibidir. Bu verileri kullanarak serpilme diyagramını çiziniz?

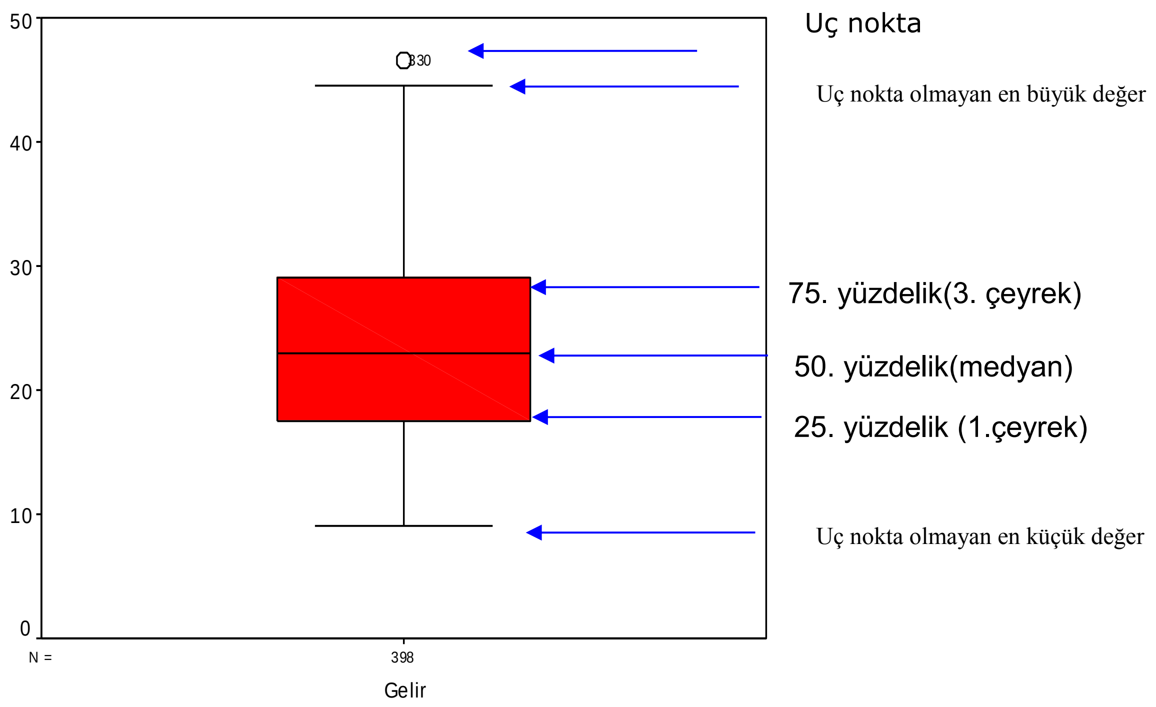
İstatistik	35	40	50	60	75
Matematik	25	50	55	40	60



Şekil 2.11.Excel'de Serpilme (Dağıtım)Grafiği

KUTU GRAFİĞİ (BOX PLOT)

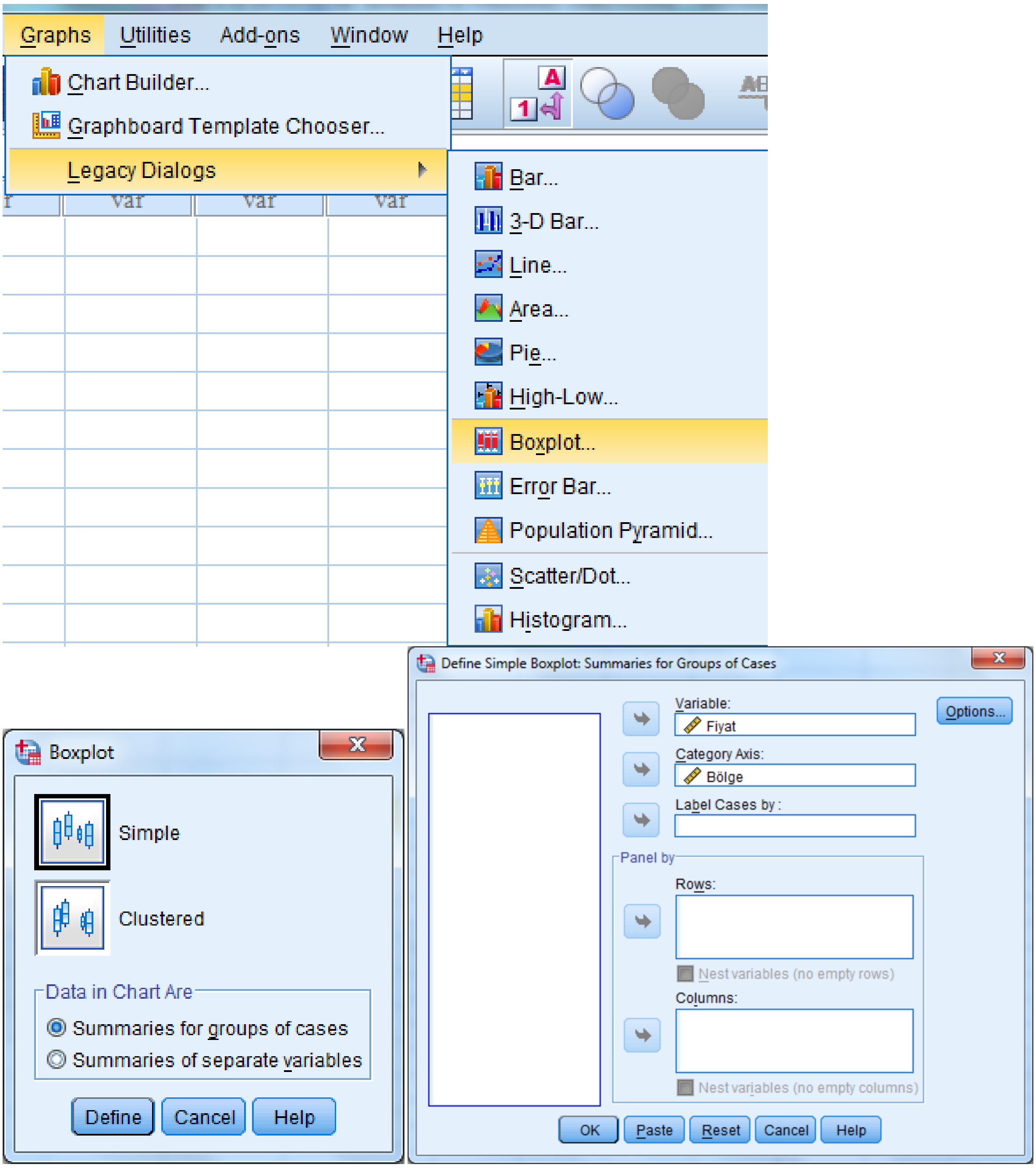
Yüzdeliklere dayanan ve tanımlayıcı istatistikleri kullanan bir grafik türüdür.

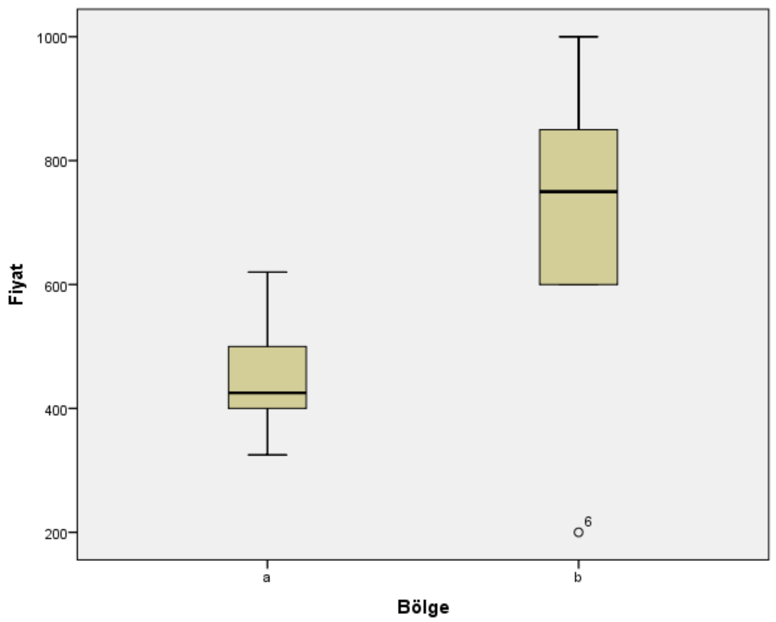


Şekil 2.12. Örnek Kutu (Box plot)Grafiği

Örnek 2.21. İki farklı bölgeye ait kiralık daire fiyatları aşağıdaki gibidir. SPSS'te kutu grafiğiniz.

	Fiyat	Bölge
1	400	a
2	425	a
3	500	a
4	620	a
5	325	a
6	200	b
7	600	b
8	750	b
9	850	b
10	1000	b


Graphs
Utilities
Add-ons
Window
Help
Chart Builder...
Graphboard Template Chooser...
Legacy Dialogs
Bar...
3-D Bar...
Line...
Area...
Pie...
High-Low...
Boxplot...
Error Bar...
Population Pyramid...
Scatter/Dot...
Histogram...
Boxplot
Simple
Clustered
Data in Chart Are
Summaries for groups of cases
Summaries of separate variables
Define
Cancel
Help
Define Simple Boxplot: Summaries for Groups of Cases
Variable:
Fiyat
Category Axis:
Bölge
Label Cases by:
Options...
Panel by
Rows:
Nest variables (no empty rows)
Columns:
Nest variables (no empty columns)
OK
Paste
Reset
Cancel
Help



Şekil 2.13. SPSS’te Kutu (Box plot)Grafiği

2.4. Örnek Problemler

1. İstatistik serilerinin grafikle gösteriminde, alanı ile ilgili sınıfın frekansın, tabanıyla ilgili sınıfın aralığına eşit, birbirine bitişik dikdörtgenlerden oluşan grafik türüne ne ad verilir?

a) Histogram b) Çubuk grafik c) Frekans eğrisi
d)Frekans poligonu e) Serpilme diyagramı

2. Bileşik serilerin grafiğine ne ad verilir?

a) Poligon b) Serpilme Diyagramı c)-den çok
d)-den az e)Histogram

3. Aşağıdaki frekans serisinin –den az ve –den çok serilerini oluşturunuz?

X	:1	2	3	4	5	6	7	8
f	:2	3	8	12	15	7	5	3

4. Aşağıdaki gruplanmış serisinde frekans (f) değerlerini bulunuz?

Sınıflar	:0-5	5-10	10-15	15-20	20-25	25-30	30-35
-den az	: 4	10	12	17	33	42	55

5. Aşağıda verilen seri için:
i) gruplanmış serisinde frekans (f) değerlerini bulunuz?
ii) yukarıdaki tabloya göre sayısal değeri 15 den küçük gözlem sayısı kaçtır? (C:25)
iii) yukarıdaki tabloya göre sayısal değeri 9 dan büyük gözlem sayısı kaçtır? (C:24)

Sınıflar	:0-3	3-6	6-9	9-12	12-15	15-18	18-21
-den çok	: 40	37	32	24	18	15	10

Cevaplar: 1-a, 2-b

III. BÖLÜM

3. ORTALAMALAR

Sık Kullanılan Bazı Matematiksel İfadeler :

{X_i: $X_1 + X_2 + \ldots + X_n$} ifadesi X değişken kümesini göstermektedir.

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + \ldots + X_n \quad ,$$

$$\sum_{i=1}^{n} c\,X_i = c\left(X_1 + X_2 + \ldots + X_n\right) = c\sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n}\left(X_i + Y_i\right) = (X_1 + Y_1) + (X_2 + Y_2) + \ldots + (X_n + Y_n) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i$$

$$\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{c} = \mathrm{c} + \mathrm{c} + \ldots + \mathrm{c} = \mathrm{nc}$$

$$\sum_{i=1}^{n} X_i Y_i = (X_1 Y_1) + (X_2 Y_2) + \ldots + (X_n Y_n)$$

3. 1. Giriş

İstatistikte birçok terimden oluşan bir sayıyı temsil ve ifadeye yeterli olan tek bir rakama ortalama denir. Ortalama aynı zamanda serinin özelliklerini de belirler. Gözlemlerin hangi nokta etrafında toplanmış olduğunu göstermesi gerektiğinden ortalama adı verilmektedir.

Bir mahalledeki ortalama gelir düzeyi bir araştırmaya göre yılda 15000€, diğerine göre ise 5000€ çıkmıştır. İki sonuçta aynı kişi ve yerden bulunmuştur. Burada ki hile “ortalama” kelimesidir. Çünkü hangi ortalama olduğu belirtilmemiştir. Eğer büyük bir sayıya ulaşmak istiyorsanız *kareli ortalama*, küçük bir sayıya ulaşmak için *harmonik ortalama* kullanılabilir. Bunun için hangi ortalamanın kullanıldığı ve araştırmanın kimleri kapsadığı sorulmalıdır.

Sayı yığınlarının kolayca anlaşılması için sayı yığınlarının en fazla yığıldığı bölgeyi tarif eden tipik değerlerin verilmesi gerekir. Bu değerler dağılışın merkezini gösterdikleri için merkezi eğilim ölçüleri olarak da bilinir. İstatistikte bir seriyi temsil etmeye yarayan tek bir rakama **ortalama** denir.

Ortalamalar, duyarlı (analitik) ortalamalar ve duyarlı olmayan (analitik olmayan) ortalamalar şeklinde iki gruba ayrılmaktadır.

3.2. Duyarlı Ortalamalar

Duyarlı ortalamalar, serinin bütün terimlerinin hesaba katıldığı ortalamadır. Duyarlı ortalamalar, aritmetik ortalama, geometrik ortalama, harmonik ve kareli ortalamaları içerir.

Tablo 3.1. Duyarlı ortalamalara ilişkin genel formüller

Duyarlı Ortalamalar	Basit Serilerde	$O_r = \sqrt[r]{\frac{\sum X_i^r}{n}}$
	Sınıflanmış Serilerde	$Or = \sqrt[r]{\frac{f_i X_i^r}{f_i}}$
	Gruplanmış Serilerde	$Or = \sqrt[r]{\frac{f_i m_i^r}{f_i}}$

Burada r'ye -∞ ile +∞ arasında değerler verilerek sonsuz ortalama bulunabilir. r=-∞ olması durumunda ortalamanın X_{min} olmasına, r=+∞ olması durumunda ise X_{max} olacağı ispatlanabilir. Yani r arttıkça ortalama büyümekte, r azaldıkça ise küçülmektedir.

3.2.1. Aritmetik Ortalama(Mean)

Aritmetik ortalama deneklerin aldıkları değerlerin toplanıp denek sayısına bölünmesiyle elde edilen değerdir. Tablo 3.1.'de r=1 alındığında aritmetik ortalama formülleri elde edilir.

Basit serilerin aritmetik ortalaması, terimlerin toplamının terim sayısına bölünmesine eşittir.

Biri gözlem değerleri, diğeri frekansları gösteren iki sütundan oluşan sınıflanmış serilerin aritmetik ortalaması hesaplanırken gözlem değerleri ile frekanslar çarpımlarının toplamı frekanslar toplamına bölünmektedir.

Gruplanmış serilerde aritmetik ortalama hesabının yapılabilmesi için öncelikle sınıf orta noktalarının (m) bulunması gerekmektedir. Sınıf orta noktalarının hesaba katılmasında gruba dahil birimlerin tamamının sınıf orta noktasında toplanmış olduğu varsayımından hareket edilmektedir.

Tablo 3.2. Aritmetik Ortalama Formülleri

Aritmetik Ortalama	**Basit Serilerde**	$\bar{X} = \frac{\sum X_i}{n}$
	Sınıflanmış Serilerde	$\bar{X} = \frac{\sum f_i X_i}{f_i}$
	Gruplanmış Serilerde	$\bar{X} = \frac{\sum f_i m_i}{f_i}$

Örnek 3.1. 20 pnömoni (zatürre) hastası için hastalık süreleri (gün) aşağıdaki şekilde bulunmuştur.

Xi: { 6,7,8,8,10,11,11,11,8,10,10,10,12,12,14,14,12,7,10,11}

N=20

$$\overline{X} = \frac{\sum_{i=1}^{N} X_i}{n} = \frac{202}{20} = 10{,}1$$

Örnek 3.2. Aşağıdaki sınıflanmış serinin aritmetik ortalamasını bulunuz?

Notlar (X_i)	:	40	60	70	80	100
Frekans(f_i)	:	5	4	5	4	2

$$\bar{X} = \frac{\sum f_i X_i}{f_i} = \frac{1310}{20} = 65{,}5$$

Örnek 3.3. Aşağıda gruplanmış olarak verilen serinin aritmetik ortalamasını bulunuz?

Sınıf sınırları	Sıklık (frekans= fi)	Sınıf Değeri (mi)	fi*mi
1.45 - 1.95	2	1.7	3,4
1.95 - 2,45	18	2,2	39,6
2,45 – 2,95	24	2,7	64,8
2,95 – 3,45	19	3,2	60,8
3,45 – 3,95	18	3,7	66,6
3,95 – 4,45	9	4,2	37,8
4,45 – 4,95	6	4,7	28,2
4,95 – 5,45	4	5,2	20,8
	100	TOPLAM	322

$$\bar{X} = \frac{\sum f_i m_i}{f_i} = \frac{322}{100} = 3{,}22$$

Soru. Aşağıda hastaların hastanede kalış süreleri verilmiştir. Buna göre ortalama hastanede kalış süresini bulunuz?

Kalış Süresi (Gün)	Frekans (f_i)	Sınıf Orta Noktası (m_i)	f_i m_i
1-5	4	3	12
6-10	10	8	80
11-15	17	13	221
16-20	8	18	144
21-25	10	23	230
26-30	4	28	112
31-35	3	33	99
Toplam	56		898

$$\bar{X} = \frac{\sum f_i m_i}{f_i} = \frac{898}{56} = 16\ gün$$

Aritmetik Ortalamanın Özellikleri

i) Aritmetik ortalamanın terim sayısı ile çarpımı seri toplamına eşittir.

Terimlerin aritm $\overline{X} = \dfrac{\sum_{i=1}^{N} X_i}{n} \Rightarrow \overline{X}n = \sum_{i=1}^{N} X_i$ nalarının toplamı sıfırdır.

$$\sum(\mathrm{X_i} - \overline{\mathrm{X}}) = \sum \mathrm{X_i} - \mathrm{n}\overline{\mathrm{X}} = \sum \mathrm{X_i} - \mathrm{n}\frac{\sum \mathrm{X_i}}{\mathrm{n}} = 0$$

ii) Terimlerin aritmetik ortalamadan sapmalarının kareleri toplamı minimumdur.

$$\sum (X_i - a)^2 = \sum [(X_i - \overline{X}) + (\overline{X} - a)]^2$$

$$= \sum (X_i - \overline{X})^2 + 2(\overline{X} - a)\underbrace{\sum (X_i - \overline{X})}_{0} + N(\overline{X} - a)^2 = \sum (X_i - \overline{X})^2 + n(\overline{X} - a)^2$$

Son eşitlikte $n(\overline{X} - a)^2 > 0$ olacağından aşağıdaki sonuç elde edilir.

$$\sum (X_i - a)^2 > \sum (X_i - \bar{X})^2$$

iii) Bir serinin bütün terimlerine aynı sayıyı eklersek (çıkarırsak) aritmetik ortalama eklenen (çıkarılan) sayı kadar arta (azalır).

$$\frac{\sum (X_i + k)}{n} = \frac{\sum X_i + nk}{n} = \overline{X} + k \quad , \quad \frac{\sum (X_i - k)}{n} = \frac{\sum X_i - nk}{n} = \overline{X} - k$$

iv) Bir serinin bütün terimlerinin aynı sayıyla çarptığımızda (böldüğümüzde) aritmetik ortalama çarptığımız (böldüğümüz) sayıyla orantılı olarak büyür (küçülür).

$$\frac{\sum kX_i}{n} = \frac{k\sum X_i}{n} = k\overline{X} \quad , \quad \frac{\sum X_i / k}{n} = \frac{1/k\sum X_i}{n} = \frac{\overline{X}}{k}$$

v) Aritmetik ortalama çok duyarlı bir ortalamadır. Çünkü serinin bütün terimleri aritmetik ortalamayı etkiler. Özellikle de **aşırı uç değerlerden** çok etkilenir ve dolayısıyla temsili olma özelliğini kaybeder.

vi) İki serinin bütün terimleri karşılıklı olarak toplanarak (çıkartılarak) elde edilen serinin aritmetik ortalaması bu serilerin aritmetik ortalamalarının toplamına (farkına) eşittir.

$$\frac{\sum (X_i + Y_i)}{n} = \frac{\sum X_i + \sum Y_i}{n} = \frac{\sum X_i}{n} + \frac{\sum Y_i}{n} = \overline{X} + \overline{Y}$$

$$\frac{\sum (X_i - Y_i)}{n} = \frac{\sum X_i - \sum Y_i}{n} = \frac{\sum X_i}{n} - \frac{\sum Y_i}{n} = \overline{X} - \overline{Y}$$

Aritmetik Ortalamanın Fayda ve Sakıncaları

Aritmetik ortalama kavram olarak basittir, hesaplanılması kolay olduğu gibi cebirsel işlemlere de elverişlidir. Bu bakımdan en çok kullanılan ortalamadır.

Aritmetik ortalama dağılımdaki terimlerden herhangi birinde meydana gelen kıymet değişikliğinden etkilenir; bu özellik aritmetik ortalama için bir üstünlük olduğu kadar, sakıncalıdır aynı zamanda. Dağılımda terim sayısının az olması durumunda olağanüstü küçük veya büyük terimler aritmetik ortalamanın değerini etkiler ve simgeleyici olmasını engeller.

Diğer taraftan dağılımın alt ve/veya üst sınırının belirsiz olması durumunda aritmetik ortalamayı hesaplamak olanaksızdır; belirsiz olan sınırlar için yapılacak kestirimler, ortalamanın kesin değerinin hesaplanılmasına olanak vermeyecektir. Bu bakımdan sözü edilen iki durumda dağılım terimlerini normal büyüklüğünün belirlenmesinde aritmetik ortalama kullanılmamalıdır.

3.2.2. Geometrik Ortalama

X_i: $\{X_1, X_2, \ldots, X_n\}$ pozitif sayılar kümesinin geometrik ortalaması, sayıların çarpımlarının n nci dereceden köküdür.

$$GO = \sqrt[n]{X_1 X_2 \ldots X_n}$$

Tablo 3.1.'de r=0 yada r→0 için limit alınarak geometrik ortalama formülleri bulunabilir.

Tablo 3.3. Geometrik Ortalama Formülleri

Geometrik Ortalama	Basit Serilerde	$logGO = \frac{\sum X_i}{n}$
	Sınıflanmış Serilerde	$logGO = \frac{\sum f_i log X_i}{\sum f_i}$
	Gruplanmış Serilerde	$logGO = \frac{\sum f_i log m_i}{\sum f_i}$

Geometrik ortalama, terimlerinin logaritmalarının aritmetik ortalamasının antilogaritmasına eşittir. Geometrik ortalama, terimlerin logaritmalarının aritmetik ortalamasının antilogaritmasına eşittir. Geometrik ortalama özellikle aynı oranda artma veya azalma eğilimi gösteren olaylarla ilgili serilere uygulanır (Örneğin, nüfus).

Örnek 3.4. X_i: { 8, 12, 25, 6 }

$GO = \sqrt[4]{8.12.15.6}$=10,95

log X_i : 0,903089 1,079181 1,397940 0,778151

$logGO = \frac{\sum X_i}{n} = 1{,}03959$ GO=10,95

Örnek 3.5. Aşağıdaki sınıflanmış serinin geometrik ortalamasını bulunuz?

X_i	f_i	log X_i	f_i log X_i
2	3	0,301030	0,903090
3	2	0,477121	0,954242
4	1	0,602060	0,602060
5	4	0,698970	2,795880
Toplam	10		5,255272

$logGO = \frac{\sum f_i logX_i}{\sum f_i} = 0{,}525527$ GO=3,35

Örnek 3.6. Aşağıda verilen gruplanmış serinin geometrik ortalamasını bulunuz?

Sınıflar	m_i	f_i	log m_i	f_i log m_i
1-3 den az	2	3	0.301030	0.903090
3-5 den az	4	3	0.602060	1.806180
5-7 den az	6	4	0.778151	03.112604
		Σ=10		□=5.821874

$logGO = \frac{\sum f_i logm_i}{\sum f_i}$=0,582187 GO=3,82

Geometrik Ortalamanın Özellikleri

I. Geometrik ortalama özellikle aynı oranda artma yada azalma eğilimi gösteren olaylara ilişkin serilere uygulanır. Örneğin nüfus çoğalması, bakteri üremesi gibi geometrik dizilerde birim zamandaki artışı bulmak için GO kullanılır.

II. Simetrik olmayan ancak logaritmaları alındığında simetrik hale dönüşen serilere geometrik uygulama uygulanabilir.

III. Serideki terimler arasında bazı değerler sıfır veya negatifse GO hesaplanamaz.

IV. Geometrik uygulama aşırı uç değerlerden aritmetik ortalamaya göre daha az etkilenir.

V. GO ≤ AO ilişkisi vardır. Bütün X_i ler eşitse GO=AO olur.

Geometrik ortalama özellikle aynı oranda artma veya azalma eğilimi gösteren olaylara ilişkin serilere uygulanır. Bu olaylar arasında öncelikle nüfus belirtilebilir. Öte yandan, aslında simetrik olmadığı halde logaritmaları alındığında simetrik hale dönüşen serilere de geometrik ortalamayı uygulamak gerekir.

Not:

Başlangıçta A kadar birey varsa, bu bireyler birim zamanda r kadar bir hızla artıyorsa, n birim zaman sonra sayıları B kadar olmuş ise **$B= A(1+r)^n$** olur. Bu formül ***bileşik faiz formülü*** olarak adlandırılır.

Ortalama artış (r) buradan hesaplanır.

$$r = \sqrt[n]{\frac{B}{A}} - 1$$

Örnek 3.7. Bir bakteri kültürü 3 günde 1000 den 4000 e çıkmış ise ortalama günlük artış hızı(r) nedir?

$$r = \sqrt[3]{\frac{4000}{1000}} - 1 = \sqrt[3]{4} - 1 = 0.587$$

Yani ortalama artış hızı= %58,7 dir.

Örnek 3.8. Bir bölgenin nüfusu 2000 yılında 500000 ölçülmüştür. Bu bölgenin yıllık nüfus artışı binde 15 ise ise 2005 yılında bu bölgenin nüfusu kaç olur.

$$B = A(1+r)^n = 500000(1+0,015)^5 = 538642$$

3.2.3. Harmonik Ortalama

X_i:{ $X_1, X_2, \ldots, X_n$ } değerlerinin harmonik ortalaması:

$$\text{Harmonik Ortalama} = \frac{1}{\frac{1}{\text{n}}\sum_{i=1}^{n}\frac{1}{X_i}}$$

Harmonik ortalama terimlerin terslerinin aritmetik ortalamasının tersidir. Tablo 3.1.'de r=-1 alınırsa harmonik ortalama formülleri elde edilir.

Tablo 3.4. Harmonik Ortalama Formülleri

Harmonik Ortalama	Basit Serilerde	$HO = \frac{n}{\sum \frac{1}{X_i}}$
	Sınıflanmış Serilerde	$HO = \frac{\sum f_i}{\sum \frac{f_i}{X_i}}$
	Gruplanmış Serilerde	$HO = \frac{\sum f_i}{\sum \frac{f_i}{m_i}}$

Örnek 3.9. Bir fabrikada üretim 4 makineyle yapılmaktadır. Bu makinelerin bir mamul için harcadıkları süreden (dk.) yararlanarak bir mamulün bu fabrikada ortalama kaç dk. üretildiğini bulunuz?

Makineler	X=Üretim Süresi (dk./parça)	1/X
I	2,5	0,4
II	2	0,5
III	1,6	0,625
IV	4	0,25
		1,775

$$\text{HO} = \frac{n}{\sum \frac{1}{X_i}} = \frac{4}{1.775} = 2,254\, dk / parca$$

Tartılı aritmetik ortalama ile çözüm:

X=Üretim Süresi (Dk./Parça)	Saatte üretilen parça (t_i)	$t_i.X_i$
2,5	60/2,5=24	60
2	60/2=30	60
1,6	60/1,6=37,5	60
4	60/4=15=15	60
	106,5	240

$$\text{TAO} = \frac{\sum t_i X_i}{\sum t_i} = \frac{240}{106.5} = 2,254\, dk.$$

Soru. X_i: { 6,8,3,5,4 } veri setinin harmonik ortalaması:

$$HO = \frac{1}{\frac{1}{5}\left(\frac{1}{6}+\frac{1}{8}+\frac{1}{3}+\frac{1}{5}+\frac{1}{4}\right)} = 4{,}65$$

Örnek 3.10. 6 öğrenci 100 TL ile farklı eczanelerden aspirin alıyorlar. Birinci öğrenci 9 adet, ikincisi 6 adet, üçüncüsü 7 adet, 4 cüsü 8 adet, 5 cisi 6 adet ve 6 cısı 8 adet aspirin alıyor. 100 Tl ile alınabilecek ortalama aspirin sayısı ne kadardır?

Fiyat=para/mal olduğundan ve para sabit ise harmonik ortalama alınır.

$$HO = \frac{6}{\frac{1}{9}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}+\frac{1}{6}+\frac{1}{8}}$$

= 7,166 yani 100 TL ile ortalama 7

Örnek 3.11. Bir otobüs firması iki şehir arasında (600 km.) 37 otobüsle seferler düzenlemektedir. Bu otobüslerin hızlarına göre dağılımı aşağıdaki gibidir. Otobüslerin ortalama hızını bulunuz?

Hız (km/saat)	Otobüs sayısı (f_i)	f_i/X_i
60	3	3/60=0,05
75	6	6/75=0,08
80	10	10/80=0,125
90	18	18/90=0,200
	37	0,455

$$\mathrm{HO} = \frac{\sum f_i}{\sum \frac{f_i}{X_i}} = \frac{37}{0.455} = 81{,}32\ km/saat$$

Soru. Aşağıda verilen sınıflanmış serinin harmonik ortalamasını bulunuz?

Xi	: 2	3	4	5
f_i	: 3	2	1	4
f_i/X_i	: 1,5	0,67	0,25	0,80

$$HO = \frac{\sum f_i}{\sum \frac{f_i}{X_i}} = \frac{10}{3{,}22} = 3{,}11$$

Örnek 3.12. Aşağıda verilen gruplanmış serinin harmonik ortalamasını bulunuz?

Sınıf	: 1-3 den az	3-5 den az	5-7 den az
m_i	: 2	4	6
f_i	: 3	3	4
f_i/m_i	: 1,5	0,75	0,67

$$HO = \frac{\sum f_i}{\sum \frac{f_i}{m_i}} = \frac{10}{2,92} = 3,42$$

Harmonik Ortalamanın Özellikleri

I. Serideki terimlerden biri sıfır ise harmonik ortalama sıfır çıkar.

II. Seri terimleri farklı işaretli olursa harmonik ortalamanın sonucu anlam taşımaz. Mesela verilerimiz -4, -2, 1,2,5 olsun. Buna göre HO=5,05 çıkar. Bu sonuç, bir ortalama maksimum değerden daha büyük bir değere sahip olamayacağı için, ortalama olarak kabul edilmez.

$$HO = \frac{1}{\frac{1}{5}\left(\frac{1}{-4} + \frac{1}{-2} + \frac{1}{1} + \frac{1}{2} + \frac{1}{5}\right)} = 5,05$$

III. HO **≤** GO ≤ AO ilişkisi vardır.

IV. HO sınırlı hallerde kullanılır. Tersine çevrildiğinde taşıyacağı anlama önem verilen oran türündeki niceliklerin ortalamasını bulmak için kullanılır. Bu niceliklere örnek olarak fiyat=para/mal, prodüktivite=iş/emek, verim=ürün/ekim alanı, hız=uzaklık/zaman verilebilir.

3.2.4. Kareli Ortalama

Kareli ortalama fiziksel uygulamalarda çok sık kullanılır. Tablo 3.1.'de r=2 alınırsa terimlerin karelerinin aritmetik ortalamasının kareköküne eşit olan kareli ortalama formülleri bulunur. Kareli ortalama, negatif değerleri de dikkate almaktadır. Kareli ortalama bazı istatistiksel işlemlerin kolaylıkla yapılmasına olanak tanır. Örneğin standart sapmanın hesabında kareli ortalamadan yararlanırlar.

Tablo 3.5. Kareli Ortalama Formülleri

Kareli Ortalama	**Basit Serilerde**	$KO = \sqrt{\frac{\sum X_i^2}{n}}$
	Sınıflanmış Serilerde	$KO = \sqrt{\frac{\sum f_i X_i^2}{\sum f_i}}$
	Gruplanmış Serilerde	$KO = \sqrt{\frac{\sum f_i m_i^2}{\sum f_i}}$

Örnek 3.13. Aşağıdaki serinin kareli ortalamasını bulunuz?

X_i :	4	5	7	8	16

$$KO = \sqrt{\frac{\sum X_i^2}{n}} = \sqrt{\frac{40}{5}} = 9,06$$

Örnek 3.14. Aşağıdaki sınıflanmış serinin kareli ortalamasını bulunuz?

X_i :	2	3	4	5
f_i :	3	2	1	4

$$KO = \sqrt{\frac{\sum f_i X_i^2}{\sum f_i}} = \sqrt{\frac{146}{10}} = 3,82$$

Örnek 3.15. Aşağıdaki sınıflanmış serinin kareli ortalamasını bulunuz?

Sınıf	:	1-3 den az	3-5 den az	5-7 den az
m_i	:	2	4	6
f_i	:	3	3	4

$$KO = \sqrt{\frac{\sum f_i X_i^2}{\sum f_i}} = \sqrt{\frac{204}{10}} = 4,52$$

Kareli Ortalamanın Özellikleri

I. Kareli ortalama negatif işaretleri de dikkate alabileceğinden HO ve GO'ya göre daha üstündür.
II. KO bazı istatistiksel işlemlerin kolaylıkla uygulanmasını mümkün kılar. Örneğin bir değişkenlik ölçüsü olan standart sapmanın hesabında kareli ortalamadan yararlanılır.
III. HO ≤ GO ≤ AO ≤ KO ilişkisi vardır.

3.2.5. Tartılı (Ağırlıklı) Ortalama

Seri terimleri veya sınıfları arasında önem farkını dikkate almak için her terime veya sınıfa önemi ile orantılı bir tartı verilerek tartılı ortalama hesaplanır.

Tablo 3.1.'de r=-1 de tartılı harmonik ortalama, r=0 için tartılı geometrik ortalama, r=1 için tartılı aritmetik ortalama ve r=2 de ise tartılı kareli ortalama formülleri bulunur.

Tablo 3.6. AO, GO, HO ve KO'nın tartılı ortalama formülleri

Basit Serilerde Tartılı Ortalamalar	$\bar{X}_t = \frac{\sum t_i X_i}{\sum t_i}$	$\log GO_t = \frac{\sum t_i \log X_i}{\sum t_i}$
	$HO_t = \frac{\sum t_i}{\sum (t_i / X_i)}$	$KO_t = \sqrt{\frac{\sum t_i X_i^2}{\sum t_i}}$
Frekans Serilerde Tartılı Ortalamalar	$\bar{X}_t = \frac{\sum t_i f_i X_i}{\sum t_i f_i}$	$\log GO_t = \frac{\sum t_i f_i \log X_i}{\sum t_i f_i}$
	$HO_t = \frac{\sum t_i f_i}{\sum (t_i f_i / X_i)}$	$KO_t = \sqrt{\frac{\sum t_i f_i X_i^2}{\sum t_i f_i}}$
Gruplanmış Serilerde Tartılı Ortalamalar	$\bar{X}_t = \frac{\sum t_i f_i m_i}{\sum t_i f_i}$	$\log GO_t = \frac{\sum t_i f_i \log m_i}{\sum t_i f_i}$
	$HO_t = \frac{\sum t_i f_i}{\sum (t_i f_i / m_i)}$	$KO_t = \sqrt{\frac{\sum t_i f_i m_i^2}{\sum t_i f_i}}$

Örnek 3.16. Bir ildeki 5 hastanede acile gelen hastaların ortalama yaşları aşağıdaki şekilde bulunmuştur.

Hastane	Hasta (ti)	Ortalama Yaş(Xi)
1	10	25
2	15	30
3	20	40
4	5	20
5	30	15

Bu ilde acile gelen hastaların ortalama yaşı nedir? Bunun için tartılı ortalama bulunur.

T_i= i nci grubun tartısı X_i= i nci grubun değeri

$$TO = \frac{\sum_{i=1}^{k} t_i X_i}{\sum_{i=1}^{k} t_i} = \frac{2050}{80} = 25,625$$

Örnek 3.17. Aşağıdaki serinin tartılı ortalamalarını bulunuz?

X_i :	2	4	5	6
t_i :	3	1	4	2

t_i	$t_i X_i$	log X_i	t_i log X_i	t_i/X_i	X_i^2	$t_iX_i^2$
3	6	0,301030	0,903090	1.50	4	12
1	4	0,602060	0,602060	0.25	16	16
4	20	0,698970	2,795880	0.80	25	100
2	12	0,778151	1,556302	0.33	36	72
∑ 10	42		5,857332	2.88		200

$$\bar{X}_t = \frac{\sum t_i X_i}{\sum t_i} = \frac{42}{10} = 4,2 \qquad HO_t = \frac{\sum t_i}{\sum t_i / X_i} = \frac{10}{2,88} = 3,47$$

$$logGO_t = \frac{\sum t_i logX_i}{\sum t_i} = \frac{5,857332}{10} = 0,585733 \quad GO_t = 3,85$$

$$KO_t = \sqrt{\frac{\sum t_i X_i^2}{\sum t_i}} = \sqrt{\frac{200}{10}} = 4,47$$

$HO_t \leq GO_t \leq AO_t \leq KO_t$

Örnek 3.18. 5 farklı klinikte kullanılan ortalama serum miktarları aşağıdaki gibidir.

Klinikler	t_i=Hasta sayısı	X_i=Serum (lt)	t_i X_i
1	3	1.5	4.5
2	8	2	16
3	5	3	15
4	4	2.5	10
5	6	4	24
Toplam	26		74

Kliniklerin genel ortalaması nedir?

$$AO_t = \frac{\sum t_i X_i}{\sum t_i} = \frac{74}{26} = 2,8\ lt.$$

Soru : Bir dersin final sınavı ara sınavına göre 3 kez fazla ağırlıklandırılmış ise, final sınavından 85 ve ara sınavlardan 70 ve 90 almış bir öğrencinin ortalama notunu bulunuz?

$$\bar{X}_t = \frac{(1\times7)+(1\times90)+(3\times85)}{1+1+3} = \frac{415}{5} = 83$$

3.3. Duyarlı Olmayan Ortalamalar

Duyarlı ortalamalarda bütün terimler veya sınıflar dikkate alınır. Hesaplamalarda bazen serinin bütün terimleri veya sınıfları dikkate alınmayabilir. Bu durumda duyarlı olmayan ortalamalar ortaya çıkar.

3.3.1. Medyan (Ortanca=Median)

Terimlerin küçükten büyüğe doğru (yada büyükten küçüğe doğru) sıralanmış bir seride tam ortaya düşen ve seriyi iki eşit kısma bölen değere medyan (ortanca) denir. Medyanın hesabı basit, sınıflandırılmış ve gruplanmış serilerde farklıdır.

Basit serilerde, terimlerin sayısı tek ise tam ortadaki değer, çift ise ortadaki iki terimin aritmetik ortalaması medyanı verir. (N+1)/2terime karşılık gelen terim medyandır.

Örnek 3.19. X_i : { 3,1,13,27,6,8,6 } gözlem değerlerinin ortancası nedir?

Sayılar büyüklük sırasına dizilirse, { <u>1, 3, 6,</u> 6, <u>8, 13, 27</u> } olur.

Ortada kalan sayı 6 olduğundan

Ortanca=6 olur.

Örnek 3.20. X_i : { 21, 9, 8, 3, 7, 9} olsun. Gözlemler büyüklük sırasına dizilirse,

X_i : { <u>3, 7</u>, 8, 9, <u>9, 21</u> }

olur. Ortada kalan iki değerin ortalaması ortancadır:

Ortanca= (8+9)/2= 8,5

Diğer bir ifade ile (N+1)/2 nci değer ortanca değerdir. Gözlem değeri çift ise sonuç şöyle bulunur. (N+1)/2 = (6+1)/2 = 3,5 yani 3 ncü ve 4 ncü gözlemlerin ortalaması ortancadır.

Sınıflandırılmış verilerden ortanca hesaplamak için önce medyan sınıfının bulunması gerekir. Bunun için ....den az eklemeli frekans (... den az Fi) bulunur ve bu kullanılarak (N+1)/2 nci gözlemin düştüğü sınıf ortanca sınıfı olarak tanımlanır.

Örnek 3.21. Aşağıda sınıflanmış olan serilerin meydanlarını bulunuz?

A Serisi			B Serisi		
X_i	f_i	$\sum f_i$	X_i	f_i	$\sum f_i$
11	2	2	13	3	3
22	3	5	24	6	9
34	4	9	37	4	13
45	2	11	48	5	18

A serisinde frekans toplamı 11 olup (N+1)/2=(11+1)/2=6. terim medyandır. Kümülatif frekans toplamlarında 6. terim (9. terim dahil) 34 değerine sahiptir. O halde A serisi için medyan 34 olur.

B serisinde frekans toplamı 18 olup (N+1)/2=(18+1)/2=9,5 terim medyandır. Ancak seride 9.5 terimi olmadığından 9 ve 10. terimlerin ortalaması medyanı verecektir. Seride 9. terim 24 ve 10. terim 37 değerine sahiptir. Dolayısıyla medyan=(24+37)/2=30,5 olur.

Gruplanmış Verilerden Ortanca Hesabı

$$\text{Medyan} = L + \frac{\frac{n}{2} - F_{(i-1)}}{f_i} \times C$$

L: Medyan sınıfının alt sınır değeri

n= Toplam gözlem sayısı

Fi-1 = Medyan sınıfından önceki sınıfların frekans toplamı

fi= Medyan sınıfının kendi sınıf frekansı

C= sınıf genişliği

Kümülatif frekanslarda N/2'nci terimi içeren sınıf medyan sınıfı kabul edilir. Medyan değeri, medyan sınıfının alt sınırından küçük ve üst sınırından büyük olamaz.

Örnek 3.22. Aşağıda verilen gruplanmış serinin meydanını bulunuz?

Sınıf sınırları	Sıklık (frekans= fi)	....den Az (Fi)
1.45 - 1.95	2	2
1.95 - 2,45	18	20
2,45 – 2,95	24	44
2,95 – 3,45	19	63
3,45 – 3,95	18	81
3,95 – 4,45	9	90
4,45 – 4,95	6	96
4,95 – 5,45	4	100
	100	

Sınıf Sınırı	2,95	....	....	3,45
Fi	45...........	50	51...........	63

Örnekte N/2=100/2=50 ‘nci terim medyandır. Dolayısıyla medyan sınıfı 2,95-3,45 sınıfıdır. Bu sınıfın alt sınırı L=2,95 medyan sınıfından önceki sınıfların frekans toplamı F_{i-1}=44, toplam frekans N=100, medyan sınıfının kendi frekansı f_i=19 ve sınıf genişliği C=0.5 olur.

$$Medyan = L + \frac{\frac{n}{2} - F_{(i-1)}}{f_i} \times C = 2,95 + \frac{\frac{100}{2} - 44}{19} \times 0,5 = 3,108$$

Örnek 3.23. Aşağıda gruplanmış serinin meydanını hesaplayınız?

Sınıflar	:	4-7 den az	7-10 dan az	10-13 den az
f_i	:	8	5	2
Toplam f_i	:	8	13	15

Toplam frekans N/2=15/2=7.5 terim medyandır. Medyan sınıfı 4-7 den az sınıfıdır. Bu sınıfın alt sınırı L=4, genişliği C=3, medya sınıfının frekansı N=8 dir. Medyan sınıfından önceki sınıf seride bulunmadığından bunun kümülatif frekansı F_{i-1}=0 kabul edilir. Buna göre medyan aşağıdaki gibi bulunur.

$$Medyan = 4 + \frac{7,5-0}{8} \times 3 = 6,81$$

Medyanın Özellikleri

I. Terimlerin medyandan mutlak sapmalarının toplamı minimumdur.

$$\sum |X_i - Medyan| = min$$

II. Basit bir sıralama ile bulunması mümkün olduğundan, medyan bir çok durumda pratiktir. Örneğin bir grup öğrencinin boy uzunluğunu teker teker ölçmeye gerek yoktur. Öğrenciler küçükten büyüğe doğru sıralanıp ortadaki öğrenci (ler) ölçülerek ortanca boy uzunluğu bulunabilir.

III. Seride açık (alt sınırı veya üst sınırı belli olmayan) sınıfların varlığı halinde medyan hesabı önem kazanır. Medyan sınıfı serinin ilk sınıfı olduğunda, sınıfın alt sınırı tahminsel olarak ele alınır.

IV. Diğer ortalamaların aksine, gruplanmış serinin medyan hesabında sınıf genişliklerinin tamamının eşit olması gerekmez.

V. Medyan serdeki anormal terimlerden etkilenmez.

Medyanı Kullanmanın Sakıncaları

I. Ortancanın standart hatası aritmetik ortalamadan daha büyüktür.

II. Ortanca üzerinde cebirsel işlemler yapılamaz.

III. Farklı alt grupların ortancaları biliniyorsa bu gruplar birleştiğinde ortanca nedir sorusu hesaplama ile bulunamaz.

Örnek 3. 24. I.maddenin doğru olduğunu bir örnekle gösteriniz?

Xi : 3 5 6 8 13

Bu serinin AO=13 ve Medyan=6 dır.

Aritmetik ortalama ve medyandan mutlak sapmalar ise aşağıdaki gibi bulunur.

$\lvert X_i - \overline{X} \rvert$	:4	2	1	1	6	Toplam=14
$\lvert X_i - Med \rvert$	:3	1	0	2	7	Toplam=13

Görüldüğü gibi medyandan mutlak sapmaların toplamı, aritmetik ortalamadan mutlak saplamaların toplamından küçüktür. Bu diğer ortalamalar için de geçerlidir.

3.3.2. Mod (Tepe Değeri)

Bir seride en çok tekrarlanan terime mod denir. En yüksek frekansa karşılık gelen X değeri Modu verir. Basit serilerde mod hesabı yapılmaz. Çünkü basit serilerde X'e karşılık gelen tüm frekanslar 1 olduğu için frekans sütunu bulunmaz.

Sınıflanmış serilerde modun belirlenmesi için frekans sütununda en yüksek frekans değerini veren X değeri bulunur.

Bir sayı kümesi içinde en fazla tekrarlanan değer o kümenin tepe değerini oluşturur.

Örnek 3.25. X_i: { 1,2, 6, 3, 7, 3,5, 6,6, 8,9 } serisinin modu nedir?

Burada en fazla tekrarlanan değer 6 olduğu için Mod= 6 olur.

Örnek 3.26. Aşağıdaki sınıflandırılmış serinin modu nedir?

X_i :	2	3	6	7
f_i :	3	6	4	5

Seride en yüksek frekans 6 olduğuna göre, buna karşı gelen değer yani 3 moddur.

Gruplanmış seride mod hesabı için aşağıdaki formül kullanılır.

$$Mod = L + \frac{d_1}{d_1 + d_2} \times C$$

d_1= Mod sınıfı frekansı – bir önceki sınıf frekansı,

d_2= Mod sınıfı frekansı – bir sonraki sınıf frekansı,

C= Mod sınıfının genişliği

L= Mod sınıfının alt sınırı

Buradan bulunacak mod yaklaşık bir değere sahip olur. Gruplanmış serilerde mod sınıfının belirlemek için, frekans sütunundaki en yüksek frekansa bakılır. En yüksek frekansa sahip sınıf mod sınıfı kabul edilir. Mod değeri, mod sınıfının alt sınırından küçük ve üst sınırından büyük olamaz.

Örnek 3.27. Bir taramada 50 kadının kanındaki (gr/lt) serum albümin değerleri aşağıdaki gibi bulunmuştur.

41 41 42 44 44 36 38 41 42 44 42 39 49 40 45 **32** 34 43 37 39

41 39 48 42 43 33 43 35 32 34 39 35 43 44 47 40 39 42 41 46

37 49 41 39 43 42 47 48 51 **52**

Bu verilere ait sıklık tablosu 6 sınıf olacak şekilde yapılsın. EnK.=32, EnB.= 52, sınıf aralığı= 4 olsun.

N=50, C=4, d_1=(17-14)=**3**, d_2=(17-7)=**10,** L=**41.5**

Sınıf Sınırları	**Frekans (fi)**	
29.5-33.5	3	
33.5-37.5	7	
37.5-41.5	14	
41.5-45.5	**17**	**Mod Sınıfı**
45.5-49.5	7	
49.5-53.5	2	
Toplam	50	

Mod=41.5+(17 - 14)*4/[(17-14)+(17-7)] =42.42

Tepe değerinin (mod) kullanışlı olabilmesi için gözlem sayısının çok fazla olması gerekir. Bazı durumlarda dağılışın birden fazla modu olabilir, çok modlu dağılışlar olabilir. Modların aynı yükseklikte olması da gerekmez. Ancak bu modların sınıf aralarının küçük değişikliği ile kaybolmayacak ayrıklıkta olması gerekir. Bu durumlarda örneğin farklı gruplardan oluştuğu anlamı çıkar.

Bazen serinin iki maksimum değeri olabilir. Bunun nedeni incelenen kütlenin homojen olmamasından ileri gelir. Örneğin kadın ve erkeklerin boy uzunluklarını gösteren seride iki maksimum nokta vardır. Biri kadınların boy uzunluğu, diğeri de erkeklerin boy uzunluğudur. Burada yapılması gereken kütleyi homojen gruplara ayırmak ve her grup için ayrı mod hesaplamaktır.

Örnek 3.28. Aşağıdaki gruplanmış serinin modunu hesaplayınız?

Sınıflar:	1-3	3-5	5-7	7-9	9-11	11-14
f_i	:2	7	18	12	18	5

Serinin en yüksek frekansı olan 18 hem 5-7 den az hem de 9-11 den az sınıfına aittir. Bu yüzden çift tepeli serinin sınıfları birleştirilir.

Sınıflar	:1-5 den az	5-9 dan az	9-13 den az
f_i	:9	30	23

C=4 d_1=30-9=21 d_2=30-23=7 L=5

$$Mod = L + \frac{d_1}{d_1 + d_2} \times C = 5 + \frac{21}{21 + 7} \times 4 = 8$$

Modun Özellikleri

I. Ortalamalar arasında mod en temsili olanıdır. Çünkü kütledeki birimleri önemli bir kısmına uyar.
II. Sınıflanmış serilerde modun tamsayı olması gerçeğin daha iyi yansıtılmasını sağlar. Örneğin bir bölgedeki ailelerin ortalama çocuk sayıları hesaplandığında kesirli bir rakam elde edilebilir. Oysa ortalama olarak mod alınırsa bu değer tam sayı çıkacaktır.
III. Mod anormal terimlerin etkisi altında kalmaz. Örneğin çok zengin bir kişinin köye taşındığını varsayalım. Bu kişinin gelir düzeyi tek ve serinin sonunda olacağından modu etkilemez.
IV. Mod uygulamada farkına varılmadan en çok başvurulan ortalamalardan biridir. Örneğin kundura ve hazır giyim eşyası üretiminde en çok satılan numaralar ve bedenler dikkate alınır. Buda mod demektir.
V. Adlandırma (nominal) ölçekli değişkenlerde mod kullanımı uygundur.

Modun Sakıncaları

I. Modun güvenirliliği azdır. Yani örnekten elde edilen mod popülasyon modundan çok farklı olabilir.
II. Ortancada olduğu gibi mod üzerinde de cebirsel işlemler yapılamaz.
III. Bazen verilerin ortalaması, ortancası olduğu halde modu olmayabilir. Bütün değerler farklı ise mod yoktur.

Soru. Hastalık nedeniyle işe gelmeyen işçilerin gelmedikleri gün sayısını gösteren frekans tablosu aşağıdaki şekilde olsun.

Sınıf Sayısı	**Gün(X_i)**	**İşçi Sayısı (f_i)**	**f_i*X_i**	**f_i**
1	0	5	0	5
2	1	8	8	13
Mod Snf.	2	10	20	23
Ortanca Snf.	3	9	27	32
5	4	6	24	38
6	5	5	25	43
7	6	4	24	47
8	7	2	14	49
9	8	1	8	50
		50	150	

Aritmetik Ortalama=150/50= 3 Ortanca= 3

Ortanca sınıfının X_i değeri doğrudan ortanca olarak alınır.

Mod=2, En yüksek sıklığa sahip sınıf mod sınıfı olduğundan bu sınıfa ait değer doğrudan mod değeri olarak alınır.

3.3.3. Kartiller (Quartiles)

Küçükten büyüğe doğru sıralanmış bir seriyi 4, 10, 100 eşit kısma bölen terimler vardır. Genel olarak **kantil** adı verilen bu değerlerden dörde bölenler **kartil (çeyreklikler),** ona bölenler **desil (ondabirlikler)** ve yüze bölenler **santil** (yüzdebirlikler) olarak adlandırılır. Kartillerin sayısı 3, desillerin 9 ve santillerin sayısı 99 dur. Medyan 2. kartile, 5. desile ve 50. santile eşittir.

Kümeye dört eşit parçaya bölen değerleri Q_1, Q_2, Q_3 ile gösterelim. Bunlar birinci, ikinci ve üçüncü yüzdelik olarak adlandırılır. Burada Q_2 medyandır.

Basit seride 1. kartil yani 1. yüzdelik (n+1)/4'üncü terimdir. 3. kartil ise 3(n+1)/4'üncü terimdir. Eğer Q_1 ve Q_3 tam veya buçukla bitiyorsa medyandaki gibi davranılır. Buna karşılık tam veya buçukla bitmeyen sayılar için buçuktan küçük küsurlar atılır, buçuktan büyük sayılar ise tam sayıya dönüştürülür.

Örnek 3.29. aşağıdaki serilerin kartillerini hesaplayınız?

X_i :11 22 34 46 57

Y_i :12 23 36 49

X serisi 1. kartil : (n+1)/4=(5+1)/4=1,5 terim yani Q_1=(11+22)/2=16,5

veya (22-11)*0,5=5,5 1.terim 11, Q_1=11+5,5=16,5

3. kartil :3(n+1)/4=3(5+1)/4=4.5 terim yani Q_3=(46+57)/2=51,5

veya (57-46)*.05=5,5 4. Terim 46 , Q_3=46+5,5=51,5

Y serisi 1. kartil : (n+1)/4=(4+1)/4=1,25 terim (23-12)*0,25=2,75
Q_1=12+2,75=14,75

3. kartil : 3(n+1)/4 =3(4+1)/4=3,75 terim (49-36)*0,75=9,75

Q_3=36+9,75=45,75

Örnek 3.30. Aşağıdaki sınıflanmış serilerin kartillerini bulunuz?

A Serisi			B Serisi		
X_i	f_i	$\sum f_i$	X_i	f_i	$\sum f_i$
11	2	2	13	3	3
22	3	5	24	6	9
34	4	9	37	4	13
45	2	11	48	5	18

A serisi : (n+1)/4=(11+1)/4=3. terim birinci kartildir. $Q_1=22$

3(n+1)/4=3(11+1)/4=9. terim üçüncü kartildir. $Q_3=34$

B serisi : (n+1)/4=(18+1)/4=4,75 terim birinci kartildir. $Q_1=24$

3(n+1)/4=3(18+1)/4=14,25 terim üçüncü kartildir. $Q_3=48$

Gruplanmış Serilerde Kartillerin Hesabı

$$Q_1 = L + \frac{\frac{n}{4} - F_{(i-1)}}{f_i} \times C \qquad Q_3 = L + \frac{\frac{3n}{4} - F_{(i-1)}}{f_i} \times C$$

.

L: Medyan sınıfının alt sınır değeri

n= Toplam gözlem sayısı

$F_{(i-1)}$ = Medyan sınıfından önceki sınıfların frekans toplamı

f_i= Medyan sınıfının kendi sınıf frekansı

C= sınıf genişliği

Kartil sınıfının belirlenmesinde yine kümülatif frekanslardan ayarlanılır. n/4'üncü terimi içeren sınıf 1. kartil sınıfı, 3N/4'üncü terimim içeren sınıf ise 3. kartil sınıfı kabul edilir.

n/4 ve 3n/4 tam sayı olmasa da formüller yine aynen kullanılır ve bu durumda kartiller yaklaşık bir değere sahip olur.

Örnek 3.31. Aşağıdaki gruplanmış serinin katillerini bulunuz?

Sınıflar	:0-2 den az	2-4 den az	4-6 dan az	6-8 den az
f_i	:4	3	1	2
Toplam f_i	:4	7	8	10

N/4=10/4=2,5'inci terim 1. kartildir. Böylece 1. kartil sınıfı 0-2 den az sınıfıdır.

L=Medyan sınıfının alt sınır değeri=0

n= Toplam gözlem sayısı=10

F_{i-1} = Medyan sınıfından önceki sınıfların frekans toplamı=0

f_i= Medyan sınıfının kendi sınıf frekansı=4

C= sınıf genişliği=2

$$Q_1 = L + \frac{\frac{n}{4} - F_{(i-1)}}{f_i} \times C = 0 + \frac{2,5-0}{4} \times 2 = 1,25$$

3N/4=30/4=7.5'inci terim 3. kartildir. Böylece 3. kartil sınıfı 4-6 dan az sınıfıdır.

L=Medyan sınıfının alt sınır değeri=4

N= Toplam gözlem sayısı=10

Fi-1 = Medyan sınıfından önceki sınıfların frekans toplamı=7

fi= Medyan sınıfının kendi sınıf frekansı=1

C= sınıf genişliği=2

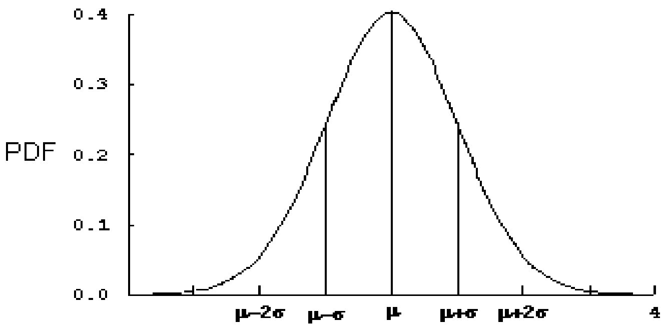
$$Q_3 = L + \frac{\frac{3n}{4} - F_{(i-1)}}{f_i} \times C = 4 + \frac{7,5-7}{1} \times 2 = 5$$

3.4. Ortalama Türünün Seçimi

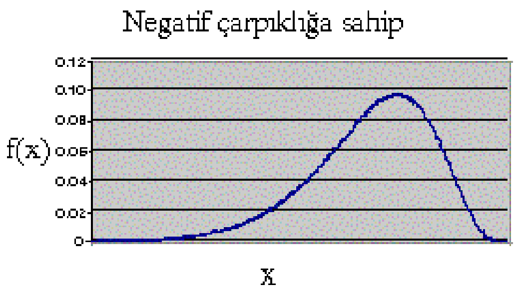
- ✓ Ortalama kıyaslama amacıyla hesaplandığında *Aritmetik Ortalama* tercih edilir. Çünkü Aritmetik Ortalama bütün terimler yada sınıflar üzerinden hesaplanan en duyarlı ortalamadır.
- ✓ Araştırmanın amacı seriyi kıyaslamayıp, seriyi temsil etmek ise yerine göre *Mod* yada *Medyan* tercih edilir.
- ✓ Terimlerin kendileri yerine oranları bizi ilgilendiriyorsa *Geometrik Ortalama* tercih edilir.
- ✓ Terimlerin tersleri ile ilgileniliyorsa *Harmonik ortalama* kullanılır.
- ✓ Sıfır veya negatif işaretli değerlere sahip serilerde *Harmonik ve Geometrik Ortalama* hesaplanamaz.
- ✓ Sınıf genişlikleri eşit olamayan gruplanmış serilerde *Medyanın* hesaplanması uygundur.
- ✓ Seri terimleri arasında önem farkı bulunduğunda *Tartılı Ortalama* uygulanır.
- ✓ Ortalama, ortanca ve mod arasında aşağıdaki genel ilişki vardır.

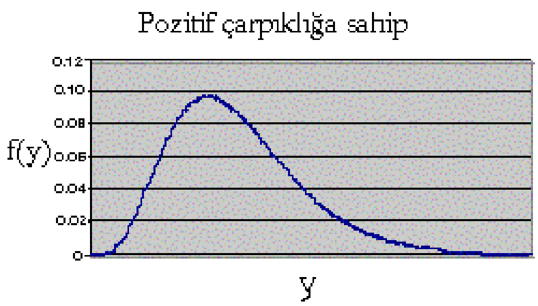
  Ortalama – Mod= 3* (Ortalama – Ortanca)

- ✓ Sıralamalı ölçümlü özelliklerde veya bütün değerlerin elde edilmesinin uzun zaman aldığı bazı durumlarda *Medyanın* kullanılması uygundur. Örneğin öğrenme davranışının incelendiği bir araştırmada bazı bireyler çok geç öğrenebilir, ortalama için bunu beklemek gerekir, *Medyan* için bunu beklemeye gerek kalmaz.



25. yüzdelik(1. çeyrek)





Simetrik eğri :

Uç nokta olmayan

Mod=Medyan=A.O.

en küçük değer

3.5. EXCEL VE SPSS'TE ORTALAMA HESABI

Örnek 3.32. Bir bölgedeki binaların yaşları aşağıdaki gibi bulunmuştur. Bu binaların ortalama yaşını duyarlı ve duyarlı olmayan ortalamalara göre Excel ve SPSS'te bulunuz.

	A
1	Bina Yaşı
2	1
3	3
4	5
5	8
6	10
7	10
8	10
9	15
10	18
11	20
12	20
13	25
14	32
15	35
16	40

Aritmetik Ortalama	=ORTALAMA(A2:A16)	16,8
Birinci Kartil	=DÖRTTEBİRLİK(A2:A16;1)	9
Üçüncü Kartil	=DÖRTTEBİRLİK(A2:A16;3)	22,5
Geometrik Ortalama	=GEOORT(A2:A16)	11,95
Harmonik Ortalama	=HARORT(A2:A16)	6,51
Ortanca	=ORTANCA(A2:A16)	15
Mod	=ENÇOK_OLAN(A2:A16)	10

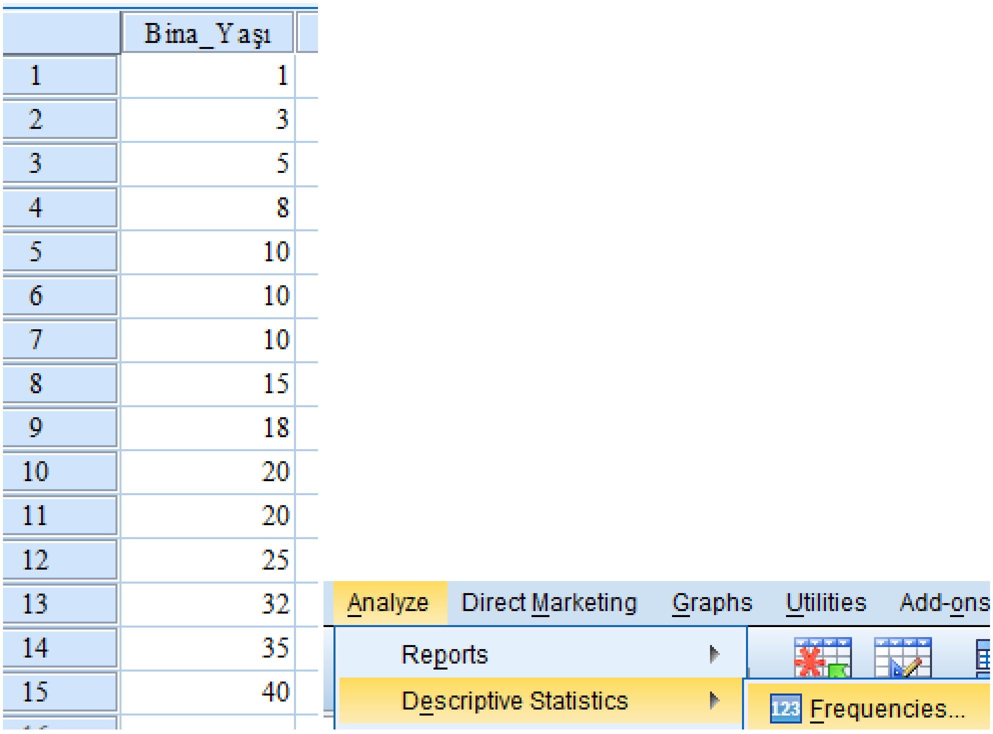
Excel’de dörttebirlik komutunun işlevi aşağıdaki gibidir:

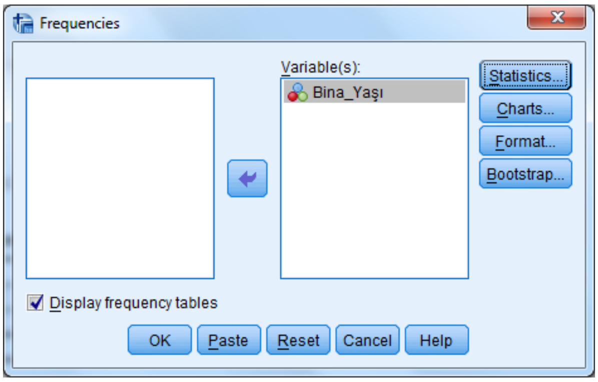
=DÖRTTEBİRLİK(A2:A16;?)

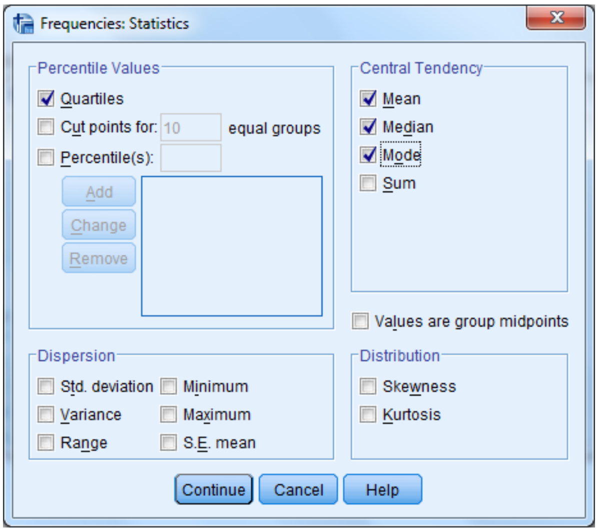
? işareti yerine aşağıdaki değerler girilerek istenen ifade bulunur.

DÖRTTEBIR DEĞERI	DÖRTTEBİRLİK SONUCU
0	En küçük değer
1	İlk dörttebirlik (25. yüzdebirlik)
2	Ortanca değer (50. yüzdebirlik)
3	Üçüncü dörttebirlik (75. yüzdebirlik)
4	En büyük değer

SPSS ÇÖZÜMÜ:



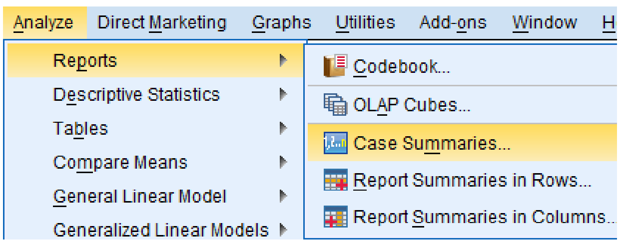


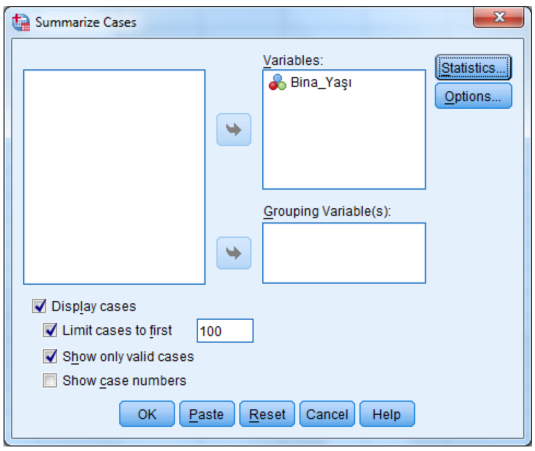


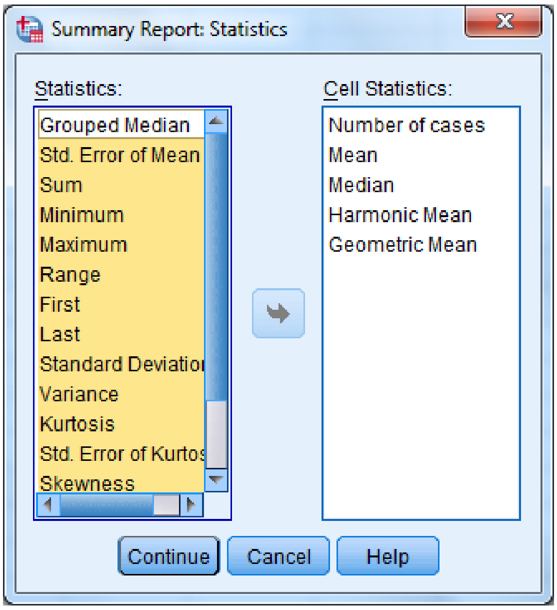
Statistics

Bina_Yaşı

N	Valid	15
	Missing	0
Mean		16,80
Median		15,00
Mode		10
Percentiles	25	8,00
	50	15,00
	75	25,00







Case Summaries[a]

		Bina_Yaşı
1		1
2		3
3		5
4		8
5		10
6		10
7		10
8		15
9		18
10		20
11		20
12		25
13		32
14		35
15		40
Total	N	15
	Mean	16,80
	Median	15,00
	Harmonic Mean	6,51
	Geometric Mean	11,95

a. Limited to first 100 cases.

3.6. ÖRNEK PROBLEMLER

1. Beş iş gününde bir banka şubesinde toplam 120 hesap açtırılmış ise günlük hesap açılma ortalaması kaçtır?

a)5 b) 12 c) 24 d) 60 e) 700

2. Bir öğrencinin istatistik dersinden I. arasınav notu 50, II. arasınav notu 60 ve final notu ise 60 dır. Dersin geçme notu için vizelerin %20 si, finalin ise %60 I alınacaktır. Buna göre bu öğrencinin başarı notu kaçtır?

a)58 b) 60 c) 64 d) 66 e) 70

3. Sınıflar	:0-5	5-10	10-15	15-20	20-25	25-30	30-35
f	: 2	5	6	10	5	2	4

Serisinin medyanı kaçtır?

a)15 b) 16 c) 17 d) 18 e) 20

4.20, 32, 25, 28, 45, 50 veri serisinin medyanı kaçtır?

a)25 b) 30 c) 32 d) 26,5 e) 28

5.2, 3, 4, 3, 2, 3, 5, 6, 7 veri serisinin modu kaçtır?

a)3 b) 2 c) 4 d) 4,5 e) 5

6.Bir işyerinde çalışan 100 işçinin almış oldukları ücretlerin aralıkları aşağıdaki tabloda verilmiştir.

İşçi Ücretleri	İşçi sayısı
500	5
750	5
1000	35
1250	25
1500	30
TOPLAM	100

İşçilerin aldıkları ücret ortalamasının mod'u nedir?

a)1250 b) 1500 c) 1000 d) 25 e) 35

Cevaplar:
1-c, 2-a, 3-c, 4-b, 5-a, 6-c

IV. BÖLÜM

4. DEĞİŞİM ÖLÇÜLERİ

Bir anakütleyi (popülasyonu) tanıtmak, başka ana kütlelerle karşılaştırabilmek için merkezi eğilim ölçüleri yanında dağılışın genişliğini, değişkenliğin büyüklüğünü gösteren bir başka tipik değerin verilmesi gerekmektedir. Bu tipik değer değişim ölçüsüdür.

Ortalamaları eşit olan seriler, değişkenlikleri veya bölüne şekilleri farklı olduğunda birbirine benzemez. Bu nedenle serileri tam olarak tanımlayabilmek için ortalamayı, değişimlerini ve bölünme şekillerini incelemek gerekir.

İstatistikte kullanılan bazı simgeler:

	Örneklem Parametresi	Anakütle Parametresi
Aritmetik ortalama	$\overline{X}$	μ
Standart sapma	s	σ
Varyans	s^2	σ^2
Gözlem sayısı	n	N
Korelasyon	r	ρ

Örnek 4.1. Üç farklı bölümde okuyan öğrencilere ait notlar aşağıdaki gibidir.

X_i: { 49,49,49,50,51,52 } ; $\overline{X}$ =AO(X)= 50

Y_i : { 35,41,50,55,58,61 } ; $\overline{Y}$ =AO(Y)= 50

Z_i: { 15,21,33,49,90,92 } ; $\overline{Z}$ =AO(Z)= 50

Bu üç değişkenin ortalaması aynı olduğu halde X, Y ve Z değişkenlerinin aldığı değerlerin en küçük ve en büyük değerlerine bakıldığında birbirlerinden çok farklıdır. X değerleri 50'nin etrafında çok yakın kümelendiği halde Z değerleri 50'den çok uzakta yer almakta, yani daha büyük değişim göstermektedir.

Bir popülasyonun bireylerinin değerleri arasındaki bu değişimin bir ölçü ile ifade edilmesi gerekir. Bu ölçülere yayılım veya değişim ölçüleri denir.

4.1. Değişim Genişliği (Range)

En basit değişim ölçüsü Değişim Genişliğidir.

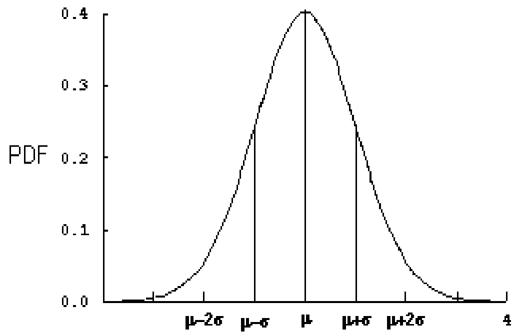
Değişim Genişliği = (En Büyük Gözlem – En Küçük Gözlem)

DG(X)= 52-49= 3

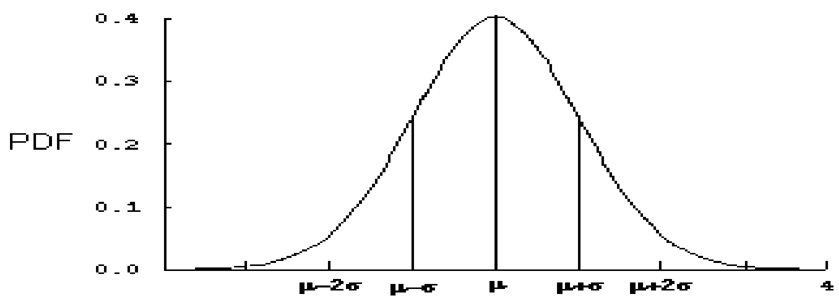
DG(Y)= 61-35= 26

DG(Z)= 92-15= 77

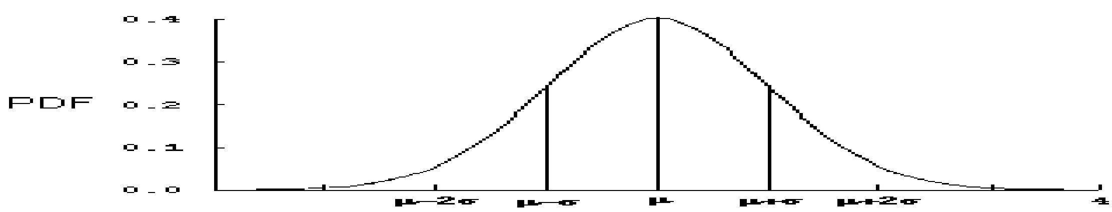
Bu dağılışların grafiklerine bakıldığında aşağıdaki şekilde bir görünüş söz konusudur.



X



Y



Z

Değişim genişliği aslında çok kullanılan bir ölçüdür. Genelde şeker hastasının en küçük ve en yüksek değerleri nedir, ateşli bir hastanın en düşük ve en yüksek ateş derecesi nedir, bunlarla ilgilenilir.

Ancak değişim genişliğinde sadece iki aşırı uç değer kullanıldığı için istikrarlı bir ölçü değildir, bilgi kaybı çoktur. Diğer bir sakıncası değişim genişliği örnek büyüklüğüne çok bağlıdır. Örnek büyüdükçe aşırı değerleri kapsama olasılığı artar. Ölçü birimleri farklı serilerin değişkenliğinin kıyaslanmasında değişim genişliğinden yararlanmak yanlış olur.

4.2. Kartiller Arası Fark (Interquartile Range)

Değişim genişliğinin serinin iki ucunda yer ala aşırı değerlerden etkilenmesi sakıncasını gidermek için *kartiller arası fark* kullanılır. Bu ölçü erinin 3. ve 1. kartilleri arasındaki farkı esas alır.

Kartiller arası fark (IQR)= Q_3-Q_1

Kartiller bütün verileri hesaba katmadığından ve Q_1'den küçük, Q_3'den büyük değerlerin değişkenliğini ihmal etmesi sakınca meydana getirir.

Bir veri kümesinin ***çeyrek sapması*** Q ile gösterilir ve aşağıdaki gibi hesaplanır.

$$Q = \frac{Q_3 - Q_1}{2}$$

Örnek 4.2. {1,3,4,6,10,12} serisinin kartiller arası farkını bulunuz?

Q_1=(n+1)/4=(6+1)/4=1.75 1.terim =1

(3-1)*0,75=1,5 Q_1=1+1,5=2,5

Q_3=3(n+1)/4=3(6+1)/4=5.25 5. terim 10

(12-10)*0,25=0,5 Q_3=10+0,5=10,5

Kartiller arası fark= Q_3- Q_1=10.5-2,5=8

4.3. Ortalama Sapma

Ortalama mutlak sapma olarak da anılan bu ölçü, terimlerin aritmetik ortalamadan mutlak sapmalarının aritmetik ortalamasıdır. Aritmetik ortalamaya göre ortalama sapma formülleri aşağıdaki gibidir:

Ortalama Sapma	**Basit Serilerde**	$O.S. = \frac{\sum \lvert X_i - \bar{X} \rvert}{n}$
	Sınıflanmış Serilerde	$O.S. = \frac{\sum f_i \lvert X_i - \bar{X} \rvert}{\sum f_i}$
	Gruplanmış Serilerde	$O.S. = \frac{\sum f_i \lvert m_i - \bar{X} \rvert}{\sum f_i}$

Ortalama sapma hesabı medyana göre yapıldığında $\overline{X}$ yerine Medyan konur.

Örnek 4.3. Aşağıdaki basit serinin ortalama sapmasını bulunuz?

X : {1,4,5,5,7,14}

$\overline{X}$ =36/6=6 $\left|X_i - \overline{X}\right|$={5,2,1,1,1,8} O.S.=18/6=3

Örnek 4.4. Aşağıdaki sınıflanmış serinin ortalama sapmasını bulunuz?

X_i	f_i	$X_i f_i$	$\left\lvert X_i - \overline{X}\right\rvert$	$f_i\left\lvert X_i - \overline{X}\right\rvert$
1	2	2	5.2	10.4
5	7	35	1.2	8.4
7	9	63	0.8	7.2
12	2	24	5.8	11.6
$\sum$	20	124		37.6

$\overline{X}$ =124/20=6,2 O.S.=37,6/20=1,88

4.4. Standart Sapma (Standard Deviation)

Kareli ortalama sapma adı verilen bu ölçü, terimlerin aritmetik ortalamadan cebirsel sapmalarının kareli ortalamasıdır. Standart sapma gözlemlerin ortalamadan ne kadar uzaklaştığını gösterir. Yani gözlemler arasında ne kadar yaygınlık olduğunu ifade eder.

Bir çalışma grubundaki her bir verinin ortalamaya göre ne kadar uzaklıkta olduğunu, bir diğer deyişle dağılımın ne yaygınlıkta olduğunu gösteren bir ölçüdür. Varyans ise belirli bir popülasyonda incelenen özelliğin ya da ölçümlerin ne genişlikte yani nasıl bir aralıkta (dar veya geniş) dağıldığının göstergesidir.

Standart sapma ***varyansın*** kare köküdür. *Varyans birimsiz ifade edildiği halde standart sapma ölçülen özelliğin birimi ile ifade edilir.* Yani milimetre yerine milimetre kare gibi veri biriminin karesinin kullanılması uygun olmadığından, varyansın yerine standart sapma kullanılır.

Ortalama etrafındaki saçılma fazla ise varyans büyük olacak, dolayısıyla standart sapma da büyük çıkacaktır. Veri seti içinde aşırı değer fazla ise bunların varyansı etkilemesi olağan karşılanabilir, ancak birkaç aşırı değerin varyansı büyütmesi olağan karşılanmaz.

Dağılımdaki maksimum ve minimum değerler arasındaki farkı dörde bölmek suretiyle, standart sapma kabaca tahmin edilebilir.

Moses’e göre standart sapma değeri örnekteki gözlem değerleri arasındaki varyasyonun derecesini tanımlayan bir istatistiktir. Değişkenin dağılışı normal dağılış gösteriyorsa, standart sapma verilerin ortalama etrafındaki dağılışını iyi bir şekilde açıklar.

$$s = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n}}$$

Standart Sapma	**Basit Serilerde**	$s = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n}}$
	Sınıflanmış Serilerde	$s = \sqrt{\frac{\sum f_i(X_i - \bar{X})^2}{\sum f_i}}$
	Gruplanmış Serilerde	$s = \sqrt{\frac{\sum f_i(m_i - \bar{X})^2}{\sum f_i}}$

	Varyans	**Standart Sapma**
Anakütle	$\sigma^2 = \frac{\sum(X_i-\mu)^2}{N}$	$\sigma = \sqrt{\frac{\sum(X_i-\mu)^2}{N}}$
Örneklem	$s^2 = \frac{\sum(X-\bar{X})^2}{n}$	$s = \sqrt{\frac{\sum(X-\bar{X})^2}{n}}$

Gözlem sayısı n<30 ise örnek verisine ilişkin standart sapma hesaplanırken paydada n yerine n-1 alınır. Çünkü elde edilen değer örneğin alındığı yığının standart sapmasını daha iyi verir. n'in büyük değerleri için (n≥30) iki tanım arasında bir fark yoktur.

$$s = \sqrt{\frac{\sum(X_i-\bar{X})^2}{n-1}}$$

Standart sapma daima ortalama sapmadan büyük çıkar. Çünkü standart sapma kareli ortalama cinsinden, ortalama sapma ise aritmetik ortalama cinsinden hesaplanan değişkenlik ölçüdür.

Gözlem değerlerinin tamamı birbirine eşit olamayacağından standart sapmanın değerleri pozitiftir.

Not: Kareli ortalamanın karesi ile Aritmetik Ortalamanın karesi arasındaki fark varyansa eşittir. $s^2 = K.O^2 - \bar{X}^2$ $s = \sqrt{\overline{X^2} - (\overline{X})^2} = \sqrt{\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2}$

Örnek 4.5. X_i:{49,49,49,50,51,52} veri setinin standart sapmasını bulunuz?

A.O.=50 , s^2= [(49-50)2+ (49-50)2+ (49-50)2+ (50-50)2+(51-50)2+ (52-50)2]/(6-1)

= (1+1+1+0+1+4)/5= 8/5=1,6

Standart sapma s veya s_x şeklinde değişken ismi ile birlikte yazılabilir.

$$s = \sqrt{Varyans} = \sqrt{s^2} = \sqrt{1.6} = 1.26$$

Örnek 4.6. Aşağıda sınıflanmış olarak verilen serilerin standart sapmasını bulunuz?

A Serisi					B Serisi				
X_i	f_i	$X_i - \overline{X}$	$(X_i - \bar{X})^2$	$f_i(X_i - \bar{X})^2$	X_i	f_i	$X_i - \overline{X}$	$(X_i - \bar{X})^2$	$f_i(X_i - \bar{X})^2$
1	2	-5,2	27,04	54,08	2	7	-1,9	3,61	25,27
5	7	-1,2	1,44	10,08	4	4	0,1	0,01	0,04
7	9	0,8	0,64	5,76	5	6	1,1	1,21	7,26
12	2	5,8	33,64	67,28	6	3	2,1	4,41	13,23
$\sum$	20			137,20		20			45,80

$s_A = \sqrt{\frac{\sum f_i(X_i - \bar{X})^2}{\sum f_i}} = \sqrt{\frac{137,2}{20}} = 2,62$ $s_B = \sqrt{\frac{45,8}{20}} = 1,51$

Örnek 4.7. Aşağıdaki gruplanmış serilenin standart sapmasını bulunuz?

Sınıflar	f_i	m_i	$m_i - \overline{X}$	$(m_i - \bar{X})^2$	$f_i(m_i - \bar{X})^2$
0-2 den az	3	1	-2,625	6,891	20,673
2-4 den az	7	3	-0,391	0,391	2,737
4-6 dan az	4	5	1,891	1,891	7,564
6-8 den az	2	7	11,391	11,391	22,782
$\sum$	16				53,756

$$s = \sqrt{\frac{\sum f_i(m_i - \bar{X})^2}{\sum f_i}} = \sqrt{\frac{53{,}756}{16}} = 1{,}83$$

4.4.1. Varyans (Variance)

Dağılış ölçüsü olarak kullanılan en uygun ölçü standart ayrılış ölçüsüdür. Standart ayrılış ölçüsü varyansın kareköküdür. Bu nedenle önce varyans hesaplanmasını görelim.

X_i: { $X_1, X_2, \ldots, X_n$} değerler kümesinin varyansı:

$$Varyans = s^2 = \frac{1}{(n-1)} \sum_{i=1}^{N} (X_i - \overline{X})^2 = \frac{1}{(n-1)} \left\{ \sum_{i=1}^{N} X_i^2 - \frac{(\sum X_i)^2}{N} \right\}$$

Örnek 4.8. X_i:{49,49,49,50,51,52} değişkeninin varyansı nedir?

$\overline{X}$ = 300/6=50

s^2= [(49-50)2+(49-50)2+ (49-50)2+ (50-50)2+(51-50)2+ (52-50)2]/(6-1)

= (1+1+1+0+1+4)/5 = 8/5=1,6

Veya ikinci formül kullanılırsa yine aynı sonuç elde edilir.

s^2= [(492+ 492+ 492+ 502+ 512+ 522)-3002 /6] /5= (15008-15000)/5=8/5=1,6

Örnek 4.9. Yeni doğan çocuğun doğum ağırlıkları (kg) ile ilgili sıklık tablosunda özetlenmiş olan veri setine ait varyansı bulunuz.

Sınıf sınırları	(m_i)	(f_i)	f_i*m_i	f_i*m_i^2
1,45 – 1,95	1,7	2	3,4	5,78
1,95 - 2,45	2,2	18	39,6	87,12
2,45 – 2,95	2,7	24	64,8	174,96
2,95 – 3,45	3,2	19	60,8	194,56
3,45 – 3,95	3,7	18	66,6	246,42
3,95 – 4,45	4,2	9	37,8	158,76
4,45 – 4,95	4,7	6	28,2	132,54
4,95 – 5,45	5,2	4	20,8	108,16
Toplam		100	322	1108,3

$$s^2 = \frac{\sum f_i(m_i - \bar{X})^2}{\sum f_i} = \frac{1}{\sum f_i - 1}\left(\sum f_i {m_i}^2 - \frac{(\sum f_i m_i)^2}{\sum f_i}\right)$$

$$= \frac{1}{100 - 1}\left(1108{,}3 - \frac{322^2}{100}\right) = 0{,}722$$

Soru : 70 erkekteki ürik asit konsantrasyon sonuçları aşağıdaki gibi bulunmuştur. Buna göre standart sapma değerini bulunuz?

Ürik Asit (mg/dl)	(m_i)	(f_i)	f_i*m_i	f_i*m_i^2
1,6-2,0	1,8	6	10,8	19,44
2,1-2,5	2,3	7	16,1	37,03
2,6-3,0	2,8	20	56,0	156,80
3,1-3,5	3,3	21	69,3	228,69
3,6-4,0	3,8	12	45,6	173,29
4,1-4,5	4,3	4	17,2	73,96
Toplam		70	215	689,21

$$s^2 = \frac{\sum f_i(m_i - \bar{X})^2}{\sum f_i} = \frac{1}{\sum f_i - 1}\left(\sum f_i {m_i}^2 - \frac{(\sum f_i m_i)^2}{\sum f_i}\right) = 0{,}419$$

Varyansın Özellikleri

I. Bir serinin bütün terimlerine aynı sayıyı eklersek (çıkartırsak) serinin varyansı değişmez.

$$\frac{\sum\left[(X_i+k)-(\overline{X}+k)\right]^2}{n}=\frac{\sum\left[X_i-\overline{X}\right]^2}{n} \qquad V(X\pm k)=V(X)$$

II. Bir serinin bütün terimlerini aynı sayılsa çarptığımızda (böldüğümüzde) varyans çarptığımız (böldüğümüz) sayının karesiyle orantılı olarak büyür (küçülür).

$$\frac{\sum\left[cX_i-c\overline{X}\right]^2}{n}=\frac{c^2\sum\left[X_i-\overline{X}\right]^2}{n} \qquad V(kX)=k^2V(X)$$

$$\frac{\sum\left[\frac{X_i}{c}-\frac{\overline{X}}{c}\right]^2}{n}=\frac{\frac{1}{c^2}\sum\left[X_i-\overline{X}\right]^2}{n}=\frac{1}{c^2}\frac{\sum\left[X_i-\overline{X}\right]^2}{n} \qquad V\left(\frac{X}{k}\right)=\frac{1}{k^2}V(X)$$

III. Birbiri ile ilişkili iki serinin bütün terimlerinin karşılıklı olarak toplanması (çıkarılması) suretiyle elde edilen serinin varyansı, bu serilerin varyanslarının toplamı ile kovaryansın 2 katının toplamına (farkına) eşittir.

$$\frac{\sum\left[(X_i-Y_i)-(\overline{X}+\overline{Y})\right]^2}{n}=\frac{\sum\left[(X_i-\overline{X})+(Y_i-\overline{Y})\right]^2}{n}$$

$$=\frac{\sum(X_i-\overline{X})^2+(Y_i-\overline{Y})^2+2(X_i-\overline{X})(Y_i-\overline{Y})}{n}$$

$$=\frac{\sum(X_i-\overline{X})^2}{n}+\frac{\sum(Y_i-\overline{Y})^2}{n}+2\frac{\sum(X_i-\overline{X})(Y_i-\overline{Y})}{n}$$

$$V(X+Y)=V(X)+V(Y)+2Cov(XY)$$

$$V(X-Y)=V(X)+V(Y)-2Cov(XY)$$

Burada **kovaryans,** seriler arasındaki ilişkinin varlığını ve yönünü (aynı veya ters) belirleyen bir ölçüdür. Kovaryans iki değişkenin birlikte değişme derecesini gösteren bir istatistiktir. İki değişkenin aynı objelere ait değerlerinin bu değişkenlere ait aritmetik ortalamadan farklarının çarpımlarının aritmetik ortalamasıdır. İki değişken aynı yönde değişiyorsa, büyük X ve Y yada küçük X ile Y ortak olaylarının meydana gelme frekansı diğer durumlardan daha büyük olur.

IV. Birbirinden bağımsız iki serinin terimlerinin karşılıklı olarak toplanması (çıkarılması) suretiyle elde edilen serinin varyansı, bu serilerin varyansları toplamına eşittir. X ve Y serileri birbirinden bağımsız olduğunda *Cov(XY)=0* olur.

$$V(X+Y) = V(X) + V(Y)$$
$$V(X-Y) = V(X) + V(Y)$$

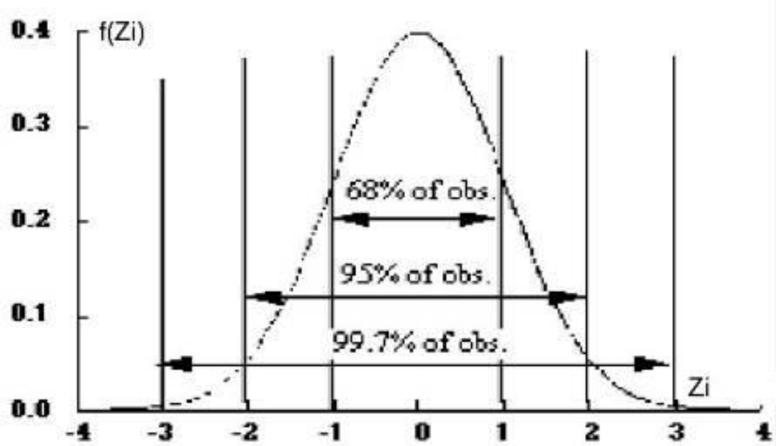
4.5. Standart Hata (Standard Error)

Standart hata örnekleme dağılımındaki ortalamaların standart sapmasıdır. Seçilecek örneklerde ortalamalar arasındaki yaygınlığı gösterir. Standart hata örnek büyüklüğünün fonksiyonudur. Böylece n attıkça hata küçülür. ***Standart hata*** Aritmetik ortalamada oluşan hatanın belirlenmesi için bulunur.

Standart hata ortalamalarla ilgilidir, deneklerle ilgili değildir. Standart hata, standart sapma gibi değişkenin dağılışı hakkında bilgi vermez.

$$S.H. = s_{\overline{X}} = \sqrt{\frac{Varyans}{n}} = \frac{s}{\sqrt{n}}$$

Standart ayrılış ölçüsü ve standart hata Normal dağılış gösteren özellikler için oldukça kullanışlıdır. Normal dağılışta bireylerin %68.27 si ortalamanın bir standart sapma solunda ve bir standart sapma sağındaki noktalar arasında bulunur. Bireylerin %95.45 i ortalamanın iki standart sapma sağında ve solunda bulunur.



örneklemlerin ortalamalarının yayılmasının ölçütüdür. Standart hata grupların ortalamalarının birbirleri ile karşılaştırılmasında kullanılır.

Ortalamanın standart hatası, ortalamanın dağılımındaki varyasyonu (değişimi) gösterir, örneklem sayısının artması ile küçülür. Standart hatanın küçük olması popülasyon parametresine ait yapılacak olan tahminler açısından ve daha dar güven aralığı sınırlar bulma açısından önemlidir.

Örneğin bir çalışma grubunun ortalamasına ait standart hata sıfır olarak bulunsun. Bu sıfır değeri, çalışma grubuna ait ortalama değerinin, grubu oluşturan örneklerde değişmediğini ve popülasyon parametresini tahmin bakımından en iyi tahmini yaptığını gösterir. **Standart hata ne kadar küçükse, popülasyon ait tahminlerimiz de o kadar isabetli olmaktadır.**

Bir serinin ortalaması verilirken standart hatası ile birlikte verilir, şöyle ki:

Aritmetik Ortalama± Standart Hata

$$\overline{X} \mp S_{\overline{X}}$$

X_i serisi için bunun ifadesi: 0±0.52 şeklindedir. Böylece ortalamaya ait değişkenlik yanında ifade edilmiş olur.Herhangi bir dağılımda gözlem değerlerinin %75'i AO ± 2 SS aralığında yer alır.

Normal dağılımda ise:

Ortalama ± 1 (standart sapma):

Verilerin % 68'inin "ortalama ± 1 (standart sapma)" kadar ortalamanın sağına-soluna yayıldığını gösterir.

Ortalama ± 2(standart sapma):

Verilerin % 95'inin "ortalama ± 2 (standart sapma)" kadar ortalamanın sağına-soluna yayıldığını gösterir.

Ortalama ± 3(standart sapma):

Verilerin % 99'unun "ortalama ± 3 (standart sapma)" kadar ortalamanın sağına-soluna yayıldığını gösterir.

Ortalama ± 1(standart hata):

Verilerin % 68 olasılıkla "± 1 (standart hata)" kadar sağa-sola yayıldığını gösterir. (Yani, bu çalışma grubu, aynı popülasyondan 100 kez tekrar tekrar oluşturulacak olursa, bunların 68'inin ortalaması, bu sınırlar arasında kalacaktır).

Ortalama ± 2(standart hata):

Verilerin % 95 olasılıkla "± 2 (standart hata)" kadar sağa-sola yayıldığını gösterir.

Ortalama ± 3(standart hata):

Verilerin % 99 olasılıkla "± 3 (standart hata)" kadar sağa-sola yayıldığını gösterir.

Not:<u>Standart Sapma mı? Standart Hata mı?</u>

Çalışma gruplarına ait veriler, sadece ilgili olduğu grubun özelliğini/özelliklerini gösteriyorsa *"ortalama ±standart sapma"* tercih edilmelidir.

Verileri birbiri ile karşılaştırarak, gruplar arasında fark olup olmadığı öğrenilmek isteniyorsa; *"ortalama ± standart hata"* kullanılması daha uygun olacaktır.

Soru. İki grup üzerinde yapılan bir çalışmada grupların yaş ve boy değerlerine bakılıyor. Eğer gruplarda AO±*SS* değerleri kullanılırsa, sadece o grubun, yaş veya boy yönünden nasıl bir dağılım gösterdiği anlaşılmış olur. Ancak gruplar birbiri ile karşılaştırılırken AO±*SH kullanılır.*

1. grup:	2. grup:
ortalama = 13.0	ortalama = 15.7
standart hata= 0.9	standart hata= 0.4

13±2.0,9 yani 11.2-14.8 (I.grup için güven sınırları % 95 olasılıkla 11.2-14.8 değerleri arasında olacaktır).

15.7±2.0,4 yani 14.9-16.5 (II. Grup için güven sınırları % 95 olasılıkla 14.9-16.5 değerleri arasında olacaktır).

4.6. Değişim (varyasyon) Katsayısı (coefficient of variation)

Bir serinin standart sapmasının aritmetik ortalamasına bölünüp 100 ile çarpılması sonucu elde edilen değere değişim katsayısı adı verilir. Değişim katsayısının ölçü birimi yoktur.

$$DK = \frac{s}{\bar{X}} \times 100$$

Değişim katsayısı farklı serilerin değişkenliklerini kıyaslamada iyi bir ölçü olabilir. Basit, sınıflanmış ve gruplanmış seriler için uygun olan bu ölçü, seriler arasındaki cins ve büyüklük farklılığını ortadan kaldırır.

Değişim katsayısı küçük olan serilerin diğerlerine göre aha az değişken olduğu söylenir. Bunun anlamı ise seri terimlerinin aritmetik ortalama etrafında daha homojen olarak dağıldığıdır.

Örnek 4.10. X_i:{49,49,49,50,51,52}serisinin değişim katsayısını bulunuz?

D.K= 1.26 *100 /50=%2,52

Değişim Katsayısının Kullanımı

A ve B gibi iki farklı popülasyondaki değişkenliği karşılaştırmak istersek doğrudan standart sapma veya standart hatalarına bakmak yanıltıcı olabilir.

Ana kütlelerin ortalamaları büyüklük olarak birbirinden çok farklı ise standart ayrılış değerlerini ortalamaların % olarak ifade ederek ölçü biriminden bağımsız bir değer elde edilir, bu değişim (varyasyon) katsayısıdır, karşılaştırmada bu kullanılır. Örneğin fillerin ağırlığı mı daha değişkendir, farelerin ağırlığı mı sorusunun cevabı için D.K. kullanılır.

Aynı şekilde özellikler farklı birimlerle ölçülüyorsa, örneğin üzerlerinde deneme yapılan farelerin kan şekerimi daha değişkendir yoksa vücut ağırlıklarımı sorusu ile karşılaşılsa bunun için varyasyon katsayılarına bakmak gerekir.

Ortalamanın sıfıra yakın olduğu durumlarda varyasyon katsayısının güvenirliliği azalır. Bu gibi durumlarda dikkatli olmak gerekir.

Örnek 4.11. Aynı şahıs üzerinde iki farklı yöntemle (X:autoanalyser, Y:Microenzymatic) ölçülen kolesterol miktarları aşağıda verilmiştir.

X: { 177,193,195,209, 226} %mg/ml   Y: {192,197,200,202,209}

Her ikisinin aritmetik ortalaması eşittir.

AO(X)= 1000/5=200   AO(Y)=1000/5=200

V(X)=(201360-200000)/4=340   s=18,44   V(Y)=(200158-200000)/4=39,5 s=6.28

DK(X)=18,44*100/200=%9,22   DK(Y)=6,28*100/200=%3,14

X yönteminin değişim katsayısı Y yöntemininkinden büyük olduğundan, X yöntemi daha değişkendir. Yani Y yöntemi daha iyidir.

Soru : Bit TV tüpü üreticisinin A ve B olmak üzere iki cins tüpü vardır. Tüplerin ortalama ömürleri $\overline{X}_A = 149$saat ve $\overline{X}_B = 187$ saat, standart sapmaları s_A=280 saat ve s_B=310 saattir. Hangi tüp daha büyük değişime sahiptir?

$$DK_A = \frac{s}{\bar{X}} \times 100 = \frac{280}{1495} \times 100 = 18{,}7$$

$$DK_B = \frac{310}{1875} \times 100 = 16{,}5$$

A tüpü daha büyük değişime sahiptir.

Sorular

1. Bir seri için kareli ortalama 10 ve aritmetik ortalama 6 bulunmuştur. Bu serinin standart sapması kaçtır? (C:8)

2. Aritmetik ortalaması 100 ve varyansı 144 olan bir serinin değişim katsayısı kaçtır? (C:12)

3. x: 25 20 35 29 19

serisi için değişim aralığı kaçtır?(C:16)

4.

x:15	20	25	30	35
f:2	4	9	5	3

Serisinin değişim aralığı kaçtır? (C:20)

5. x: 3,5,45,75 serisinin geometrik ortalaması kaçtır? (C:15)

6. Bir seride Mod=6, A.O.=12 ve Medyan=9 bulunmuştur. Bu serinin dağılım için ne söylenebilir?

V.BÖLÜM

5. MOMENTLER VE ASİMETRİ ÖLÇÜLERİ

5.1. Momentler

Terimlerin sıfırdan sapmalarının $(X_i - 0)$ veya aritmetik ortalamadan sapmalarının $(X_i - \overline{X})$ çeşitli kuvvetlerinin aritmetik ortalamalarının tamamına **moment** adı verilir. Momentler sıfır etrafındaki momentler ve aritmetik ortalama etrafındaki momentler olmak üzere ikiye ayrılır. Sapmaların derecesi (r) ile gösterilir ve r momentin derecesini belirler.

Sıfıra Göre Momentler	Basit Serilerde	$m_r = \frac{\sum X_i^r}{n}$
	Sınıflanmış Serilerde	$m_r = \frac{\sum f_i X_i^r}{\sum f_i = n}$
	Gruplanmış Serilerde	$m_r = \frac{\sum f_i m_i^r}{\sum f_i = n}$

Burada r=1 ise birinci moment $m_1 = \overline{X}$ yani aritmetik ortalamayı verir.

Aritmetik Ortalamaya Göre Momentler	Basit Serilerde	$\mu_r = \frac{\sum (X_i - \bar{X})^r}{n}$
	Sınıflanmış Serilerde	$\mu_r = \frac{\sum f_i (X_i - \bar{X})^r}{\sum f_i}$
	Gruplanmış Serilerde	$\mu_r = \frac{\sum f_i (m_i - \bar{X})^r}{\sum f_i}$

Bu serilerde r=1 için $\mu_1 = 0$, r=2 için $\mu_2 = \sigma^2$ yani varyanstır.

Sıfıra göre momentler bilindiğinde, aritmetik ortalamaya göre momentler bulunabilir. Bu şekilde elde edilen bağıntılara **König Teoremi** adı verilir.

$$\mu_2 = m_2 - m_1^2$$

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3$$

$$\mu_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4$$

Örnek 5.1. Aşağıdaki basit serinin sıfıra ve aritmetik ortalamaya göre momentlerini bulunuz?

X_i	X_i^2	X_i^3	X_i^4	$X_i - \overline{X}$	$(X_i - \overline{X})^2$	$(X_i - \overline{X})^3$	$(X_i - \overline{X})^4$
1	1	1	1	-5	25	-125	625
3	9	27	81	-3	9	-27	81
4	16	64	256	-2	4	-8	16
6	36	216	1296	0	0	0	0
10	100	1000	10000	4	16	64	256
12	144	1728	20736	6	36	216	1296
∑ 36	306	3036	32370	0	90	120	2274

Sıfıra göre momentler:

m_1=36/6=6 m_2=306/6=51 m_3=3036/6=506 m_4=32370/6=5395

Aritmetik Ortalamaya Göre Momentler:

μ_1=0/6=0 μ_2=90/6=15 μ_3=120/6=20 μ_4=2274/6=379 veya

[illegible]

[illegible]

[illegible]

Örnek 5.2. Aşağıdaki sınıflanmış serinin sıfıra ve aritmetik ortalamaya göre momentlerini bulunuz?

X_i	f_i	X_i^2	X_i^3	X_i^4	f_iX_i	$f_iX_i^2$	$f_iX_i^3$	$f_iX_i^4$
2	3	4	8	16	6	12	24	48
3	6	9	27	81	18	54	162	486
4	4	16	64	256	16	64	256	1024
6	7	36	216	1296	42	252	1512	9072
∑	20				82	382	1954	10630

$(X_i - \overline{X})$	$(X_i - \overline{X})^2$	$(X_i - \overline{X})^3$	$(X_i - \overline{X})^4$
-2,1	4,41	-9,261	19,4481
-1,1	1,21	-1,331	1,4641
-0,1	0,01	-0,001	0,0001
1,9	3,61	6,859	13,0321

$f_i(X_i - \bar{X})$	$f_i(X_i - \bar{X})^2$	$f_i(X_i - \bar{X})^3$	$f_i(X_i - \bar{X})^4$
-6,3	13,23	-27,783	58,3443
-6,6	7,26	-7,986	8,7846
-0,4	0,04	-0,004	0,0004
13,3	25,27	48,013	91,2247
∑ 0	45,80	12,240	158,3540

Sıfıra göre momentler:

m_1=82/20=4,1 m_2=382/20=19,1 m_3=1954/20=97,7 m_4=10630/20=531,5

Aritmetik Ortalamaya Göre Momentler:

μ_1=0/20=0 μ_2=45,8/20=2,29 μ_3=12,24/20=0,612

μ_4=158,354/20=7,9177

Örnek 5.3. Aşağıdaki serinin sıfıra ve aritmetik ortalamaya göre momentlerini bulunuz?

Sınıflar	f_i	m_i	m_i^2	m_i^3	m_i^4
0-2 den az	2	1	1	1	1
2-4 den az	1	3	9	27	81
4-6 dan az	8	5	25	125	625
6-8 den az	5	7	49	343	2401

$f_i m_i$	$f_i m_i^2$	$f_i m_i^3$	$f_i m_i^4$	$(m_i - \overline{X})$	$(m_i - \overline{X})^2$	$(m_i - \overline{X})^3$	$(m_i - \overline{X})^4$
2	2	2	2	-4	16	-64	256
3	9	27	81	-2	4	-8	16
40	200	1000	5000	0	0	0	0
35	245	1715	12005	2	4	8	16
∑ 80	456	2744	17088				

$f_i(m_i - \overline{X})$	$f_i(m_i - \overline{X})^2$	$f_i(m_i - \overline{X})^3$	$f_i(m_i - \overline{X})^4$
-8	32	-128	512
-2	4	-8	16
0	0	0	0
10	20	40	80
∑ 0	56	-96	608

Sıfıra göre momentler:

m_1=80/16=5 m_2=456/16=28,5

m_3=2744/16=171,5 m_4=17088/16=1068

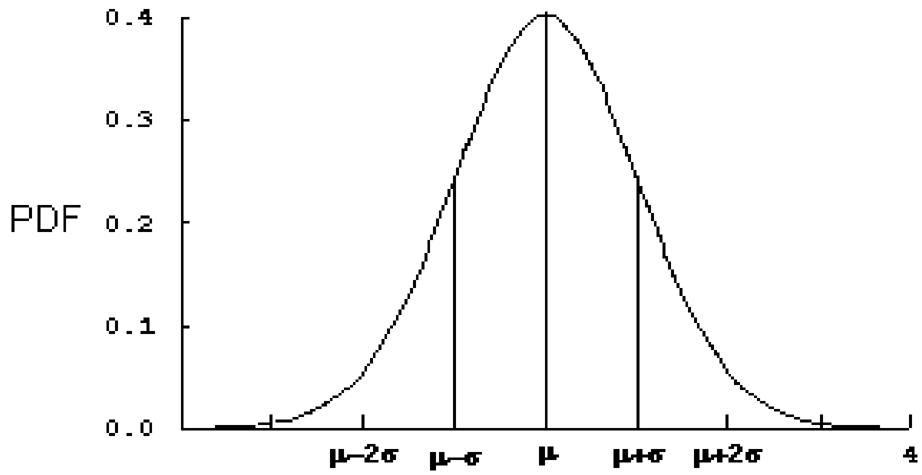
Aritmetik Ortalamaya Göre Momentler:

μ_1=0/16=0 μ_2=56/16=3.5

μ_3=-96/16=-6 μ_4=608/16=38

5.2. Çarpıklık (Skewness)

Çarpıklık bir dağılımın simetrik olmayış veya simetriklikten ayrılma derecesidir. Yani çarpıklık, bir dağılımın ortalaması etrafındaki asimetri derecesini belirtir.

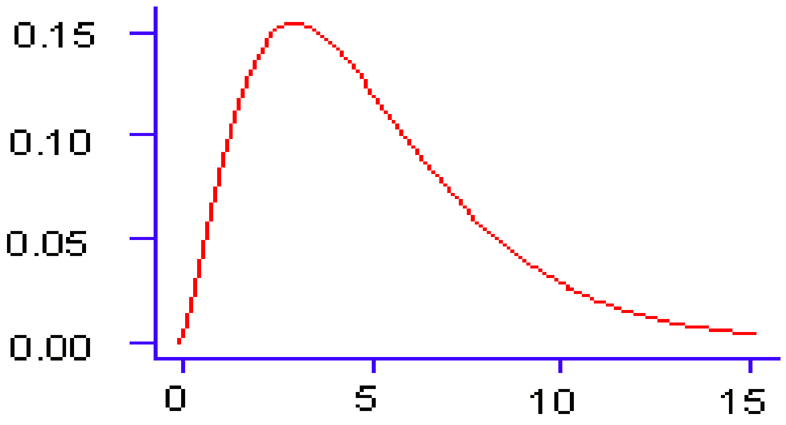


Simetrik yada çan şeklinde eğri

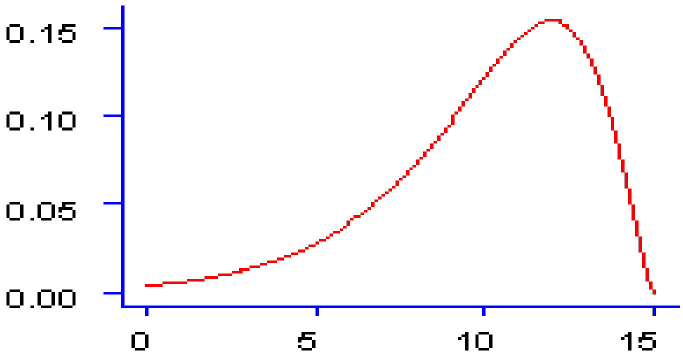
Simetrik yada çan şeklindeki frekans eğrileri, merkezdeki maksimumdan eşit uzaklıkta yer alan gözlemlerin aynı frekansa sahip olduğunu gösterir. Yani frekansların serinin maksimum noktası etrafında simetrik olarak dağılması seriye simetrik seri adını kazandırır. Normal eğri buna örnektir. Normal eğride serinin tam ortasında maksimuma ulaşan frekanslar, sonra hızla düşmeye başlar.

Simetrik seri **sivri, basık ve normal seri** gibi olabilir. Normal eğrinin özelliği grafikte maksimum noktanın belirli bir yüksekliğe sahip olmasıdır. Bu yüksekliğin normalin üstüne çıkması durumunda sivri seri, normalin altına düşmesi durumunda ise basık seriden söz edilir.

Simetrik seri belli bir yüksekliğe sahip olmadığında normal olmaktan çıkar. Ayrıca simetrik seri sağa veya sola doğru eğilim gösterdiğinde de simetriklik özelliğini kaybeder. Frekanslar serinin tam ortasında değil de, ortadan önceki bir noktada yığıldığında "**sağa çarpık seri"**, ortadan sonraki bir noktada yığıldığında ise "**sola çarpık seri"** den söz edilir.



Pozitif (sağa) çarpık eğri



Negatif (sola) çarpık eğri

Çarpıklık= $\alpha_3 = \dfrac{\mu_3}{\sqrt{\mu_2^3}} = \dfrac{\mu_3}{\sigma^3}$

Seri simetrik ise α_3=0 dır. Ancak α_3=0 ise serinin mutlaka simetrik olması gerekmez. α_3>0 ise seri sağa (pozitif) çarpık, α_3<0 ise sola (negatif) çarpıktır. $|\alpha_3| > 0,5$ ise genelde çarpıklık kuvvetli kabul edilir.

Örnek 5.4. Örnek 5.2.'deki verilerin çarpıklığını bulup, yorumlayınız?

μ_2=2.29 , μ_3=0.612

$$\alpha_3 = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{0.612}{\sqrt{2,29^3}} = 0,177 > 0$$

Çarpıklık pozitif olduğundan seri sağa çarpık (eğik), fakat çarpıklık zayıftır.

Not: Paket programlar ve Excelde çarpıklık katsayısı aşağıdaki formülle yaklaşık olarak hesaplanır.

$$ÇK = \alpha_3 = \frac{n}{(n-1)(n-2)} \sum \left(\frac{X_i - \overline{X}}{s} \right)^3$$

Soru. Aşağıdaki veri seti için çarpıklık katsayısını bulup, sonucu yorumlayınız?

X	
1	-1,64
3	-0,35
4	-0,10
6	0,00
10	0,84
12	2,83
TOPLAM	1,57
AO	6,00
ss	4,24

$$\alpha_3 = \frac{n}{(n-1)(n-2)} \sum \left(\frac{X_i - \overline{X}}{s} \right)^3$$

$$= \frac{6}{(6-1)(6-2)} 1,57 = 0,47 > 0$$

Seri hafif sağa çarpık bir seridir.

Pearson Çarpıklık Ölçüsü

$$Ç_{P1} = \frac{\overline{X} - Mod}{s} \qquad Ç_{P2} = \frac{3(\overline{X} - Medyan)}{s}$$

Pearson çarpıklık ölçülerinden birincisi ($Ç_{P1}$) simetrik serilerde sıfır, sola çarpık serilerde negatif ve sağa çarpık serilerde pozitif çıkar. Aşırı çarpık serilerde (asimetri-simetrik olmayan) AO-Mod farkı AO-Medyan farkının yaklaşık 3 katı kadardır. Mod hesabının zor olduğu durumlarda $Ç_{P2}$ kullanılır. Çarpıklık ölçüsü değerleri 1'e yakınsa serinin çarpık olduğu (simetrik olmadığı) söylenebilir.

Bowley Çarpıklık Ölçüsü

$$Ç_B = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}$$

Bowley çarpıklık ölçüsü kartillere göre hesaplanan bir çarpıklık ölçüsüdür. Simetrik bir seride (Q_2-Q_1) ile (Q_3-Q_2) farklarının eşit olması beklenir. Birinci fark daha büyükse seri sola çarpık, ikinci fark daha büyükse seri sağa çarpık bir dağılım gösterir. Bowley çarpıklık ölçüsü aralığında değer alır. Çarpıklık katsayısı 0'a yakınsa seri simetrik bir seridir. +1'e yakınsa sağa çarpık, -1'e yakınsa seri sola çarpık seridir denilebilir.

Örnek 5.5. Aşağıdaki veri seti için Pearson, ve Bowley çarpıklık katsayısını bulup, sonucu yorumlayınız?

X: 3,4,5,2,3,4,5,6,4,7

AO	4,3
SS	1,49
MOD	4
MEDYAN	4
Q_1	3
Q_3	5,25

$$Ç_{P1} = \frac{\overline{X} - Mod}{s} = \frac{4,3-4}{1,49} = 0,2$$

$$Ç_{P2} = \frac{3(\overline{X} - Medyan)}{s} = \frac{3(4,3-4)}{1,49} = 0,6$$

$$Ç_B = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)} = \frac{(5,25-4)-(4-3)}{(5,25-4)+(4-3)} = 0,11$$

Tüm değerler pozitif olduğundan seri hafif sağa çarpık bir seridir denilebilir.

5.3. Basıklık (Kurtosis)

Basıklık bir dağılımın sivrilik derecesidir ve çoğunlukla normal dağılıma göre ele alınır. Basıklık, normal dağılımla karşılaştırıldığında, bir dağılımın göreceli dikliğini ya da düzlüğünü verir. Pozitif basıklık, görece dik bir dağılımı belirtir. Negatif basıklık görece düz bir dağılımı belirtir. Normal seri simetrik serinin özel bir şeklidir. Bir serinin normal olabilmesi için hem simetrik olması (α_3=0) hem de normal bir yüksekliğe sahip olması gerekir.

α_4 simgesiyle gösterilen ve momentlere dayanan basıklık ölçüsü aşağıdaki gibi bulunur:

$$\alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}$$

Not: *Bir serinin normal seri olup olmadığını çarpıklık ve basıklık katsayılarına bakarak anlayabiliriz. Normal seride α_3=0 ve α_4=3 dır.*

Örnek 5.6. Örnek 5.2.'deki verilerin basıklığını bulup, yorumlayınız?

μ_2=2,29 , μ_4=7,9177 $\alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{7,9177}{2,29^2} = 1,51 < 3$

Seri 3'den küçük olduğundan basık bir seridir.

Örnek 5.7. Örnek 5.3.'deki verilerin basıklığını bulup, yorumlayınız?

μ_2=3.5 , μ_4=38 $\alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{38}{3,5^2} = 3,1 > 3$

Seri 3'den büyük olduğundan sivridir.

Not: *Bir serinin normal seri olup olmadığını çarpıklık ve basıklık katsayılarına bakarak anlayabiliriz. Normal seride α_3=0 ve α_4=3'tür. Buna normallik testi denir.*

Not: Paket programlar ve Excelde basıklık katsayısı aşağıdaki formülle yaklaşık olarak hesaplanır.

$$B = \alpha_4 = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{X_i - \overline{X}}{s} \right)^4 \right] - \frac{3(n-1)^2}{(n-2)(n-3)}$$

B=0 ise dağılım simetrik ve normal dağılıma uyar.

B<0 ise dağılım basık

B>0 ise dağılım sivridir.

Soru. Aşağıdaki veri seti için Basıklık katsayısını bulup, sonucu yorumlayınız?

X: 3 4 5 2 3 4 5 6 4 7

AO=4,3 s=1,49 Mod=4 Medyan=4

Q1=3 Q3=5,25

$$B = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{X_i - \overline{X}}{s} \right)^4 \right] - \frac{3(n-1)^2}{(n-2)(n-3)}$$

$$= \left[\frac{10 \times 11}{9 \times 8 \times 7} \left\{ \left(\frac{3-4,3}{1,49} \right)^4 + \left(\frac{4-4,3}{1,49} \right)^4 + \ldots + \left(\frac{7-4,3}{1,49} \right)^4 \right\} \right] - \frac{3 \times 9^2}{8 \times 7}$$

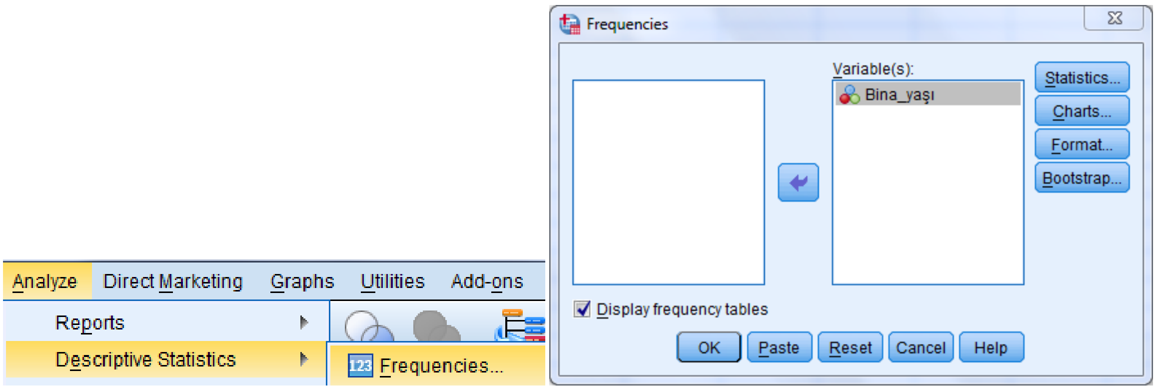
$= -0,15$ ***Seri hafif basıktır.***

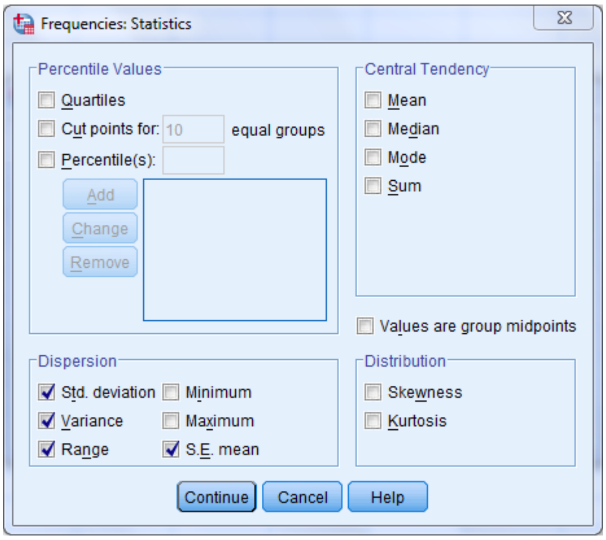
5.4. Excel Ve Spss Uygulaması

Örnek 5.8. Aşağıdaki veri seti için değişim ölçülerini Excel ve SPSS'te hesaplayınız.

	A
1	Bina Yaşı
2	1
3	3
4	5
5	8
6	10
7	10
8	10
9	15
10	18
11	20
12	20
13	25
14	32
15	35
16	40

Aritmetik Ortalama	=ORTALAMA(A2:A16)	16,80
Değişim Genişliği	=MAK(A2:A162)-MİN(A2:A16)	39,00
Birinci Kartil	=DÖRTTEBİRLİK(A2:A16;1)	9,00
Üçüncü Kartil	=DÖRTTEBİRLİK(A2:A16;3)	22,50
Kartiller Arası Fark	=D4-D3	13,50
Ortalama Sapma	=ORTSAP(A2:A16)	9,65
Standart Sapma	=STDSAPMA(A2:A16)	11,92
Varyans	=VAR(A2:A16)	142,03
Standart Hata	=D7/KAREKÖK(BAĞ_DEĞ_SAY(A2:A16))	3,08
Değişim Katsayısı	=(D7/D1)*100	70,94





Statistics

Bina_yaşı

N	Valid	15
	Missing	0
Std. Error of Mean		3,077
Std. Deviation		11,918
Variance		142,029
Range		39

5.5. Örnek Problemler

1. Bir seri için kareli ortalama 10 ve aritmetik ortalama 6 bulunmuştur. Bu serinin standart sapması kaçtır?

a) 8 b) 10 c)12 d) 14 e)15

2. Aritmetik ortalaması 100 ve varyansı 144 olan bir serinin değişim katsayısı kaçtır?

a) 8 b) 10 c)12 d) 14 e)15

3. x: 25 20 35 29 19

serisi için değişim aralığı kaçtır?

a) 8 b) 10 c)12 d) 14 e)16

4.

x:15	20	25	30	35
f:2	4	9	5	3

Serisinin değişim aralığı kaçtır?

a) 8 b) 10 c)15 d) 20 e)22

Cevaplar: 1-a, 2-c, 3-e, 4-d

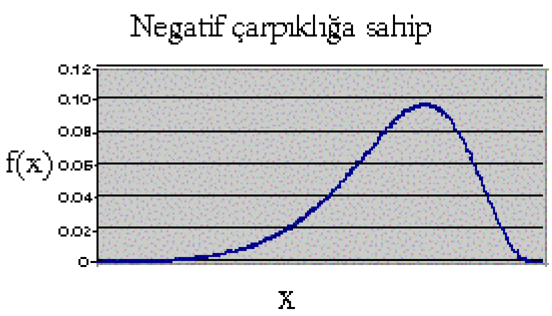
VI.BÖLÜM

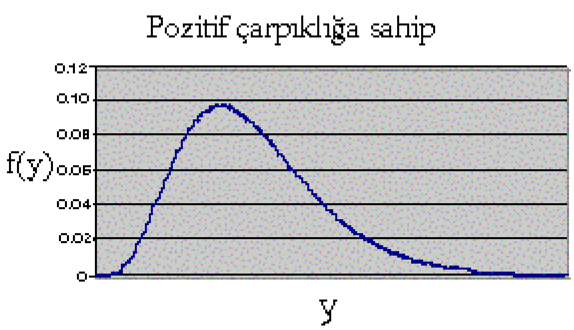
6. NORMAL DAĞILIM

İstatistikte uygulanan birçok analiz ve tahmin yöntemleri X değişkeninin Normal dağılış göstermesi halinde geçerlidir. Bu nedenle dağılışın Normal olamadığı hallerde yapılan analiz geçerliliğini kaybeder. Dolayısıyla bu dağılış istatistikte önemli bir yere sahiptir.

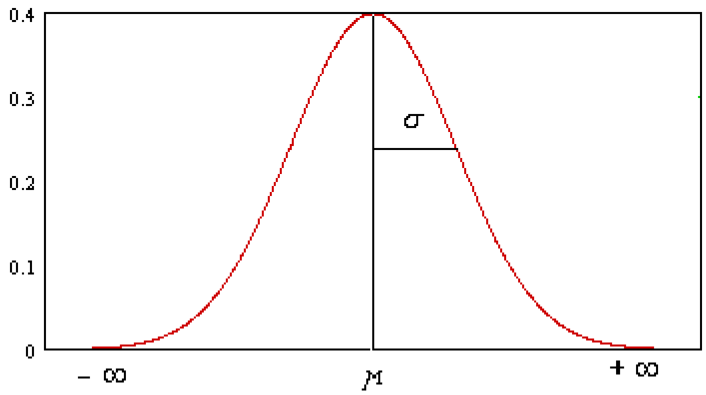
Eldeki değişken sürekli ve tek modlu ise Normal dağılış göstermese de bazı dönüşümlerle dağılışı Normal dağılışa yaklaştırmak mümkün olabilir. Bunun için √X, log(X), 1/X gibi dönüşümler kullanılabilir.

1809 tarihinde Alman matematikçi Gauss tarafından bulunmuştur.





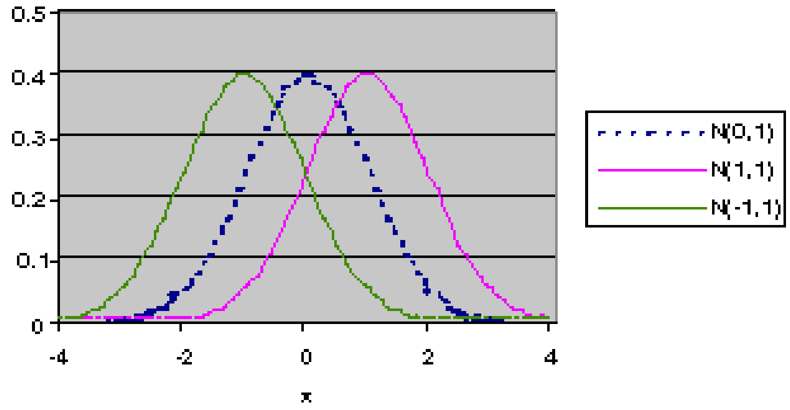
- *Normal dağılım simetrik bir dağılımdır. Bu simetriklik sağa veya sola çarpık olacak şekilde bozulabilir. Sağda uzun kuyruk varsa pozitif çarpık, solda uzun kuyruk varsa negatif çarpık olur.*
- *Normal dağılış eğrisi simetrik bir eğridir.*



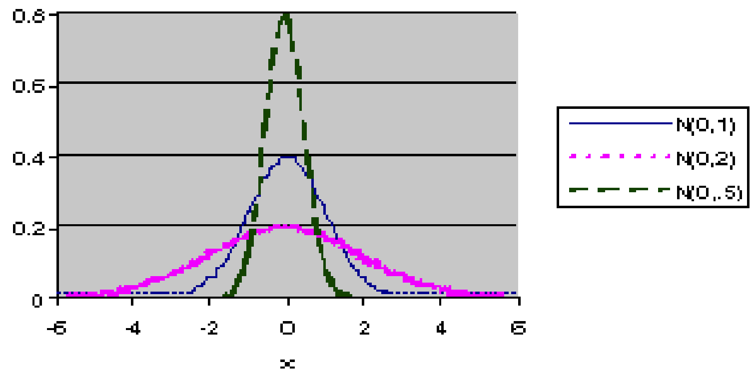
X ~N(μ, σ²) ise Olasılık Fonksiyonu

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_i-\mu}{\sigma}\right)^2} \quad , -\infty < X_i < +\infty$$

Ortalaması Farklı Normal Dağılışlar



Variyansı Farklı Normal Dağılışlar

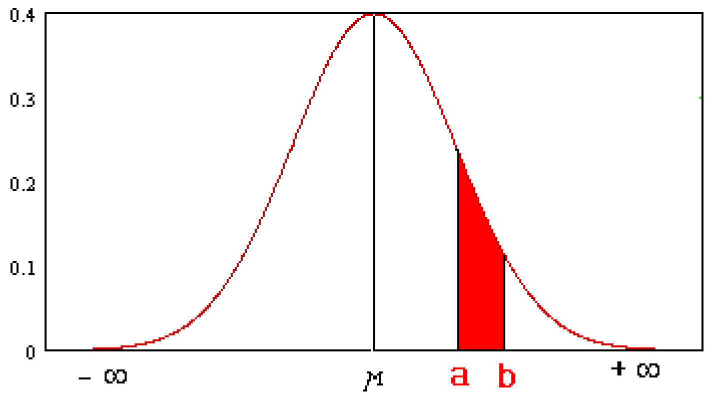


$$F(X) = \int_{+a}^{+b} f(x) \cdot dx$$

- Bu eğrinin altında kalan alan 1'e eşittir. Normal dağılışla ilgili olasılıklar hesaplanırken integral almak gerekir.

$$F(X) = \int_{-\infty}^{+\infty} f(x) \cdot dx = 1$$

- X' in a ile b arasında olma olasılığı:



- İntegrali her μ ve her farklı σ^2 için almak oldukça zordur. Bu nedenle X değişkeni standardize edilerek Standart Normal dağılış elde edilir.
- X'i standardize etmek için her X değerinden ortalaması çıkarılarak standart sapmasına bölünür.

$$Z_i = \frac{X_i - \mu}{\sigma} \approx N_Z(\mu = 0, \sigma^2 = 1)$$

- Bu değişkenin olasılık fonksiyonu:

$$f(Z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{Z^2}{2}} \quad , -\infty < Z_i < +\infty$$

Örnek 6.1.

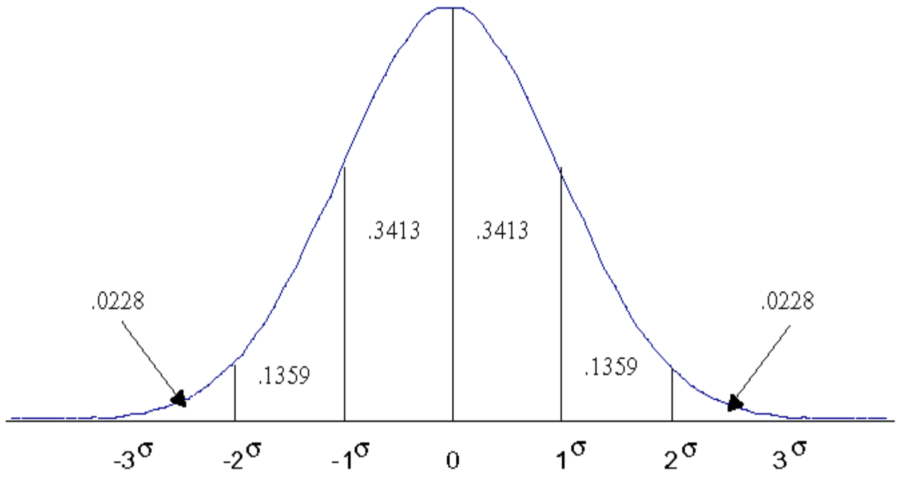
1. P(1.96<Z<+∞)=P(Z>1.96)= 0.025 olduğu görülür.
2. P(Z<-1.96) =?, simetri nedeniyle,
   P(Z<-1.96)=P(Z>1.96)= 0.025 olur.
3. P(-1.96<Z<1.96)= 1- [P(Z<-1.96)+P(Z>1.96)]
   =1 – 2*P(Z>1.96)= 1- 2*0.025= 1- 0.05= 0.95 olur.

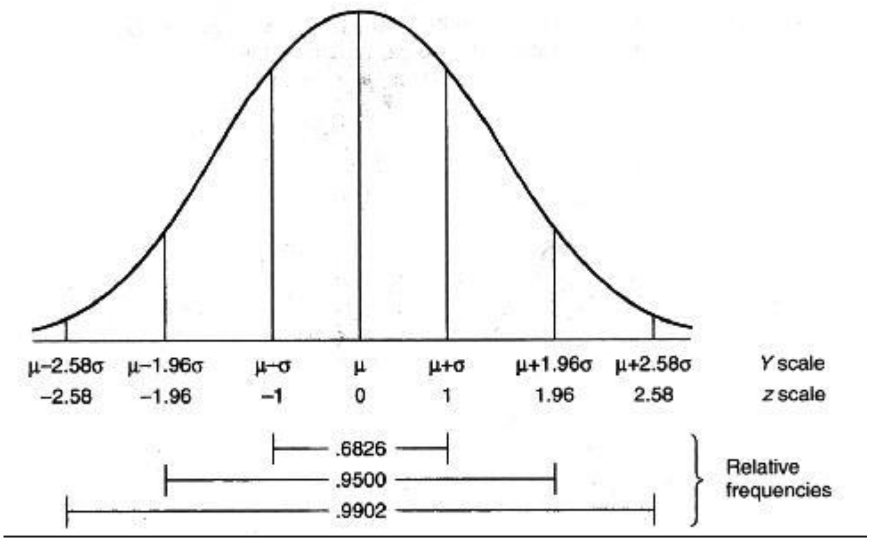
- P(Z<1.96)=1-P(Z>1.96)=1-0.025=0.975
- P(1.2<Z<2)=P(Z<2)-P(Z<1.2)=0.9772-0.8849=0.0923
- P(Z<-0.48)= P(Z>0.48)= 0.3156
- P(-2<Z-1.5)=P(1.5<Z<2)=P(Z<2)-P(Z<1.5)
  =0.9772-0.9332=0.044
- P(Z>-1.85)= 1- P(Z>1.85)=1-0.0322=0.9678

veya P(Z<1.85)=0.9678

- P(- 0.56<Z<0.84)= 1- P(Z<-0.56)-P(Z>0.84)
  =1-P(Z>0.56)-P(Z>0.84)
  = 1-0.2877-0.2005=0.5118

- Normal dağılışın 1,2 ve 3 standart sapma değerleri





Standart normal bölünme eğrisinin verildiği grafik incelendiğinde Z±1, Z±2 ve Z±3 apsislerine karşı gelen ordinatlar arasında kalan alanlar, bütün alanın sırasıyla %68.27 , %95.45 ve %99.73 e eşittir. Bu alanlar aşağıdaki gibi bulunmuştur:

$$P(-1 < Z < 1) = P(-1 < Z < 0) + P(0 < Z < 1) = 0,34134 + 0,34134 = 0,6827$$

$$P(-2 < Z < 2) = P(-2 < Z < 0) + P(0 < Z < 2) = 0,47725 + 0,47725 = 0,9545$$

$$P(-3 < Z < 3) = P(-3 < Z < 0) + P(0 < Z < 3) = 0,49865 + 0,49865 = 0,9973$$

Ayrıca bütün birimlerin %68,27'si ($\mu \pm \sigma$) aralığında, %95,45'i ($\mu \pm 2\sigma$) aralığında ve %99,73'ü ($\mu \pm 3\sigma$) aralığındadır.

Olasılık Biliniyorsa Z değerinin Bulunması

P(Z> a)= 0,063 ise a=?,

Tablonun içinden 0.063 bulunarak buna karşılık gelen sol kenardaki ve üstteki değer okunur. a=1,53 olur.

- X~N(5,9) ise P(X>a)=0,1587 a=?, önce standart Normal'e dönüşüm yapılır. Sonra tablo kullanılır.

Z		0.03
		↑
1.5	←	0.063

$$P\left(\frac{X\text{-}\mu}{\sigma} < \frac{a-\mu}{\sigma}\right) = P\left(Z > \frac{a-5}{3}\right) = 0{,}1587$$

olasılığı 0.1587 olan Z tablo değeri 1' dir. Bu nedenle $\frac{a-5}{3}=1$ çözülürse, $a = 3+5 = 8$ olur

Örnek 6.2. Klinik semptomlara göre felcin sebebini ortaya koymak zordur. Bunun için anjio (angiogram) yapmak gerekir. Bunun da belirli riskleri vardır. Bu test yerine beyindeki CBF (cerabralbloodflow) değerini ölçmek teşhise yardımcı olmaktadır. Toplumda CBF in ortalamasının 75 ve standart sapmasının 17 olduğu bilinmektedir. CBF değeri 40 altına düşerse tehlikeli bölgeye girilmiş demektir. Bu toplumdan tesadüfen çekilen bir şahsın 40' ın altında CBF değerine sahip olma olasılığı nedir?

X~ N(75, 17^2) olduğuna göre, P(X < 40)=?

P(X<40)= P[Z<(40-75)/17]=P(Z<-2.06] =P[>2,06]

= 0,0197

yani tesadüfen çekilen şahsın CBF inin 40 ın altında bulunması olasılığı % 1,97 dir.

Örnek 6.3. Bir şahsın DBP (Diastolic(küçük) kan basıncı) 90 ile 100 mm Hg civarında ise bu kişi yüksek tansiyon sınırındadır denir. 35-44 yaş grubundaki erkeklerde DBP değerleri Normal dağılış gösterir, ortalaması 80 ve variyansı 144 dür. Bu toplumdan tesadüfen seçilen bir şahsın yüksek tansiyon sınırında olma olasılığı nedir?

X: (DBP değerleri) ~N(80, 144), P(90<X<100)=?

$$P(90 < X < 100) = P(\frac{90-80}{\sqrt{144}} < \frac{X-\mu}{\sigma} < \frac{100-80}{\sqrt{144}})$$

$$P(\frac{10}{12} < Z < \frac{20}{12}) = P(0.83 < Z < 1.67)$$

$$P(Z > 0.83) - P(Z > 1.67) = 0.2033 - 0.0475$$

$$= 0.1555, \text{ yani } \%15.55 \text{ dir.}$$

Örnek 6.4."Glaucoma" göz içi basınçla ilgili bir hastalıktır. Toplumda göz içi basıncın değerinin 16 mm Hg ortalama ve 3 mm Hg standart sapma ile Normal dağılış gösterdiği bilinmektedir. Eğer basınç değerleri 12 mm Hg ile 20 mm Hg arasında ise sağlıklı kabul edilmektedir. Buna göre Toplumun % kaçı bu değerler arasında bulunmaktadır?

X~ Normal dağılış, μ=16, σ=3, X~ N(16,9)
P(12 $\leq$ X $\leq$ 20)=P[(12-16)/3 $\leq$ (X- μ)/ $\sigma \leq$ [(20 -16)/3]
P[-4/3 $\leq$ Z $\leq$ 4/3]= P(-1,33 $\leq$ Z $\leq$ 1,33)
= 1-2. P(Z>1,33)= 1- 2*0,0918=0,8164
yani %81,64 sağlıklıdır.

Örnek 6.5. İstatistik final sınavında aritmetik ortalama 72, standart sapmada 15 bulunmuştur. a) 60 b)93

puan alan öğrencilerin standart puan derecelerini bulunuz?

$$Z = \frac{X-\mu}{\sigma} = \frac{60-72}{15} = -0,8 \qquad Z = \frac{93-72}{15} = 1,4$$

Örnek 6.6. Bir fabrikada üretilen akülerin dayanma süreleri normal bölünmekte olup, ortalama dayanma süresi 800 saat ve standart sapma 72 saat olarak hesaplanmıştır. Üretimden alınacak herhangi bir akünün 764-872 saat arasında dayanma olasılığını bulunuz?

$$Z = \frac{X-\mu}{\sigma} = \frac{764-800}{72} = -0,50 \qquad P(-0,50 < Z < 0) = 0,1915$$

$$Z = \frac{872-800}{72} = 1,00 \qquad P(0 < Z < 1) = 0,3413$$

$$P(-0,50 < Z < 1) = P(-0,50 < Z < 0) + P(0 < Z < 1)$$
$$= 0,1915 + 0,3413 = 0,5328$$

Örnek 6.7. 300 öğrenciye uygulanan zeka testi sonuçları normal dağılım göstermektedir. Puanların ortalaması 100 ve varyansı 144 ise;

a) 105 puandan aşağı olan öğrenci sayısını,

b) 90 ile 110 puan arasında not alan öğrenci sayılarını bulunuz?

a) $$Z = \frac{X-\mu}{\sigma} = \frac{105-100}{12} = 0.42$$

$$P(Z < 0.42) = 0.6628 \qquad 0,6628 \times 300 = 199 kişi$$

b) $$Z_{alt} = \frac{X-\mu}{\sigma} = \frac{90-100}{12} = -0,83 \quad Z_{ust} = \frac{110-100}{12} = 0,83$$

$$P(-0,83 < Z < 0.83) = 0,5934 \qquad 0,0,5934 \times 300 = 178 kişi$$

Örnek 6.8. 300 öğrenciye uygulanan zeka testi sonuçları normal dağılım göstermektedir. Puanların ortalaması 100 ve varyansı 144 ise;

a) 105 puandan aşağı olan öğrenci sayısını,

b) 90 ile 110 puan arasında not alan öğrenci sayılarını bulunuz?

a) $$Z = \frac{X-\mu}{\sigma} = \frac{105-100}{12} = 0.42$$

$$P(Z < 0.42) = 0.6628 \qquad 0,6628 \times 300 = 199 kişi$$

b) $$Z_{alt} = \frac{X-\mu}{\sigma} = \frac{90-100}{12} = -0,83 \quad Z_{ust} = \frac{110-100}{12} = 0,83$$

$$P(-0,83 < Z < 0.83) = 0,5934 \qquad 0,0,5934 \times 300 = 178 kişi$$

Örnek 6.9. Bir gazetecinin her gün dağıttığı gazeteler için sarf etmiş olduğu zaman (dk) normal dağılım göstermekte olup, günlük ortalama zaman sarfiyatı 12 dk ve standart sapması ise 2 dk.dır. Buna göre 1 yıl boyunca gazete dağıtımının;

a) 17 dk. dan fazla sürdüğü gün sayısını,

b) 10 dk. dan az sürdüğü gün sayısını,

c) 9 ile 13 dk. arasında gerçekleşen gün sayısını

bulunuz?

a) $Z = \dfrac{X-\mu}{\sigma} = \dfrac{17-12}{2} = 2,5$

$P(X > 17) = P(Z > 2,5) = 0.0062 \qquad 0,0062 \times 365 \sim 2\,gün$

b) $Z = \dfrac{10-12}{2} = -1$

$P(X < 10) = P(Z < -1) = P(Z > 1) = 0,1587 \qquad 0,1587 \times 365 \sim 58\,gün$

$Z_{alt} = \dfrac{9-12}{2} = --1,5 \quad Z_{ust} = \dfrac{13-12}{2} = 0,5$

$P(9 < X < 13) = P(-1,5 < Z < 0.5) = 0,6247 \qquad 0,6247 \times 365 = 228\,gün$

Örnek 6.10. Bir tarlada yetiştirilen buğday bitkisinin uzunlukları normal dağılım göstermekte olup, ortalaması µ ve standart sapması 6 cm dir. Aynı zamanda ürünlerin %4,78'inin 82 cm den büyük olduğu bilinmektedir. Buna göre yığın ortalamasının kaç cm olduğunu tahmin ediniz?

$X \sim N(\mu, 36) \quad ve \quad P(X > 82) = 0,0478$

$P(X > 82) = P\left(\dfrac{X-\mu}{\sigma} > \dfrac{82-\mu}{6}\right) = 1 - P(Z < \dfrac{82-\mu}{6}) = 0,0478$

$P(Z < \dfrac{82-\mu}{6}) = 0,9522 \qquad P(1,67) = 0,9522$

$\dfrac{82-\mu}{6} = 1,67 \Rightarrow \mu = 72\,cm$

Örnek 6.11. X sürekli değişkeni normal dağılıma sahip olup, ortalaması 100 ve varyansı σ^2 dir. P(X<106)=0,8849 olduğuna göre standart sapma değerini bulunuz**?**

$$X \sim N(100, \sigma^2) \quad ve \quad P(X < 106) = 0,8849$$

$$P(X < 106) = P\left(\frac{X-\mu}{\sigma} < \frac{106-100}{\sigma}\right) = 0,8849$$

$$P(Z < \frac{6}{\sigma}) = 0,8849 \qquad P(1,2) = 0,8849$$

$$\frac{6}{\sigma} = 1,2 \Rightarrow \sigma = 5$$

Soru. Bir tavuk çiftliğindeki yumurtaların ağırlıkları normal dağılım göstermektedir. Yapılan ölçümler neticesinde yumurtaların %5'inin 85 gr.dan büyük, %72'sinin ise 65 gr.dan küçük olduğu gözlenmiştir. Yumurta ağırlıklarının ortalamasını ve standart sapması bulunuz?

μ=53,86gr σ=18,87gr

6.1. PUAN DÖNÜŞÜMLERİ

Bir gözlem sonucunda elde edilen ve üzerinde herhangi bir düzenleme yapılmamış ölçme sonuçları ham veri ya da ham puan olarak isimlendirilir. Ham verilerin anlaşılması ve yorumlanması güçtür. Bu nedenler ölçme sonuçları üzerinde bazı düzeltme ve dönüşümler uygulanarak anlaşılması ve yorumlanması daha kolay bir şekle dönüştürülebilir. Puan dönüşümü bunlardan birisidir.

Puan dönüşümü; ham verilerin karakteristiği bilinen tipik puanlara dönüştürülmesi işlemidir.

i) Standart Z Puanı:

Veri analizi yaparken alınan verilerin hatasız biçimde karşılaştırılabilmesi için aritmetik ortalama ve standart sapmadan yararlanılır. Aritmetik ortalama ve standart sapmanın aynı olduğu gruplarda karşılaştırma yapmak kolaydır ancak Aritmetik ortalaması ve standart sapması farklı olan dağılımların aynı aritmetik ortalama ve standart sapmaya sahip dağılım haline dönüştürülmesi ve sağlıklı karşılaştırma yapılabilmesi için verilerin standartlaştırılması gerekir.

Z puan dönüşümü ham verilerin Z puanlarına dönüştürülmesinde kullanılan doğrusal bir puan dönüşümüdür. Z puanları evrende normal dağılım gösteren bir olasılık dağılımına sahiptir. Bu nedenle evrende süreklilik gösteren ve normal dağılıma sahip özelliklere uygulanabilir.

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

ii) **Standart T Puanı:**

İşlevi Z puanı ile aynıdır. Yani verileri belli bir standarda getirip karşılaştırmak için kullanılır. Z puanı ile farkı ise şudur: Z puanında 0 olarak kabul edilen aritmetik ortalama T puanında 50 kabul edilir Z puanında 1 kabul edilen standart sapma T puanında 10 kabul edilir. Bu düzenleme ile veriler Z puanındaki negatif ve kesirli olabilen ifadelerden kurtularak pozitif ve tam sayı olarak ifade edilebilir.

Bir grup verinin ya da puanın ortalama ve standart sapması bilindiğinde bu puanların her biri ayrı ayrı T puanına dönüştürülebilir.

T puanı şu şekilde bulunur:

$$T = 50 + \frac{10(x-\bar{x})}{s}$$

T puanları Z puanından farklı olarak örneklem ortalaması ve örneklem standart sapması ile tanımlanmaktadır. Bunun medeni T puanlarının genellikle küçük örneklemlere yönelik istatistiksel işlem ve analizlerde kullanılmasıdır.

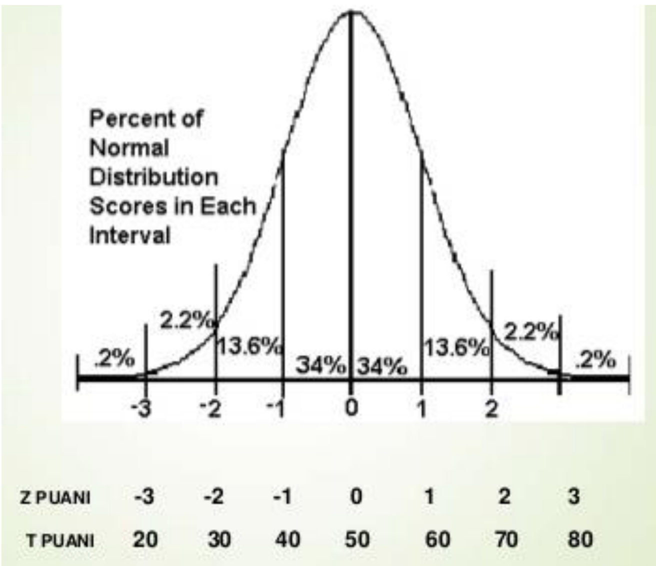
$$T = 50 + 10.Z$$

Z ve T puanları hipotezlerin belli güven aralıklarında doğru olup olmadığını anlamamıza da yardımcı olur. Çoktan seçmeli sorulardan oluşan Bilim Sınavı cevap kâğıtları ÖSYM'de optik okuyucu ile okunarak, adayların iki testin her birindeki sorulara verdikleri doğru ve yanlış cevaplar ayrı ayrı toplanacak, doğru cevap sayısından yanlış cevap sayısının dörtte biri çıkarılarak ham puanlar elde edilecektir. Bu ham puanlar, her test için ayrı olmak üzere, ortalaması 50, standart sapması 10 olan standart puanlara dönüştürülecektir. Standart puanlar kullanılarak, tıp fakültesi mezunu adaylar için Ağırlıklı Klinik Tıp Bilimleri Puanı (K) ve Ağırlıklı Temel Tıp Bilimleri Puanı (T) olmak üzere iki ayrı puan hesaplanacaktır. Tıp fakültesi dışındaki fakültelerden mezun adaylar için ise yalnız T Puanı hesaplanacaktır.

Z ve T puanları hipotezlerin belli güven aralıklarında doğru olup olmadığını anlamamıza da yardımcı olur.

H_0: Test edilen konu olay test konusu farklılık yaratmamıştır.

H_1: Test edilen şeyin önceki durum ile sonraki durum arasında fark yaratacağını ifade eder.



Soru:

Aşağıdaki tabloda 5 testin standart sapması ve aritmatik ortalaması ile bir öğrencinin bu testlerden aldığı puanlar verilmiştir.

Test	$\overline{X}$ (Testin ortalaması)	S_x (Standart sapma)	X (Öğrencinin notu)
1	18	2	27
2	12	3	26
3	24	3	30
4	32	4	34
5	26	4	36

i) Bu öğrencinin 3. Testteki Z puanı kaçtır.

$$Z = \frac{X - \overline{X}}{Sx} = \frac{30\text{-}24}{3} = 2$$

ii) Bu öğrencinin 3. Testteki T puanı kaçtır.

T=50+10Z
50+10*2 = 70

iii) Bu öğrenci hangi testte daha başarılıdır.

1. 4,5
2. 4,6
3. 2
4. 3
5. 2,5

en yüksek olan 2. test

6.2. Binom Dağılımına Normal Dağılım Yaklaşımı

Eğer X parametreleri n ve p olan Binom dağılışı gösteriyorsa, $P(a \leq X \leq b)$ olasılığı hesaplanırken, N(np, npq) normal dağılış eğrisinin (a-1/2) ve (b+1/2) bölgesi altında kalan alanla yaklaşık tahmini yapılabilir.

Eğer a=b olursa o zaman (a-1/2) ve (a+1/2) bölgesindeki alan hesaplanır.

Eğer $P(X \geq a)$=? Olasılığı sorulursa, düzeltme yapılırken, a<n.p ise a+1/2, değilse a-1/2 sınırı alınır.

Örnek 6.12. X~B(25,0.4), $P(7 \leq X \leq 12)$=? Bulunmak isteniyor. Normal yaklaşımla;

$$\mu = n.p = 25*0.4 = 10 \qquad \sigma = \sqrt{npq} = \sqrt{25*0.4*0.6} = 2.45$$

$$P(7 \leq X \leq 12) = P[(6.5-10)/2.45 \leq Z \leq (12.5-10)/2.45]$$
$$= P(-1.43 \leq Z \leq 1.02) = 1 - [P(Z>1.43)+P(Z>1.02)]$$
$$= 1 - 0.0764 - 0.1539 = 0.7679$$

Örnek 6.13.

Bir araştırmaya göre Ankara'da yaşayan erişkinlerin % 50'sinin en az bir kredi kartı bulunmaktadır. Bu gruptan rassal seçilen 30 erişkinden 19 tanesinde en az bir kredi kartı bulunması olasılığı nedir?

Bu soruda istenen olasılık değeri binom formülü kullanılarak bulunabilir. Çünkü örnekte,

$n = 30, \; p = 0.5; \; x = 19, \; q = 1 - p = 1 - 0.5 = 0.5$ ve $n - x = 30 - 19 = 11$

dir. Bu değerler formülde yerine konacak olursa,

$$P(x = 19) = \binom{30}{19}(0.5)^{19}(0.5)^{11} = 0.050$$

değeri elde edilir. Ayrıca bu örnekte $n\,p = 30\,(0.5) = 15$ ve $n\,q = 30\,(0.5) = 15$ olduğu için, istenen olasılık değeri normal dağılım yardımıyla da elde edilebilmektedir. Bu yaklaşımda üç aşama izlenmektedir.

Aşama 1. Binom dağılımı için ve 'nin hesaplanmasıdır.
Normal dağılımın kullanılabilmesi için dağılımın ortalama ve standart sapmasının bilinmesi gerekir. Binom dağılımı için bu parametreler,

$$= n\,.\,p = 30(0.5) = 15$$

$$= \sqrt{n\,.\,p\,.\,p} = \sqrt{(30)(0.5)(0.5)} = 2.7386$$

SÜREKLİLİK İÇİN DÜZELTME FAKTÖRÜ

Normal dağılımın binom dağılımına yaklaşımını sağlamak için n denemede x başarılı sonuç sayısına ± 0.5 değeri eklenir.

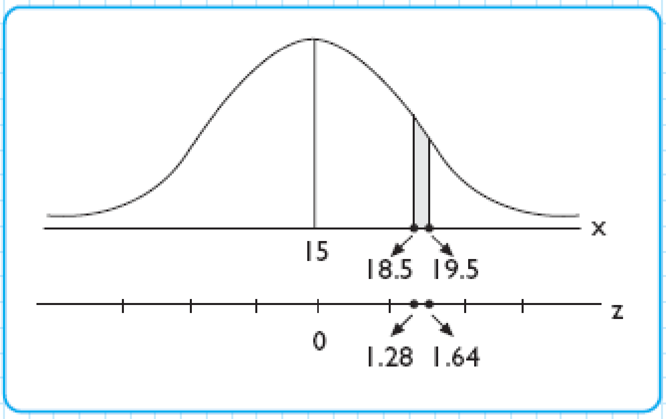
Verilen örnekte, x = 19 olduğu için, süreklilik düzeltmesi sonucunda 18.5 ve 19.5 değerleri elde edilir. Bu durumda, binom dağılımında P (x = 19)olasılık değeri yerine, normal dağılımda $P(18.5 \leq x \leq 19.5)$ olasılık değeri bulunacaktır.

Aşama 3. Normal dağılım kullanılarak istenen olasılığın hesaplanmasıdır.
Daha önceki örneklere benzer biçimde standart normal dağılım tablosundan yararlanabilmek için, x sınır değerlerine karşılık gelen z değerlerinin bulunması gerekir.

$$x = 18.5 \text{ için } z = \frac{18.5 - 15}{2.7386} = 1.28$$

$$x = 19.5 \text{ için } z = \frac{19.5 - 15}{2.7386} = 1.64$$

$P(0 \leq z \leq 1.64) = 0.4495$;
$P(0 \leq z \leq 1.28) = 0.3997$
$P(1.28 \leq z \leq 1.64) = 0.4495 - 0.3997 = 0.0498$



Normal dağılım yaklaşımı sonucunda elde edilen (yaklaşık) olasılık değeriyle binom formülünden elde edilmiş olan kesin olasılık değerleri arasında (0.0509 – 0.0498 = 0.0011) çok küçük bir fark bulunmaktadır ve bu fark da ihmal edilebilecek düzeydedir. Süreklilik düzeltmesi, hep normal dağılım yaklaşımının kullanılmasında uygulanmaktadır. Yukarıda eşitlik durumda süreklilik verilmişti. Ancak; bazen binom dağılımında istenen olasılık bir aralık olabileceği gibi, eşitsizlik durumları da olabilmektedir. Örneğin $P(7 \leq x \leq 12)$ olasılık değerinin normal dağılım yaklaşımında aranan olasılık değeri $P(6.5 \leq x \leq 12.5)$, $P(x \geq 9)$ için $P(x \geq 8.5)$ ve $P(x \leq 10)$ içinse $P(x \leq 10.5)$ olarak bulunmaktadır.

Örnek 6.14

Yapılan bir pazar araştırması neticesinde, çamaşır makinesi kullanan ev hanımlarından % 63'ünün yerli malı çamaşır makinesini tercih ettikleri bulunmuştur. Bu gruptan, rassal seçilen 100 ev hanımından, 55 – 60 tanesinin, yerli malı çamaşır makinesi tercih etme olasılığını bulunuz.

İki sonuçlu (binom) bu deneyde,

$$n = 100 \;; \quad p = 0.63 \,, \quad q = 1 - p = 1 - 0.63 = 0.37$$

dir ve istenen olasılık P (55 < x < 60) 'dir. Burada $n\,p > 5$ ve $n\,q > 5$ olması nedeniyle istenen olasılık değeri, normal dağılım yaklaşımıyla bulunabilir. Ancak; burada aranacak olasılık P (54.5 ≤ x ≤ 60.5) biçimindedir.
İlk olarak ve değerleri hesaplanacak olursa,

$$= n\,p = 100\,(0.63) = 63$$

$$= \sqrt{n\,p\,q} = \sqrt{100\,(0.63)\,(0.37)} = 4.8280$$

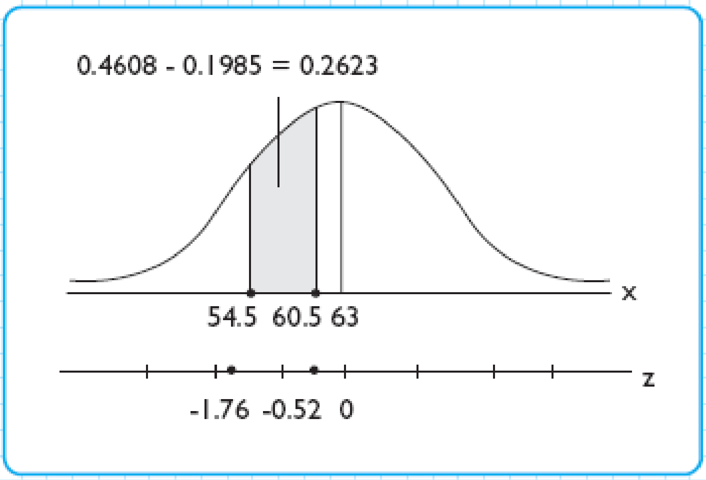
biçimindedir. Daha sonra gerekli z değerleri hesaplanır.

$$x = 54.5 \text{ için } z = \frac{54.5 - 63}{4.8280} = -1.76$$

$$x = 60.5 \text{ için } z = \frac{60.5 - 63}{4.8280} = -0.52$$

Bu değerlerin kullanımı sonucunda ortalamanın solunda yer alan iki alan bulunur ve büyük alandan küçük alanın çıkartılması sonucunda istenen olasılık değerine ulaşılır.

P (–1.76 ≤ z ≤ 0) = 0.4608
P (–0.52 ≤ z ≤ 0) = 0.1985
P (–1.76 ≤ z ≤ – 0.52) = 0.2623



Örnek 6.15.

18 yaşın üzerindeki nüfusu hedef alan bir kamuoyu araştırması sonucunda, milli piyangodan ikramiye çıkacağına inananların oranı % 54 olarak bulunmuştur. Bu kitleden rassal seçilen 100 kişiden 60 ya da daha fazla kişinin piyangodan ikramiye çıkacağına inanması olasılığını bulunuz.

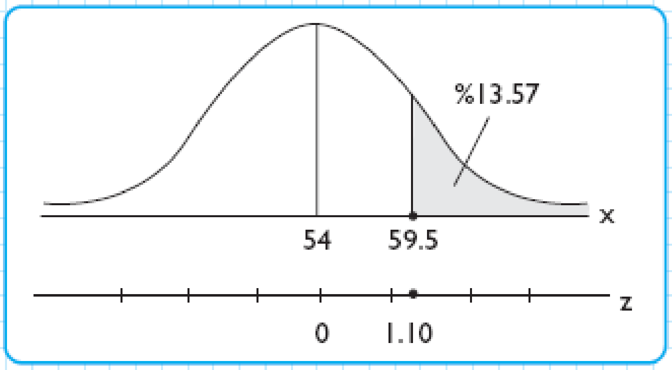
Yukarıdaki örneklerde olduğu gibi, binom deneyine uyan bu deney de, normal dağılım yaklaşımıyla çözülebilir.

$n = 100\ ;\ \ p = 0.54\ ,\ \ q = 1 - p = 1 - 0.54 = 0.46$

$= n\,p = 100\,(0.54) = 54$

$= \sqrt{n\,p\,q} = \sqrt{100\,(0.54)\,(0.46)} = 4.9840$

Bu değerlerden yararlanarak P (x ≥ 59.5)olasılık değeri standart normal dağılım tablosundan elde edilir.



$x = 59.5$ için $z = \dfrac{59.5 - 54}{4.9840} = 1.70$

$P\,(x \geq 59.5) = P\,(z \geq 1.10) = 0.5 - P\,(z < 1.10) = 0.5 - 0.3643 = 0.1357$

6.3. Poisson Dağılımına Normal yaklaşım

Eğer $X \sim P(\lambda)$ ise , $P(a \leq X \leq b)$ olasılığı hesaplanırken,
$X \sim N(\lambda, \lambda)$ normal yaklaşımı kullanılır. Genel de $\lambda >= 5$ olduğu durumlarda bu yaklaşım kullanılır.

Örnek 6.16. Bir kutusunda belli bir alandaki A bakterisi sayısı Poisson dağılımı göstersin. $\mu=10$ olsun. Tesadüfen seçilen bir kutudaki bakteri sayısının 20 veya daha fazla olması olasılığı nedir?

$X \sim P(10)$, bu normal yaklaşımla tahmin edilecekse Poisson dağılışı için $\mu=\sigma^2$ olduğundan

$Y \sim N(10,10)$ alınır.

$P[X \geq 20] = P[Y \geq 19.5] = P[Z \geq (19.5-10)/\sqrt{10}] = P(Z \geq 3.0) = 0.0013$

Örnek 6.17. Bir firma her hafta ortalama olarak belli bir üründen 42 tane sipariş almaktadır. Firmanın elinde hafta başında bu üründen 55 tane varsa, o hafta 55 üründen daha fazla alınması olasılığı
a) Poisson dağılımı ile b) Poisson dağılımının normale yaklaşımı ile bulunuz?

$$X \sim P(\lambda) \quad \lambda = 42$$

$$P(X \geq 55) = \sum_{x=55}^{\infty} \frac{e^{-42} 42^x}{x!} = 1 - \sum_{x=0}^{54} \frac{e^{-42} 42^x}{x!}$$

$$P(X \geq 55) 1 - P(X < 55) = 1 - P\left(Z < \frac{55,5-42}{\sqrt{42}}\right)$$
$$= 1 - P(Z < 2,08) = 1 - 0,9812 = 0,0188$$

Z TABLOSU

Areas Under the Normal Curve

Z	Cum p	Tail p	Z	Cum p	Tail p	Z	Cum p	Tail p	Z	Cum p	Tail p	Z	Cum p	Tail p
0.00	0.5000	0.5000	0.40	0.6554	0.3446	0.80	0.7881	0.2119	1.20	0.8849	0.1151	1.60	0.9452	0.0548
0.01	0.5040	0.4960	0.41	0.6591	0.3409	0.81	0.7910	0.2090	1.21	0.8869	0.1131	1.61	0.9463	0.0537
0.02	0.5080	0.4920	0.42	0.6628	0.3372	0.82	0.7939	0.2061	1.22	0.8888	0.1112	1.62	0.9474	0.0526
0.03	0.5120	0.4880	0.43	0.6664	0.3336	0.83	0.7967	0.2033	1.23	0.8907	0.1093	1.63	0.9484	0.0516
0.04	0.5160	0.4840	0.44	0.6700	0.3300	0.84	0.7995	0.2005	1.24	0.8925	0.1075	**1.64**	**0.9495**	**0.0505**
0.05	0.5199	0.4801	0.45	0.6736	0.3264	0.85	0.8023	0.1977	1.25	0.8944	0.1056	1.65	0.9505	0.0495
0.06	0.5239	0.4761	0.46	0.6772	0.3228	0.86	0.8051	0.1949	1.26	0.8962	0.1038	1.66	0.9515	0.0485
0.07	0.5279	0.4721	0.47	0.6808	0.3192	0.87	0.8078	0.1922	1.27	0.8980	0.1020	1.67	0.9525	0.0475
0.08	0.5319	0.4681	0.48	0.6844	0.3156	0.88	0.8106	0.1894	1.28	0.8997	0.1003	1.68	0.9535	0.0465
0.09	0.5359	0.4641	0.49	0.6879	0.3121	0.89	0.8133	0.1867	1.29	0.9015	0.0985	1.69	0.9545	0.0455
0.10	0.5398	0.4602	0.50	0.6915	0.3085	0.90	0.8159	0.1841	1.30	0.9032	0.0968	1.70	0.9554	0.0446
0.11	0.5438	0.4562	0.51	0.6950	0.3050	0.91	0.8186	0.1814	1.31	0.9049	0.0951	1.71	0.9564	0.0436
0.12	0.5478	0.4522	0.52	0.6985	0.3015	0.92	0.8212	0.1788	1.32	0.9066	0.0934	1.72	0.9573	0.0427
0.13	0.5517	0.4483	0.53	0.7019	0.2981	0.93	0.8238	0.1762	1.33	0.9082	0.0918	1.73	0.9582	0.0418
0.14	0.5557	0.4443	0.54	0.7054	0.2946	0.94	0.8264	0.1736	1.34	0.9099	0.0901	1.74	0.9591	0.0409
0.15	0.5596	0.4404	0.55	0.7088	0.2912	0.95	0.8289	0.1711	1.35	0.9115	0.0885	1.75	0.9599	0.0401
0.16	0.5636	0.4364	0.56	0.7123	0.2877	0.96	0.8315	0.1685	1.36	0.9131	0.0869	1.76	0.9608	0.0392
0.17	0.5675	0.4325	0.57	0.7157	0.2843	0.97	0.8340	0.1660	1.37	0.9147	0.0853	1.77	0.9616	0.0384
0.18	0.5714	0.4286	0.58	0.7190	0.2810	0.98	0.8365	0.1635	1.38	0.9162	0.0838	1.78	0.9625	0.0375
0.19	0.5753	0.4247	0.59	0.7224	0.2776	0.99	0.8389	0.1611	1.39	0.9177	0.0823	1.79	0.9633	0.0367
0.20	0.5793	0.4207	0.60	0.7257	0.2743	1.00	0.8413	0.1587	1.40	0.9192	0.0808	1.80	0.9641	0.0359
0.21	0.5832	0.4168	0.61	0.7291	0.2709	1.01	0.8438	0.1562	1.41	0.9207	0.0793	1.81	0.9649	0.0351
0.22	0.5871	0.4129	0.62	0.7324	0.2676	1.02	0.8461	0.1539	1.42	0.9222	0.0778	1.82	0.9656	0.0344
0.23	0.5910	0.4090	0.63	0.7357	0.2643	1.03	0.8485	0.1515	1.43	0.9236	0.0764	1.83	0.9664	0.0336
0.24	0.5948	0.4052	0.64	0.7389	0.2611	1.04	0.8508	0.1492	1.44	0.9251	0.0749	1.84	0.9671	0.0329
0.25	0.5987	0.4013	0.65	0.7422	0.2578	1.05	0.8531	0.1469	1.45	0.9265	0.0735	1.85	0.9678	0.0322
0.26	0.6026	0.3974	0.66	0.7454	0.2546	1.06	0.8554	0.1446	1.46	0.9279	0.0721	1.86	0.9686	0.0314
0.27	0.6064	0.3936	0.67	0.7486	0.2514	1.07	0.8577	0.1423	1.47	0.9292	0.0708	1.87	0.9693	0.0307
0.28	0.6103	0.3897	0.68	0.7517	0.2483	1.08	0.8599	0.1401	1.48	0.9306	0.0694	1.88	0.9699	0.0301
0.29	0.6141	0.3859	0.69	0.7549	0.2451	1.09	0.8621	0.1379	1.49	0.9319	0.0681	1.89	0.9706	0.0294
0.30	0.6179	0.3821	0.70	0.7580	0.2420	1.10	0.8643	0.1357	1.50	0.9332	0.0668	1.90	0.9713	0.0287
0.31	0.6217	0.3783	0.71	0.7611	0.2389	1.11	0.8665	0.1335	1.51	0.9345	0.0655	1.91	0.9719	0.0281
0.32	0.6255	0.3745	0.72	0.7642	0.2358	1.12	0.8686	0.1314	1.52	0.9357	0.0643	1.92	0.9726	0.0274
0.33	0.6293	0.3707	0.73	0.7673	0.2327	1.13	0.8708	0.1292	1.53	0.9370	0.0630	1.93	0.9732	0.0268
0.34	0.6331	0.3669	0.74	0.7704	0.2296	1.14	0.8729	0.1271	1.54	0.9382	0.0618	1.94	0.9738	0.0262
0.35	0.6368	0.3632	0.75	0.7734	0.2266	1.15	0.8749	0.1251	1.55	0.9394	0.0606	1.95	0.9744	0.0256
0.36	0.6406	0.3594	0.76	0.7764	0.2236	1.16	0.8770	0.1230	1.56	0.9406	0.0594	**1.96**	**0.9750**	**0.0250**
0.37	0.6443	0.3557	0.77	0.7794	0.2206	1.17	0.8790	0.1210	1.57	0.9418	0.0582	1.97	0.9756	0.0244
0.38	0.6480	0.3520	0.78	0.7823	0.2177	1.18	0.8810	0.1190	1.58	0.9429	0.0571	1.98	0.9761	0.0239
0.39	0.6517	0.3483	0.79	0.7852	0.2148	1.19	0.8830	0.1170	1.59	0.9441	0.0559	1.99	0.9767	0.0233

Z	Cum p	Tail p	Z	Cum p	Tail p	Z	Cum p	Tail p	Z	Cum p	Tail p
2.00	0.9772	0.0228	2.40	0.9918	0.0082	2.80	0.9974	0.0026	3.20	0.9993	0.0007
2.01	0.9778	0.0222	2.41	0.9920	0.0080	2.81	0.9975	0.0025	3.21	0.9993	0.0007
2.02	0.9783	0.0217	2.42	0.9922	0.0078	2.82	0.9976	0.0024	3.22	0.9994	0.0006
2.03	0.9788	0.0212	2.43	0.9925	0.0075	2.83	0.9977	0.0023	3.23	0.9994	0.0006
2.04	0.9793	0.0207	2.44	0.9927	0.0073	2.84	0.9977	0.0023	3.24	0.9994	0.0006
2.05	0.9798	0.0202	2.45	0.9929	0.0071	2.85	0.9978	0.0022	3.25	0.9994	0.0006
2.06	0.9803	0.0197	2.46	0.9931	0.0069	2.86	0.9979	0.0021	3.26	0.9994	0.0006
2.07	0.9808	0.0192	2.47	0.9932	0.0068	2.87	0.9979	0.0021	3.27	0.9995	0.0005
2.08	0.9812	0.0188	2.48	0.9934	0.0066	2.88	0.9980	0.0020	3.28	0.9995	0.0005
2.09	0.9817	0.0183	2.49	0.9936	0.0064	2.89	0.9981	0.0019	3.29	0.9995	0.0005
2.10	0.9821	0.0179	2.50	0.9938	0.0062	2.90	0.9981	0.0019	3.30	0.9995	0.0005
2.11	0.9826	0.0174	2.51	0.9940	0.0060	2.91	0.9982	0.0018	3.31	0.9995	0.0005
2.12	0.9830	0.0170	2.52	0.9941	0.0059	2.92	0.9982	0.0018	3.32	0.9995	0.0005
2.13	0.9834	0.0166	2.53	0.9943	0.0057	2.93	0.9983	0.0017	3.33	0.9996	0.0004
2.14	0.9838	0.0162	2.54	0.9945	0.0055	2.94	0.9984	0.0016	3.34	0.9996	0.0004
2.15	0.9842	0.0158	2.55	0.9946	0.0054	2.95	0.9984	0.0016	3.35	0.9996	0.0004
2.16	0.9846	0.0154	2.56	0.9948	0.0052	2.96	0.9985	0.0015	3.36	0.9996	0.0004
2.17	0.9850	0.0150	**2.57**	**0.9949**	**0.0051**	2.97	0.9985	0.0015	3.37	0.9996	0.0004
2.18	0.9854	0.0146	2.58	0.9951	0.0049	2.98	0.9986	0.0014	3.38	0.9996	0.0004
2.19	0.9857	0.0143	2.59	0.9952	0.0048	2.99	0.9986	0.0014	3.39	0.9997	0.0003
2.20	0.9861	0.0139	2.60	0.9953	0.0047	3.00	0.9987	0.0013	3.40	0.9997	0.0003
2.21	0.9864	0.0136	2.61	0.9955	0.0045	3.01	0.9987	0.0013	3.41	0.9997	0.0003
2.22	0.9868	0.0132	2.62	0.9956	0.0044	3.02	0.9987	0.0013	3.42	0.9997	0.0003
2.23	0.9871	0.0129	2.63	0.9957	0.0043	3.03	0.9988	0.0012	3.43	0.9997	0.0003
2.24	0.9875	0.0125	2.64	0.9959	0.0041	3.04	0.9988	0.0012	3.44	0.9997	0.0003
2.25	0.9878	0.0122	2.65	0.9960	0.0040	3.05	0.9989	0.0011	3.45	0.9997	0.0003
2.26	0.9881	0.0119	2.66	0.9961	0.0039	3.06	0.9989	0.0011	3.46	0.9997	0.0003
2.27	0.9884	0.0116	2.67	0.9962	0.0038	3.07	0.9989	0.0011	3.47	0.9997	0.0003
2.28	0.9887	0.0113	2.68	0.9963	0.0037	3.08	0.9990	0.0010	3.48	0.9997	0.0003
2.29	0.9890	0.0110	2.69	0.9964	0.0036	3.09	0.9990	0.0010	3.49	0.9998	0.0002
2.30	0.9893	0.0107	2.70	0.9965	0.0035	3.10	0.9990	0.0010	3.50	0.9998	0.0002
2.31	0.9896	0.0104	2.71	0.9966	0.0034	3.11	0.9991	0.0009			
2.32	0.9898	0.0102	2.72	0.9967	0.0033	3.12	0.9991	0.0009	3.60	0.9998	0.0002
2.33	**0.9901**	**0.0099**	2.73	0.9968	0.0032	3.13	0.9991	0.0009	3.70	0.9999	0.0001
2.34	0.9904	0.0096	2.74	0.9969	0.0031	3.14	0.9992	0.0008	3.80	0.9999	0.0001
2.35	0.9906	0.0094	2.75	0.9970	0.0030	3.15	0.9992	0.0008	3.90	1.0000	0.0000
2.36	0.9909	0.0091	2.76	0.9971	0.0029	3.16	0.9992	0.0008			
2.37	0.9911	0.0089	2.77	0.9972	0.0028	3.17	0.9992	0.0008			
2.38	0.9913	0.0087	2.78	0.9973	0.0027	3.18	0.9993	0.0007			
2.39	0.9916	0.0084	2.79	0.9974	0.0026	3.19	0.9993	0.0007			

Z Tablosu

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2297	0.2266	0.2236	0.2207	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1563	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1094	0.1075	0.1057	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0352	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0076	0.0073	0.0071	0.0070	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0042	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0014	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

Z	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998

KAYNAKLAR

1. Serper, Ö. (2000). Uygulamalı İstatistik I,II. Ezgi Kitabevi, Bursa.
2. Tekin, V.N. (2006). İstatistiğe Giriş, Seçkin Yayıncılık, Ankara.
3. Spigel, MR. and Stephens LJ. (1999). İstatistik, Schaum's Outlines, Çeviri editörleri: Alptekin Esin ve Salih Çelebioğlu, Nobel Dağıtım.
4. Çil, B.(2004). İstatistik, Detay Yayıncılık, Ankara.
5. Şenesen, Ü. (2002). Matematiksel İstatistik, Literatür Yayınları, İstanbul.
6. Akdeniz, F.(1984). Olasılık ve İstatistik, Ankara Ünv. Basımevi, Ankara.
7. Ünver, Ö. ve Gamgam, H.(2006). Uygulamalı istatistik Yöntemler, Seçkin yayıncılık, Ankara.
8. Çömlekçi, N.(1994). Temel İstatistik, Bilim Teknik Yayınevi, İstanbul.
9. Işık A. (2006). *Uygulamalı İstatistik I-II*, Beta Yayınları, İstanbul.