İSTATİSTİK-1

ORTAK DERS

DR. ÖĞR. ÜYESİ LEYLA İŞBİLEN YÜCEL

İSTANBUL ÜNİVERSİTESİ AÇIK VE UZAKTAN EĞİTİM FAKÜLTESİ
ORTAK DERS

İSTATİSTİK-1

ORTAK DERS

DR. ÖĞR. ÜYESİ LEYLA İŞBİLEN YÜCEL

ÖNSÖZ

Teknolojinin hayatımızın her alanına girmeye başlamasıyla her birimiz adeta veri ve bilgi bombardımanı altında kalmış bulunuyoruz. Üstelik bu bilgi kirliliği denilebilecek durum, ancak ve ancak bilginin doğru yönetilebilmesiyle aşılabilecek bir durumdur. Verinin doğru okunması, amaçlarımıza hizmet edip edemeyeceğinin ortaya konması, kısacası bu veri deryasından kendimize nasıl fayda sağlayacağımızı bize gösterecek olan başlıca kaynak istatistik okur yazarlığıdır. Dünyada istatistiğin kullanımının her geçen gün bireyler ve firmalar tarafından giderek daha da fazla kullanılması istatistiği yükselen bir değer haline getirmiştir. Örneğin ASA (American Statistical Association) 2013 yılını dünya İstatistik yılı ilan etmiştir. North Carolina Üniversitesi'nden Marie Davidian'ın 19 Aralık 2013'te JASA (Journal of American Statistical Association) dergisinde yayınlanan "The International Year of Statistics: A Celebration and A Call to Action" adlı makalesinde istatistiğin dünyada yükselen bir değer olmasından bahsetmektedir. İstatistiğin doğru kullanımı ve gücü hakkında son yıllarda en büyük katkıyı 2012 Amerika devlet başkanlığı seçimlerinden önce yaptığı tutarlı tahminlerle Nate Silver yapmıştır. "The Signal and the Noise: Why So Many Predictions Fail But Some Don't?" adlı kitabı bir anda en çok satanlar arasına girerek, tüm siyasilerin ve halkın ilgi odağı haline gelmiştir.

Ülkemizde çok uzak değil, bundan 10-15 yıl önce istatistik denilince akla, futboldaki özet sonuçlar veya sokakta yapılan anketler gelmekteydi. Oysa günümüzde tıpkı dünyada olduğu gibi istatistik hayatın her alanına girmiş ve etkin bir biçimde kullanılmaya başlanmıştır.

Yrd. Doç. Dr. Leyla İşbilen Yücel

İÇİNDEKİLER

ŞEKİLLER LİSTESİ

[1] http://upload.wikimedia.org/wikipedia/commons/1/16/Poisson_pmf.svg

1. İSTATİSTİĞE GİRİŞ, TEMEL KONULAR VE KAVRAMLAR, İSTATİSTİĞİN TARİHTEN GÜNÜMÜZE GELİŞİMİ

Bu Bölümde Neler Öğreneceğiz?

- İstatistik nedir?
- İstatistiğin temel konuları (betimsel ve çıkarımsal istatistik)
- İstatistiğin tarihçesi
- Tarihten günümüze önemli istatistik çalışmaları
- TÜİK hakkında genel bilgiler
- Günümüzde istatistik

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
İstatistiğe giriş ve temel konular, istatistiğin tarihsel gelişimi, ülkemizde istatistik kurumları	İstatistik biliminin tanıtılması, amacı ve kullanım alanları ve katkıları	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kelimeler

- İstatistik
- Tesadüfi olay
- Betimsel istatistik
- Çıkarımsal istatistik
- Status-devlet
- Bernoulli
- Gauss
- Laplace
- Moivre
- Adolphe Quotelet
- Karl Pearson
- TÜİK

1. 1. İstatistik Nedir?

Tanım 1: İstatistik, belirli bir amaç için veri toplama, tablo ve grafiklerle özetleme, sonuçları yorumlama, sonuçların güven düzeylerini açıklama, örneklemlerden elde edilen (hesaplanan) sonuçları kitle için genelleme (çıkarsama), karakteristikler arasındaki ilişkileri araştırma, çeşitli konularda geleceğe ilişkin projeksiyon yapma, deney tasarlama ve düzenleme v.b. kapsayan bir yöntemler topluluğudur.[2]

Tanım 2: İstatistik verileri işler ve özetler, araştırmacılara yol gösterir. Değişkenler arasındaki ilişkileri incelememize, tahminler ve öngörüler yapmamıza, doğru karar vermemize yarar.[3]

Tanım 3: Belirli amaç ya da amaçlar doğrultusunda gözlenen yığın olaylardan sayısal verilerin işlenerek, ilgili olayların oluşturduğu yığınların bilimsel olarak incelenmesinde kullanılan teknik ve yöntemler bilimi olarak tanımlanabilir.[4]

Tanım 4: İstatistik; gözlem ve ölçme sonucunda elde edilen sayısal (rakamsal) verilerin tablolar veya grafikler halinde sunulması ve bunların karar alma sürecinde nasıl kullanılacağına ilişkin yöntemler sunar. İstatistiğe konu olacak olaylar gözlemlenebilir ve ölçülebilir olmalıdır. İstatistikler belirli ölçümler yapılarak elde edilmişlerdir yani istatistikler rastgele bulunmuş sayılar değildirler. Belirli bir kitleyi oluşturan birimlerin gözlemlenmesiyle elde edilen, ölçülebilen veya sayılabilen nicel bilgileri ifade etmektedirler. Bu nedenle; logaritma cetveli, telefon rehberi, milli piyango sonuçları birer istatistik değildirler.

Tanım 5: İstatistik; olasılığa dayalı yani sonucun baştan kesin olarak bilinmediği rastlantısal (tesadüfî) durumlar üzerinde çalışırken, belirsizlik altında karar vermeyi sağlar.

Tanım 6: Verilerin derlenmesi, işlenmesi ve yorumlanmasına yönelik yöntemlerin tümü. (İşleme ve yorum verileri temsil edebileceği düşünülen kuramsal dağılım ve modellere geçişi de içermek üzere matematiksel istatistik kapsamına girer)[5]. İstatistiksel araştırmalara birkaç örnek verecek olursak, örneğin, bir araştırmacı, hamile iken diyetisyen kontrolünde iyi ve kaliteli beslenmenin doğacak bebeğin genel sağlık ölçümlerinde, düzensiz ve bilinçsiz beslenen bir annenin bebeğine kıyasla olumlu bir etkisi var mıdır? Fazla kilolu olmanın yaşam süresi üzerinde gerçekten olumsuz etkileri var mıdır? Sigara kanser ilişkisinin boyutu nelerdir? Kanser olmayı tetikleyen diğer faktörler neler olabilir ve risk sıralaması nasıldır?

---

[2] *(Özer Serper, Uygulamalı İstatistik 1 -2)*
[3] *(Necmi Gürsakal, Betimsel İstatistik, syf. 6)*
[4] *(Murat Açıköğretim Yayınları, İstatistik)*
[5] *(Büyük Larousse, Cilt: 10, Syf: 5872)*

1.2. İstatistiğin Temel Konuları

İstatistik konuları bakımından ikiye ayrılır; Betimsel (tasviri) istatistik ve Çıkarımsal istatistik.

1.2.1. Betimsel İstatistik

Verilerin toplanması, yorumlanıp özetlenmesi gibi konularla ilgilenir. Tanımsal istatistik, verilerin bir takım sayısal ve(ya) grafiksel yöntemlerin kullanımı ile bilgiye dönüştürülmesini ve karar verme sürecinde kullanılmasını sağlar.[6]

İstatistik, geçmişi ve içinde bulunulan durumu tanımlayarak bir veri kümesine ilişkin özet değerler ve grafikler ortaya koyduğunda betimsel istatistik (descriptive statistics) adını alır. Özetleme ve verilerin yoğunlaştırılması diğer bir deyişle verilerin hacimleri azaltılarak kullanım değerlerinin arttırılması betimsel istatistiğin konusuna girer. Tablolar ve grafikler yardımı ile verilerin özetlenmesi ve çok sayıda sayıdan oluşan bir veri grubunun "ortalama" gibi tek bir sayıya indirgenmesi yine bu alan için geçerlidir. Kısaca betimsel istatistik bir veri kümesinde bulunan bilgiyi sayısal ve grafiksel yöntemleri kullanarak özetler ve sunar.[7]

Betimsel amaçlı istatistik kitledeki tüm birimlerden ilgili değişken ya da değişkenler için veri toplandığında bunları kullanarak kitleyi özetlemeyi (betimlemeyi) amaçlar. Bu ise frekans dağılımı oluşturularak, grafikler çizilerek ya da parametreler (kitle ortalaması ve varyansı) hesaplanarak yapılır.[8]

1.2.2. Çıkarımsal İstatistik

Tümevarımsal amaçlı istatistik kitle(ler)den rastgele seçilen örneklem(ler)den toplanan verileri kullanarak kitle(ler) nin parametrelerini tahmin etmeyi veya parametrelerle ilgili olan savların doğru olup olmadığının araştırılmasını amaçlar. Yanlı örneklem gözlendiğinde belirsizlikler içeren kitle hakkındaki önermelere geçmek için geliştitilen süreçle istatistiksel sonuç çıkarır.[9]

Çıkarımsal istatistik, araştırma sürecinin betimsel istatistiğin bıraktığı yerden devam ettirilmesi suretiyle, çalışmanın amacına uygun bir şekilde daha ileri tekniklerle (olasılık başta olmak üzere) karar verme ve kitleye dair çıkarsamalar yapma sürecidir. Bu çıkarsamalar örnekleme dayalı olarak yapıldıkları için belirli bir hata düzeyinde ifade edilirler. (Kitle **N** adet birimden oluşan teorik topluluktur, aslında tüm birimlere ulaşmak olasıdır ancak zaman ve

---

[6] Onur Özsoy, İktisatçılar ve İşletmeciler İçin İstatistik, Excel Uygulamalı, 2010, Ankara, Syf: 4.
[7] Necmi Gürsakal, Betimsel İstatistik, Syf:20.
[8] Fikri Akdeniz, Olasılık ve İstatistik, Syf: 288.
[9] Fikri Akdeniz, Olasılık ve İstatistik, Syf: 288.

maliyet kısıdından dolayı genellikle istatistiksel çalışmalar kitleyi iyi bir biçimde temsil eden daha küçük alt gruplar yani örneklemler üzerinden yürütülür.

Örneklemin büyüklüğü **n** ile gösterilir. O halde N büyüklüğündeki bir kitleden n büyüklüğünde tane mümkün örneklem çekilebilir($C_n^N = \frac{N!}{(N-n)!\,n!}$). Bu örneklemler büyüklük yani hacim olarak birbirinin aynı fakat içerdikleri birimler bakımından birbirinden farklıdırlar, kombinasyonun doğası gereği seçilen birim yerine konmamaktadır, bu nedenle mümkün örneklem sayısı hesaplanırken permütasyon yerine kombinasyon kullanılmaktadır.

Örnekleme dayanan çıkarımlar mutlak doğru olamayacağından **"olasılık"** kelimesi kullanılarak ifade edilir. Çıkarımsal istatistiğin genel olarak kullandığı yöntem "tümevarım" dır. Önceleri betimsel amaçlar taşıyan istatistik, 17-18 ve 19. yy lar boyunca Bernoulli, Gauss, Laplace, Moivre gibi ünlü matematikçilerin olasılık teorisini geliştirmeleriyle çıkarımsal istatistiğin temellerini atmışlardır.[10]

****Her şey sabit, belirli ve kontrol altında olsaydı, yani değişkenlik olmasaydı istatistiğe gerek kalmazdı!!!*

Yukarıdaki cümle istatistiğin özünü ifade etmektedir. Her şey bir ihtiyacı gidermek için ortaya çıkıp kabul gördüğüne göre, istatistik de, gerçek hayatı olduğu gibi ele alabilmeyi sağlamıştır. Olasılık teorisinin katkılarıyla belirsizliği ölçebilmeyi, geleceğe dair tahminler ve çıkarımlar yapabilmeyi olanaklı kılmıştır. Yaşamın belirsizlikleri insanoğlunu her zaman 2 kere 2'nin dört etmeyeceği gerçeğiyle karşı karşıya bırakmıştır. Özellikle sosyal bilimlerde olayları salt pozitif bilimlerle açıklamak yetersiz kalmaktadır. İstatistik böyle durumlarda ölçümlemeyi sağlayabilen yöntem ve teknikler bütünü olarak öne çıkmaktadır.

1.3. İstatistiğin Tarihçesi

"İstatistik kelimesi Almanya'da Achenwall tarafından 1748'de kullanılmaya başlanmışsa da, istatistik çalışmaları çok eski çağlara kadar uzanır. Tacitus, imparatorluk zenginlikleri üzerinde Augustus'un geniş bir anket yaptığını yazar. Böylece askerler, gemiler her çeşit gelir kaynakları sayılmış ve kamu gelirleri tespit edilmiştir.

Bütün Ortaçağ boyunca ve 17. Yy'a kadar istatistik sadece tasviri olarak kalmıştır. Bu alanda iki okul ortaya çıktı: Göttingen Üniversitesi profesörü ve Almanlar tarafından istatistiğin babası olarak bilinen Achenwall'in temsil ettiği tasviri okul ve bazı sosyal olayların yaklaşık düzenliliğinden öngörüler ve kanunlar çıkarmayı deneyen siyasi matematikçiler okulu. 18. Yy'da Fransız Desparcieux ve İsviçreli Wargentin toplum olaylarını öngörmenin pratik önemini göseren ve çok gelişmekte olan sigorta sanayinin hareket noktasını teşkil eden ilk ölüm tablolarını (mortalite tablosu) düzenlemişlerdi.

Fransa'da ilk resmi istatistik bürosu 17. Yy'ın sonlarında kurulmuş ve içişleri bakanı Chaptal 1801'de genel nüfus sayımını gerçekleştirmişti. Türkiye'de ilk resmi istatistik bürosu 1933'te kuruldu.

[10] Özer Serper, Uygulamalı İstatistik 1- 2.

Jacques Bernoulli ve özellikle Laplace, tasviri sayısal bilgilere matematik bilgileri ilave ederek sonuçların değerlendirilmesinde olasılık hesap imkânlarının kullanılmasını araştırdılar. Birincisi ünlü "büyük sayılar kanunu" nu ortaya attı, ikincisi ise "analitik olasılık teorisi" adlı eserinde karmaşık sebepli doğal olayların incelenmesinde olasılık hesaplarından sağlanacak faydaları açıkça belirtti. Quotelet, çevresinde bütün diğer insanların yer alacağı hayali bir "ortalama insan" bulmak maksadıyla bu yöntemin uygulanışını insanların fiziki, ahlaki ve düşünsel özelliklerine uyguladı." [11]

"Quotelet bütün sosyal olayları birey etrafında toplamıştır. Daha ileri giderek istatistiği bütün sosyal alanlara hatta ahlaki olaylara da yaymıştır. İlk "sosyal kriminolog" olarak adlandırılmaktadır. 1836 yılında Quotelet, "Toplum suçu hazırlar, Suçlu ise ancak bir araçtır" demiştir. Bugün Quotelet modern istatistiğin kurucusu olarak kabul edilmekle birlikte, ileri sürdüğü sistem ve vardığı sonuçlar reddedilmektedir. Quotelet 'ten sonraki gelişmeler istatistik yöntembiliminde matematiğin daha yaygın bir şekilde uygulanmasını sağlamış ve matematiksel istatistik disiplininin meydana çıkmasına yol açmıştır.

İstatistik; Türklere de yabancı bir bilim dalı değildir. Selçuklular döneminde İran'da, İlhanlılar döneminde Hint'te nüfus sayımları düzenlenmişti. Öte yandan Osmanlı İmparatorluğu'nun yönetim sistemi nüfus ve arazi hakkında düzenli biçimde bilgi toplamasını gerektiriyordu. Bu nedenle imparatorluk zamanında 30-40 yıl gibi aralıklarla nüfus sayımları ve arazi yazımları yapılmaktaydı. Bu istatistiksel işlemler sırf vergi toplanması amacıyla yapıldığı için sadece vergi mükellefleri göz önünde bulunduruluyordu. İmparatorluğun gerileme devrinde bu sayımların arkası kesilmiş, modern tekniklere dayanan istatistikler de 20. Yy'e kadar ülkeye girememiştir. Gerçi bu gibi istatistiklerin düzenlenmesi yolunda 19. Yy'de bazı çalışmalar yapılmıştır. Örneğin orduyu modernleştirmek amacıyla askere alınabilecek erkeklerin sayısı öğrenilmek istenmiş ve böylece ilk sayım 1831'de gerçekleştirilmiştir. Ancak bu sayım yalnız erkekleri dikkate aldığı gibi ülkenin bütünü de kapsamamıştır. Nitekim sayım yapılan yerler Rumeli ve Anadolu sancakları ve kasabalarıdır. Dolayısıyla bu sayım bugünkü anlamıyla gerçek bir nüfus sayımı olmaktan uzak kalmıştır. 1844'te yapılan ikinci nüfus sayımının amacı ise kimlik belgesi verilecek vatandaşların belirlenmesi idi. Bu nedenle kadınlar da erkeklerle birlikte sayılmıştır. Ne var ki askerlik korkusuyla birçok erkek sayım dışında kalmıştır.

19. Yy'de ülke nüfusu hakkında bilgi verecek sonucu sayım 1874'te başlamış ve araya giren Rus savaşı nedeniyle 1884'e kadar sürmüştür. Nüfus kütüklerini düzenleme ve halka nüfus tezkeresi verme amaçlarıyla başlatılan bu sayımın uygulanmasında da bazı aksaklıklar görülmüştür. Birçok teknik hata içeren ve kısmi nitelikteki bu sayımlar yanında ölüm ve doğumlar gibi nüfus hareketlerini suçları ithalat ve ihracatı belirlemek amacıyla çalışmalar yapılmıştır. Ne var ki bu çabalar da sonuçsuz kalmıştır." [12]

---

[11] *Meydan Larousse, Cilt: 6, Syf: 511-512)*

[12] *(Özer Serper, Uygulamalı İstatistik 1 – 2)*

1.3.1. Tarihten Günümüze Öne Çıkan İstatistiksel Çalışmalar

- Antikçağ'da Çin'de, Mısır'da, Yunanistan'da nüfus ve bu nüfusun maddi yaşam koşullarına ilişkin sayısal verilere sahip olma gereksinimi izlerine rastlanmıştır.

- Roma'da düzenli nüfus sayımlarına rastlanır. Roma'da önce 5 yıl, sonra da 10 yıl ve 15 yıl gibi düzenli aralıklarla yapılan ve "census" adı verilen bu sayımlar 600 yıl kadar sürmüştür. Her Romalı kendi ve babasının adını, servetini, arazisini, kölelerinin sayısını bildirmek zorundaydı.[13]

- 1662'de Graunt erkek çocuklarının doğumlarının kız çocuklarının doğumlarına oranı gibi değişmezlikleri ortaya çıkardı.

- 1742'de Edmond Halley çağdaş aktüer (hayat sigortacılığı) çalışmalarının temelini oluşturan bir hayat tablosu yaptı.

- 1767'de Süssmilch erkek doğum oranı ve 20 yaşına dek erkek çocuk oranındaki gelişme üzerine önemli çalışmalar yayınladı.

- Pierre Simon De Laplace (1749-1827) "Çözümsel Olasılık Kuramı" (1812) adlı eserinde, nedenlerinin tümünü bilip tek tek çözümlenemeyecek kadar karmaşık olan doğal olayların incelenmesinde bu kuramdan sağlanacak yararları ortaya koydu. Önceleri betimsel amaçlar taşıyan istatistik, 17-18 ve 19. yy boyunca Bernoulli, Gauss, Laplace, Moivre gibi ünlü matematikçilerin olasılık teorisini geliştirmeleriyle çıkarımsal istatistiğin temellerini atmışlardır.[14]

- Adolphe Quotelet (1796-1874) yöntemin uygulama alanını, canlı varlıkların antropometrik, psikolojik ve toplumsal açıdan incelenmelerini sağlamıştır. Onun girişimiyle 1885'te Londra'da kurulan Uluslararası İstatistik Enstitüsü'nün öncüsü olan ilk uluslararası istatistik kongresi 1853'te Brüksel'de toplanmıştır.

- Karl Pearson (1857-1936) Biyoistatistik'i buldu. Bu yıllarda istatistik ile iktisadın arasındaki ilişkilerin fark edilmesiyle ekonometri doğmuş oldu. [15]

[13] Özer Serper, Uygulamalı İstatistik 1-2.
[14] A.g.e.
[15] *(Büyük Larousse, Cilt: 10, Syf: 5873)*

1.4. Türkiye İstatistik Kurumu Hakkında Genel Bilgiler

"Cumhuriyet döneminde istatistik işlerine büyük ölçüde önem verilmeye başlanmıştır. Nitekim 1926'da Merkezi İstatistik Dairesi kurulmuştur. Kurumun adı adeta yazboz tahtası yapılarak 1930'da İstatistik Umum Müdürlüğü, 1945'te İstatistik Genel Müdürlüğü, 1952'de İstatistik Umum Müdürlüğü, 1960'ta İstatistik Genel müdürlüğü şekline çevrilmiş ve 1962'de Devlet İstatistik Enstitüsü adında karar kılınmıştır. 2005 yılında ise Türkiye İstatistik Kurumu (TUİK) kurulmuş olup, kurum istatistik konseyi ve Türkiye İstatistik Kurumu başkanlıklarından oluşmuştur. TUİK Başkanlığı 5429 sayılı Türkiye İstatistik Kanunu'nun uygulanması sağlamak ve kanunla kendisine verilen görevleri yerine getirmek üzere kurulmuş olup, merkez ve taşra örgütlerine sahiptir. Merkez örgütü ana hizmet danışma ve yardımcı hizmet birimlerinden oluşmaktadır. Taşra örgütü ise bölgesel düzeydeki tüm istatistiksel faaliyetleri yürütmek ve yerel birimlerle koordinasyonu sağlamakla görevli bölge müdürlüklerini kapsamaktadır."[16]

"Enstitü gerekirse Ankara dışında geçici veya sürekli bürolar kurma ve mahalli muhabirler kullanma yetkisine sahiptir. Görevleri: her çeşit çalışmalara ilişkin istatistikleri düzenler, istatistikle ilgili olarak kamu kuruluşları arasında koordinasyon saplar, çözüm ve incelemeler yapar, düzenlenen istatistikleri yayınlar, istatistik yıllığı çıkarır, sayımlar ve anketler hazırlar ve sonuçlarını yayınlar, kamuoyu çalışmaları yapan bilimsel kurumların çalışmalarında yardım eder, yurt içi ve yurt dışı istatistik eğitim işleri yapar."[17]

Kurumun adı 2005 yılında Türkiye İstatistik Kurumu (TUİK) olarak değiştirilmiştir ve halen bu isimle anılmaktadır.

1.5. Günümüzde İstatistik

ASA (American Statistical Association) 2013 yılını dünya İstatistik yılı ilan etmiştir. North Carolina Üniversitesi'nden Marie Davidian'ın 19 Aralık 2013'te JASA (Journal of American Statistical Association) dergisinde yayınlanan "The International Year of Statistics: A Celebration and A Call to Action" adlı makalesinde istatistiğin dünyada yükselen bir değer olmasından bahsetmektedir.

Son yıllarda gazetelerde istatistikle ilgili başlıklar atılmıştır, birkaç örnek vermek gerekirse;

-"For today's graduate, just one word: Statistics", The New York Times, 2009.

-"The allure of the statistics field grows", The Boston Globe, 2012

-"What are the odds that stats would be this popular?", The New York Times, 2012.

-"The numbers guy", The Wall Street Journal, 2013.

[16] A.g.e.

[17] *Meydan Larousse, Cilt: 6, Syf: 512.*

Öne çıkan yazılar bunlar olmakla beraber, istatistik biliminin yükselişi çok sayıda makaleye konu olmuştur.

İstatistiğin doğru kullanımı ve gücü hakkında son yıllarda en büyük katkıyı 2012 Amerika devlet başkanlığı seçimlerinden önce yaptığı tutarlı tahminlerle Nate Silver yapmıştır. “The Signal and the Noise: Why So Many Predictions Fail But Some Don’t?” adlı kitabı bir anda en çok satanlar arasına girerek, tüm siyasilerin ve halkın ilgi odağı haline gelmiştir.

Ülkemizde çok uzak değil, bundan 10-15 yıl önce istatistik denilince akla, futboldaki özet sonuçlar veya sokakta yapılan anketler gelmekteydi. Oysa günümüzde tıpkı dünyada olduğu gibi istatistik hayatın her alanına girmiş ve etkin bir biçimde kullanılmaya başlanmıştır.

Bölüm Soruları

1. İstatistik nedir?
2. Betimsel istatistik nedir?
3. Çıkarımsal istatistik nedir?
4. İstatistiğin kullanım amaçları nelerdir?
5. İstatistiğin doğuşu ve tarihsel gelişimi, dünyadaki son durumu ile ilgili gelişmeler nelerdir?

2. İSTATİSTİKSEL ÇIKARSAMA SÜRECİ VE İSTATİSTİKTE TEMEL KAVRAMLAR

Bu Bölümde Neler Öğreneceğiz?

- İstatistiksel çıkarsama süreci
- İstatistikte temel kavramlar
- Değişken türleri
- İstatistikte hata türleri

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Çıkarımsal istatistik süreci, temel kavramlar ve hata türleri	İstatistiksel çıkarsamanın yapılışı ve temel kavramların öğrenilmesi	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

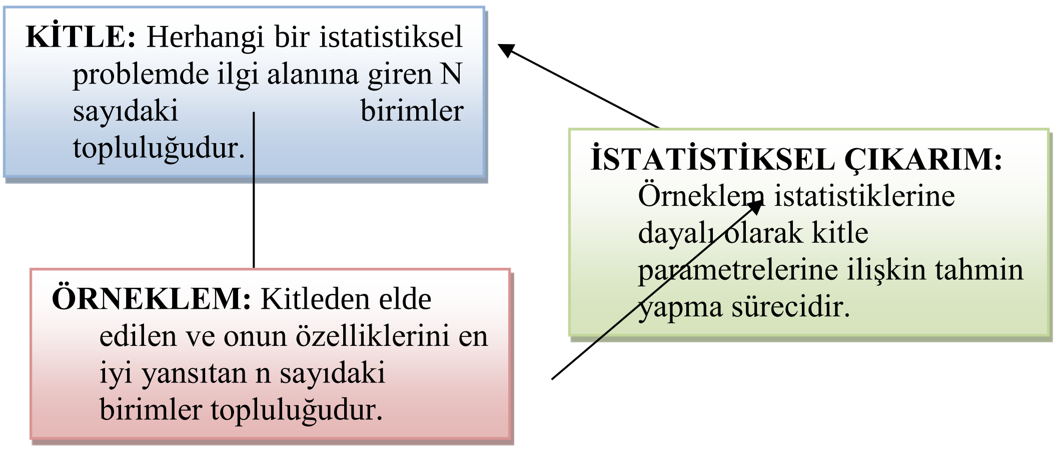
Anahtar Kelimeler

- İstatistiksel çıkarsama
- Kitle
- Parametre
- Örneklem
- İstatistik
- Birim
- Değişken
- Nominal veri
- Ordinal veri
- Rastsal hata
- Sistematik hata

2.1. İstatistiksel Çıkarsama Süreci

Örneklemden elde edilen bilgilere dayalı olarak kitleyle ilgili tahmin yapma ve karar verme sürecidir. İstatistiksel çıkarım istatistiğin temel amaçlarından birisidir. Yöneticiler veya karar verme mekanizmaları istatistiksel çıkarım yöntemleri ile aldıkları kararları ve geleceğe dönük olarak yaptıkları tahminleri test ederler. İstatistiksel çıkarım yöntemlerini kullanan yöneticiler firmanın gelecekte üretim düzeyinin ne olacağını, önceki dönemlere ait verileri inceleyerek ve mevcut piyasa koşullarını göz önünde bulundurarak tahmin edebilirler. Yapılan tahminlere dayalı olarak firmanın yönetim, pazarlama, tedarik ve reklam gibi fonksiyonları ve diğer stratejilerine ilişkin kararlar alınabilir. İstatistiksel çıkarım süreci aşağıdaki gibidir:

Şekil.2.1. İstatistiksel Çıkarım Süreci



2.2. İstatistikte Temel Kavramlar

Kitle:

Populasyon, yığın ve anakitle (anakütle) şeklinde de adlandırılan kitle, herhangi bir istatistiksel problemde ilgi alanına giren konu ile ilgili birimlerin tamamının oluşturduğu topluluktur. Büyüklüğü "N" ile gösterilir yani kitle N tane birimden oluşmaktadır.

Bir diğer tanım; kitle, hakkında bilgi edinilmek istenen ve biçimsel homojenliğe sahip (belirli bir tanıma uyan) kollektif olay niteliğindeki birimlerin oluşturduğu topluluğa denir.

Parametre:

Kitle ile ilgili ölçümlere parametre denir. Örneğin kitleye ilişkin varyans (σ^2), standart sapma (σ) ve ortalama (μ) birer parametredir. Bir kitle için bu gibi bilgiler tek bir değerden ibaret olduğundan, parametreler sabit bir sayı gibi düşünülebilirler. Parametrelerin hesaplanabilmesi için kitledeki her bir birimin hesaba alınması yani tamsayım yapılması gerekir. Ancak bu durum çoğu kez tercih edilmez, zira hem maddi olanaklar bakımından hem de zaman sorunu göz önüne alındığında, parametreler çoğu kez bilinmeyen, ancak tahmin edilen değerlerdir.

Örneklem:

Kitleden elde edilen ve kitlenin özelliklerini en iyi yansıtan "n" adet birimden oluşan birimler topluluğudur. Kitlenin olası alt kümelerinden her birine örneklem denir. Araştırmalarda kitleyi oluşturan bütün birimlere ulaşmak bazen imkânsız çoğu zaman ise masraflı olabilir. Bu kitleyi iyi temsil edebilen bir alt küme seçilerek incelenir. Bu alt kümeden hareketle kitle için tahminlerde bulunuruz. İşte kitlenin bu alt kümesine "örneklem" adı verilir. Ekonometri bölümü 1. Sınıf öğrencilerinin ağırlıklarını ölçmek istediğimizde, tüm öğrencilerin kilosunu öğrenmek yerine rastsal olarak bir örneklem çekerek de yaklaşık bir değere ulaşabiliriz. Şüphesiz, parametrenin yansız bir tahminin yapmak yani gerçek değerine çok çok yakın bir istatistik elde etmek için seçeceğimiz örneklemin kitleyi çok iyi temsil etmesi ve yeterli büyüklükte olması gerekmektedir. Söz gelimi, 500 öğrencinin ortalama ağırlığını tahmin etmekten bahsediyorsak, 10 öğrencinin ağırlığını ölçerek böyle bir tahmin yapmak uygun olmayacaktır. İlerleyen derslerde ideal örneklem büyüklüğünün nelere bağlı olarak belirlendiğini ve kaç olması gerektiğini öğrenmiş olacaksınız.

İstatistik (örneklem istatistiği):

Örneklemle ilgili tanımsal ölçüme denir. Örneğin örneklem varyansı, standart sapması ve örneklem ortalaması birer istatistiktir. Aynı kitle içinde bir örneklemden diğerine geçildiğinde istatistiklerin aldıkları değerler de değişmektedir. N büyüklüğündeki bir kitleden n hacimli mümkün tüm örnekler çekilse ki bu örneklemlerin sayısı N'in n'li kombinasyonu kadar sayıdadır, her biri için birer ortalama, varyans, vb. hesaplanabilir. Parametreler tek bir değer alırlarken, istatistikler örneklemden örnekleme değişen farklı değerler alabilmektedirler. Bu durumu kısaca bir örnekle açıklayalım. N=5 büyüklüğünde bir kitle olsun ve birimler aşağıdaki gibi olsun:

1 3 5 7 9

Kitle ortalaması, bütün bu birimlerin toplamının N'e yani 5'e bölünmesiyle elde edilir ve bu da 5'tir. Şimdi bu kitle ortalamasını örneklemler üzerinden tahmin etmeye çalışalım. n=3 büyüklüğünde mümkün tüm örneklemlerin sayısı 5'in 3'lü kombinasyonu kadardır yani 10 adettir.

Örneklemler	Örneklem Ortalaması
1 3 5	3
1 3 7	3,66
1 3 9	4,33
1 5 7	4,33
1 5 9	5
1 7 9	5,66
3 5 7	5
3 5 9	5,66
3 7 9	6,33
5 7 9	7

Örneklem ortalamalarının ortalaması yani mümkün tüm örneklemlerden hesaplanan örneklem ortalamalarının ortalamasını hesaplarsak 4,997 elde ederiz ki bu da kitle ortalamasına çok yakın bir değerdir. Bu örneği vermekteki amacımız parametrenin örneklemler yoluyla nasıl tahmin edildiğine dair çok temel ve basitçe bir açıklama ihtiyacından ileri gelmektedir. Gerçekte böylesi az sayıda birimden oluşan bir kitle söz konusu olduğunda zaten örnekleme yapılmaksızın her birime kolayca ulaşılabileceği için doğrudan doğruya parametreler hesaplanabilir. Bu örnekten anlamamız gerekenler şöyle özetlenebilir: N büyüdükçe kitle birimlerine ulaşmak maddi olarak ve de zaman kısıdı açısından tercih edilmez, örnekleme yapılarak parametre tahmini yapılır. Elbette mümkün tüm örneklemlerin çekilmesi söz konusu değildir, zaten böyle bir imkân olsa tüm kitleye ulaşılmış demektir ve tamsayım söz konusudur. Gerçekte, kitleye dair bir çıkarsama yapılacağı zaman bir tane örneklem çekilir. Ancak bu örneklemin çekilme yöntemleri ve hacminin ne kadar büyüklükte olması gerektiği hususları önemlidir. Bu konulara ileride ayrıntısıyla değineceğimizi belirterek, şimdilik basitçe şunları söylemekte yarar vardır.

Örneklemin kitleyi en iyi şekilde temsil etmesi gerekir, yani hanelerin aylık kira ödemeleriyle ilgili bir çalışma yapılacaksa, örneğin İstanbul için düşünürsek, sadece Beşiktaş, Florya, Şişli'yi örneklem olarak seçersek İstanbul'u temsil etmiş olmayız. Ya da okulların başarı oranlarının hesaplanmak istendiği bir araştırmada sadece gelişmiş şehir merkezlerindeki okulları örnekleme alıp kırsal kesimi yok sayarsak bu durumda da gerçekten oldukça uzak, taraflı sonuçlar elde etmiş oluruz. Örneklemin temsili olmasının yanında, büyüklüğü de önemlidir. Örneğin 10 bin nüfuslu bir semtte araştırma yaparken 100 kişilik bir örneklem kullanmamalıyız. Çünkü bu rakam kitleyi temsil etmekten uzaktır.

Örneklem büyüklüğünün hesaplanması için çalışmanın başında belirlenen, araştırmacı tarafından kararlaştırılan belirli bir hata düzeyine göre örneklem hacmi hesaplanmaktadır. Basitçe ifade etmek gerekirse, büyük bir kitlenin görece büyük bir örneklemle temsil edilmesi gerekir. Örneğin 500 büyüklüğündeki bir kitle için 30 hacimli bir örneklem yeterli sayılabilirken, 10 bin hacimli bir kitle için örneğin 500 birimlik örneklem gerekebilir. Bu bağlamda, ülkemizde yapılan eğilim anketlerine dair açıklamaları değerlendirirken, ankete kaç kişinin katıldığını yani n örneklem hacmini göz önüne almak ve sonuçlara ancak yeterli örneklem üzerinden ulaşıldı ise güven duymak gerekmektedir. Türkiye'de çeşitli konulara ilişkin yapılan araştırmaların sonuçları medyada sık sık yer almaktadır. Fakat genellikle sadece bulgulara yönelik açıklama yapılırken, örneklemin kaç kişiden oluştuğuna değinilmemektedir.

Örneğin toplumun yarısının bayanlardan oluştuğunu düşünürsek bu rakam kabaca 40 milyondur. 20 milyonu çocuklardan oluşsa geriye 20 milyon yetişkin bayan kalır. Bu durumda Türkiye'deki kadınlara ilişkin bir araştırma yapılacağını düşünürsek, söz gelimi 500 kişi 1000 kişi gibi örneklemler ile yapılan araştırmalara pek fazla itibar etmemek gerekmektedir. İstatistiksel araştırmanın temeli örneklemedir. Doğru, güvenilir ve yeterli veri yoksa en gelişmiş teknikler, modellemeler uygulansa dahi, kitleye dair doğru ve geçerli çıkarsama yapılmış olmaz. Gerçeğe yakın ve tutarlı sonuçlar elde etmek, doğrudan doğruya verinin kalitesine bağlıdır.

Birim:

Ölçülmeye ve (ya) sayılmaya elverişli olan canlı veya cansız tüm varlıklar istatistik birim olarak kabul edilebilir. Kişiler, hayvanlar, bitkiler, evler, arabalar, olaylar (evlenme, boşanma, ölüm, deprem, trafik kazaları, seller, çığ düşmeleri) v.b. hakkında ölçüm veya sayım yapılabileceği için istatistiksel birimdirler. Ancak, korku, hayal, rüya, v.b. ölçülemeyen, sayılamayan soyut varlıklar istatistiksel birim değildirler ve istatistiksel çalışmalara konu olamazlar.

Birimler 3 grupta incelenebilirler:[18]

Sürekli olanlar	**Doğal olanlar**	**Gerçek olanlar**
İnsan, bina, şirket, v.b.	*İnsan, kitap,v.b.*	*Ev,arsa, bisiklet,v.b.*
Ani olanlar	**Doğal olmayanlar**	**Varsayıma dayananlar**
doğum, evlenme, boşanma,v.b.	*arsa, pasta,v.b.*	*örneklem*

Değişken:

Birimlerin araştırmaya konu olan herhangi bir özelliğine değişken denir. Birimlerimiz her bir öğrenci olduğunda bunların boy ölçüleri, ağırlık ölçüleri birer değişkendir, öğrenciden öğrenciye değişir. Ailelerin gelir düzeyleri, sahip oldukları çocuk sayısı bir değişkendir. Değişkenlerin alabilecekleri her bir değere ise "şık" denir. Örneğin medeni hal bir değişken olarak ele alındığında evli, bekâr, dul, vb. değerlerinin her biri birer şıktır.

Değişkenler;

a. Sayısal (nicel, kantitatif) değişkenler

Boy, ağırlık, alan, hacim, vb.

b. Sayısal olmayan (nitel, kalitatif) değişkenler

[18] *(Murat Açıköğretim Yayınları, ss: 1-2)*

Cinsiyet, meslek, göz rengi, memleketi, vb. olarak ikiye ayrılır.

Sayısal değişkenler de;
a.Sürekli
b.Kesikli

olmak üzere ikiye ayrılır. Sürekli değişkenler ölçmeye tabidirler. Kesikli değişkenlerde ise sayma söz konusudur. Ailedeki çocuk sayısı, bir caddedeki evlerin sayısı, bir öğrencinin sahip olduğu kitapların sayısı, vb. bunlar hep sayılabilen yani kesikli değişkenlerdir. Değişkendir çünkü kişiden kişiye, aileden aileye farklı değerler almaktadır. Sürekli değişkenlere örnek verecek olursak, bir kutunun ağırlığı, bir insanın boy uzunluğu, bir arazinin yüz ölçümü, vb. sürekli değişkenlerdir. Özetle, eğer sayısal bir değişken sayılabiliyorsa kesiklidir, ölçülebiliyorsa süreklidir.

Nitel veriler kendi aralarında ikiye ayrılırlar; nominal veri ve ordinal veri.

Nominal veri; kategori ifade eder, Kategoriler arasında bir sıralama veya üstünlük yoktur. İstatistiksel araştırmalara sayısal kodlama yapılarak sayısallaştırılmak suretiyle dâhil edilirler. Örneğin;

Medeni hal: evli (1), bekâr (2), ayrılmış (3)
Göz rengi: siyah (1), yeşil (2), kahverengi (3),vb.
Cinsiyet: Kadın (1), erkek (2)
Meslek: öğretmen (1), bankacı (2), avukat (3), sanatçı (4), vb.

Ordinal veri; verinin sıralanması şeklindedir. Kategoriler arasında bir üstünlük sıralaması söz konusudur.

Örneğin;
Zayıf (1), orta (2), iyi (3), çok iyi (4)
Eğitim durumu: ilkokul (1), ortaokul (2), lise (3), üniversite (4), yüksek lisans (5), doktora (6)

Örnek:

a) Ekonometri bölümünde okuyan öğrencilerin evlerinden fakültelerine olan uzaklıklar

Bir ölçüm söz konusu olduğu için sürekli nicel veridir.

b) Bir okuldaki öğretmenlerin cinsiyetleri

Kategori belirten, nominal veridir.

c) 2014 yılında İstanbul'a düşen kar yağış miktarı
Bir ölçüm söz konusu olduğu için sürekli nicel veridir.
d) Son 5 yılda Tunceli-Erzincan karayoluna düşen çığ sayısı
Sayma işlemine dayandığı için kesikli nicel veridir.
e) Göz rengi
Kategori belirten, nominal veridir.
f) Bir evin tavan yüksekliği
Bir ölçüm söz konusu olduğu için sürekli nicel veridir.
g) Bir ilacın kana karışma süresi
Bir ölçüm söz konusu olduğu için sürekli nicel veridir.
h) Mezun olunan okul, askeri rütbeler, akademik ünvanlar
Sıralama söz konusu, ordinal veridir.

2.3. İstatistikte Hata Kavramı

Genel olarak iki tür istatistik hatadan bahsedilebilir;

1. Rastsal (tesadüfi) hata
2. Sistematik hata

Rastsal hata adı üzerinde, bilinçsizce yani herhangi bir kasıt olmadan yapılan hatalardır. Bu tür hatalar örneklem büyüdükçe artı ve eksi yönde birbirini götürerek etkisiz hale gelecektir. Bu nedenle rastsal hatalar genellikle göz ardı edilirler. Örneğin bir anket uygulamasında kişinin cinsiyeti "kadın" olduğu halde "erkek" şeklinde giriş yapılmış olabilir. Aynı anketör bir başka anket formunda tam tersi bir işaretleme yapmış olabilir veya bir başka anketör onun yaptığının tersi bir veri girişi yapmış olabilir. Yani bu ve benzeri bazen bir yönde bazen de tersi yönde yapılan hatalar **gözlem sayısı arttıkça denkleşerek toplamda göz ardı edilebilmektedirler**, böyle hatalara rastsal (tesadüfi) hata denir ve sakıncası yok denecek kadar azdır. 16. yy'da Bernoulli ve diğer matematikçiler tarafından yapılan "büyük sayılar kanunu", gözlem sayısı arttıkça olayların rastsal nedenlerin etkisinden kurtulduğunu ortaya koymaktadırlar.

Sistematik hata ise bazen kasıtlı olarak yapılan bazen ise kullanılan teçhizatın bozukluğundan kaynaklanabilen, hep aynı yönde seyreden ve örneklem büyüse de tolere edilemeyen hatta bu şekilde daha da artan yapıdaki hatalardır. Sonuçların objektifliğini etkileyeceği için **kesinlikle düzeltilmeleri gerekir**. Örneğin hükümet karşıtı bir anket uygularken soruları hep muhaliflere yöneltmek sonuçların yanlı ve tek yönlü çıkmasına neden olacaktır. Sistematik hatalar bazen de kullanılan teçhizatın bozukluğundan kaynaklanabilir. Örneğin bozuk bir klavyenin "h" harfini devamlı olarak "g" olarak basması, bir tartının ayarının bozuk olması nedeniyle hep 15 gr eksik tartması, vb. Verilen örneklerden de anlaşılacağı üzere, sistematik hatalar yapılan istatistiksel çalışmanın en başından kalitesiz olmasına neden olur. İstatistiksel çalışmaların en temel ihtiyacı öncelikle doğru ve sağlam verilerdir.

Bölüm Soruları

1. Örneklem ve Ana kütle arasındaki fark nedir?
2. Tamsayım nedir?
3. İstatistiksel çıkarım sürecini anlatınız.
4. İstatistik ve parametre kavramlarını açıklayınız.
5. Sürekli ve ani birimlere örnek veriniz.
6. Doğal ve doğal olmayan birimlere örnek veriniz.
7. Gerçek olan ve varsayıma dayanan birimlere örnek veriniz.
8. Nicel, Nitel, Kesikli, Sürekli değişkenlere birer örnek veriniz.
9. Nominal veri ne demektir? Örnek veriniz.
10. Ordinal veri ne demektir? Örnek veriniz.
11. İstatistikte kaç tür hata vardır? Örnek vererek açıklayınız.

3. VERİLERİN DERLENMESİ, DÜZENLENMESİ VE SUNUMU

Bu Bölümde Neler Öğreneceğiz?

- Veri kaynakları
- Verilerin elde edilmesi
- Anket
- Verilerin düzenlenmesi
- Grafiklerle sunum

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
İstatistiksel verinin elde edilmesi, düzenlenmesi ve grafifklerle sunumu	Betimsel istatistik çalışmalarının temelleri öğrenilecek	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kelimeler

- Birincil veri
- İkincil veri
- Veri kaynakları
- Anket
- Dizi
- Range (açıklık)
- Nokta diyagramı
- Çizgi diyagramı
- Histogram
- Sıklık poligonu
- Box-Whisker grafiği
- Daire (pasta) grafiği

3.1. Verilerin Derlenmesi

Veriler, istatistik çalışmanın temelini oluşturmaktadır. İstatistiksel bir analizin doğruluğu ve güvenilirliği öncelikle verinin kalitesine (çalışmanın amacına uygun ve doğru bir şekilde toplanmış güvenilir verilere) bağlıdır.

3.1.1. Veri Kaynakları

Verilerin elde edilme süreçleri çeşitlilik göstermektedir. Şayet veriyi direkt olarak kaynağından sağlama imkanı var ise böyle elde edilen verilere ***birincil veri*** denir. Çeşitli kurum ve kuruluşların (DPT-Devlet Planlama Teşkilatı, SGK, TOBB, v.b.) yayınladıkları haftalık, aylık, yıllık, v.b. bültenlerden elde edilen veriler de birincil veri kapsamındadır. Mesleki kuruluşlar (Mimarlar Odası, Makine Mühendisleri Odası, v.b.), medya kuruluşlarının ya da araştırma firmalarının yayınladıkları veriler ise daha ziyade ***ikincil veri*** kapsamındadır. Uluslar arası veri kaynaklarının başında ise OECD (Organization for Economic Cooperation and Develeopment), WB (World Bank), BM (UN: United Nations), IMF (International Money Fund) gelmektedir.

3.1.2.Verilerin Elde Edilmesi

İstatistiksel araştırmaların yapılma nedenleri genellikle şu şekilde sıralanabilir:

1. Durum tespiti
2. Geleceğe dönük projeksiyon yapma

Durum tespitinden kasıt, bir kurumun, firmanın, v.b. içinde bulunduğu durumun bilimsel olarak ortaya konmasıdır. Her geçen gün artan rekabet şartları, eskilerin tecrübe, basiret ve sezgiye dayalı yönetim anlayışlarını yetersiz kılmıştır. Artık yapılması gereken böylesi bir ortamda firmanın ne durumda olduğu, hangi şartlarda rekabet etmekte olduğu ve gelecekteki hedeflerine ulaşabilmesi için bugünden neler yapılması gerektiğini bilimsel olarak ortaya koymaktır. Bunun da olmazsa olmaz tek şartı elbette çeşitli ölçümler yapmaktır. Yani veri toplayıp istatistiksel analizler yapmaktır. Firmanın mevcut şartlarının incelenmesi ve ortaya konması durum tespiti, geleceğe dönük yapılması gerekenlerin ortaya konması ise, projeksiyon yapıldığı anlamına gelir.

Verinin toplanması denince ilk akla gelen ve en çok kullanılan yöntem ***anket*** yöntemidir. Anketlerin uygulanma şekilleri farklılık arz eder, örneğin anketörler vasıtasıyla yüzyüze görüşme yaparak olabileceği gibi, e-posta, posta v.b. yöntemlerle de anket formları göndermek suretiyle veri elde edilebilir. Ancak posta yoluyla yapılan anketlere büyük olasılıkla geri dönüş yapılmamaktadır veya en iyi ihtimalle yarım yamalak cevaplanmış, bir sürü cevapsız soru bırakılmış anket formlarıyla karşılaşılmaktadır. Bu nedenle en sağlam veri toplama şekli yüzyüze görüşme yaparak anket yapmaktır denebilir. Gelişmiş ülkelerde posta yoluyla anket cevpalama oranı oldukça yüksektir, her ne konuda olursa olsun bir anket muhakkak bir iyileştirme çabasına hizmet etmektedir, böyle düşünüldüğünde, anketi cevaplamanın ne kadar önemli olduğu anlaşılabilir. Ülkemizde ise bırakın postayla anket cevaplamayı, bazılarının, beş dakikalığına yolunu çeviren anketörleri bile azarladıklarını görmekteyiz. Halbuki artık siyasi partiler bile araştırma şirketlerine anketler yaptırarak halkın nabzını tutmakta, neredeyse seçim sonuçlarıyla aynı denecek yakınlıkta tespitler yapmaktadırlar. Bu sayede politik planlamalar daha gerçekçi yapılabilmekte, kaynaklar daha etkin (gereksiz israf olmadan) kullanılabilmektedir. Örneğin daha küçük bir örnek verelim, bir deterjan fabrikası anket yaptırıyor olsun. Bu ankete ne kadar fazla katılım olur ve dürüstçe cevaplar verilirse, firma yöneticisi de bu doğrultuda üretime geçecek, böylece milli servetimiz heba olmayacaktır. Bir anketten ne olur ki? dememek gerekir. Sonuçta gökten para yağmadığına göre, batan bir tek fabrika bile olsa, bu durum, sadece o fabrika sahibinin iflası gibi görülmemelidir. Orada çalışan binlerce insan işsiz kaldığında, onların aileleri, akrabaları, komşuları, v.b. zaman içinde dalga dalga bu durumdan etkileneceklerdir. O halde işini zamanında doğru düzgün yapmaya çalışan, halkın fikirlerine değer veren ve kaynaklarını etkin bir biçimde kullanmak isteyen herkese bunu bir yardım talebi gibi kabul ederek yardım etmeliyiz. Bakış açımız gerçekten de böyle olmalıdır, bu sadece basit bir örnektir ama binlercesi düşünülebilir.

Anketin öneminden bahsettikten sonra, anketin nasıl uygulanması gerektiğinden bahsedelim. Adım adım özetlersek;

1. Anket konusunun ve amacının belirlenmesi
2. Anketin uygulanacağı birimlerin belirlenmesi (genel olarak herkes mi, yoksa sadece bayanlar mı, ya da sadece 20-25 yaş arası gençler mi, v.b.)
3. Örnekleme yöntemine karar verilir. (bu konu örneklemenin konusudur ve çeşitli yöntemler vardır, örneğin basit rastgele örnekleme, sistematik örnekleme, tabakalı örnekleme, küme örneklemesi, vb.)
4. Çerçevenin belirlenmesi (çerçeve anket uygulanacak birimlerin listesi gibi düşünülebilir, böylece çalışmanın sınırları çizilmiş olur, örneğin anket bir ilçede uygulanacaksa her mahalleden kaç kişiye anket uygulanacağını belirlemek çerçeveyi çizmek anlamına gelir)
5. Anket yüzyüze görüşme yöntemi ile uygulanacaksa, bu işi en iyi ve doğru şekilde yapabilecek, konuyla ilgili bilgilendirilmiş anketörlerin seçilmesi. (örneğin inşaat malzemeleri üreten bir firma, yapı malzemelerinin tercihi konusunda anket yaptırmak istediğinde, malzemelerin özelliklerini ve ne işe yaradığını bilmeyen konudan bihaber anketörlere görev vermek sizce ne kadar doğru olur?)
6. Anketin sahada uygulanması ve verilerin toplanması. (anketin uygulanması başlı başına uzmanlık isteyen bir durumdur. Öyle ki, anketörün tutum ve tavrı, soru sorma stili, ses tonu, giydiği kıyafete kadar özenle tespit edilmelidir. Örneğin bir diş macunu firmasının anketinde, 32 dişi pırıl pırıl parlayan çok güzel dişlere sahip bir anketöre bu işi yaptırırsanız,

belki haftada bir diş fırçalayan biri, günde üç kez diş fırçalamakta olduğunu beyan edebilir, çünkü anketörün görünümü karşısında ezilmiş ve kendini kötü hissettiği için de maalesef yanlış beyanda bulunmuştur.)

Verilerin anket yoluyla toplanması genel olarak bu biçimdedir. Örnekleme konusu ilerleyen haftalarda ele alınacaktır.

3.2. Verilerin Düzenlenmesi

Aşağıdaki örnek veriyi kullanarak verilerin düzenlenmesi ve özetlenmesi konularını ele alacağız:

Örnek Veri:

19	13	5	4	1	9	9	7	11	11
12	1	1	17	13	18	5	2	12	14
16	18	20	19	19	14	17	15	2	3
6	3	12	16	13	13	11	10	10	8
10	4	2	5	15	10	14	10	1	7

Örnek veri seti, 1 ile 20 arasında değer alan 50 adet gözlemden oluşmaktadır.

Dizi:

Bir veri setinin dizi haline getirilmesi demek, ya küçükten büyüğe, ya da büyükten küçüğe sıralamaktır. Genellikle küçükten büyüğe sıralama yapılır. Diziye bakarak en küçük değer ile en büyük değer arasındaki farkı belirleyerek, böylece gözlemlerin hangi aralıkta dağıldıklarını görebiliriz. Dizideki en küçük değere Xmin, en büyük değere Xmax, aralarındaki farka ise "açıklık" yani "range", "değişim aralığı" da denebilir.

1	1	2	2	3	3	4	4	5	5
5	6	6	6	7	7	7	7	8	8
8	9	9	9	9	10	10	10	10	10
10	11	11	11	11	12	12	13	13	13
14	14	15	16	16	17	18	18	19	20

Xmin=1 Xmax=20 range=20-1=19

3.3. Grafikler

3.3.1. Basit Serinin Sunumu İçin Nokta Diyagramı

Verinin değişim aralığını göstermek üzere ölçeklendirilmiş bir doğru üzerinde her bir gözlemin bir nokta ile temsil edildiği bir grafiktir.

Şekil.3.1. Nokta Diyagramına Bir Örnek

```
                  *
                  *
              *  *  *   *
          * *  * * *  *   *      *
    *  * *  * *  *  * * *  *   *   *   *   *     *      *
    *  * *  * *  *  * * *  *   *   *   *   *  *  *   *  *  *   *
______________________________________________________________________

    1  2  3  4  5  6  7  8  9  10  11  12  13 14  15 16 17 18 19 20
```

Gözlem değerlerinin daha çok hangi değerler etrafında toplandıklarını, ne yönde ve ne büyüklükte dağıldıklarını ve de uç değerleri görmek açısından faydalı bir düzenleme olan nokta diyagramında aynı değere sahip birden fazla gözlem söz konusu olduğunda, noktalar bu değer üzerinde dikey olarak sıralanırlar. Örnek verinin nokta grafiğine bakıldığında verinin daha çok sol tarafta toplandığını, gözlemlerin büyük bölümünün 5-11 aralığına düştüğünü görebiliyoruz.

Örnek: Aşağıdaki verinin nokta diyagramını çizerek yorumlayınız.

103	98	111	106	107	112	107	114
106	100	104	96	100	105	118	121
117	101	110	109	107	99	145	101
109	111	105	101	108	102	129	108

Önce veri setini bir dizi haline getirmeliyiz, yani küçükten büyüğe doğru sıralamalıyız:

96	98	99	100	100	101	101	101
102	103	104	105	105	106	106	107
107	107	108	108	109	109	110	111
111	112	114	117	118	121	129	145

```
        *    *

       **   ***** *

   *  ** ************  *   **   *       *            *

_____________________________________________________________________

   95   100          110          120          130          140    145
```

Nokta diyagramına baktığımızda verilerin 107 dolayında yaklaşık olarak simetrik dağıldıklarını, 129 ve 145 uç değerlerini görüyoruz.

3.3.2. Frekans Serisinin Sunumu İçin Çizgi Diyagramı

Veri kümesine bakarak verideki her bir değerin kaç defa tekrar ettiğini görmek mümkündür. Grafiksel bir yola başvurmaksızın, her bir gözlem değerini frekansıyla (sıklığıyla) birlikte gösteren bir tablo düzenlersek, frekans (sıklık) dağılımını elde etmiş oluruz. Böyle serilerin grafik sunumunda çizgi diyagramı kullanılır.

Örnek veri setimizin frekans dağılımını oluşturalım:

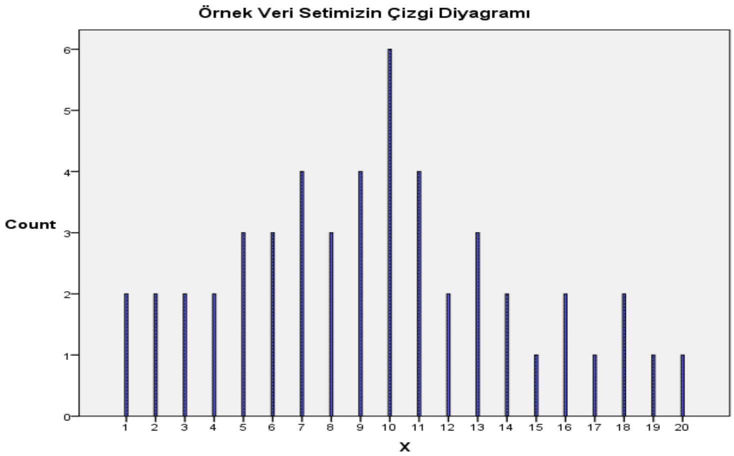
X_i (değerler)	f_i (frekanslar yani sıklıklar)
1	2
2	2
3	2
4	2
5	3
6	3
7	4
8	3
9	4
10	6
11	4
12	2
13	3
14	2
15	1
16	2
17	1
18	2
19	1
20	1

Toplam $\sum_{i=1}^{20} f_i = n = 50$

Sıklıkların toplamı yani $\sum_{i=1}^{k} f_i$, gözlem sayısı n'e eşittir. Burada k, sınıf sayısıdır. İleride sınıflanmış seri kavramını göreceğiz. Buradaki sınıftan kasıt, farklı değerler olarak anlaşılmalıdır. Yani şimdilik gerçek anlamda bir sınıftan bahsetmiyoruz.

Frekans dağılımının grafiksel gösteriminde çoğunlukla çizgi diyagramına başvurulur.

Şekil.3.2. Çizgi Diyagramına Bir Örnek



3.3.3. Sınıflanmış Serinin Sunumu İçin Histogram

Gözlem sayısı çoğaldığında, her bir verinin çizgi diyagramında olduğu gibi tek tek gösterilmesi güçleşmektedir. Bu nedenle, büyük veri seti söz konusu olduğunda verilerin sınıflandırılmaları yoluna gidilir. Verinin belli sınıflara bölünerek bu aralıklara karşılık gelen frekanslarıyla birlikte gösterildiği tabloya, sınıflandırılmış frekans dağılımı denir. Sınıflandırılmış frekans dağılımını oluşturmak için izlenmesi gereken adımlar şöyledir:

1. Verinin değişim aralığı (açıklığı- range'i) hesaplanır.

2. Değişim aralığı eşit uzunlukta alt aralıklara yani sınıflara bölünür. Sınıfların üst sınırı ile alt sınırı arasındaki fark o sınıfın sınıf genişliğini verir. Sınıf orta noktası, sınıfın alt sınırı ile üst sınırının ortalamasıdır. Sınıf sayısını bulmak için aşağıdaki formüllerden herhangi biri kullanılabilir;

k, sınıf sayısı ve n, veri setindeki gözlem sayısını ifade etmek üzere;

$$k = \sqrt{n}$$

$$k = 1 + 3.3\log n$$

$$k = 5\log n$$

Değişim aralığı, sınıf sayısı ve sınıf genişliği arasındaki ilişki şöyledir;

$$sınıfgenişliği = \frac{range}{k}$$

3. Her bir sınıf aralığına kaç gözlemin düştüğü sayılır. Bu sayılar o sınıfın frekansını verir.

Örnek: Örnek veri setimizi sınıflandırılmış frekans serisi haline getirelim.

İlk olarak verinin değişim aralığını buluyoruz:
Xmax=20
Xmin=1
Range=19

Sonra, k sınıf sayısını buluyoruz:

k=1+3.3log(50)=6.6 bulduk, fakat k tamsayı olmak zorundadır, o halde k=7 alalım.

Sınıf genişliği= 19/7=2.714 bunu da 3'e yuvarlayalım.

İlk sınıfı bulmak için, Xmin değerine sınıf genişliğini ekleyelim. Böylece ilk sınıfın üst sınırını da bulmuş olacağız.

Sınıflar	Sıklıklar
1-4 den az	6
4-7 den az	8
7-10 dan az	11
10-13 den az	12
13-16 dan az	6
16-19 dan az	5
19-22 den az	2

Toplam=50

Histogram:

Histogram, dikey eksende sıklıkların, yatay eksende sınıf aralıklarının bulunduğu, yan yana dikdörtgen kutucuklardan oluşan bir grafik türüdür. Her bir kutu bir sınıfa tekabül eder ve kutunun alanı, ilgili sınıfın sıklığını verir.

Kutunun alanı=taban*yükseklik

Burada taban, sınıf genişliğidir, yani ilgili sınıfın üst sınırı ile alt sınırı arasındaki farktır. Bu değeri "h" ile ifade edersek, yükseklik şuna eşittir:

Kutunun yüksekliği = f_i/h'tır.

Bu durumda kutunun alanı=h*(f_i/h) = f_i'dir yani ilgili sınıfın sıklığı (ilgili sınıftaki birim sayısı) dır.

Histogram çizerken şu adımlar izlenir:

1. Sınıf sayısı (k) belirlenir. Hatırlanacağı üzere k'ye elde etmek için çeşitli formüllerimiz vardı.

$$k = \sqrt{n}$$
$$k = 1 + 3.3\log n$$
$$k = 5\log n$$

2. Sınıf aralığının genişliği belirlenmeli. Peki bu işlem nasıl yapılır? Bir çubuk düşünelim. Bu çubuğun boyu "**range"** olsun yani **"Xmax-Xmin"** demek istiyoruz. k'ya yani sınıf sayısını da bulduğumuza göre, bu çubuğu kaç parçaya böleceğimiz belli demektir. *Çubuğu k tane parçaya böleceğiz*. O halde **"range/k"** bize çubuğu böldüğümüz k tane parçanın bir tanesinin uzunluğunu verecektir. Bu uzunluk sınıf genişliğidir, yani sınıfın üst sınırı ile alt sınırı arasındaki farktır yani histogramda ilgili sınıfa ait kutunun taban uzunluğudur.

Örnek: Örnek veri setimizin histogramını çizelim.

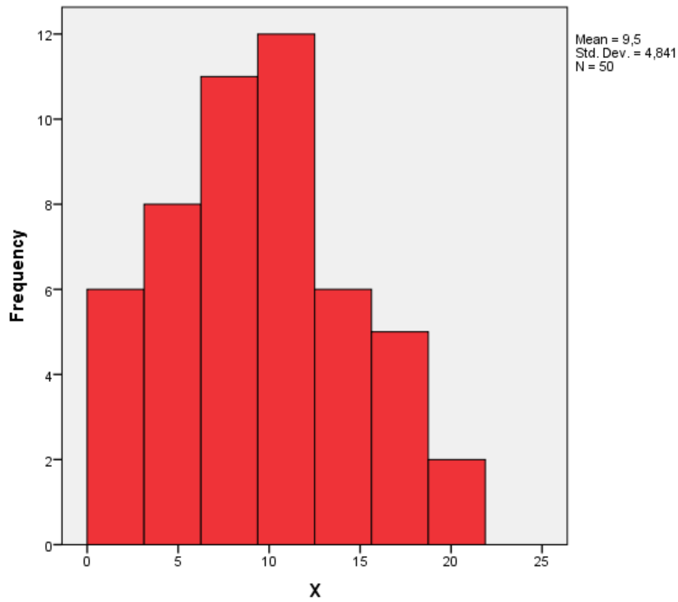
k=7 alarak, sınıf genişliği 2.714 (bu değeri range/k'dan bulduk) olarak alınsaydı sınıflar aşağıdaki gibi olurdu:

Sınıflar	Sıklıklar
1-3,714	6
3,714-6,428	8
6,428-9,142	11
9,142-11,856	10
11,856-14,57	7
14,57-17,284	4
17,284-20	4

Son sınıfın üst sınırı, 19.998 olarak bulduk ancak bu tamamen virgülden sonra 3 basamakla yetinmemizden kaynaklandı, şayet en azından 5 basamak alsaydık, 20'yi kapsayan bir değer bulacaktık ama böyle olduğunu bildiğimiz için son sınıfın üst sınırını "20" diyerek kapattık. Görüldüğü gibi sınıf genişliğini yuvarlamadan da sınıflandırılmış sıklık dağılımını elde edebiliriz. Gözlemlerimiz tamsayı oldukları için gözlem değerleri sınıf sınırlarıyla çakışmadı. Sıklıklar yazılırken örneğin ilk sınıfın üst sınırı 3.714 olduğundan, bu sınıfa sadece

"1,2,3" dâhil edildi, 4 tabii ki alınmadı. 2. Sınıfın sıklıkları yazılırken "4,5 ve 6" alında, tabii ki 7 alınmadı, v.b.

Şekil.3.3. Histograma Bir Örnek

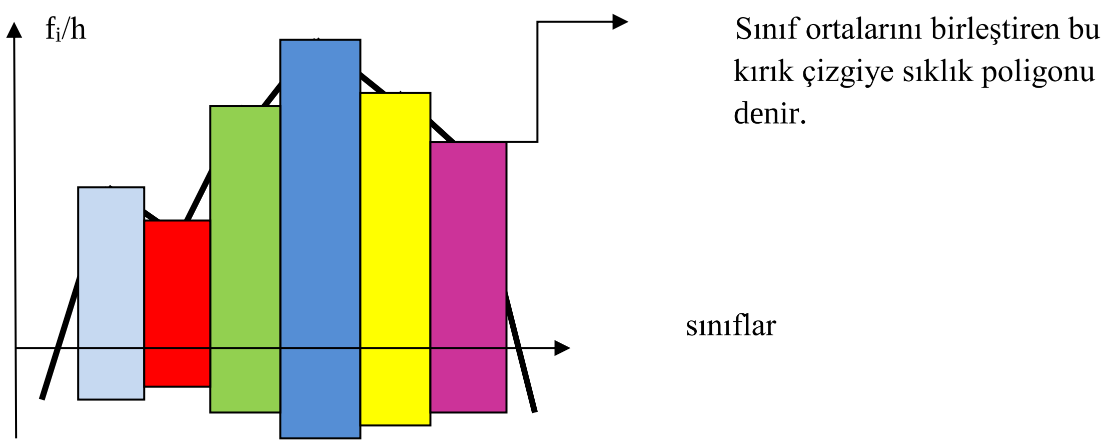


Bu uygulamadan da görüldüğü üzere, bir frekans serisini sınıflandırılmış frekans serisine dönüştürmenin birden fazla yolu vardır (yuvarlama yaparak veya yuvarlama yapmadan yahut k'nın formül seçiminden kaynaklı farklı sınıf sayıları ortaya çıkabilir, vb.). *Dikkat edilmesi gereken hususlar ise, veri setindeki tüm gözlemlerin sınıflar tarafından kapsanması, hiçbir verinin açıkta kalmamasıdır. Ayrıca, her bir veri sadece bir sınıfa ait olmalıdır, bir veri birden fazla sınıfta yer alamaz.*

3.3.3.1. Sıklık Poligonu

Sıklık poligonu frekans poligonu olarak da adlandırılır. Histogramda dikdörtgen kutuların üst kısmında sınıf ortalarının birleştirilmesiyle elde edilir.

Şekil.3.4. Sıklık Poligonuna Bir Örnek

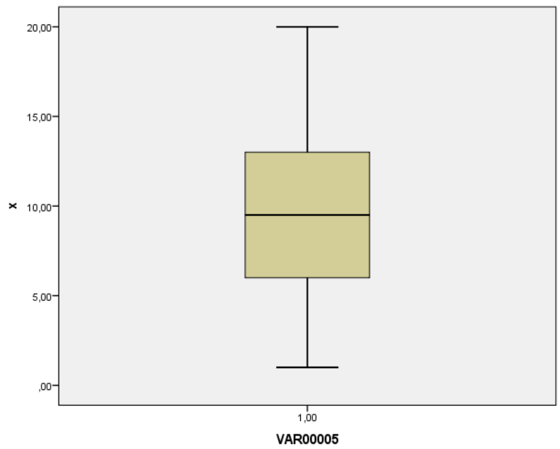


3.3.4. Box-Whisker Grafiği

Bu grafik türünde beş adet istatistiksel özet bir arada görülür. Bunlar; Xmax, Xmin, Q1 (1. Çeyrek yani küçükten büyüğe sıralanmış bir seriyi %25'ten kesen değer), Q2 (2. Çeyrek, asıl adı medyandır, küçükten büyüğe sıralanmış bir seriyi %50'den kesen değer), Q3 (3. Çeyrek yani küçükten büyüğe sıralanmış bir seriyi %75'ten kesen değer).

Aşağıdaki şekilde örnek veri setimizin SPSS'te çizilmiş Box-Whisker grafiği yer almaktadır.

Şekil.3.5. Box-Whisker Grafiğine Bir Örnek



3.3.5. Daire (pasta) Grafiği

Daire şeklindedir. Bu grafiği çizmek için her bir sınıfın sıklığının toplam sıklık içindeki oranı bilinmelidir. Toplam sıklık 360^0 olmak üzere, sınıflar sıklıkları oranında birer dilimle temsil edilirler.

Örnek: İstanbul'da 10 milyon yetişkin bulunduğunu varsayalım. Bu kişilerden 3 milyonu ilk okul mezunu, 4 milyonu ortaokul mezunu, 2 milyonu lise mezunu, 1 milyonu üniversite mezunu ise daire grafiğini çiziniz.

Çözüm:
Oran orantıyla sınıfların kaç derecelik açıyla temsil edilmesi gerektiğini hesaplayacağız.

10 milyon kişi $\qquad$ 360^0 ise

3 milyon kişi $\qquad$ α

$3*360=10*\alpha \qquad \alpha=108^0$

10 milyon kişi $\qquad$ 360^0 ise

4 milyon kişi $\qquad$ α

$4*360=10*\alpha \qquad \alpha=144^0$

10 milyon kişi $\qquad$ 360^0 ise

2 milyon kişi $\qquad$ α

$2*360=10*\alpha \qquad \alpha=72^0$

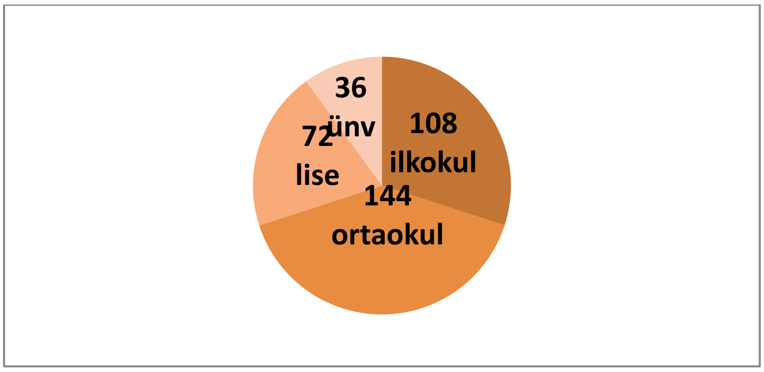
10 milyon kişi $\qquad$ 360^0 ise

1 milyon kişi $\qquad$ α

$1*360=10*\alpha \qquad \alpha=36^0$

Örneğe ilişkin pasta diyagramı aşağıdaki gibidir:

Şekil.3.6. Daire Grafiğine Bir Örnek



Yukarıdaki daire grafiğine bakıldığında, 144 derecelik kısım ortaokul mezunlarına, 108 derecelik kısım ilkokul, 72 derecelik kısım lise, 36 derecelik kısım üniversite mezunlarına aittir.

Bölüm Soruları

1. Birincil veri kaynaklarına örnekler veriniz.
2. İkincil veri kaynaklarına örnekler veriniz.
3. İstatistiksel araştırma yapmanın gerekçeleri neler olabilir?
4. Anketin önemi ve uygulanmasındaki adımları kısaca belirtiniz.
5. Bir frekans (sıklık) dağılımında toplam sıklıklar bize neyi verir?
6. Çizgi diyagramı ne tür serilerde kullanılır?
7. Histogram ne tür serilerde kullanılır?
8. Histogramda sınıf genişliği nasıl elde edilir?
9. Box-Whisker grafiğinide ortadaki çizgi ne anlama gelir?
10. Pasta (daire) grafiğinde dilimlerin büyüklüğü nasıl hesaplanır?

4. İSTATİSTİKTE ÖNEMLİ NOTASYONLAR

Bu Bölümde Neler Öğreneceğiz?

- Notasyonlar
- Faktöriyel hesabı
- Yuvarlama işlemi
- İstatistikte sıklıklar bakımından seri türleri

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
İstatistiksel hesaplamalarda önemli notasyonlar	İstatistiksel hesaplamalarda kullanacağımız önemli notasyonlar, faktöriyel hesabı ve yuvarlama işlemlerinin öğrenilmesi	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kelimeler

- Toplam sembolü $\sum$
- Ortalama bulma
- Ortalamadan sapmalar serisinin elde edilişi
- Faktöriyel hesabı
- Sayıların yuvarlanması

4. 1. İstatistikte Önemli Notasyonlar

İstatistiksel analiz yapmanın yolu bir takım hesaplamalar yapmaktan geçer. Bu hesaplamalarda kullanılan bir takım önemli notasyonların hatırlanması, ileriki haftalarda yapacağımız istatistiksel analizlere temel teşkil etmekte ve büyük önem taşımaktadır. Matematik konularından aşina olduğumuz bu notasyonlar çok basit gibi görünse de, istatistik hesapların hassas oluşu ve en küçük bir işlem hatasının bile hatalı sonuç bulmaya dolayısıyla karar alma ve yorum yapma sürecinde gerçeği yansıtmayan sonuçlarla bizi karşı karşıya bırakmaktadır.

4.1.1 Σ (Toplam Sembolü)

Kullanacağımız en temel notasyon "Σ" sembolüdür. Bu sembol, birimleri tek tek toplamak anlamına gelir. Yani toplama operatörüdür. X bir değişken olmak üzere, şayet $\sum X$ şeklinde bir ifade görürseniz, X'in tüm gözlem değerlerini toplayacağımızı anlamalıyız.

N hacimli bir kitleden n büyüklüğünde rastsal bir örneklem çekmiş olalım. Bu örneklemin ortalamasını bulmak için yapmamız gereken işlem, tüm gözlem değerlerini toplayarak, sonucu, birim sayısına yani n'e bölmektir. Yaptığımız bu basit işlem formül olarak $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ şeklinde ifade edilir. Burada $\bar{X}$, örneklem ortalamasıdır. Üstteki çizgi ilgili değişkenin yani üstüne konduğu değişkenin ortalaması olduğunu göstermektedir. Ortalama formülünün payında bulunan $\sum_{i=1}^{n} X_i$ ifadesi, i=1'den n'e kadar tüm X değerlerini toplayınız. Daha sonra ortalamayı hesaplamak için bu sonucu toplam gözlem sayısına yani toplam birim sayısına yani n'e böleriz.

Örnek: Bir ailedeki çocukların boy uzunlukları aşağıdaki gibidir. Çocukların boy ortalamasını hesaplayınız.

Sedef: 142 cm, Ömer: 156 cm, Mira: 168 cm, Murat: 172 cm

Burada gözlem sayımız (birim sayımız) ailedeki çocuk sayısıdır yani n=4'tür. X'in sağ alt köşesinde yer alan **"i operatörü"** sıra belirtmektedir. Burada X_1= 142 cm (Sedef'in boyu), X_2=156 cm (Ömer'in boyu), X_3=168 cm (Mira'nın boyu), X_4= 172 cm (Murat'ın boyu). O halde ailedeki dört çocuğun boy ortalaması;

$$\bar{X} = \frac{\sum_{i=1}^{4} X_i}{4} = \frac{638}{4} = 159.5 \text{ cm}$$ olarak bulunur.

4.1.2. $\sum \boldsymbol{X^2}$ işleminin uygulanışı

Burada yapılması gereken işlem, X'in tüm değerlerinin karelerini alarak bunları toplamaktır. Sıklıkla yapılan bir hata şudur, X'leri toplayıp elde edilen sonucun karesini almaktır. Yani;

$\sum X^2 \neq (\sum X)^2$

Yapmamız gereken X'lerin tek tek karelerini almak, sonra bu kareleri toplamaktır.

Örnek: Aşağıda verilen veri kümesinde $\sum X^2$'yi hesaplayınız.

5 14 87 101

$\sum X^2 = (5^2 + 14^2 + 87^2 + 101^2) = 17991$'dir.

Halbuki sıkça karıştırılan $(\sum X)^2 = (5 + 14 + 87 + 101)^2 = 207^2 = 42849$'dur. Aradaki farka dikkat etmek gerekir.

4.1.3. $\sum(\boldsymbol{X} - \overline{\boldsymbol{X}})$ Notasyonu

Bu formülde anlatılmak istenen, öncelikle veri kümesinin ortalamasını bulmak, sonra da her bir X gözlem değerinden ortalamayı çıkarmaktır, en sonunda da bu farkları toplamaktır. Yani ortalamadan saptırılan değerlerin toplamı demektir. Aslında formüller adeta konuşmaktadır, bize sırasıyla ne yapmamız gerektiğini söylemektedir. $\sum(\boldsymbol{X} - \overline{\boldsymbol{X}})$ ifadesi bize, X'lerden ortalama değerlerini tek tek saptır (yani fark al), sonra da bu değerleri topla demektedir.

Örnek: Aşağıdaki veri kümesi için $\sum(X - \bar{X})$ 'i hesaplayınız.

12 21 33 45 59

İlk yapacağımızşey beş gözlemden oluşan bu veri kümesinin ortalamasını bulmaktır.

$$\bar{X} = \frac{\sum_{i=1}^{5} X_i}{5} = \frac{12+21+33+45+59}{5} = \frac{170}{5} = 34$$

Sonra X'leri tek tek ortalamadan saptıracağız. Bu işlemi bir sütun açarak yukarıdan aşağıya doğru sırayla yazmak gerekir. Yan yana yazmak, ileride daha çok sayıda veriyle karşılaşınca, işlemleri içinden çıkılmaz bir hale sokar. Şimdiden bu el alışkanlığını kazanmak ileride bize kolaylık sağlayacaktır. O halde ortalamadan sapmalar serisini oluşturalım:

X_i	$X_i - \bar{X}$	
12	-22	bu değeri, 12-34'ten bulduk.
21	-13	
33	-1	
45	11	
59	25	

Uyarı!!! Bir seride ortalamadan sapmaların toplamı daima sıfırdır. Eğer sıfır bulmadıysanız işlem hatası yapmışsınız demektir.

Yukarıdaki örnekte ortalamadan sapmaların toplamı;

$\sum(X - \bar{X})$ = (-22 -13 -1 +11 +25) = 0'dır.

4.1.4. e^x Notasyonu

e sayısı matematikte bir sabittir ve yaklaşık değeri e=2.718'dir. Dönem sonunda kesikli dağılımlar konusuna geldiğimizde "Poission Dağılımı" ile ilgili olasılık hesaplarken e'yi sık sık kullanacağız.

$e^0 = 1$
$e^{-2} = 0,1353$
$e^5 = 148,413$ v.b.

4.2. Faktöriyel Hesabı

"!" ile gösterilen bu notasyon, geriye doğru tüm rakamların çarpılması anlamına gelir.

0! =1
1! =1
2! = 2.1 =2

3! = 3.2.1 =6
4! = 4.3.2.1=24
……
30! = 30.29.28……5.4.3.2.1= $2{,}652528598 \times 10^{32}$

4.3. Yuvarlama İşlemi

Hesap makinelerinde virgülden sonra çok sayıda rakam gösterilir. Bu kadar çok sayıyla baş etmenin yolu yuvarlama yapmaktır. Virgülden sonra kaç basamak ilerletmek gerektiğini zaman içerisinde uygulamalar yapa yapa deneyimleyip karar vereceğiz, ancak tecrübelere dayanarak, virgülden sonra üç basamak yürütmenin isabetli sonuçlar verdiğini söyleyebiliriz. Elbette bu mutlak bir kural değildir.

Yuvarlama işlemi yuvarlanacak basamağın sağında kalan sayıya bağlı olarak gerçekleştirilir. Eğer bu rakam 5 ten küçükse yuvarlanacak basamak olduğu gibi bırakılır, eğer 5'ten büyük ve eşitse bir üst sayıya yuvarlanır. Birkaç örnek verelim:

2.426889

Şimdi bu sayıyı virgülden sonra 3 basamak ilerletmek suretiyle kullanacağımızı düşünelim. Bu durumda virgülden sonra 3. Basamakta yer alan rakama ve sağındaki ilk rakama bakarak yuvarlama yapacağız. Virgülden sonraki 3. Basamakta yer alan sayı "6" dır, 6 'nın sağında ise "8" bulunmaktadır. "8" rakamı 5'ten büyük olduğuna göre 6'yı bir üst rakama yuvarlarız, yani bu sayının yuvarlanmış hali 2.427'dir.

1.58912458 yuvarlanmış hali 1.589'dur. 9'un sağındaki rakam 5'ten küçük olduğu için 9 rakamı aynen kalır.

2.45726911111178 yuvarlanmış hali 2.457'dir.

4.4. İstatistikte frekanslar (sıklıklar) Bakımından Seri Türleri

Üç tür seriden bahsedebiliriz; basit seri, sıklık serisi, sınıflandırılmış seri.

4.4.1. Basit Seri:

verilerin küçükten büyüğe veya büyükten küçüğe sıralandığı seri türüdür, başka hiçbir kuralı yoktur.

Örnek 1: 5 8 17 25 69
Örnek 2: 100 80 54 51

4.4.2. Sıklık Serisi:

Gösterimi yan yana iki sütun şeklindedir. Sol baştaki sütunda X değerleri, sağdaki sütunda ise sıklıkları yani frekansları temsil eden f_i kolonu bulunur.

Örnek:

X_i	f_i
2	2
5	1
9	3

Bu seri aslında 2 2 5 9 9 9 basit serisinin toplulaştırılmış ve düzenli yazımından başka bir şey değildir. Solda X'in aldığı değerler, sağdaki sıklık kolonunda ise bu değerlerden kaçar tane olduğu belirtilir.

4.4.3. Sınıflanmış Seri:

Verilerin belirli kurallara göre toplulaştırılarak sınıflar halinde sunumuna sınıflanmış seri denir.

Örnek:

Sınıflar	f_i
5-8	5
8-11	7
11-14	2
14-17	1

Bu seriye bakıldığında 5-8 sınıfında yani bir sayı doğrusu gibi düşünürsek 5 ile 8 arasında değer alan 5 adet X gözlem değerinin olduğunu görürüz. 8 ile 11 arasında 7 adet veri, 11 ile 14 arasında 2 adet veri, 14 ile 17 arasında 1 tane veri olduğunu görürüz. Genel olarak, 5 ile 17 arasında değer alan 15 adet verimiz var demektir. Burada dikkat edilmesi gereken husus, verilen aralıkta yani sınıfta kaçar tane veri olduğunu biliyoruz ama bu değerlerin tam olarak kaç olduğunu bilmiyoruz. Örneğin ilk sınıfı ele alalım ve sayı doğrusu üzerinde anlatmaya çalışalım:

X1 X2 X3 X4 X5

5

8

X1 X2 X3 X4 X5

5

8

Yani verilen sınıf aralığında kaç gözlem olduğunu biliyoruz ama bunların tam olarak hangi değeri aldıklarını bilmiyoruz.

Uyarı!!! Sınıflanmış seride sınıfın üst sınırını dahil değil gibi düşünelim. Örneğin ilk sınıf 5-8 sınıfında 5'ten 8'e kadar olan değerler kastedilmektedir ama 8 dahil değildir, yani [5,8) aralığıdır. Bir başka gösterim –den az gösterimidir, sınıflar şu şekilde sıralanır, 5-8 den az, 8-11'den az, v.b.

Bölüm Soruları

1. Aşağıdaki veri kümesi için istenenleri hesaplayınız.

4 7 11 15 23

a. $\sum X = ?$

b. $\sum X^2 = ?$

c. $(\sum X)^2 = ?$

d. $\overline{X} = ?$

e. $\sum(\mathbf{X} - \overline{\mathbf{X}}) = ?$

f. $\sum(\mathbf{X} - \overline{\mathbf{X}})^2 = ?$

Çözüm:

a. $\sum X = 60$

b. $\sum X^2 = 16 + 49 + 121 + 225 + 529 = 940$

c. $(\sum X)^2 = (60)^2 = 3600$

d. $\overline{X} = \frac{4+7+11+15+23}{5} = \frac{60}{5} = 12$

e. $\sum(\mathbf{X} - \overline{\mathbf{X}}) = ?$

X_i	$X_i - \overline{X}$
4	-8
7	-5
11	-1
15	3
23	11

$$\sum(\mathbf{X} - \overline{\mathbf{X}}) = (\mathbf{-8 - 5 - 1 + 3 + 11}) = \mathbf{0}$$

f. $\sum(\mathbf{X} - \overline{\mathbf{X}})^2 = ?$

X_i	$(X_i - \bar{X})$	$\sum(X - \bar{X})^2$
4	-8	64
7	-5	25
11	-1	1
15	3	9
23	11	121

$\sum(X - \bar{X})^2 = 64+25+1+9+121 = 220$

5. MERKEZİ EĞİLİM ÖLÇÜLERİ

Bu Bölümde Neler Öğreneceğiz?

- Merkezi eğilim ölçüleri
- Duyarlı ortalamalar
- Duyarsız ortalamalara giriş

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Merkezi eğilim ölçüleri	Duyarlı ortalamaların öğrenilmesi, duyarsız ortalamalara giriş	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kelimeler

- Merkezi eğilim ölçüleri
- Analitik (duyarlı) ortalamalar
- Analitik olmayan (duyarsız) ortalamalara giriş
- Aritmetik ortalama
- Aritmetik ortalamanın özellikleri
- Tartılı aritmetik ortalama
- Kareli ortalama
- Geometrik ortalama
- Harmonik ortalama

5. 1. Merkezi eğilim ölçüleri

Merkezi eğilim ölçüleri bir seri hakkında bilgi edinmemizi ve çeşitli kıyaslamalar yapmamıza olanak sağlayan ölçütlerdir. Verilerin daha çok hangi değerlere eğilimli olduklarını bulmak için merkezi eğilim ölçülerini kullanırız. Bu ölçüler analitik olan (duyarlı) ve analitik olmayan (duyarsız) ortalamalardır. Analitik ortalamalar; aritmetik ortalama, tartılı ortalama, kareli ortalama, harmonik ortalama ve geometrik ortalamadır. Analitik olmayan ortalamalar ise mod, medyan ve kartillerdir.

5.1.1. Analitik (Duyarlı) Ortalamalar

Analitik ortalamalara “duyarlı” denmesinin nedeni; veri kümesindeki tüm gözlem birimlerini hesaba katmasıdır. Bu hesaplama yöntemi, özellikle uç değerler (uç değer bir diğer ifade ile sapan değer, serideki değerlerden çok farklı olan değerlerdir) söz konusu olduğunda ortalamayı uç değerlere doğru çekerek doğru tahmin yapmayı zorlaştırırlar. Bu tip durumlarda ya uç değerler hesaplamadan dışlanır, ya da analitik olmayan ortalama türlerinden faydalanılır. Bu duruma en güzel örnek, bir sınıfın ortalama notunu hesaplarken en yüksek ve en düşük değerin dışlanarak hesaba alınmamasıdır.

Örneğin örneklemimiz aşağıdaki gibi olsun ve yukarıdaki duyarlı olma durumunu aritmetik ortalamayı kullanarak basitçe anlatmaya çalışalım:

20 23 31 42 55

Bu örneklemin aritmetik ortalaması tüm değerlerin toplamının gözlem sayısı n’e bölünmesiyle 34,2 olarak hesaplanır. Ortalamaların taşıması gereken en önemli özellik, gözlemleri iyi bir şekilde temsil etmesidir (ortalama aslında serinin tek bir rakamla özetlenmesidir) ve bulduğumuz bu değer örneklemdeki değerleri iyi kötü temsil etmektedir.

Şimdi örnekleme bir uç değer ekleyelim, yeni örneklemimiz şöyle olsun:

20 23 31 42 55 1000

Bu örneklemin aritmetik ortalaması 195,166’dır. Görüldüğü gibi uç değer ortalamayı kendine doğru çekmiş, bu durum ne küçük gözlemleri ne de uç değeri temsil edebilecek bir ortalamanın hesaplanmasına neden olmuştur. İşte böyle durumlarda, tüm gözlemleri hesaba katmayan analitik olmayan yani bir diğer adıyla duyarsız ortalamalar tercih edilmelidir.

5.1.1.1. Aritmetik Ortalama ($\bar{X}$, μ)

Örneklemde $\bar{X}$, kitlede μ ile gösterilir. Veri kümesindeki tüm gözlemlerin toplamının toplam gözlem sayısına bölümüdür.

Basit seride:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Sıklık serisinde:

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i X_i}{n}$$

Sınıflanmış seride:

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i m_i}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i m_i}{n}$$

k, sınıf sayısıdır. m_i ise i. sınıfın orta noktasıdır (mid point)

Kitlede Gösterimi:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} \qquad \mu = \frac{\sum_{i=1}^{k} f_i X_i}{N} \qquad \mu = \frac{\sum_{i=1}^{k} f_i m_i}{N}$$

Örnek:
Aşağıda bir öğrencinin 5 dersten aldığı notlar verilmiştir. $\bar{X}$ =?

4,5,7,8,10

$\bar{X}$ =6.8

Örnek:
Bir kavşakta bir ayda meydana gelen trafik kazalarının dağılımı aşağıdaki gibidir. Günlük ortalama kaza sayısını hesaplayınız.

X_i	f_i
0	14
1	7
2	4
3	2
4	3
	$\sum f_i = 30$

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i X_i}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i X_i}{n} = 1{,}1$$

Günlük ortalama kaza sayısı yaklaşık olarak 1'dir.

Örnek:

Aşağıdaki serinin aritmetik ortalamasını hesaplayınız.

Sınıflar	f_i	m_i	$f_i m_i$
1-3	4	2	8
3-5	8	4	32
5-7	3	6	18
	$\sum f_i = 15$		$\sum f_i m_i = 58$

Sınıf orta noktaları sınıfın alt sınır ile üst sınırının ortalamasıdır, yani ikiye bölümüdür. $m_1 = (1+3)/2 = 2$ v.b.

Burada biraz duralım, sınıfları tek tek ele alalım. İlk sınıfımız 1-3 aralığıdır ve bu aralıkta 4 gözlem vardır. Fakat biz bu 4 değerin tam olarak ne olduklarını bilmiyoruz, sadece [1,3) aralığında olduğunu biliyoruz değil mi? O halde standart bir ortalama bulabilmek için hepimizin ortak bir yaklaşım sergilemesi gerekir. Yani şunu anlatmak istiyoruz, diyelim ki Ali'ye göre X1= 1, X2= 1.2, X3=2.8, X4=2.9 olsun. İkinci sınıftaki değerleri de şöyle olsun; X5= 3.1, X6= 4, X7= 4.1, X8= 4.3, X9= 4.4, X10= 4.6, X11= 4.8, X12= 4.9 olsun. Son sınıftakiler de şöyle olsun; X13= 5.2, X14= 5.9, X15= 6.8

Şimdi Ali'nin serisinin ortalaması tüm bu değerlerin toplamının toplam gözlem sayısına yani 15'e bölümüdür. Ali'nin serisini açıkça bir daha yazalım:

1	1.2	2.8	2.9	3.1	4	4.1	4.3	4.4
4.6	4.8	4.9	5.2	5.9	6.8			

Ortalamayı bulalım:

$$\bar{X} = \frac{\sum_{i=1}^{15} X_i}{15} = 60/15=4$$

Umut'un serisi de şöyle olsun:

1.1	1.3	1.7	2.4	3	4	4.2	4.5	4.6
4.7	4.8	4.9	5	7	7.9			

Bu ortalamayı bulalım:

$$\bar{X} = \frac{\sum_{i=1}^{15} X_i}{15} = 61,1/15=4.0733333 \sim 4.073 \text{ bulduk.}$$

Demek ki herkes bu aralıkta sonsuz tane değer ataması yapabilir ve bunun sonucunda da herkesin bulacağı ortalama değeri farklı olacaktır. İşte bu durumun önüne geçebilmek için sınıf orta noktalarını yani m_i'leri X_i gibi düşünüyoruz, yani sanki o değerlerin hepsi sınıf orta değerini alıyormuş gibi varsayıyoruz, aksi halde aynı ortalama değerini bulamayız ve herkes kendine göre bir değer bulur.

Sınıflanmış serinin basit seri hali şöyledir:

2	2	2	2	4	4	4	4
4	4	4	4	6	6	6	

Ortalamayı bulalım;

$$\bar{X} = \frac{\sum_{i=1}^{k} f_i m_i}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} f_i m_i}{n} = 3,8666666 \sim 3,867$$

Aritmetik Ortalamanın Özellikleri

1. En çok bilinen ve en sık kullanılan ortalamadır. Her türlü veride hesaplanabilir ve tek bir değer alır. Sıfır içeren serilerde harmonik ortalama hesaplanamaz, geometrik ortalama hesaplanabilir ama anlamsız olur. Seride negatif değer var ise, geometrik ortalama hesaplanamaz. Bazı serilerde birden fazla sayıda mod bulunabilir. **Ancak aritmetik ortalama her türlü veri seti için hesaplanabilir ve tek bir tanedir.**

2. $\bar{X}$ duyarlı bir ortalamadır. Çünkü bütün verileri hesaba alır. Extrem (uç, outlier) değerler olduğunda bu değerlerden etkilenmesi, aritmetik ortalamanın zayıf yönüdür.

Böyle durumlarda $\bar{X}$ yerine, duyarsız ortala da denilen, sıralamaya dayanan ve uç değerlerden etkilenmeyen medyan tercih edilebilir.

3. Bir seride $\bar{X}$ 'dan sapmaların toplamı sıfırdır.

$$\sum_{i=1}^{N}(X_i - \mu) = 0$$

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

4. Açık uçlu dağılımlarda hesaplanamaz. Açık uçlu dağılıma bir örnek verelim:

Sınıflar	f_i	m_i	$f_i m_i$
1-3	4	2	8
3-5	8	4	32
5-…	**3**	**?**	**?**
	$\sum f_i = 15$		

Son sınıfın üst sınırı açık bırakılmış, yani oradaki değerin ne olduğu bilinmeden sınıf ortası bulunamayacağı için aritmetik ortalama da hesaplanamaz. Sınıf genişlikleri eşit olmak zorunda değildir, yani biz kendiliğimizden oranın “7” olduğunu iddia edemeyiz.

5.1.1.2. Tartılı Ortalama (Ağırlıklı Ortalama)

Bazen veride yer alan bazı gözlemlere diğerlerinden daha fazla önem verilir. Böyle durumlarda ortalama hesaplanırken bu özel değerlere verilen önemi yansıtmak için belirli tartılar tahsis edilir.

Örnek: İstatistik dersine ilişkin vize notunun ağırlığı finalin üçte biri olsun. Öğrenci vizeden 80, finalden 60 almış ise, bu öğrencinin başarı notu tartılı aritmetik ortalama ile hesaplanır. Tartılı aritmetik ortalamanın formülü şöyledir:

$$\bar{X} = \frac{\sum t_i X_i}{\sum t_i}$$

Örneğimize dönersek;

$t_1=1$ vizenin tartısı

$t_2=3$ finalin tartısı (bir anlamda önemi)

$$\bar{X} = \frac{(1*80)+(3*60)}{1+3} = 65$$

Örnek: Aşağıdaki tabloda bir öğrencinin ders notları ve derslerin ağırlıkları yer almaktadır (yani her ders aynı öneme sahip değildir) öğrencinin not ortalamasını hesaplayınız.

Ders	Ağırlık	Puan	Ağırlıklı Puan
Fizik	2	70	140
Kimya	2	80	160
Matematik	2	85	170
Türkçe	1	90	90
Tarih	1	65	65

$$\bar{X} = \frac{\sum t_i X_i}{\sum t_i}$$

= (140 + 160 + 170 + 90 + 65) /(2 + 2 + 2 + 1 + 1) = 625/8= 78.125 ‘tir.

Normal ortalaması (yani her dersin eşit öneme sahip olduğu durumda);

$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$ = (70+80+85+90+65)/ 5= 390/5=78’dir.

5.1.1.3. Kareli Ortalama

Kareli ortalama serideki değerlerin karelerinin toplamının gözlem sayısına bölünerek karekökünün alınmasıyla hesaplanır.

Basit Seride:

KO=$\sqrt{\frac{\sum X^2}{n}}$

Sıklık Serisinde:

KO=$\sqrt{\frac{\sum fi.X^2}{n}}$

Sınıflanmış Seride:

KO=$\sqrt{\frac{\sum fi.mi^2}{n}}$

Örnek: Aşağıdaki serinin kareli ortalamasını hesaplayınız.

Sınıflar	f_i	m_i^2	$f_i m_i^2$
1-3	4	4	16
3-5	8	16	128
5-7	3	36	108

KO=$\sqrt{\frac{\sum fi.mi^2}{n}} = \sqrt{\frac{16+128+108}{15}} = 4.09878 \sim 4.099$

5.1.1.4. Geometrik Ortalama

Geometrik ortalama tüm verileri hesaba katan duyarlı bir ortalamadır. Genellikle oransal olarak artan serilerde bu artışı düzlemek (smooth etmek) amacıyla tercih edilir. Tüm gözlemlerin çarpılarak, gözlem sayısı kadar dereceden kök alınmasıyla hesaplanır.

Geometrik Ortalama ≤ Aritmetik Ortalama

$X_1, X_2, \ldots X_n$ veri kümesinin geometrik ortalaması;

$$GO = \sqrt[n]{X_1 X_2 \ldots X_n} = (X_1 X_2 \ldots X_n)^{1/n} = \left(\prod_{i=1}^{n} X_i\right)^{1/n}$$

$$\log GO = \frac{1}{n}(\log X_1 + \log X_2 + \ldots + \log X_n)$$

Basit seride:

$$\log GO = \frac{\sum_{i=1}^{n} \log X_i}{n}$$

Sıklık serisinde:

$$\log GO = \frac{\sum_{i=1}^{k} f_i \log X_i}{n}$$

Sınıflanmış seride:

$$\log GO = \frac{\sum_{i=1}^{k} f_i \log m_i}{n}$$

Örnek: Aşağıdaki serinin geometrik ortalamasını bulunuz.

X_i
4
5
7
8
16

$$GO = \sqrt[5]{4*5*7*8*16} = 7.09$$

$$\log GO = \frac{1}{5}(\log 4 + \log 5 + \log 7 + \log 8 + \log 16)$$

$\log X_i$
0,602
0,699
0,845
0,903
1,204

$$\sum_{i=1}^{5} \log X_i = 4,253$$

$$\log GO = \frac{1}{5} * 4,253 = 0,8506$$

$$GO = 10^{0,8506} = 7,09$$

Örnek: Aşağıdaki sıklık serisinin geometrik ortalamasını hesaplayınız.

X_i	f_i	$\log X_i$	$f_i \log X_i$
2	3	0,301	0,903
3	2	0,477	0,954
4	1	0,602	0,602
5	4	0,698	2,795

$$\sum f_i = 10 \qquad \sum f_i \log X_i = 5,255$$

$$\log GO = \frac{\sum_{i=1}^{k} f_i \log X_i}{n}$$ =5,255/10=0,5255

GO=$10^{0,5255}$=3,35

Örnek: Aşağıdaki sınıflanmış serinin geometrik ortalamasını hesaplayınız.

Sınıflar	f_i	m_i	$\log m_i$	$f_i \log m_i$
1-3	3	2	0,301	0,903
3-5	3	4	0,602	1,806
5-7	4	6	0,778	3,112

$$\sum f_i = 10 \qquad \sum_{i=1}^{k} f_i \log m_i$$ =5,821

$$\log GO = \frac{\sum_{i=1}^{k} f_i \log m_i}{n} = 0,5821$$

GO=$10^{0,5821}$=3,82

5.1.1.5. Harmonik Ortalama

Oransal olarak belirtilebilen değişkenlerin ortalamalarının hesaplanmasında harmonik ortalama kullanılır. Sıfır değerli ya da negatif işaretli değişkenler olduğunda hesaplanamaz.

Basit seride:

$$HO = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

Sıklık serisinde:

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

Sınıflanmış seride:

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

Örnek:

X_i	$\frac{1}{X_i}$
4	0,25
5	0,2
7	0,14
8	0,12
16	0,06

$$\sum_{i=1}^{k} \frac{1}{X_i} = 0,77$$

$$HO = \frac{5}{0,77} = 6,49$$

Örnek: Harmonik ortalamasını hesaplayınız.

X_i	f_i	$\frac{f_i}{X_i}$
2	3	1,5
3	2	0,67
4	1	0,25
5	4	0,80

$$\sum f_i = 10 \qquad \sum_{i=1}^{k} \frac{f_i}{X_i} = 3,22$$

HO=10/3,22=3,11

Örnek: Harmonik ortalamasını hesaplayınız.

Sınıflar	f_i	m_i	$\frac{f_i}{m_i}$
1-3	3	2	1,5
3-5	3	4	0,75
5-7	4	6	0,67

$$\sum f_i = 10 \qquad \sum_{i=1}^{k} \frac{f_i}{m_i} = 2,92$$

HO=10/2,92=3,42

5.1.2. Analitik Olmayan (Duyarsız) Ortalama

Analitik olmayan ortalamalara "duyarsız" denmesinin nedeni, özellikle uç değerlerin olduğu durumlarda bu gibi değerleri göz ardı etmesi, tüm gözlemleri hesaba almaması ve sıralamaya dayanmasından kaynaklanmaktadır. Bu sıralamanın ne anlama geldiğini ilerideki sayfalarda göreceğiz.

Bölüm Soruları

1. Duyarlı ve duyarlı olmayan ortalama ne demektir? Örnek veriniz.
2. Aşağıda bir sınıftaki öğrencilerin aldıkları notların dağılımı mevcuttur. Sınıfın ortalama başarısı nedir?

Sınıflar	Öğrenci sayısı
0-20	10
20-40	15
40-60	20
60-80	5
80-100	2

3. Bir öğrencinin vizeden aldığı not 50 ve finalden aldığı not 52 ‘dir. Bu sınıfta geçme notu 60 ise öğrenci ilgili dersten geçebilir mi? (Vize ağırlığı %30)
4. $\sum_{i=1}^{n}(X_i - \bar{X}) = 0$ ifadesini 4 gözlemden oluşan bir seriden yola çıkarak doğrulayınız.
5. Toplam frekans değeri 15 olan bir tane sınıflandırılmış, bir de sıklık serisi oluşturarak bu serilerde aritmetik ortalama, kareli ortalama, geometrik ortalama ve harmonik ortalama değerlerini hesaplayarak bu 4 değer arasındaki sıralamayı belirtin.

6. ANALİTİK OLMAYAN (DUYARSIZ) ORTALAMALAR VE ÇARPIKLIK İNCELEMESİ

Bu Bölümde Neler Öğreneceğiz?

- Duyarsız ortalama türleri (mod, medyan, kartiller)
- Duyarlı ve duyarsız ortalamalar yoluyla çarpıklık ölçümü

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Duyarsız ortalama türleri, çarpıklık ölçümleri	Duyarlı ve duyarsız ortalamaları kullanarak çarpıklık ölçümleri yapılacaktır	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kelimeler

- Medyan
- Mod
- Asimetri
- Kartiller

6.1. Analitik Olmayan (Duyarsız) Ortalamalar

Analitik olmayan ortalamalara “duyarsız” denmesinin sebebi, duyarlı ortalamalarda olduğu gibi bütün gözlemleri hesaba katmaması ve daha ziyade sıralamaya önem vermesindendir. Başlıca duyarsız ortalamalar medyan, moddur. Kartiller de bu kapsamda ele alınabilir, kartiller yani dörde bölenler (quartile) serinin eğilimini (verilerin hangi noktada yığılım yaptığını,v.b.) ölçmeye yarayan, sıralamaya dayanan duyarsız ölçümlerdir.

6.1.1 Medyan

Medyan, duyarsız bir ortalama türüdür. Sıralamayı esas alır. Bir seride küçükten büyüğe (ya da tersi) sıralanmış veriyi tam ortadan ikiye böler. Bu nedenle bazen **“ortanca”** olarak da adlandırılır. Tüm değerleri hesaba katan analitik ortalamalar, şayet veride extrem (uç) değerler varsa seriyi temsil kabiliyetlerini kaybederler. Aritmetik ortalamaya kıyasla daha tutarlı bir sonuç elde edilir. Açık sınıf aralıklı veri setlerinde merkezi eğilim ölçüsü olarak kullanılabilir. Her bir veri seti için bir tek medyan söz konusudur.

Örneğin;

1 2 4 6 8 10 100

serisinde medyan tam ortadaki değer yani “6” dır. Aritmetik ortalaması ise 18.71 olarak elde edilir ki bu değer serideki “100” verisinden dolayı ortalama şişmiş ve temsili olma özelliğini yitirmiştir. İşte duyarsız ortalamaların önemi böyle durumlarda ortaya çıkmaktadır.

Basit seride medyan bulurken önce gözlem sayısı n’e bakarız. Eğer n tekse, medyan $(\frac{n+1}{2})$. terimdir. Örneğin;

1 2 4 7 11

Veri kümesinde n=5, medyan 3. Terim yani X_3=4’tür.

Şayet n çift sayı ise, medyan, $\frac{n}{2}$. ile $(\frac{n}{2}+1)$. terimin ortalamasıdır. Aslında kısaca medyan şöyle de hesaplanabilir; gözlem sayısı ister tek ister çift sayı olsun $\frac{n+1}{2}$. terim medyandır, yani seriyi ortadan ikiye bölen değerdir.

Örneğin;

1 2 4 7

Veri kümesinde n=4, medyan $X_{2.5}$ 'uncu terimdir yani X_2 ile X_3'ün ortalamasıdır. Bu serinin medyanı 2 ile 4'ün ortalaması yani 3'tür.

Örnek: Aşağıdaki sıklık serisinin medyanını hesaplayınız.

X_i	f_i	$\sum f_i$
12	5	5
34	4	9
40	6	15
68	3	18
85	2	20
	$\sum f_i = 20$	

n=20, çift sayı, medyan $X_{10.5}$'uncu terimdir yani X_{10} ile X_{11}'in ortalamasıdır. Birikimli sıklık kolonuna bakıyoruz, X_{10} ile X_{11}'i kapsayan değer "40" tır. Yani medyan 40'tır.

Medyan=$X_{10.5}$ =(40+40)/2=40

Birikimli Sıklık Kolonu Nasıl Oluşturulur?

Sıklık kolonu yani f_i 'lerin yukarıdan aşağıya doğru toplanarak oluşturulan kolona birikimli sıklık kolonu denir. Özellikle medyan hesaplarken birim sayısı çok olduğunda seriyi açmak ve medyanı görmek kolay olmayacağından, birikimli sıklık kolonundan yararlanmaktayız. Aşağıdaki örneğe bakınız:

X_i	f_i	$\sum f_i$	
12	5	5	X1, X2, X3, X4, X5
34	4	9	X6, X7, X8, X9
40	6	15	X10, X11, X12, X13,X14, X15
68	3	18	X16, X17, X18
85	2	20	X19, X20

Bu seride medyan $\frac{20+1}{2}$ =10.5'inci terim olduğuna göre bize X10 ve X11 lazım. Birikimli sıklık kolonuna bakarak bunların değerlerini tespit edip ortalamasını alacağız. X10 = 40 ve X11=40 olduğunu görüyoruz. Bu durumda medyan değerimiz 40'tır.

Seriyi açarsak daha iyi anlaşılacaktır:

12 12 12 12 12 34 34 34 34 **40** **40** 40
40

40	40	68	68	68	85	85		
							X_{10}	**X_{11}**

Medyan= $\frac{40+40}{2}=40$

Sınıflanmış seride medyan

$$med = Q_2 = L + \frac{\frac{n}{2} - F}{f_{med}} * c$$

Q_2, medyanın "ikinci bölen" gösterimidir. Medyan da bir kartildir, bu gösterim oradan gelmektedir.

Formülde yer alan ifadelerin şu anlama gelmektedir:

L: medyan sınıfının alt sınırı (tam ortadaki terimi içeren sınıf medyan sınıfıdır)

c: sınıf genişliği

F: medyan sınıfından önceki sınıfların sıklıklarının toplamı (ya da bir diğer deyişle, medyan sınıfından bir önceki sınıfın birikimli sıklık değeri)

f_{med}: medyan sınıfının frekansı

Örnek: Medyanı hesaplayınız.

Sınıflar	f_i	$\sum f_i$
0-2	12	12
2-4	9	21
4-6	3	24
6-8	6	30

n=30, çift sayı, medyan $X_{15.5}$ yani X_{15} ile X_{16}'yı kapsayan sınıf medyan sınıfıdır. Bu durumda medyan sınıfı "2-4" sınıfıdır.

$$med = Q_2 = 2 + \frac{15-12}{9} * 2 = 2,67$$

Örnek: Medyanı hesaplayınız.

Sınıflar	f_i	$\sum f_i$
2-7	5	5
7-12	12	17
12-17	21	38

17-22	6	44
22-27	4	48
27-32	2	50
		$\sum f_i$ =50

n=50 çift sayı, X_{25} ile X_{26} yı kapsayan sınıf medyan sınıfıdır. Yani “12-17” sınıfı medyan sınıfıdır.

$$med = Q_2 = 12 + \frac{25-17}{21} * 5 = 13,9$$

6.1.2. Mod

Seride en çok tekrarlanan değere “mod” denir. Duyarsız bir ortalamadır. Genellikle kategorik değişkenlerle ilgili (cinsiyet, meslek, v.b.) uygulamalarda tercih edilir. Bir de, serilerin asimetri durumlarını incelerken yığılmanın yönünü bulmada kullanılır.

Basit seride modun elde edilişi:

10 10 11 12 13 13 13 13 17 20

Seride en çok tekrar eden değer 13 olduğundan, Mod=13’tür.

Sıklık serisinde mod:

Sıklık serisinde mod bulurken yapacağımız şey, fi sıklık kolonuna bakarak en büyük değere karşılık gelen X değerini tespit etmektir.

Örnek:

X_i	f_i	
10	2	
11	1	
12	1	
13	$\longrightarrow$ 4	sıklığı en fazla olan X değeri moddur, bu serinin modu 13’tür.
17	1	
20	1	

Sınıflanmış seride modun bulunuşu:

Sınıflanmış seride mod aşağıdaki formül ile hesaplanır.

$$\text{mod} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} * c$$

L: mod sınıfının alt sınırı
c: sınıf genişliği
Δ_1: mod sınıfının sıklığı ile mod sınıfından bir önceki sınıfın sıklığının farkı
Δ_2: mod sınıfının sıklığı ile mod sınıfından bir sonraki sınıfın sıklığının farkı

Örnek: Modu hesaplayınız.

Sınıflar	f_i	
2-7	5	
7-12	12	
12-17	**21**	→ mod sınıfı
17-22	6	
22-27	4	
27-32	2	

$$\text{mod} = 12 + \frac{9}{9+15} * 5 = 13,88$$

6.1.3. Kartiller (Dörde Bölenler, Çeyrek Bölenler, Quartiles)

Küçükten büyüğe ya da büyükten küçüğe sırlanmış bir seriyi %25'ten bölen değer Q1 yani 1. Kartildir. %50'den (tam ortadan) bölen değer Q2 yani medyandır. Seriyi %75'ten bölen değer Q3 yani 3. kartildir.

Basit Seride Kartillerin Bulunuşu:

Q1 : $\frac{n+1}{4}$. terim   Q2 : $\frac{n+1}{2}$. terim   Q3: $3(\frac{n+1}{4})$. terim

Örnek: Kartilleri hesaplayınız.

1 5 8 11 14 20 54

n= 7

Q1 , $\frac{7+1}{4} = 2$. terimdir, yani küçükten büyüğe sıralı olarak verilen bu seriyi %25'ten bölen değer 5'tir. Q1 = 5'tir.

Q2 , $\frac{21+1}{4} = 4$. terimdir, yani küçükten büyüğe sıralı olarak verilen bu seriyi %50'den bölen değer 11'dir. Q2 = medyan = 11'dir.

Q3 ,3($\frac{7+1}{4}$) = 6 . terimdir, yani küçükten büyüğe sıralı olarak verilen bu seriyi %75'ten bölen değer 20'dir. Q3 = 20'dir.

Örnek: Kartilleri hesaplayınız.

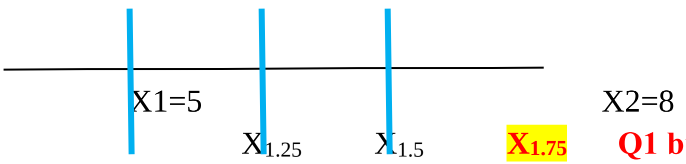
5 8 11 14 20 54

n= 6

Uyarı!!! n çift sayı olduğunda kartilleri hesaplamak gerekir, n tek iken kartiller hemen seriden görülüp teşhis edilebilir, n çift olduğunda bulacağımız kartiller genellikle seride olmayan yani bizim hesapladığımız değerlerdir.

Q1 , $\frac{6+1}{4} = 1,75$. terimdir, yani küçükten büyüğe sıralı olarak verilen bu seriyi

%25'ten bölen değer X1 ile X2 arasında yalnız X2'ye daha yakındır.



Q1 bu değerdir!

Peki $X_{1.75}$ 'i nasıl hesaplayacağız?

Aradaki mesafeyi dörde bölüp ya üç parçayı X1'e ekleyerek, ya da bir parçayı X2'den çıkararak bulacağız. X2-X1= 8 – 5 = 3 birim diyelim, bu durumda bir parçanın uzunluğu 3/4 = 0.75'tir. O halde $X_{1.25}$ = 5 + 0.75 = 5.75'tir. $X_{1.5}$ = 5 + 0.75+0.75 = 6.5

$X_{1.75}$ = 5+ 0.75 + 0.75 + 0.75 = 7.25'tir. Q1 yani birinci çeyrek 7.25'tir.

6.2. Serinin Asimetrisinin (çarpıklığının) "Aritmetik ortalama – Mod – Medyan" Kapsamında Açıklanması

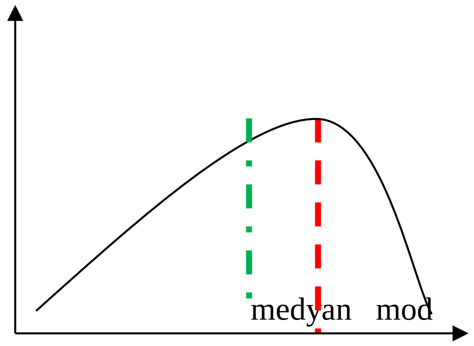
Medyan, ortanca değer olduğundan, bu üç ortalama ele alındığında **daima ortada yer alır.** Aritmetik ortalama ve modun yer değiştirmesine bağlı olarak serinin çarpıklığı yorumlanır. Yani,

$\overline{X}$ =medyan=mod ise seri simetriktir.

$\overline{X} \leq$ medyan $\leq$ mod biçiminde bir sırlama var ise; yığılma sağdadır, seri sola çarpıktır.

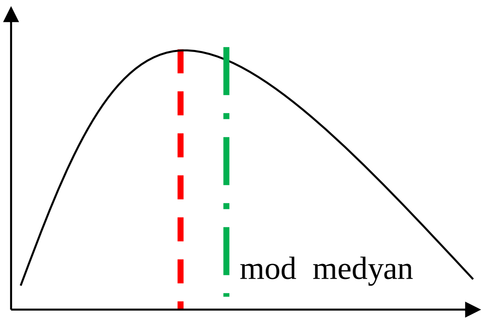
Sırlama mod $\leq$ medyan $\leq \overline{X}$ biçiminde ise; yığılma soldadır, seri sağa çarpıktır.

Şekil.6. 1. Sola çarpık serinin grafiği ($\overline{X}$ $\leq$ medyan $\leq$ mod)



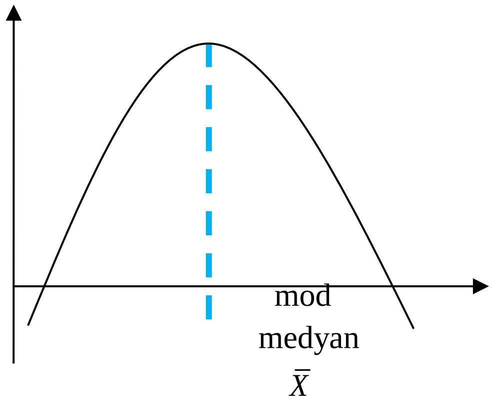
Yukarıdaki şekilde aritmetik ortalamanın yerini belirlemeye ihtiyaç yoktur. Çarpıklığın tespitinde modun medyana göre konumunun bilinmesi yeterlidir. Mod medyanın sağında ise, artimetik ortalama mecburen medyanın solunda yer alır ve seri sola çarpıktır.

Şekil.6. 2. Sağa çarpık serinin grafiği (mod $\leq$ medyan $\leq$ $\overline{X}$)



Yığılma soldaki değerlerde, seri sağa çarpıktır.

Şekil.6.3. Simetrik serinin grafiği ($\overline{X}$ =medyan=mod)



Simetrik seride ortalama = medyan = mod'dur yani bu üç değer hepsi aynı yerdedir, çakışıktır.

Bölüm Soruları

1. Mod ve medyanın aritmetik ortalamadan farkı nedir?

2. Bir basit seri oluşturarak, aritmetik ortalama-kareli ortalama-geometrik ortalama-harmonik ortalama değerlerini hesaplayın ve Harmonik ortalama< Geometrik Ortalama<Aritmetik ortalama< Kareli Ortalama eşitsizliğini doğrulayınız.

3. Bir sınıftaki öğrencilerin istatistik dersinden aldıkları notların dağılımı şöyledir:

Notlar	**Öğrenci sayısı**
0-25	18
25-50	20
50-75	9
75-100	2

Notların dağılımından yola çıkarak mod değerini hesaplayınız.

4. Yukarıda serinin medyan değeri nedir?

5. Serideki Aritmetik ortalama, mod, medyan değerlerinden yola çıkarak asimetrisi için nasıl bir yorum yaparsınız?

6. Hangi ortalama türü birden fazla değer alabilir?

7. Medyan hangi durumlarda aritmetik ortalamaya tercih edilir?

7. MERKEZİ DEĞİŞİM ÖLÇÜLERİ

Bu Bölümde Neler Öğreneceğiz?

- Merkezi değişim ölçüleri
- Chebyshev Teoremi

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Merkezi değişim ölçüleri, Chebyshev teoremi	Serilerin yayılımlarını incelemek amacıyla merkezi değişim ölçüleri öğrenilecektir.	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kelimeler

- Sapma
- Varyans
- Standart sapma
- Sheppard düzeltmesi
- Değişim katsayısı
- Chebyshev teoremi

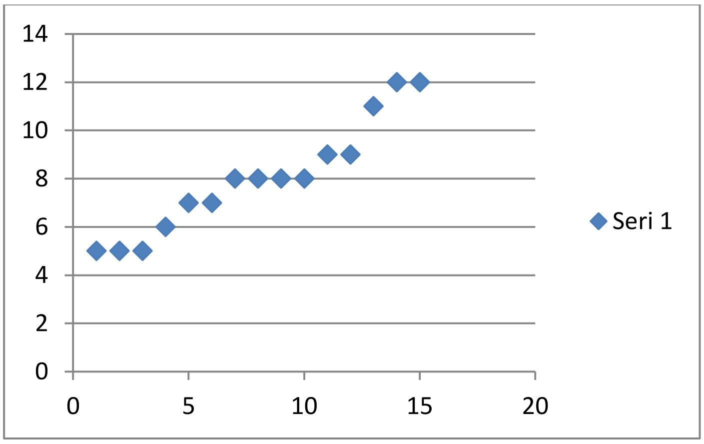
7. 1. Merkezi Değişim Ölçüleri

Bir serideki gözlemlerin (birimlerin) birbirinden ya da herhangi bir ortalama değerden uzaklıklarının çeşitli ölçümlerine merkezi değişim ölçüleri denir. Bir seriyi özetlemede ortalamalr tek başına yeterli bir veri değildirler. Ortalamaların yanısıra değişim ölçülerine de ihtiyaç vardır. Örneğin aşağıda ortalamaları birbiriyle aynı ama dağılımları farklı üç seri göreceksiniz:

Şekil.7.1. Seri 1

Seri1: 5 5 5 6 7 7 8 8 8 8 9 9 11 12 12

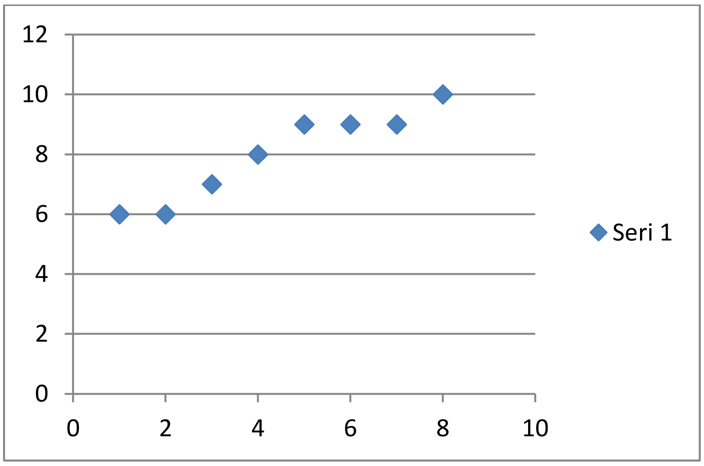
$\bar{X}_1 = 8$



Şekil.7. 2. Seri 2

Seri2: 6 6 7 8 9 9 9 10

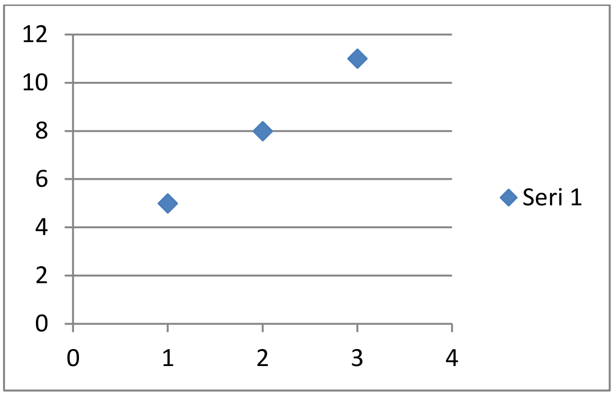
$\bar{X}_2 = 8$



Şekil.7. 3. Seri 3

Seri3: 5 8 11

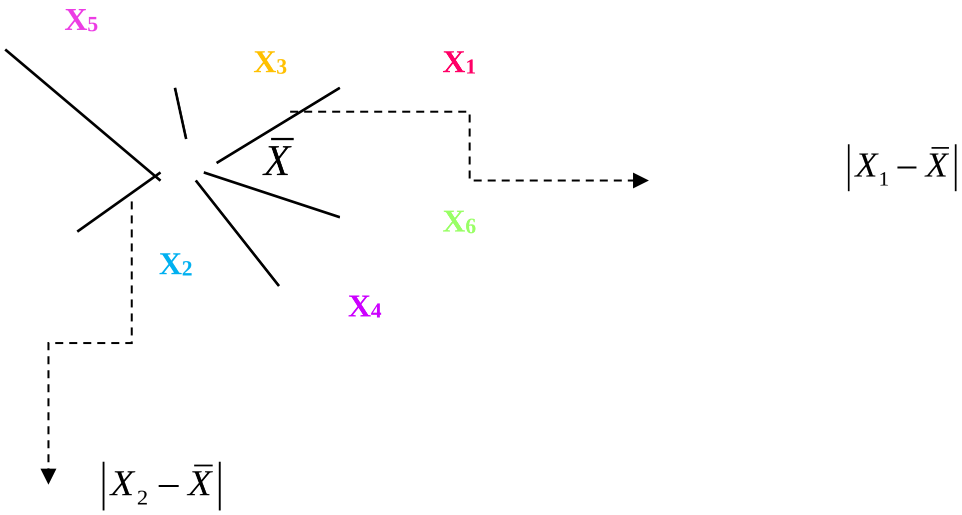
$\bar{X}_3 = 8$



Her üçünün de ortalaması birbiriyle aynı olmasına rağmen, veri sayıları, değişimleri, dağılımları birbirinden farklıdır. Öyleyse artık değişim ölçülerinin neler olduklarına ve nasıl hesaplandıklarına ilişkin açıklamalarımıza geçebiliriz.

Verilerin dağılımını ölçmede, gözlem değerlerinin ortalamadan sapmalarını kullanabiliriz. Bu sapmalar yani $|X_i - \bar{X}|$ ne kadar büyükse, X_i gözlemi ortalamadan o denli uzakta demektir.

Şekil.7. 4. Birimlerin Ortalamadan Uzaklıkları (sapmaları)



7.1.1. Varyans

Şekil. 7.4'e bakalım. Sizce bu sapmalar ne anlama geliyor ve nasıl ölçülecek?

Mesela, $\bar{X}$ 'dan sapmalar ile ölçülebilir mi? Hayır ölçülemez, çünkü ortalamadan sapmaların toplamı daima sıfırdır.

$$\sum(X_i - \bar{X}) = 0$$

Öyleyse biz de sapmaların karelerini alıp işe başlarız yani aritmetik ortalama değil de, kareli ortalamayı kullanırız. Sapmaların kareli ortalaması bulunduğunda, istatistik teorisinde ve uygulamalarda en yaygın kullanılan dağılma ölçüsü olan **standart sapmayı** yani **σ**'yı elde ederiz. Standart sapmanın karesine yani **σ^2**'ye ise **varyans** denir. Varyans değişimin ölçüsüdür.

Varyans kısa yoldan şöyle hesaplanabilir:

$$\sigma^2 = KO^2 - \bar{X}^2$$

7.1.2. Standart Sapma

Standart sapma, varyansın kareköküdür.
Standart sapmanın normal yoldan bulunuşu aşağıdaki gibidir:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}}$$ kitlenin standart sapması

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$ örneklemin standart sapması

σ^2 kitle varyansıdır, s^2 ise örneklem varyansıdır.

7.1.2.1.Sıklık Serisinde Standart Sapma Hesabı

Kitlede:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} f_i (X_i - \mu)^2}{\sum f_i}}$$

Örneklemde:

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

7.1.2.2. Sınıflanmış seride standart sapma hesabı

Kitlede:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} f_i (m_i - \mu)^2}{\sum f_i}}$$

Örneklemde:

$$s = \sqrt{\frac{\sum_{i=1}^{n} f_i (m_i - \bar{X})^2}{n-1}}$$

Sheppard düzeltmesi:

Sınıfları temsilen sınıf ortalarının gelişigüzel tam orta nokta olarak kabul edilmesi (aslında belki verilerin sınıftaki dağılımı tam ortalarda olmayabileceği halde), standart sapma hesabında hataya neden olur. Bunu Sheppard düzeltmesiyle gidermek mümkündür.

$$\sigma^* = \sqrt{\sigma^2 - \frac{c^2}{12}}$$

σ^* düzeltilmiş standart sapmadır. c, ortak sınıf genişliğidir. Standart sapma bulunduktan sonra bu düzeltme yapılırsa, sınıf ortalarını tam orta nokta kabul etmekten kaynaklanan hatanın giderilmiş olduğu düşülmektedir.

Varyansın hesaplanmasıyla ilgili birkaç örnek yapalım:

Örnek: Aşağıdaki basit seride varyans ve standart sapmayı hesaplayınız.

4 5 8 15 18 22

Bu veri kümesini **kitle** gibi düşünürsek kullanmamız gereken formül şudur:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}}$$

Öncelikle kitle ortalamasını yani µ'yü hesaplayalım. Basitçe tüm gözlemleri toplayarak gözlem sayısına bölüyoruz.

$$\mu = \frac{4+5+8+15+18+22}{6} = 12$$

sonra formüle bakıyoruz, dediğimiz gibi formüller konuşurlar ve bize sırasıyla ne yapmamız gerektiğini söylerler. Şimdi kitle ortalaması µ'yü 12 olarak hesapladık. Jher bir gözlem değerini tel tek bu ortalama değerden saptıracağız. Bunun için bir kolon oluşturalım:

Xi- µ
-8
-7
-4
3
6
10

Sonra bu değerlerin tek tek karelerini alıp toplayacağız ve bulduğumuz toplamı n'e yani 6 'ya bölerek varyansı bulmuş olacağız.

$\underline{(Xi- \mu)^2}$

64
49
16
9
36
100

$\sigma^2 = \frac{64+49+16+9+36+100}{6} = \frac{274}{6} = 45.667$

Standart sapma ise varyansın karekökü olduğuna göre;

σ = 6.758'dir.

Bu veri kümesini **örneklem** gibi düşünürsek kullanmamız gereken formül şudur:

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

s^2 = 274/ 5 = 54.8

s = 7.403 olarak elde edilir.

Uyarı!!! örneklem varyansında paydada (n-1) kullandığımız için, kitle varyansına göre bir miktar daha büyük bir değer buluyoruz. Böyle yapmamızın nedeni, örneklem üzerinden çalışıldığında bir miktar hataya göz yummak demektir. Halbuki kitle üzerinden yapılan hesaplarda yani parametre hesaplarken tamsayım söz konusu olduğu için böyle düzeltmelere ve önlemlere gerek yoktur, zira hesaplanan değerler kesindir.

Örnek: Aşağıdaki sıklık serisinde varyansı hesaplayınız.

Xi	fi
4	2
6	1
8	4
10	3

Bu seri aslında şu basit serinin toplulaştırılmış halidir:

4	4	6	8	8	8	8	10	10	10

Şimdi varyansı bulalım. İlk yapmamız gereken şey ortalamayı bulamaktır. İsterseniz yine bu veri kümesini bir örneklem gibi kabul edelim;

<u>fiXi</u>
8
6
32
30

$$\bar{X} = \frac{\sum_{i=1}^{4} fiX_i}{10} = \frac{76}{10} = 7.6$$

Formülümüz şöyle; $s = \sqrt{\dfrac{\sum_{i=1}^{n} f_i (X_i - \bar{X})^2}{n-1}}$

Şimdi tüm X değerlerini ortalamadan saptıracağız:

$X_i - \bar{X}$

4-7.6=-3.6
6-7.6=-1.6
8-7.6=0.4
10-7.6=2.4

Şimdi bu farkların karelerini alalım:

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

12.96
2.56
0.16
5.76

Şimdi bu karesel ifadeleri sıklıklarla ağırlıklandıracağız yani çarpacağız;

$fi\ (X_i-\bar{X})^2$

2*(12.96)= 25.92
1*(2.56)= 2.56
4*(0.16)= 0.64
3*(5.76)=17.28

Artık hepsini toplayıp (n-1)'e bölebiliriz ve böylece örneklem varyansını bulmuş oluruz:

$$s^2 = \frac{25.92+2.56+0.64+17.28}{9} = \frac{46.4}{9} = 5.156$$

s= 2.27'dir.

Varyansın Formülünü İnceleyelim

Varyans kelimesi "variation" yani değişim anlamına gelmektedir. İstatistiksel olarak yorumlarsak, bir veri kümesindeki gözlemlerin ortalamadan uzaklıklarının bir ölçüsüdür diyebiliriz. Varyansı büyük bir seride birimlerin ortalamadan farklı (ortalamaya uzak) değerlere sahip olduğunu görürüz. Varyansın küçük olması ise gözlemlerin ortalama dolayında dağılım gösterdiği durumlarda meydana gelir. Bu anlatılanlardan şunu anlamalıyız, eğer bir serideki gözlem birimleri ortalamadan uzakta çok farklı değerler alıyorlarsa bu serinin varyansı (değişimi, yayılımı), ortalama dolayında bir dağılıma sahip olan bir veri kümesinin varyansına göre daha büyüktür. Formülün paydasında "n" vardır, yani birim sayısı arttıkça varyans küçülmektedir, varyansın yani değişimin küçülmesi ise ileriki konularda anlatacağımız parametre tahminlerinin daha isabetli daha az hata ile yapılmasını sağlayacak bir unsurdur.

7.1.3. Değişim Katsayısı

Bir seriyi özetlemede sadece ortalamanın yeterli olamayacağından bahsetmiştik. Değişim katsayısı, ortalama ile varyansı birlikte kullanarak serilerin değişimlerini ölçmeyi ve farklı türden olsalar bile kıyaslama imkânı sağlayan bir ölçüttür. Değişim katsayısının formülü aşağıdaki gibidir:

$$DK = \frac{\sigma}{\mu} * 100 \quad \text{veya,} \quad DK = \frac{\sigma}{\overline{X}} * 100$$

Değişim katsayısı ölçü birimlerinden bağımsızdır. Oransal yapısından dolayı, pay ve paydasında aynı cins ve büyüklükten değerler birbirini götüreceği için basit seri, sıklık serisi ve sınıflanmış seriler arasındaki cins ve büyüklük farklılığını ortadan kalkar.

DK'sı küçük olan serilerin, diğerlerine göre daha az değişken olduğu (daha homojen yani ortalamaya yakın dağılıma sahip) birimlerden oluştuğu söylenir. Çünkü standart sapma küçüldükçe DK da küçülür. Standart sapmanın küçük olması ise ancak ortalama etrafındaki saçılımın $\overline{X}$ 'ya doğru çekilmesiyle mümkündür birimler $\overline{X}$ 'dan uzaklaştıkça standart sapma da büyümektedir.

Örnek: İki hisse senedi olsun. Bu hisselerin 1 aylık ortalama getirileri ve standart sapmaları aşağıdaki gibi ise, siz bir portföy yöneticisi olarak **riski seven** bir yatırımcıya hangi hisse senedini tavsiye edersiniz?

$\mu_1=70$ $\sigma_1=6$
$\mu_2=25$ $\sigma_2=3$

$$DK_1 = \frac{6}{70} * 100 = 8,57$$

$$DK_2 = \frac{3}{25} * 100 = 12$$

DK sı büyük olanı yani 2. hisse senedini tercih etmelidir. Şayet riski sevmeyen yatırımcı olsaydı, ona değişimi görece daha az olan ilk hisse senedini almasını tavsiye ederdik. Borsada riskli senetlerin getirisi de kaybı da büyüktür. Ama riski seven yatırımcı "risk yoksa getiri de yoktur" mantığına sahip olduğu için bu onun tercihidir. Riski düşük olan senetler her ne kadar güvenli olsa da, maalesef getirisi de düşüktür. Elbette bu bir tercih meselesidir.

7.2. Chebyshev Teoremi

Ortalama ve standart sapmanın birlikte kullanıldığı bir durum da, verilerin yüzde kaçının hangi aralıkta bulunduğunu ölçmeye yarayan Chebyshev Teoremi'dir. Ortalaması μ, standart sapması σ olan bir veri kümesindeki gözlemlerin $\mu \pm k\sigma$ aralığına düşenlerin oranı en az $1-\frac{1}{k^2}$ 'dir. **Burada k, 1'den büyük bir sayıdır**.

Örnek: Bir lisedeki öğrencilerin IQ ortalaması 105, standart sapması 9'dur. Öğrencilerin en az % kaçı 85,2 ile 124,8 arasında IQ'ya sahiptir?

$$\mu - k\sigma = 85,2$$
$$\mu + k\sigma = 124,8$$

Buradan k=2,2 elde edilir.

$$1-\frac{1}{(2,2)^2} = 0,79$$

Yorum: Öğrencilerin en az %79'u verilen aralıkta IQ skoruna sahiptir.

Bölüm Soruları

1. 12, 15, 4, 8,4 rakamlarından oluşan serinin varyans ve standart sapma değerlerini hesaplayınız.

2. Risk almak istemeyen bir yatırımcı için hangi hisse senedini önerirsiniz?

   A hisse senedi ortalama getiri 14 , standart sapması 9

   B hisse senedi ortalama getiri 12 standart sapması 4

3. Aşağıdaki sınıflanmış seride varyans, standart sapma ve değişim katsayısını hesaplaynız:

Sınıflar	fi
10-20	2
20-30	4
30-40	7
40-50	2

4. Yukarıdaki soru için düzeltilmiş varyansı hesaplayınız. (Sheppard düzeltmesi)

5. Sheppard düzeltmesi ne tür serilerde yapılır? Neden ihtiyaç duyulur?

6. Aritmetik ortalaması 10, varyansı 11 olan bir seride değişim katsayısı yüzde kaçtır?

7. Aşağıdaki sıklık serisinde değişim katsayısını hesaplayınız.

Xi	fi
10	4
11	2
14	3
16	1

8. STANDART DEĞİŞKEN, ASİMETRİ İNCELEMELERİ, TOPLANMA ORANI

Bu Bölümde Neler Öğreneceğiz?

- Standart değişken
- Verilerin normal eğri altındaki dağılımları
- Çarpıklık (asimetri) incelemeleri
- Toplanma oranı

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Standart değişken, verilerin nromal eğri altındaki dağılımları ve yığılmaları, asimetri incelemeleri, Toplanma oranı	Standart değişken yoluyla verilerin normal eğri altındaki dağılımları incelenecek, verilerin yığılmaları yani asimetrisine bakmak öğrenilecektir. Son olarak da adil dağılımın ölçümü demek olan Toplanma oranı yani Gini katsayısının ölçülmesine değinilicektir.	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kelimeler

- Standart değişken
- Simetrik seri
- Sola çarpık seri
- Sağa çarpık seri
- Pearson çarpıklık katsayısı
- Bowley çarpıklık katsayısı
- Toplanma oranı

8. 1. Standart Değişken

Standart değişken z ile gösterilir.

Kitlede;

$$z = \frac{X_i - \mu}{\sigma}$$

Örneklemde;

$$z = \frac{X_i - \overline{X}}{s}$$

X_i : her hangi bir gözlem değeri
μ:kitlenin aritmetik ortalaması
σ: kitlenin standart sapması
$\overline{X}$: örneklemin aritmetik ortalaması
s: örneklemin standart sapması

z bize, X_i'nin ortalamadan kaç standart sapma uzaklaştığını gösterir.

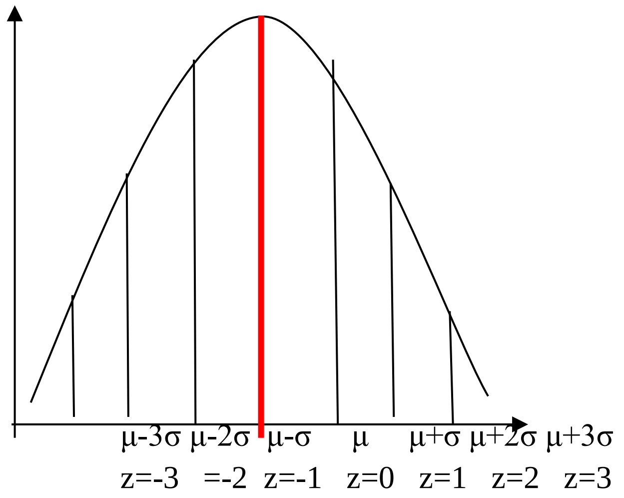
z=0 ise $\overline{X} = \mu$

z>0 ise $\overline{X} > \mu$

z<0 ise $\overline{X} < \mu$

Mesela z=2 bulmak, X_i'nin ortalamadan büyük olduğunu ve ortalamadan 2 σ uzakta olduğunu gösterir.

Şekil.8.1. Verilerin Normal Eğri Altındaki Dağılımları



Verinin yaklaşık %68'i, $\mu \pm \sigma$ yani z=-1 ile z=1 arasında yer alır. %95'i $\mu \pm 2\sigma$ yani z=-2 ile z=2 arasında yer alır. %99'u ise $\mu \pm 3\sigma$ yani z=-3 ie z=3 arasında yer alır.

Örnek: Yabancı uyruklu bir işçi etnik ayrımcılık yapıldığından ötürü kendisinin aynı işi yapmakta olduğu diğer çalışanlardan daha düşük gelir elde ettiğini düşünmektedir. Bu iddiasını test etmek amacıyla bilgi toplamış, kendisiyle aynı işi yapan diğer çalışanların yıllık ortalama gelirlerinin 40000 tl ve s=4000 tl olarak hesaplamıştır. Kendisinin yıllık ortalama geliri 36000 tl olduğuna göre, sizce iddiasında haklı mıdır?

$$z = \frac{36000-40000}{4000} = -1$$

Ortalamadan σ kadar daha aşağıda bir gelir elde etmektedir. İddiasında haklıdır.

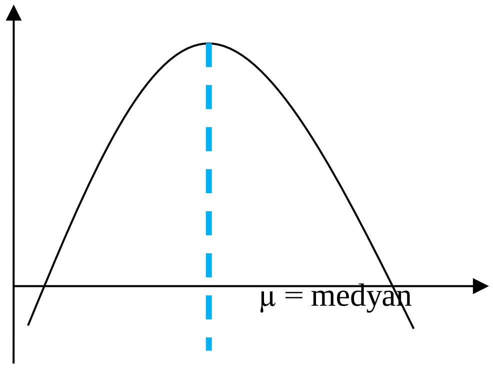
Şimdi artık öğrendiğimiz duyarlı, duyarsız ortalamalar ve standart sapmayı kullanarak serinin asimetrisi (çarpıklığı) yani verilerin yığılma yönlerini bulmayı ve bu yığılmanın miktarını ölçmeye geçebiliriz.

8.2. Pearson Çarpıklık Katsayısı

$$Pearson = \frac{3(\mu - medyan)}{\sigma}$$

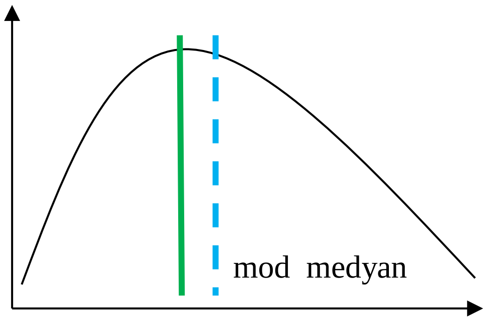
P=0 ise seri simetrik bir dağılıma sahiptir. Çarpıklık sıfırdır. Formüle bakarsanız P'nin sıfır olması, ortalama ve medyanon birbirine eşit olmasına bağlıdır. Bu iki değer eşit olduğunda biliyoruz ki seri simetriktir.

Şekil.8.2. Simetrik Seri



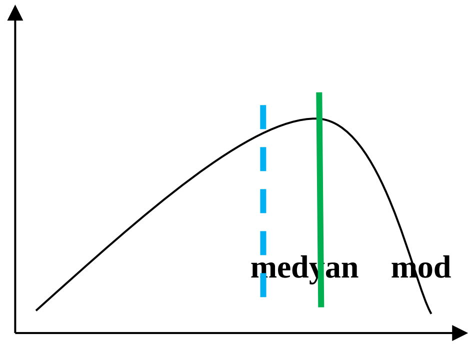
P>0 ise sağa çarpıktır. P'nin pozitif olması ortalamanın medyandan büyük olması yani medyanın sağında olması demektir. O halde mod (seride en yüksek sıklığa sahip değer, serinin yığılma yaptığı değer) medyanın solunda yer alır, bu durumda pozitif asimetri söz konusudur, yani seri sağa çarpıktır.

Şekil.8.3. Sağa Çarpık Seri



P<0 ise sola çarpıktır. P'nin negatif olması ortalamanın medyandan küçük olması yani medyanın solunda olması demektir. O halde mod (seride en yüksek sıklığa sahip değer, serinin yığılma yaptığı değer) medyanın sağında yer alır, bu durumda negatif asimetri söz konusudur, yani seri sola çarpıktır.

Seri.8.4. Sola Çarpık Seri



Örnek: $\overline{X}$ =100, σ=4, medyan=103 ise Pearson çarpıklık ölçütünü hesaplayarak serinin asimetrisi hakkında yorum yapınız.

P= $\frac{3*(100-103)}{4}$ = -2.25 > 0 , sağa çarpık seri

8.3. Bowley Çarpıklık Katsayısı

Bowley asimetri ölçütü, kartilleri kullanarak ölçüm yapar. Formülü şöyledir:

$$Bowley = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

$$-1 \leq Bowley \leq 1$$

B=0 ise seri simetriktir, B>0 ise sağa çarpık, B<0 ise sola çarpıktır.

Örnek:
Q_1=90
Q_2=medyan=120
Q_3=125 ise Bowley çarpıklık katsayısını hesaplayarak serinin asimetrisi hakkında yorum yapınız.

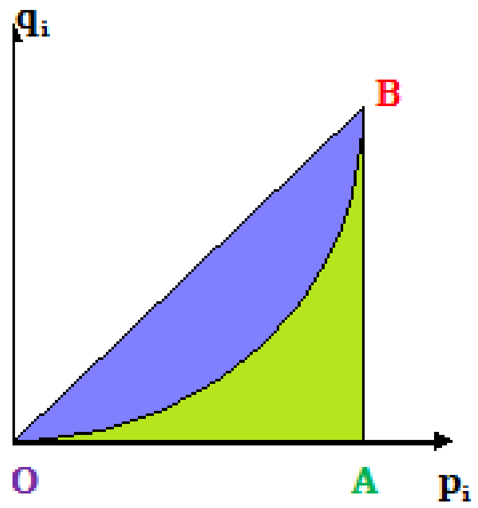
B = $\frac{125+90-2*120}{125-90}$ = -0.714 < 0 , sola çarpık seri.

8.4. Toplanma Oranı (Gini Katsayısı)

Toplanma oranı ölçü birimine sahip değildir. $0 \leq TO \leq 1$'dir. Yukarıdaki şekle bakınız, burada 45^0 lik doğrunun hemen altındaki eğri ile arasında kalan mor renkli alanın, OAB ikizkenar üçgeninin alanına oranı bize toplanma oranını vermektedir. *Toplanma oranı neticede bir alan olduğundan, daima pozitif değer almaktadır. Negatif değer alamaz!!!*

TO'nun sıfıra yakın çıkması yani 45^0 lik doğru ile alttaki eğrinin arasındaki alanın küçülmesi adil bölüşümü gösterir.

Şekil. 8.5. Toplanma Oranı (Gini Katsayısı)



$$TO = \frac{\sum_{i=1}^{k-1}(p_i - q_i)}{\sum_{i=1}^{k-1} p_i}$$

P_i: ilgili sınıfın birikimli sıklığı/$\sum f_i$

q_i: ilgili sınıfın birikimli $f_i X_i$ değeri/$\sum f_i X_i$

1'e yakın çıkması ise söz konusu alanın giderek büyümesi, 45^0 lik tam eşitlik doğrusundan uzaklaşıldığını ve adaletsiz bir bölüşüm olduğunu gösterir. OAB eğrisi tam adaletsiz bölüşümü simgeler. Tam eşit bölüşümde OB doğrusu ile eğri çakışıktır, yani aradaki alan sıfırdır.

Küreselleşen ekonomide, Türkiye’nin de içinde olduğu gelişmekte olan ekonomilerde (emerging markets) gelir dağılımının adaletsiz olduğu bilinmektedir. Birleşmiş Milletlerin 2006 yılında yayınladığı Gelir Dağılımı Eşitsizliği Raporun’a göre, araştırmaya konu 122 ülkenin Gini katsayıları ortalaması 0.40 olarak gerçekleşirken, Türkiye’nin gini katsayısı dünya ortalamasının üzerinde kalarak 0.42 seviyesinde olduğu belirtilmektedir. Türkiye’deki gelir dağılımı, dünyanın diğer 45 ülkesinden daha adil, 76 ülkesinden daha adaletsiz olduğu görülmektedir.

En Adaletsiz Gelir Dağılımına Sahip On Ülke(Gini Katsayıları En Büyük Olan On Ülke)

1. Namibya Gini Katsayısı: %70,7	2. Lesotho Gini Katsayısı: %63,2	3. Botsvana Gini Katsayısı: %63
4. Siearra Leone Gini Katsayısı: %62,9	5. Afrika Gini Katsayısı: %61,3	6. Bolivya Gini Katsayısı: %60,6
7. Brezilya Gini Katsayısı: %59,7	8. Güney Afrika Gini Katsayısı: %59,3	9. Şili Gini Katsayısı: %57,1
10. Paraguay Gini Katsayısı: %56,8		

Dünyanın en fakir %20′lik kesimin dünyadaki gelirin ortalama olarak %6,3′ünü aldığı görülmektedir. Aynı verilere göre Türkiye’de en fakir %20′lik kesimin gelirden aldıkları payların ortalaması %5,8 olarak gerçekleşmiştir. Dolayısıyla, Türkiye’nin en fakir %20’sinin gelirden aldığı pay, dünya ortalaması olan %6,3′ten düşük olduğu görülmektedir.

En Adil Gelir Dağılımına Sahip On Ülke (Gini Katsayıları En Düşük Olan On Ülke)

1. Danimarka Gini Katsayısı: %23,2	2. Macaristan Gini Katsayısı: %24,4	3. Belçika Gini Katsayısı: %25
4. İsveç Gini Katsayısı: %25	5. Çek Cumhuriyeti Gini Katsayısı: %25,4	6. Slovakya Gini Katsayısı: %25,8
7. Norveç Gini Katsayısı: %25,8	8. Bosna Hersek Gini Katsayısı: %26,2	9. Özbekistan Gini Katsayısı: %26,8
10. Finlandiya Gini Katsayısı: %26,9		

Ve Türkiye 76. Sırada, Gini katsayısı %42’dir. Dünyanın 122 ülkesinde en zengin %20′lik dilimlerin gelirden aldıkları payların ortalaması %46,7 düzeyinde gerçekleşmiştir. Aynı verilere göre Türkiye’de en zengin %20′lik dilimlerin gelirden aldıkları payların ortalaması yüzde %47,7 olduğu görülmektedir. Türkiye’nin en zengin %20’sinin gelirden aldığı pay, dünya ortalaması olan %46,7′den yüksek olarak gerçekleşmiştir. Dolayısıyla Türkiye’de gelir dağılımı adaletsizliği, dünya ortalama gelir dağılımı adaletsizliğinden yüksek olduğu görülmektedir.

Bölüm Soruları

1. Bir öğrencinin istatistik ve matematik derslerinden aldığı notlar sırasıyla 75 ve 82'dir. İstatistik dersindeki notların ortalaması 57 ve standart sapması 9'dur. Matematik dersinde ise ortalama 70 ve standart sapma 16 ise öğrenci hangi dersten daha başarılıdır?

2. Bir sınıflandırılmış sıklık serisi oluşturarak asimetri için Bowley ve Pearson ölçülerini hesaplayınız.

3. Milli geliri görece daha düşük olan bir ülkenin Gini katsayısının, daha gelişmiş ve zengin bir ülkenin Gini katsayısından daha düşük olabilmesi durumunu nasıl açıklarsınız? (örneğin ABD'nin en düşük Gini katsayıları listesinde ilk 10'da bulunmaması)

4. Aşağıdaki sıklık serisi için Pearson ve Bowley asimetri ölçümlerini hesaplayarak yorumlayınız:

Xi	fi
10	4
11	2
14	3
16	1

5. Bir şirketin yedi günlük kar miktarları aşağıdaki gibidir. Asimetrisini ölçünüz.

-10	-5	20	30	35	45	55

9. MOMENTLER

Bu Bölümde Neler Öğreneceğiz?

- Sıfıra göre ve ortalamaya göre momentler
- Momentler yoluyla çarpıklık ve basıklık ölçümleri
- König Teoremi

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Sıfıra göre ve ortalamaya göre momentler	Momentler yoluyla çarpıklık ve basıklık ölçümlerini öğreneceğiz.	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kelimeler

- Sıfıra göre momentler
- Ortalamaya göre momentler
- Çarpıklık
- Basıklık
- König Teoremi

9.1. Momentler

Moment, bir rastlantı değişkeninin sıfıra göre veya ortalamaya göre çeşitli kuvvetlerinin beklenen değeridir.

Sıfıra göre demek, sıfırdan sapmaların kuvvetlerinin beklenen değeri demektir. Ortalamaya göre moment ise, ortalamadan sapmaların kuvvetlerinin ortalamalarıdır.

9.1.1.Sıfıra Göre Momentler (basit moment)

$$M_R = \frac{\sum X_i^R}{N} \qquad M_R = \frac{\sum f_i X_i^R}{\sum f_i} \qquad M_R = \frac{\sum f_i m_i^R}{\sum f_i}$$

Sıfıra göre 1. Moment aritmetik ortalamaya eşittir.

$$M_1 = \frac{\sum X_i}{N} = \overline{X}$$

Sıfıra göre 2. Moment, kareli ortalamanın karesine eşittir.

$$M_2 = \frac{\sum X_i^2}{N} = KO^2$$

Sıfıra göre 3. Moment aşağıdaki gibi hesaplanır:

$$M_3 = \frac{\sum X_i^3}{N}$$

Sıfıra göre 4. Moment aşağıdaki gibi hesaplanır:

$$M_4 = \frac{\sum X_i^4}{N}$$

9.1.2.Ortalamaya Göre Momentler

$$\mu_R = \frac{\sum (X_i - \bar{X})^R}{N} \qquad \mu_R = \frac{\sum f_i (X_i - \bar{X})^R}{\sum f_i}$$

$$\mu_R = \frac{\sum f_i (m_i - \bar{X})^R}{\sum f_i}$$

Basit seride ortalamaya göre 1. Moment sıfıra eşittir.

$$\mu_1 = \frac{\sum (X_i - \bar{X})^1}{N} = 0$$

Ortalamaya göre 2. Moment varyansı verir.

$$\mu_2 = \frac{\sum (X_i - \bar{X})^2}{N} = \sigma^2$$

Ortalamaya göre 3. Moment aşağıdaki gibidir:

$$\mu_3 = \frac{\sum (X_i - \bar{X})^3}{N}$$

Ortalamaya göre 4. Moment aşağıdaki gibidir:

$$\mu_4 = \frac{\sum (X_i - \bar{X})^4}{N}$$

Momentler temel olarak şunu ölçerler:

1. momentler; merkezi konum ölçüsü (ortalama)
2. momentler; varyans
3. momentler; çarpıklık
4. momentler; basıklık.

9.2. Momentler Yoluyla Çarpıklık Ölçümü

Çarpıklık ve basıklık ölçüleri bir dizideki gözlem değerlerinin saçılımının (dağılımının) şeklini ortaya koyarlar. Bu ölçüler yorumlanırken normal dağılımın özellikleri referans alınır. Çünkü Normal dağılım eğrisi tam simetrik ve normal (ideal) bir basıklığa sahiptir. Çarpıklık ölçüsü serinin frekans (sıklık) dağılımının simetrik dağılımdan uzaklaşma derecesini gösterirken, basıklık ölçüsü gözlemlerin normal dağılım eğrisinin ne kadar altında ya da üstünde dağılıma sahip olduğunu gösteren ölçülerdir.

Momentler yoluyla çarpıklık ve basıklık aşağıdaki gibi hesaplanır. Çarpıklık katsayısı α_3'tür.

$$\alpha_3 = \frac{\mu_3}{\sigma^3}$$

$\alpha_3 < 0$ ise sola çarpık

Veya

$$\text{Çarpıklık} = \frac{3(\bar{X} - \text{Medyan})}{s}$$

$$\text{Çarpıklık} = \frac{(\bar{X} - \text{Mod})}{s}$$

Burada s, standart sapmadır. Şayet;

$\alpha_3 = 0$ ise simetrik. Simetrik dağılım gösteren serilerde merkezi eğilim ölçüleri, dağılımın tam ortasında yer alır. Gözlemlerin %50 'si merkezi eğilim ölçüsünden büyük, diğer yarısı ise küçüktür.

$\alpha_3 > 0$ ise sağa çarpıktır.
$\alpha_3 < 0$ ise sola çarpıktır.

9.3. Momentler Yoluyla Basıklık Ölçümü

Basıklık katsayısı α_4 'tür.

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{\mu_4}{(\text{var } yans)^2}$$

$\alpha_4 < 3$ ise normalden basık

$\alpha_4 = 3$ ise normal basıklıkta veya sivrilikte (ideal sivrilik, mesela normal dağılımın basıklığı 3'tür)

$\alpha_4 > 3$ ise normalden sivri

Örnek: Aşağıdaki serinin sıfıra göre ve ortalamaya göre momentlerini hesaplayınız. Çarpıklık ve basıklığını araştırınız.[19]

Basit momentler (yani sıfıra göre momentler) aşağıdaki gibidir:

	Xi	Xi^2	Xi^3	Xi^4
	1	1	1	1
	3	9	27	81
	4	16	64	256
	6	36	216	1296
	10	100	1000	10000
	12	144	1728	20736
toplam	36	306	3036	32370
ortalama	6	51	506	5395
basit momentler	**M1**	**M2**	**M3**	**M4**

Ortalamaya göre momentler aşağıdaki gibidir:

	Xi- Xort	(Xi-Xort)^2	(Xi-Xort)^3	(Xi-Xort)^4
	-5	25	-125	625
	-3	9	-27	81
	-2	4	-8	16
	0	0	0	0
	4	16	64	256
	6	36	216	1296
toplam	0	90	120	2274
ortalama	0	15	20	379
basit momentler	µ1	µ2	µ3	µ4

$$\mu_2 = \sigma^2 = 15$$

$$\sigma = \sqrt{15} = 3{,}872$$

19 http://www.google.com.tr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CBoQFjAA&url=http%3A%2F%2Fberkayvaz.files.wordpress.com%2F2013%2F10%2Folasilik-ve-istatistik-i_4-hafta.pptx&ei=87bHU7SgI8e6ygPSr4KQDg&usg=AFQjCNEIvmpReOsRZr_GidFPtudI0VAv7g&bvm=bv.71198958,d.bGQ (Öğr. Gör. Berk Ayvaz, İTİCÜ Mühendislik ve Tasarım Fakültesi Endüstri Mühendisliği Bölümü)

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{20}{3{,}872^3} = 0{,}34 > 0 \quad \text{sağa çarpık seri}$$

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{379}{(15)^2} = 1{,}68 < 3 \quad \text{normalden basıktır}$$

Örnek: Aşağıdaki serinin sıfıra göre ve ortalamaya göre momentlerini hesaplayınız. Çarpıklık ve basıklığını araştırınız.[20]

Sıfıra göre momentler aşağıdaki gibidir:

	Xi	fi	fiXi	Xi^2	fiXi^2	Xi^3	fiXi^3	Xi^4	fiXi^4
	2	3	6	4	12	8	24	16	48
	3	6	8	9	54	27	162	81	486
	4	4	16	16	64	64	256	256	1024
	6	7	42	36	252	216	1512	1296	9072
toplam		20	82		382		1954		10630
ortalama			**4,1**		**19,1**		**97,7**		**531,5**
			M1		**M2**		**M3**		**M4**

[20] A.g.e.

Ortalamaya göre momentler aşağıdaki gibidir:

	Xi-Xort	i(Xi- Xort)	Xi- Xort)^2	i(Xi- Xort)^2	Xi- Xort)^3	i(Xi- Xort)^3	Xi- Xort)^4	i(Xi- Xort)^4
	-2,1	- 6,3	4 ,41	1 3,23	- 9,261	- 27,783	1 9,4481	5 8,3443
	-1,1	- 6,6	1 ,21	7 ,26	- 1,331	- 7,986	1, 4641	8, 7846
	-0,1	- 0,4	0 ,01	0 ,04	- 0,001	- 0,004	1 E-04	0, 0004
	1,9	1 3,3	3 ,61	2 5,27	6 ,859	4 8,013	1 3,0321	9 1,2247
to plam		0		4 5,8		1 2,24		1 58,354
or talama		**0**		**2 ,29**		**0 ,612**		**7, 9177**
		μ 1		**μ 2**		**μ 3**		**μ 4**

$$\mu_2 = \sigma^2 = 2,29$$

$$\sigma = \sqrt{2,29} = 1,51$$

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{0,612}{1,51^3} = 0,177 > 0 \quad \text{sağa çarpık}$$

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{7,9177}{(2,29)^2} = 1,52 < 3 \ \text{Normalden basık bir seri}$$

Örnek: Aşağıdaki serinin sıfıra göre ve ortalamaya göre momentlerini hesaplayınız. Çarpıklık ve basıklığını araştırınız.[21]

Sıfıra göre momentler aşağıdaki gibidir:

sınıflar	i	i	fimi	mi^2	fimi^2	mi^3	fimi^3	mi^4	fimi^4
-2			2	1	2	1	2	1	2
-4			3	9	9	27	27	81	81
-6			40	25	200	125	1000	625	5000
-8			35	49	245	343	1715	2401	12005
		toplam	80		456		2744		17088
		ortalama	5		28,5		171,5		1068
			M1		M2		M3		M4

Ortalamaya göre momentler aşağıdaki gibidir:

sınıflar	i	mi	mi-Xort	fi(mi-Xort)	(mi-Xort)^2	fi(mi-Xort)^2	(mi-Xort)^3	fi(mi-Xort)^3	fi(mi-Xort)^4
-2		1	4	-8	16	32	64	-128	512
-4		3	2	-2	4	4	8	-8	16
-6		5		0	0	0		0	0
-8		7		10	4	20		40	80
		toplam		0	24	56	64	-96	608
		ortalama		0		3,5		-6	38

21 A.g.e.

$$\mu_2 = \sigma^2 = 3,5$$

$$\sigma = \sqrt{3,5} = 1,87$$

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{-6}{1,87^3} = -0,917 < 0$$

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{38}{(3,5)^2} = 3,102 > 3$$

Seri, normalden sivri ve sola çarpık bir dağılıma sahiptir.

σ^2'nin sağlaması:

$$\sigma^2 = KO^2 - \bar{X}^2 = M_2 - M_1^2 = 28,5 - 25 = 3,5$$

Örnek: Aşağıdaki gruplandırılmış serinin Pearson çarpıklık katsayısını moda dayalı olarak hesaplayıp yorumlayınız.

X_i	f_i
1-3	1
3-5	2
5-7	4
7-9	3

Çözüm: Serinin modu şöyle hesaplanır;

$$\text{Mod} = L + \frac{\Delta_1}{\Delta_1+\Delta_2} * c = 5 + \frac{2}{2+1} * 2 = 6.33$$

Moda dayalı Pearson çarpıklık ölçütü;

$$\text{Çarpıklık} = \frac{(\bar{X} - mod)}{s} = \frac{5.8-6.33}{1.87} = -0.28$$

Yorum: Seri hafif bir biçimde sola çarpıktır.

9.4. König Teoremi

König teoremi, basit momentlerden ortalamaya göre momentlere geçişi sağlar. İki moment türü arasındaki ilişkiyi aşağıdaki gibi açıklamaktadır:

$$\mu_2 = M_2 - M_1^2$$

$$\mu_3 = M_3 - 3M_1M_2 + 2M_1^3$$

$$\mu_4 = M_4 - 4M_1M_3 + 6M_1^2M_2 - 3M_1^4$$

Bölüm Soruları

1. Sıfıra göre ve aritmetik ortalamaya göre birinci, ikinci, üçüncü ve dördüncü momentler (sadece uygun olanlar) hangi ölçütlere karşılık gelir?

2. Aşağıda yer alan seriden hareketle sıfıra göre 1. 2. 3. ve 4. moment değerlerini hesaplayınız.

Notlar	Öğrenci sayısı
0-20	4
20-40	12
40-60	8
60-80	6
80-100	2

3. Yukarıdaki serinin aritmetik ortalamadan sapmalara göre ilk üç momentini hesaplayınız.

4. König teoremini bu seri için doğrulayınız.

5. Aşağıdaki seride $\alpha 3$ ve $\alpha 4$ hesaplayarak yorumlayınız:

Sınıflar	fi
50-60	1
60-70	3
70-80	7
80-90	4
90-100	2

10. OLASILIĞA GİRİŞ VE TEMEL KAVRAMLAR

Bu Bölümde Neler Öğreneceğiz?

- Olasılığa giriş ve temel kavramlar

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Olasılığa giriş ve temel kavramlar	Olasılığın temel kavramları öğrenilecektir.	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

AnahtarKelimeler

- Rastsal deneme
- Örneklem uzayı
- İhtimal oranı (odds ratio)
- Rastlantı değişkeni
- Toplama kuralı
- Ayrık olaylar
- Çarpma kuralı
- Bağımsız olaylar
- Permütasyon
- Kombinasyon
- Stirling Yaklaşımı

10.1. Olasılığa Giriş – Temel Kavramlar

Bir olayın olasılığı, olayın ne ölçüde ortaya çıkabileceğini ölçen sayısal bir değerdir. Yani, olayın olma şansının ölçüsüdür.

Olasılık; [0,1] aralığında değer alır. Bir A olayının olasılığı P(A)=0 ise, bu olayın olma ihtimali sıfırdır yani imkânsızdır. P(A)=1 ise, bu olayın olma ihtimali %100'dür yani 1'dir ve kesindir. Bir olasılık değerinin 1'den büyük veya negatif bir değer alması söz konusu olamaz.

Rastsal Deneme (deney):

Farklı sonuçları olan bir olayın gözlenmesi sürecidir. Örneğin bir paranın atılması, bir zar atılması, bir iskambil destesinden kart çekilmesi, v.b.

Örneklem Uzayı (sampling space):

Bir deneyin mümkün tüm sonuçlarını kapsayan ve "S" ile gösterilen kümeye örneklem uzayı denir. Olayın mümkün tüm sonuçlarına e1,e2,….en dersek;

S={e1, e2,…en}

P(S)=1

$$\sum_{i=1}^{n} P(e_i) = 1$$

Örneklem uzayının eleman sayısı n(S);

$$n(S) = mümküntümsonuçlar^{tekrarsayısı}$$

Herhangi bir olayın olasılığı daima [0,1] aralığında değer alır, negatif ya da 1'den büyük olasılık yoktur.

0≤P(A)≤1

Örneklem Uzayının oluşturulmasına ilişkin birkaç örnek verelim;
**Zar atma deneyinde mümkün sonuçlar 6 tanedir;
S={1,2,3,4,5,6}
n(S)=6^1=6
**Para atma deneyinde mümkün iki sonuç vardır, ya yazı gelir ya da tura gelir;
S={Y,T}
n(S)=2^1=2

**2 para atma deneyinde örneklem uzayı şöyledir;
S={YY,YT,TY,TT}

Burada $n(S) = mümküntümsonuçlar^{tekrarsayısı}$ formülünü hatırlayalım. İki para peş peşe ya da aynı anda atıldığında mümkün iki sonuç vardır, yapılan deneme sayısı yani atış sayısı 2 olduğu için;

n(S)=2^2=4'tür.

***Bir futbol maçında örneklem uzayı;
S={kazanma, beraberlik, galibiyet}
***A olayı: 2 para atma deneyinde sadece 1 yazı gelme olsun, A olayının örneklem uzayı;
A={YT,TY}

$$P(A) = \frac{2}{4}$$

***B olayı: sadece 2 tura gelmesi olsun, B olayının örneklem uzayı;
B={TT}

$$P(B) = \frac{1}{4}$$

*** C olayı: en az 1 yazı gelmesi, C olayının örneklem uzayı;
C={YT,TY,YY}

$$P(C) = \frac{3}{4}$$

Odds Ratio (İhtimal oranı):

Bazen bir olayın meydana gelme olasılığı, ihtimal(şans) ifadeleriyle tanımlanır. Mesela; A olayı, yarın havanın iyi olma ihtimali 3'e 1'dir dersek, P(A)=3/4'tür.

B olayı, Ali'nin yarınki sınavdan geçme şansı 4'e 1'dir şeklinde tanımlanmış olsun, bu durumda P(B)=4/5'tir.

Rastlantı değişkeni (rassal değişken):

Alacağı değerler bir deneme sonucu ortaya çıkan değişkenlere rastlantı değişkeni denir. Deney sona ermeden alacağı değerleri bilemeyiz, bu nedenle adı tesadüfi (rastlantısal) değişkendir. Mesela; bir maçtaki gol sayısı, bir marketin günlük karı, bir kişinin kilosu, bir zarın atılması, bir paranın atılması, v.b.

Rastlantı değişkeni bir başka şekilde şöyle ifade edilebilir; örneklem uzayındaki örneklem birimlerine reel değerler atayan fonksiyona rassal değişken denir. Rastlantı değişkenlerini genellikle X ile, aldığı değerleri ise x ile gösteririz. X rastlantı değişkeninden bahsederken, kısaltarak X r.d. deriz. X rd bir fonksiyondur demiştik, tanımlanmasına göre kesikli veya sürekli rd olabilir. S örneklem uzayının elemanları sonlu sayıda veya sayılabilir sonsuzlukta ise X rd kesiklidir. Şayet bir aralıkta değer alıyor ve ölçüm sonucu değer alıyor ise bu durumda sürekli rd dir.

İki zar attığımızı düşünelim. Burada X kesikli bir rd dir. Örneklem uzayındaki mümkün tüm sonuçların sayısı n(S)= 6^2'dir.

S={ 1,1.1,2.1,3.1,4……4,1.4,2….6,5.6,6} şeklinde 36 birimden oluşur ve 2 zar attığımızda bunlardan biri gelecektir.

"X rd: üste gelen yüzlerin toplamı 12 olsun" şeklinde tanımlandığında bu koşulu sağlayan tek bir çift vardır, o da (6,6) gelmesidir. O halde P(x=12)=1/36'dır. Gördüğünüz gibi X rd. , S'deki bir örneklem birimini reel bir sayıya bağladı yani ona reel bir değer atanmış oldu.

10.1.1. Toplama Kuralı

İki ayrık olay (ayrık olay aynı anda olması mümkün olmayan, kesişimleri boş küme olan olaylardır) düşünelim, biri n1 farklı şekilde, diğeri n2 farklı şekilde gerçekleşebiliyor olsun. Bu işlemlerden biri veya diğeri n1+n2 farklı şekilde gerçekleşebilir.

Örnek: Bir zar atalım. Kaç farklı şekilde tek ya da çift gelebilir?

Çözüm:

1	3	5	n1=3
2	4	6	n2=3

n1+n2=6 farklı şekilde tek ya da çift gelebilir.

Örnek: Bir torbada 2 Siyah, 4 Kırmızı ve 5 Mavi bilye olsun. Rastgele çekilecek bir bilye için, bu bilye kaç farklı yoldan siyah ya da kırmızı ya da mavi olabilir?

Çözüm:

2 siyah	n1 =2
4 kırmızı	n2=4
5 mavi	n3=5

n1+n2+n3= 11 farklı yoldan çekilebilir.

Ayrık Olaylar:

Aynı anda gerçekleşmesi mümkün olmayan olaylara denir. Örnekler:

*** Bir öğrencinini bir dersten hem geçmesi hem kalması (ya geçer ya kalır)
*** Havanın hem karlı hem de 40 derece olması
*** Beşiktaş –Fenerbahçe maçının berabere yada Beşiktaş'ın şampiyonluğuyla bitmesi, v.b.

A ve B ayrık olaylar olmak üzere, P(AUB) = P(A)+P(B)'dir. Olayların kesişimleri boş kümedir.

Örnek: Bir öğrencinin bir dersten geçme ihtimali (önceki yılların başarı oranı göz önünde bulundurularak) %54, kalma ihtimali ise %46'dır. Bu öğrencinin geçme veya kalma ihtimali nedir?

Çözüm:	A: geçme	B:kalma
	P(A) = 0.54	P(B) = 0.46

A ve B ayrık olaylardır çünkü aynı anda ortaya çıkamazlar, o halde P(AUB) = 0.54 +0.46=1'dir.

10.1.2. Çarpma Kuralı

İki olay düşünelim. İlk olay n1 farklı şekilde, diğeri n2 farklı şekilde meydana geliyor olsunlar. Bu iki olay birlikte n1*n2 farklı şekilde ortaya çıkabilirler.

Örnek: Bir zarı 2 kez atalım. Ortaya çıkacak sonuçlar kaç farklı şekilde gerçekleşebilir?

Çözüm:

n1=6
n2= 6

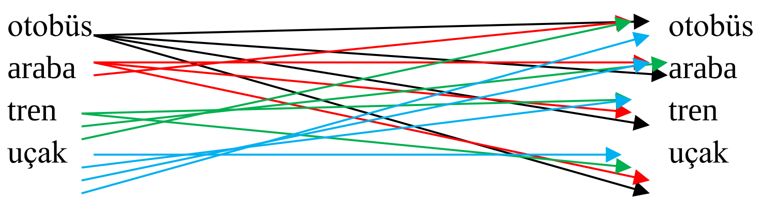
n1*n2 =36 farklı şekilde atış yapılabilir.

Örnek: İstanbul'dan Ankara aktarmalı Erzurum'a gidecek bir yolcu, bu yolculuğu aşağıdaki bilgiler dâhilinde kaç farklı şekilde yapabilir?

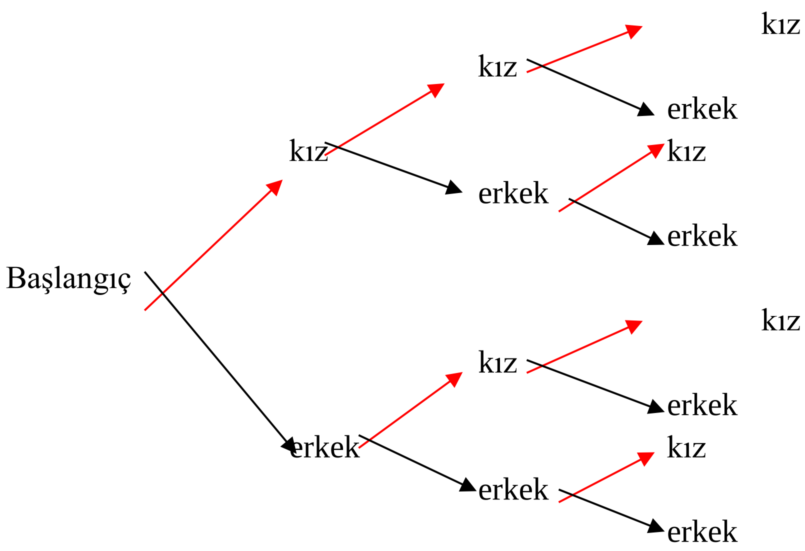
	otobüs		otobüs	
İstanbul	araba	**Ankara**	araba	**E**rzurum
	tren		tren	
	uçak		uçak	

Çözüm: n1=4 , n2=4olduğuna göre; 4*4 = 16 farklı şekilde gidebilir. Aşağıda bu durumun ağaç diyagramı ile gösterimini göreceksiniz:

İstanbul'dan Ankara'ya **Ankara'dan Erzurum'a**



Örnek: Üç çocuklu bir ailede çocukların muhtemel cinsiyetlerinin kaç farklı şekilde gerçekleşebileceğini hesaplayarak ağaç diyagramı ile gösteriniz.



İlk olay 1. Çocuğun olmasıdır ve n1=2 farklı şekilde olabilir. 2. Olay 2. Çocuğun olmasıdır ve bu da yine n2=2 farklı şekilde olabilir. 3. Çocuk da y az kız ya da erkek olacağı için n3=2'dir. Bu durumda olayın toplamı yani üç çocuğun mümkün cinsiyetleri 2*2*2= 8 farklı şekilde olabilir. Şöyle ki, ağaç diyagramından takip edersek, ilk çocuğun kız olması durumundan başlayalım, muhtemel sıralamalar şöyledir: KKK, KKE, KEK, KEE. İlk çocuğun erkek olması durumunda sıralamalar; EKK, EKE, EEK, EEE. Toplamda sekiz farklı sıralama olabilir.

Bağımsız Olaylar:

Bir A olayının olması bir diğer B olayının olmasını etkilemiyorsa böyle olaylar bağımsız olaylardır. A ve B bağımsız olaylar ise;

P(A∩B)= P(A)*P(B) olmalıdır.

Örnek: İlk atışta tura, 2. de Yazı gelmesi?

Çözüm: $\frac{1}{2} * \frac{1}{2}$

Örnek: İki zar atıldığında birinin 2, diğerinin 6 gelmesi?

Çözüm: $\frac{1}{6} * \frac{1}{6}$

Örnek: Bir A noktasından B noktasına 3 farklı şekilde, B'den C'ye 6 farklı şekilde gidilebilmektedir. Peki A'dan C'ye kaç farklı şekilde gidilebilir?

Çözüm: 3*6=18 farklı şekilde gidilebilir.

10.2. Permütasyon

n tane nesnenin r tanesinin veya tamamının sıralama sayısıdır. Diziliş önemlidir. Formülü aşağıdaki gibidir:

$$_nP_r = \frac{n!}{(n-r)!}$$

Örnek: 4 kişi 4 koltuğa kaç farklı şekilde oturabilir?

Çözüm: $P\binom{n}{r} = P\binom{4}{4} = \frac{4!}{(4-4)!}$=4*3*2*1=24 farklı şekilde oturabilirler.

Örnek: 7 kitaptan 3'ü, 3 boşluğa kaç farklı şekilde yerleştirilebilir?

Çözüm: $P\binom{7}{3} = \frac{7!}{(7-3)!}$= 210 şekilde

Örnek: 5 kişi bir yuvarlak masa etrafına kaç farklı şekilde oturabilirler?

Çözüm: Yuvarlak masada yan yana gelme durumundan dolayı özel bir formül vardır, (n-1)! Şeklinde hesaplanır. O halde çözüm 4! Dir.

Örnek: 2 kırmızı, 4 sarı, 8 yeşil mandal bir ipe kaç farklı şekilde sıralanabilir?

Çözüm: bu da yine aynı türden nesnelerin permütasyonu başlığı altında özel bir durumdur ve formülü;

$\frac{n!}{n1!n2!n3!\ldots nk!}$ ‘dir.

n= 14 , n1=2, n2=4, n3=8 olduğundan, sorumuzun cevabı;

$\frac{14!}{2!4!8!}=$ 45045’tir.

Örnek: a, b, c harflerinin 2’li permütasyonları kaç tanedir, dizilişleri yazınız.

Çözüm: $P\binom{3}{2}=6$ tane diziliş vardır. Bunlar ;

ab ac bc ba ca cb

Bu dizilişlere dikkatle bakınız, az önce permütasyonun tanımını yaparken dizilişin önemli olduğundan bahsetmiştik. Bu örnek tam da ne demek istediğimizi anlatan bir örnektir, öyle ki aslında **ab** dizilişi ile **ba** dizilişi aslında aynı şey gibi görünüyor değil mi? Çünkü aynı birimlerden oluşuyorlar, ama permütasyon mantığına göre aynı şey değiller, işte bu yüzden diziliş önemlidir. Kombinasyonda ise aynı ögelerden oluşan dizilişler birbirinin aynısı kabul edilir ve diziliş önemli değildir.

10.3. Kombinasyon

Sıralamaya bakılmaksızın n nesnenin r tanesinin seçilmesidir.

$${}_nC_r=\frac{n!}{r!(n-r)!}$$

Örnek: a, b ve c’nin 2’li kombinasyonları kaç tanedir, yazınız

Çözüm:

$C\binom{3}{2}=3$ tanedir ve bunlar;

ab   ac   bc

Örnek: 52’lik bir desteden 3 kart kaç farklı şekilde çekilebilir?

Çözüm: $C\binom{52}{3} = 22100$

Örnek: 5 evli çift arasından 4 kişilik bir grup kaç farklı şekilde seçilebilir?

Çözüm: $C\binom{10}{4} = 210$

Örnek: 3 evli çiftten 2 kadın 1 erkek olacak şekilde 3 kişilik bir grup kaç farklı şekilde seçilebilir?

Çözüm: 3 kadın $\downarrow$ 2 kadın   3 erkek $\downarrow$ 1 erkek

$\binom{3}{2} * \binom{3}{1} = \frac{3!}{(3-2)!} * \frac{3!}{(3-1)!} =$ 18 şekilde.

Örnek: 5 evli çift arasından 4 kişilik bir grup karı koca aynı grupta olmayacak şekilde kaç farklı biçimde seçilebilir?

Çözüm: 5 kadın, 5 erkek var, karı-koca aynı grupta olmayacak şekilde 4 kişilik gruplar oluşturacağız, eşlere isimler verelim daha anlaşılır olsun:

Hazal	Arda
Selin	Mert
Meltem	Atakan
Selda	Burak
Ayten	Mustafa

Kombinasyonlar aşağıdaki gibidir. Peki bu rakamlar ne demek oluyor biraz açalım:

$$\binom{5}{4}+\binom{5}{1}\binom{4}{3}+\binom{5}{2}\binom{3}{2}+\binom{5}{3}\binom{2}{1}+\binom{5}{4}$$

4 kişilik grubun tamamı kadınlardan oluşursa	4 kişilik grubun 1'i kadın 3'ü erkek olabilir, 5 kadından birini seçiyoruz, diğer 3'ünü ise, seçilen kadının eşi hariç, kalan 4 erkekten seçiyoruz	4 kişilik grubun 2'si kadın 2'si erkek olabilir, 5 kadından 2'sini seçiyoruz, diğer 2'sini ise, seçilen kadınların eşleri hariç, kalan 3 erkekten seçiyoruz	4 kişilik grubun 3'ü kadın 1'i erkek olabilir, 5 kadından 3'ünü seçiyoruz, diğer bir kişiyi ise, seçilen kadınların eşleri hariç, kalan 2 erkekten seçiyoruz	4 kişilik grubun tamamı erkeklerden oluşursa
Örneğin; Selin, Meltem, Ayten, Selda	Hazal, Mert, Atakan, Mustafa (bu grupta Hazal'ın eşi Arda olamaz, o nedenle 4'ün 3'lüsünü alıyoruz.	Meltem, Selda, Arda, Mustafa	Hazal, Selin, Ayten, Burak	Mert, Atakan, Burak, Mustafa

10.4. Stirling Yaklaşımı

n çok büyük olduğunda n! 'i hesaplamak için kullanılır.

$n! \cong \sqrt{2\Pi n}n^n e^{-n}$ ya da;

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

e=2,718
π=3,14

Örnek: n=20 ise, n!=?

$$20! \cong \sqrt{2\Pi 20}\,20^{20} e^{-20} = 2.427 * 10^{18}$$

Hesap makinesiyle hesaplanan ise; 2,432*10^18

Bölüm Soruları

1. Permütasyon ve kombinasyonun farkı nedir?

2. Bir para 5 kez peş peşe atıldığında örneklem uzayının eleman sayısı kaç olur?

3. 7 çocuklu bir ailede çocukları çocukların cinsiyet sıralamasının EKKKKKK (ilk çocuk erkek, diğerleri kız) olma ihtimali nedir?

4. 52'lik bir desteden peş peşe çekilen iki kartın ikisinin de sinek olma ihtimali nedir?

5. Sayısal lotoda 6'lı bilmenin olasılığı nedir?

6. "ERGENE" kelimesinin harfleri kendi aralarında kaç değişik şekilde sıralanabilir?

7. Bir okulda 50 öğretmen ve 380 öğrenci vardır. 5 öğretmen ve 50 öğrenciden oluşan 55 kişilik bir komisyon kaç farklı şekilde oluşturulabilir?

8. "Malatya" kelimesinin harflerine ilişkin permütasyonların sayısı nedir?

9. Hilesiz bir para 2 kez atılıyor. En az 1 tura gelme olasılığı nedir?

10. 20 kalemden 16'si sağlam, 5'i bozuktur. 4 kalemi rastgele seçtiğinizde; her ikisinin de sağlam olması olasılığı nedir?

11. Bir zar üç defa atılıyor. En az birinde 3 gelmesi olasılığı nedir?

12. Atılan bir zarın 2 veya 5 gelmesi olasılığı nedir?

11. ÇEKİLİŞLERLE İLGİLİ BAĞIMLILIK VE BAĞIMSIZLIK KAVRAMLARI, OLASILIĞIN KLASİK VE FREKANS TANIMI, KOŞULLU OLASILIK, OLASILIK FONKSİYONU, DAĞILIM FONKSİYONU

Bu Bölümde Neler Öğreneceğiz?

- Çekilişlerle ilgili bağımlılık ve bağımsızlık kavramları
- Olasılığın tanımları
- Koşullu olasılık
- Olasılık fonksiyonu
- Dağılım fonksiyonu

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
İadeli ve iadesiz çekilişlerde bağımlılık bağımsızlık durumlarının incelenmesi, olasılığın farklı tanımları, koşullu olasılık, kesikli rd nin olasılık fonksiyonu ve dağılım fonksiyonu	İadeli ve iadesiz çekilişlerde bağımlılık bağımsızlık durumlarının incelenmesi, olasılığın farklı tanımları, koşullu olasılık, kesikli rd nin olasılık fonksiyonu ve dağılım fonksiyonunun öğrenilmesi.	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kelimeler

- İadeli çekiliş
- İadesiz çekiliş
- Bağımsızlık
- Bağımlılık
- Olasılığın klasik tanımı
- Olasılığın frekans tanımı
- Koşullu olasılık
- Olasılık fonksiyonu
- Dağılım fonksiyonu

11.1. Çekilişlerle İlgili Bağımlılık Ve Bağımsızlık Kavramları

Çekilişler iadeli ise denemeler bağımsızdır, iadesiz ise bağımlıdır. Top çekme deneyinde çekilen toplar yerine iade edildiğinde her çekiliş sanki ilk çekilişmiş gibi olur yani olasılıklar değişmez, eğer çekilen top yerine konmazsa her defasında top eksileceği için her çekilişte olasılıklar değişir yani olaylar bağımlı hale gelirler.

Örnek: Bir torbada 10 tane top olsun. 3 mavi, 4 turuncu, 3 kırmızı renkte olsunlar. Peş peşe 3 top çekilsin. Bu topların iadeli ve iadesiz durumları ayrı ayrı ele alarak , MTK renkte olması ihtimallerini hesaplayınız.

Çözüm: Önce iadeli durumu ele alalım. İlk topun mavi olması ihtimali 3/10'dur. İkincinin turuncu olması 4/10, üçüncünün kırmızı olması 3/10'dur. Yani sorulan olasılık şöyle bulunur:

$$\frac{3}{10} * \frac{4}{10} * \frac{3}{10} = 36/1000 = 0.036$$

İadesiz durumda,

$$\frac{3}{10} * \frac{4}{9} * \frac{3}{8} = 0.05$$

Uyarı!!! Burada olasılıkları neden toplamıyoruz da çarpıyoruz sizce? Bunun sebebi, olayın ancak peş peşe 3 top çektikten sonra tamamlanmasıdır. Tıpkı bir şehirden bir başka şehre aktarmalı olarak giderken kullanabileceğimiz ulaşım araçları sayısını çarpmamız gibi. Toplama yapmak için olayların her basamağının birbirinden ayrı ve tamamlanmış olması yani ayrık olması gerekir.

11.2. Olasılığın Klasik ve Frekans Tanımı

Olasılık

Klasik yaklaşım

Frekans yaklaşımı

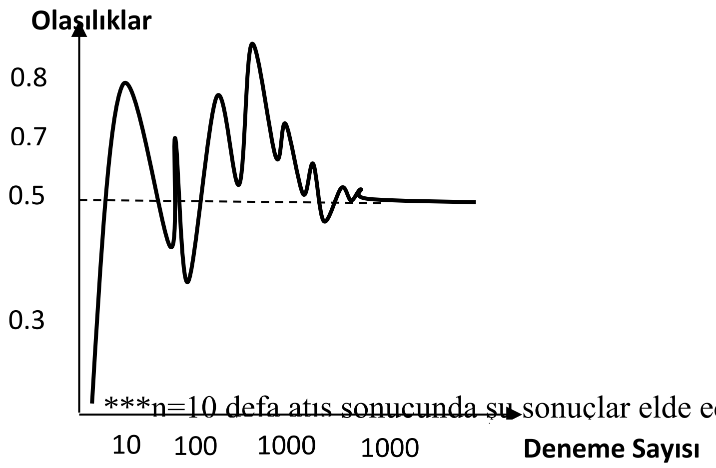
11.2.1. Klasik yaklaşım

Bir örneklem uzayında n tane mümkün sonuç olsun. A olayı S örneklem uzayının bir alt kümesi olmak üzere, A'nın k tane elemanı olsun. Bu durumda klasik yaklaşıma göre P(A)=k/n'dir.

11.2.2. Frekans yaklaşımı

Bir paranın Yazı gelme olasılığı bilindiği üzere %50'dir. Ancak, diyelim ki bir parayı 10 defa atsak 5 defa Yazı 5 defa Tura gelecek diye bir zorunluluk yoktur. Onu da yazı veya dokuzu tura biri yazı, vb. çok farklı sayıda örneklemler ortaya çıkabilir. Fakat, deneme sayısı arttığında yani 500 defa, 1000 defa, 10000 defa atış yapılsa, olasılık bir noktadan sonra stabil hale gelir ve bu noktada olasılık, klasik yaklaşımın öngördüğü 0.5'tir. Bu durum aşağıdaki şekilde anlatılmak istenmiştir:

Şekil.11.1. Olasılığın Klasik ve Frekans Yaklaşımı



***n=10 defa atış sonucunda şu sonuçlar elde edilmiş olsun:

TTYYYYYTYY

Bu örneklemde P(Yazı)=7/10, P(Tura)=3/10'dur.

***Bir başka örneklem şöyle olsun:

YYYYTYYYYT

Bu örneklemde P(Yazı)=8/10, P(Tura)=2/10'dur.

TTYTTTTTTT

Bu örneklemde P(Yazı)=1/10, P(Tura)=9/10'dur.

n=1000 defa olsun,

Örneğin P(Yazı)=468/1000, P(Tura)=532/1000'dir.

Halen klasik yaklaşımın öngördüğü olasılığa ulaşılamadığına göre, deneme sayısı daha da arttırılmalıdır.

Söz gelimi n=5000 olsa, P(Yazı)=2500/5000, P(Tura)=2500/5000 olacaktır. (Tabii kesin bir durum değil, belki de olasılığın stabil (durağan) hale gelmesi 10000 defa deneme gerektirebilir)

Bazen de, bir olayın mümkün sonuçları eşit olasılığa sahip olmayabilir. Mesela atılan zar hileli ise ve çoğunlukla "2" geliyorsa, bu durumda zarın üst yüzüne gelen sayıların olasılıklarını belirlemek için çok sayıda deneme yapmak gerekir. Örneğin hileli bir para düşünelim, bu durumda P(Y)=P(T)=0.5 olması beklenemez. İşte böyle durumlarda;

P(A)=(n denemede A'nın ortaya çıkma sayısı)/n 'dir.

Örnek: Hileli bir para 1000 kez atılıyor. Sonuçlar kaydediliyor. 320 defa Tura, 680 defa Yazı geliyor. Bu durumda P(Yazı)=0.68, P(Tura)=0.32'dir.

Örnek: Bir paranın 3 kez atılması deneyinde B olayı; en az bir kez Tura gelmesi olsun. P(B)=?

n(S)=2^3=8 olası sonuç vardır.
S={YYY,YYT,YTY,TYY,TTY,TYT,YTT,TTT}

P(B)=7/8'dir. Çünkü bahsedilen durumu sağlayan 7 sonuç vardır.

C olayı en fazla 2 Yazı gelme olayı ise P(C)=?

Yani 0,1 ya da 2 yazı gelmesi demektir.
S={YYY,YYT,YTY,TYY,TTY,TYT,YTT,TTT}
P(C)=7/8'dir.

11.3. Koşullu Olasılık

$$P(A/B)=\frac{P(A\cap B)}{P(B)}$$

Yukarıdaki koşullu olasılığın sözel ifadesi şudur; B olayının gerçekleştiği biliniyorken, A olayının meydana gelme olasılığı. Sağ tarafta kalan olay, koşulu yaratan olaydır. Şayet A ile B olayları bağımsız iseler bu defa koşullu olasılık;

$$P(A/B)=\frac{P(A\cap B)}{P(B)}=\frac{P(A)*P(B)}{P(B)}$$ olur,

P(B)'ler birbirini götürür ve; $P(A/B)=P(A)$ olur.

Örnek:

A ve B iki tedavi yöntemidir. Tedavilerin kilolarına göre tasnif edilmiş insanlardaki başarı olasılıkları aşağıdaki tabloda özetlenmiştir:

	Zayıf	**Normal**	**Kilolu**	**Obez**	**Toplam**
A	0,05	0,10	0,15	0,05	0,35
B	0,05	0,20	0,25	0,15	0,65
Toplam	0,10	0,30	0,40	0,20	1,00

a) A yöntemiyle başarı olasılığı nedir?
P(A)=0,35

b) B yöntemiyle başarı olasılığı nedir?
P(B)=0,65

c) Seçilen bir kişinin kilolu olduğu biliniyorken, A yöntemiyle tedavi olma olasılığı nedir?
$P(A/kilolu)= P(A\cap kilolu)/P(kilolu)= 0.15/0.4=0.375$

d) Seçilen bir kişinin zayıf olduğu biliniyorken B yöntemiyle tedavi olma olasılığı nedir?
$P(B/zayıf)= P(B\cap zayıf)/P(zayıf)=0.05/0.10 = 0.5$

e) A yöntemiyle tedavi olduğu biliniyorken bu kişinin obez olma olasılığı?
$P(obez/A)= P(obez\cap A)/P(A)= 0.05/0.35 =0.143$

f) $P(normal/B)= P(normal\cap B)/P(B)= 0.2/0.65 =0.308$

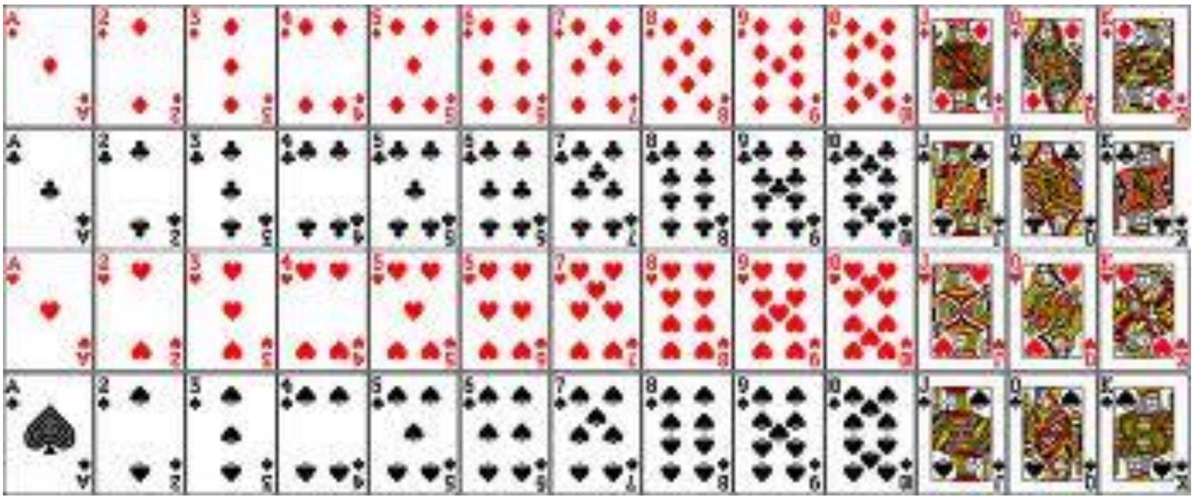
g) $P(A/zayıf)= P(A\cap zayıf)/P(zayıf)= 0.05/0.10 = 0.5$

h) $P(B/normal)= P(B\cap normal)/P(normal) = 0.20/ 0.30 = 0.667$

Örnek: 52'lik desteden rastgele bir kart çekiliyor. Bu kartın kırmızı olduğu biliniyor iken As olması ihtimali nedir?

Çözüm: "A: kırmızı olması", "B:As olması" şeklinde tanımlanıyor, buna göre P(B/A) olasılığı soruluyor.

Şekil. 11.2. İskambil Destesinde Bulunan Kağıtlar



Toplam 52 kart vardır. 4 gruptan oluşur; sinek ve maça (siyah renkliler), kupa ve karo (kırmızı renkliler) A harfiyle gösterilen kâğıdın özel adı "As" tır. Her gurupta birer tane bulunur, 1 yerine geçer. Her grupta toplam 13 kâğıt vardır. J: Vale (11 demektir), Q : Kız (12 demektir), K: Papaz (13 demektir). Yukarıdaki şekilde yukarıdan aşağı doğru gruplar karo, sinek, kupa ve maçadır.

Örneğimize dönersek; P(A)=26/52
P(B) = 4/52
P(A∩B)= 2/52

O halde P(B/A)= $\frac{2/52}{26/52}$ = 2/26'dır.

Uyarı!!! Burada A ile B'nin kesişimini niçin 2/26 değil de 2/52 aldık? Çünkü B olayı As olması şeklinde tanımlanmıştır. Dolayısıyla sadece kırmızı asların bulunduğu grubu esas alamayız, sinek ve maça asları da as olduğuna göre destenin tümünü esas almalıyız.

Örnek: P(Karo/sekiz)=?

Çözüm: (1/52)/(4/52)=0,25

Örnek: maça veya papaz?

Çözüm: P(maça U papaz)=P(maça)+P(papaz)-(maça∩papaz) = (13/52)+(4/52)-(1/52)

Örnek: P(karo/ papaz)=1/4

11.4. Kesikli Rd'nin Olasılık Fonksiyonu

Kesikli bir rd nin olasılık dağılımına olasılık fonksiyonu denir. X'in alabileceği değerlerin bu değerleri alma olasılıklarıyla birlikte yazılmış haline olasılık fonksiyonu denir. Herhangi bir fonksiyonun kesikli rd için olasılık fonksiyonu olabilmesi için örneklem uzayındaki tüm birimlerin olasılıklarının [0,1] aralığında olması ve tüm olasılıkların toplamının 1'e eşit olması gerekmektedir.

Örnek: Bir olayın aşağıda gösterildiği gibi 4 mümkün sonucu olsun. X rd kesikli olduğuna göre, P(e4)=?

P(e1)=0.3
P(e2)=0.4
P(e3)=0.2
Çözüm: P(e4)=1-(0.3+0.4+0.2)=0.1

Örnek: İki zar atılması deneyinde X rd. Üste gelen yüzlerin çarpımı şeklinde tanımlı olsun. X rd'nin olasılık fonksiyonunu yazınız.

Çözüm: Öncelikle S örneklem uzayını oluşturup X rd nin aldığı değerleri bulacağız. Sonra bu değerleri alması olasılıklarını bularak, iki sütun haline, solda x değerleri sağda ise P(x)'ler olmak üzere X rd'nin olasılık fonksiyonunu yazacağız.

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

x	P(X=x)
1	1/36
2	2/36
3	2/36
4	3/36
5	2/36
6	4/36
8	2/36
9	1/36
10	2/36
12	4/36
15	2/36
16	1/36
18	2/36
20	2/36
24	2/36
25	1/36
30	2/36
36	1/36

X 'in olasılık fonksiyonu bu şekildedir, solda x değerleri, sağda da bu değerleri hangi olasılıklarla almakta olduğunu yazdığımızda X'in olasılık fonksiyonunu yazmış oluruz. Ancak burada bir sağlama yapalım, olasılık fonksiyonu olma şartlarından biri, her bir olasılığın [0,1] arasında olmasıydı, baktığımızda bu koşulun sağlandığını görüyoruz. Peki $\sum P(x) = 1$ mi? Evet, toplam olasılık 1'e eşittir, o halde yaptığımız işlem doğrudur.

11.5. Dağılım Fonksiyonu:

X rd'nin dağılım fonksiyonu F(X) ile gösterilir. X rd nin x'e eşit veya daha küçük olması olasılığıdır. Bu birikimli olasılığa dağılım fonksiyonu denir.

$F(X)= P(X \leq x) = \sum_{xi \leq x} f(xi)$

Örnek: Hilesiz bir zar atılsın. X rd üste gelen sayılar olsun. F(X)'i yazınız.

Çözüm:

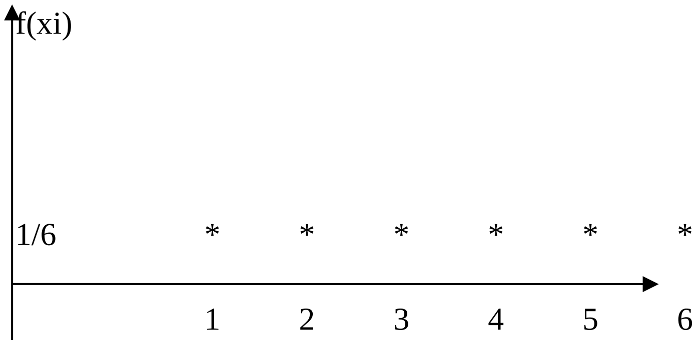
Dağılım fonksiyonu, olasılık fonksiyonunun birikimli halidir, o nedenle öncelikle f(x)'i oluşturmalıyız.

S={1,2,3,4,5,}

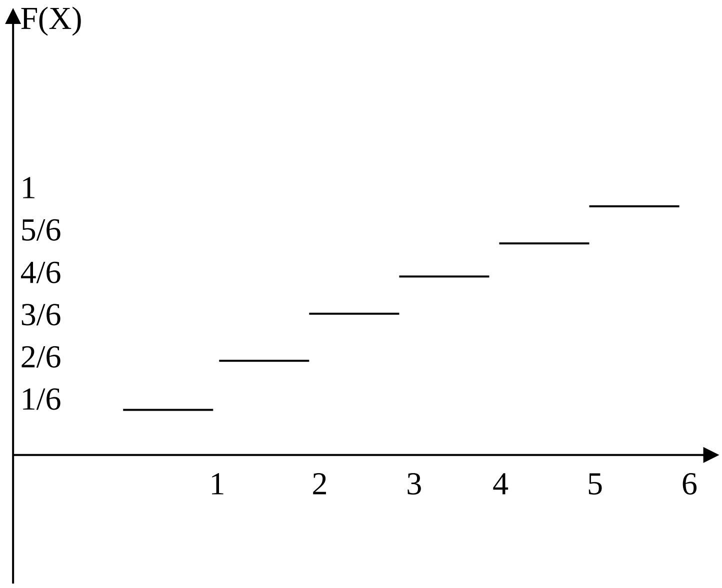
x_i	1	2	3	4	5	6
$f(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6

$$F(X)=\begin{cases} 0 & x_i < 1 \\ 1/6 & 1 \leq x_i < 2 \\ 2/6 & 2 \leq x_i < 3 \\ 3/6 & 3 \leq x_i < 4 \\ 4/6 & 4 \leq x_i < 5 \\ 5/6 & 5 \leq x_i < 6 \\ 1 & 6 \leq x_i \end{cases}$$

Şekil. 11.3. Olasılık Fonksiyonunun Grafiği



Şekil.11.4. Dağılım Fonksiyonunun Grafiği



Örnek:

x_i	-3	0	2	3
$f(x_i)$	0.2	0.1	0.4	c

a. c=?
b. P(X>0)=?
c. P(X=-2)=?
d. FX)=?

Çözüm:

a. f(xi)=P(X=xi) ve toplam f(xi)=1 olmak zorundadır, o halde c= 1-(0.2+0.1+0.4) = 0.3'tür.
b. P(X>0)= P(2)+P(3) = 0.4 + 0.3 =0.7
c. P(X=-2) = 0

d.

$$F(X)=\begin{cases} 0 & x_i < -3 \\ 0.2 & -3 \le x_i < 0 \\ 0.3 & 0 \le x_i < 2 \\ 0.7 & 2 \le x_i < 3 \\ 1 & 3 \le x_i \end{cases}$$

Bölüm Soruları:

1. Bir torbada 4 Sarı, 5 Yeşil, 3 Mavi top vardır. Peş peşe 3 top çekiliyor. İadesiz çekiliş şartı ile P(SYM)=?

2. Bir torbada 4 Sarı, 5 Yeşil, 3 Mavi top vardır. Peş peşe 3 top çekiliyor. İadeli çekiliş şartı ile P(YYM)=?

3. Bir iskambil destesinden rastgele bir kart çekiliyor. Kartın Kız olduğu biliniyorken maça olması ihtimali nedir?

4. Bir iskambil destesinden rastgele bir kart çekiliyor. Kartın siyah olduğu biliniyorken 4'lü olması ihtimali nedir?

5. iskambil destesinden bir kart çekiliyor:

a. P(As U Kupa) b. P(Karo U Sinek) c. P(6/ Siyah) d. P(Sinek U Kız)

12. OLASILIK YOĞUNLUK FONKSİYONU, BEKLENEN DEĞER, VARYANS, MENDEL'İN OLASILIK DENEMELERİ

Bu Bölümde Neler Öğreneceğiz?

-Sürekli rd'nin olasılık yoğunluk fonksiyonu (o.y.f)
- Beklenen değer
- Varyans
- Mendel'in olasılık denemeleri

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Olasılık yoğunluk fonksiyonu, beklenen değer, varyans, Mendel'in olasılık denemeleri	Sürekli rd'nin olasılık yoğunluk fonksiyonu, beklenen değer, varyans öğrenilecek, son olarak da Mendel'in genetik çaprazlamaya dayanan olasılık denemelerinden bahsedilecektir.	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kelimeler

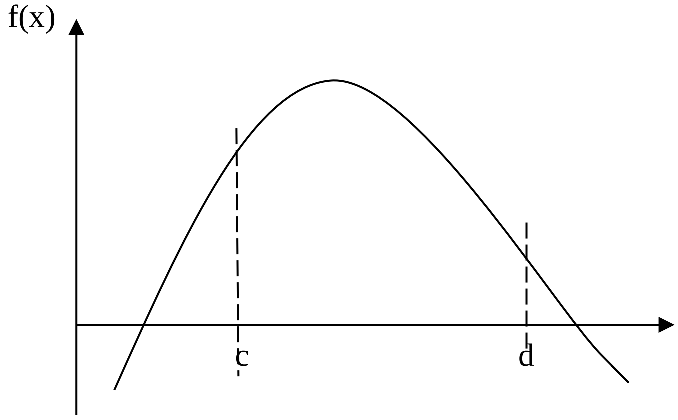
- Olasılık yoğunluk fonksiyonu
- Beklenen değer
- Varyans
- Mendel’in olasılık denemeleri
- Çiçek renklerinin genetik çaprazlanması
- Göz renklerinin genetik çaprazlanması
- Kan gruplarının genetik çaprazlanması

12.1. Sürekli Rd’nin Olasılık Yoğunluk Fonksiyonu

X bir sürekli rd ise ve aşağıdaki koşular sağlanıyorsa, f(x) fonksiyonuna olasılık yoğunluk fonksiyonu denir; (*kesikliyi P(x), sürekliyi f(x) ile gösteriyoruz*)

$\int_{-\infty}^{\infty} f(x)dx = 1$ ve $f(x) \geq 0$ ise, f(x) bir o.y.f’dur.

Şekil. 12.1. Sürekli Rd’nin Olasılık Yoğunluk Fonksiyonunun Grafiği



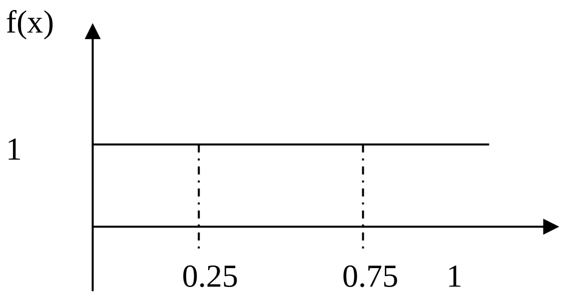
Şekilde görülen eğri f(x) o.y.f’dur. c ile d arasındaki alan, X rd’nin [c,d] aralığında değer alması olasılığını verir. (integral almak alan bulmaktır)

P(c<x<d)= $\int_{c}^{d} f(x)dx = F(d) - F(c)$

Uyarı!!! sürekli rd’de sınırların dahil olup olmaması sonucu değiştirmez. Yani P(a<X<b) = P(a≤X<b) = P(a≤X≤b) ,v.b.

Örnek: f(x)=1 , 0 <X<1 ise;
a. P(0.25 < X<0.75)=?
b. P(X>0.25)=?

Çözüm:



a. $P(0.25<x<0.75)= \int_{0.25}^{0.75} 1dx = (0.75 - 0.25) = 0.5$

b. $P(X>0.25)= \int_{0.25}^{1} 1dx = (1 - 0.25) = 0.75$

Örnek: $f(x)=a/x^3$, $10<X<15$ ve f(x) bir o.y.f olduğuna göre a=?

Çözüm:

$\int_{10}^{15} \left(\frac{a}{x^3}\right) dx = 1$ olmalıdır.

$a(\int_{10}^{15} \left(\frac{1}{x^3}\right) dx) = 1$

$$\frac{x^{-2}}{-2}\Big|_{10}^{15} = \frac{1}{a}$$

(10'dan 15'e kadar belirli integral)
Buradan a = 360 buluruz.

12.2. Beklenen Değer

Bir rd nin beklenen değeri E(X) ile gösterilir (expected value). Aslında genel anlamda "ortalama" demek olan bu ölçüm, X rd nin aldığı değerlerin, bu değerleri alması olasılıklarıyla tek tek çarpılarak toplanmasıyla elde edilir.

X kesikli bir rd olmak üzere;

$E(X) = \sum_{i=1}^{n} xiP(xi) = x1P(X = xi) + x2P(X = x2) + \cdots ... + xnP(xn)$

Örnek: Homojen (hilesiz) bir zar atılsın. X rd üste gelen sayılar ise E(X)=?

Çözüm:

xi	1	2	3	4	5	6
P(X=xi)	1/6	1/6	1/6	1/6	1/6	1/6

$$E(X) = 1*\frac{1}{6} + 2*\frac{1}{6} +3*\frac{1}{6} +4*\frac{1}{6} +5*\frac{1}{6} +6*\frac{1}{6} = 3.5$$

Örnek: Bir para 3 kez atılsın. X rd üste gelen yazıların sayısı olsun. E(X)=?

Çözüm:

n(S) = 2^3 =8 mümkün sonuç vardır.

S={ YYY, YYT, YTY, TYY, YTT, TTY, TYT, TTT}

Yazıların sayısı: 0 1 2 3

xi	0	1	2	3
f(xi)	1/8	3/8	3/8	1/8

$$E(X)= 0*\frac{1}{8} + 1*\frac{3}{8} +2*\frac{3}{8} +3*\frac{1}{8} = 1.5$$

12.3. Varyans

Bir X rd'nin varyansı beklenen değerler cinsinden şu şekilde hesaplanır:

$$V(X)= E(X^2)-[E(X)]^2$$

Örnek: Bir zar atma deneyinde X rd üste gelen sayılar olsun. V(X)=?

Çözüm: X rd nin olasılık fonksiyonu;

xi	1	2	3	4	5	6
P(X=xi)	1/6	1/6	1/6	1/6	1/6	1/6

$$E(X) = 1*\frac{1}{6} + 2*\frac{1}{6} +3*\frac{1}{6} +4*\frac{1}{6} +5*\frac{1}{6} +6*\frac{1}{6} = 3.5$$

$$E(X^2) = 1^2 *\frac{1}{6} + 2^2 *\frac{1}{6} +3^2 *\frac{1}{6} +4^2 *\frac{1}{6} +5^2 *\frac{1}{6} +6^2 *\frac{1}{6} = 91/6$$

$$V(X) = \frac{91}{6} - (3.5)^2 = 2.917$$

Uyarı!!!

$$\mathbf{E(X^2) \neq [E(X)]^2}$$

Örnek: X rd nin olasılık fonksiyonu aşağıdaki gibidir:

xi	1	3	7
f(xi)=P(X=xi)	2/5	1/5	2/5

a. E(X)=?
b. V(X)=?
c. E(2X+1)=?

Çözüm:

a. $E(X) = 1*\frac{2}{5} + 3*\frac{1}{5} + 7*\frac{2}{5} = 19/5 = 3.8$

b. $V(X) = E(X^2) - [E(X)]^2$

$E(X^2) = 1^2 *\frac{2}{5} + 3^2 *\frac{1}{5} + 7^2 *\frac{2}{5} = 109/5 = 21.8$

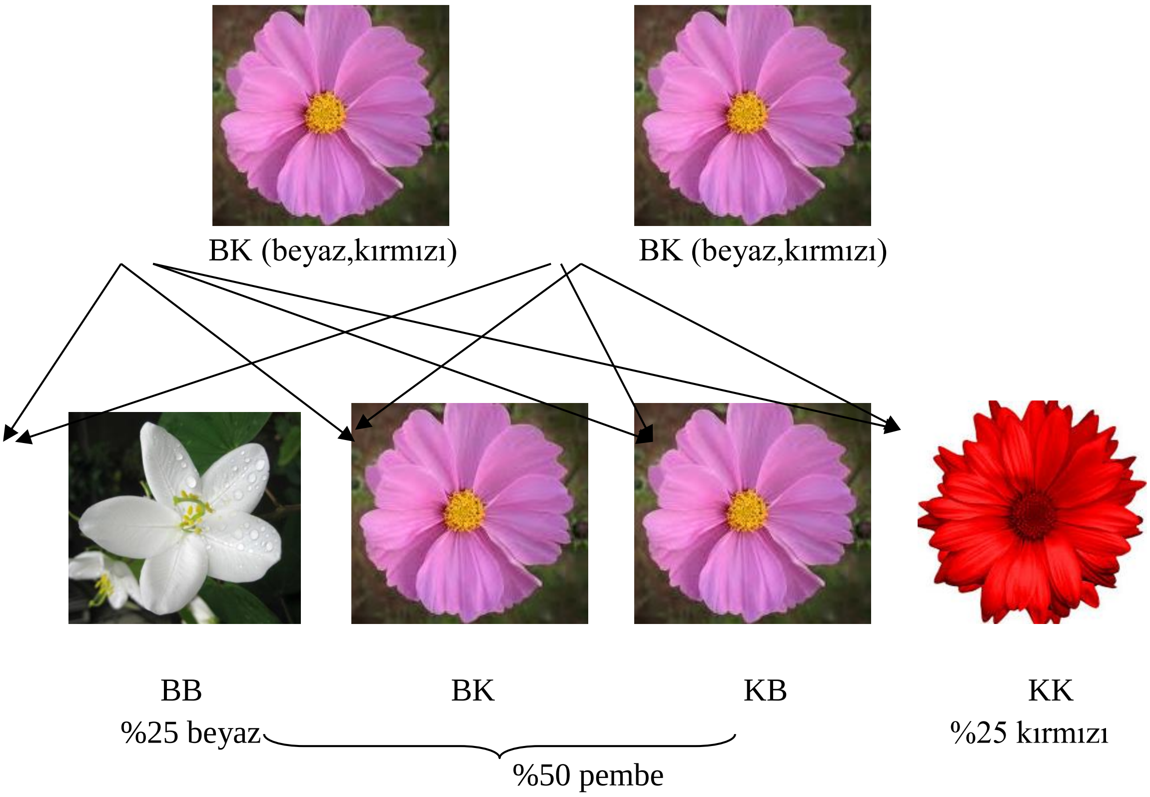
$V(X) = 21.8 - (3.8^2) = 7.36$

c. $E(2X+1) = 2E(X)+1 = 2*(3.8)+1 = 8.6$

12.4. Mendel'in Olasılık Denemeleri

Gregor Mendel, biyolojiye olasılık teorisini uygulayarak genetik biliminin kurucusu olmuştur. Meşhur bezelye deneyinin yanı sıra, pek çok genetik deneyi yapmıştır. Bunlardan sadece 3'ünü ele alacağız. Çiçeklerle ilgili olan deney, kan grubu deneyi ve göz rengi deneyi.

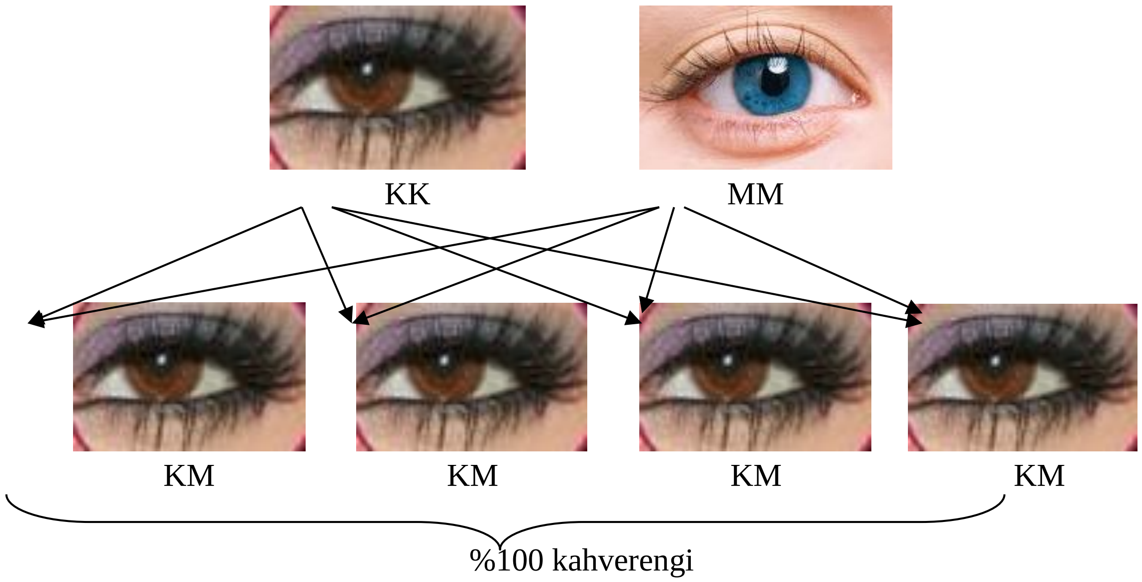
12.4.1. Çiçek Deneyi



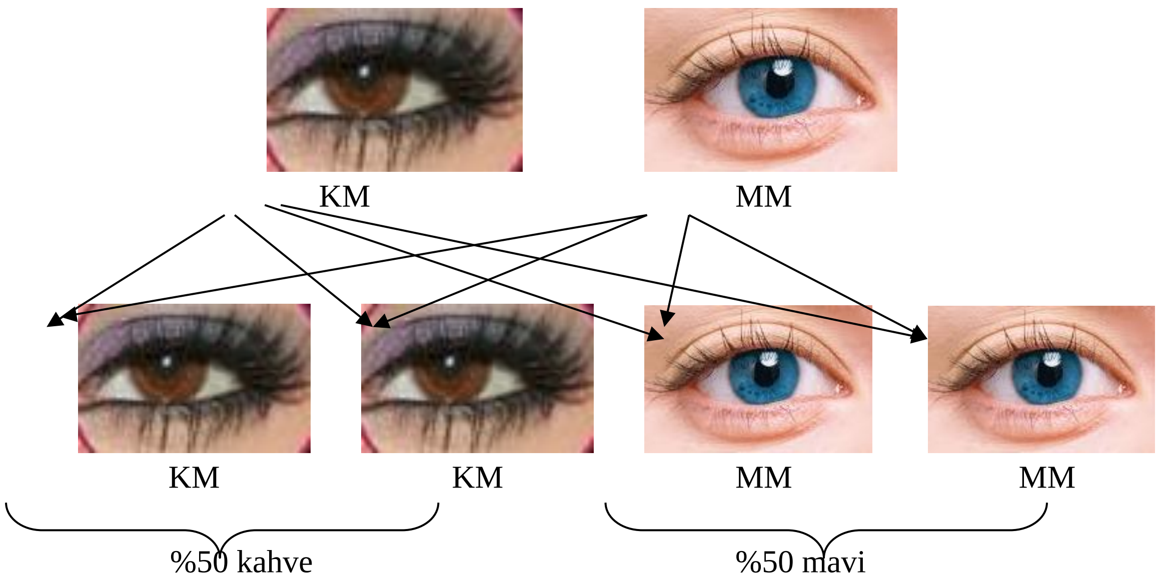
12.4.2. Göz Rengi Deneyi

Mendel, renkli gözün saf gen olduğunu ve çekinik özellikte olduğunu, koyu renk gözün ise saf veya karma gen olabileceğini ve baskın özellikte olduğunu bulmuştur. Yani siyah veya kahverengi göz rengi, yeşil veya mavi renge baskındır. Renkli gözlü bir insanın gözlerindeki genler saf gendir yani ya yeşildir ya da mavidir. Şayet koyu renk gen olsaydı, koyu renk gen baskın gelirdi ve renkli göz ortaya çıkmazdı. Fakat koyu renkli bir göz, saf gen olabileceği gibi, içinde çekinik olarak renkli gen barındırıyor olabilir. Yani iki renkli gözlü insanın çocukları %100 ihtimalle renkli olur yani bu durum kesindir. Fakat iki siyah gözlü (ya da kahverengi gözlü)iki inanın çocukları, sahip olmuş olabilecekleri renkli genlerin çaprazlanması sonucunda renkli gözlü olabilirler. Bu durum aşağıda örneklerle açıklanacaktır.

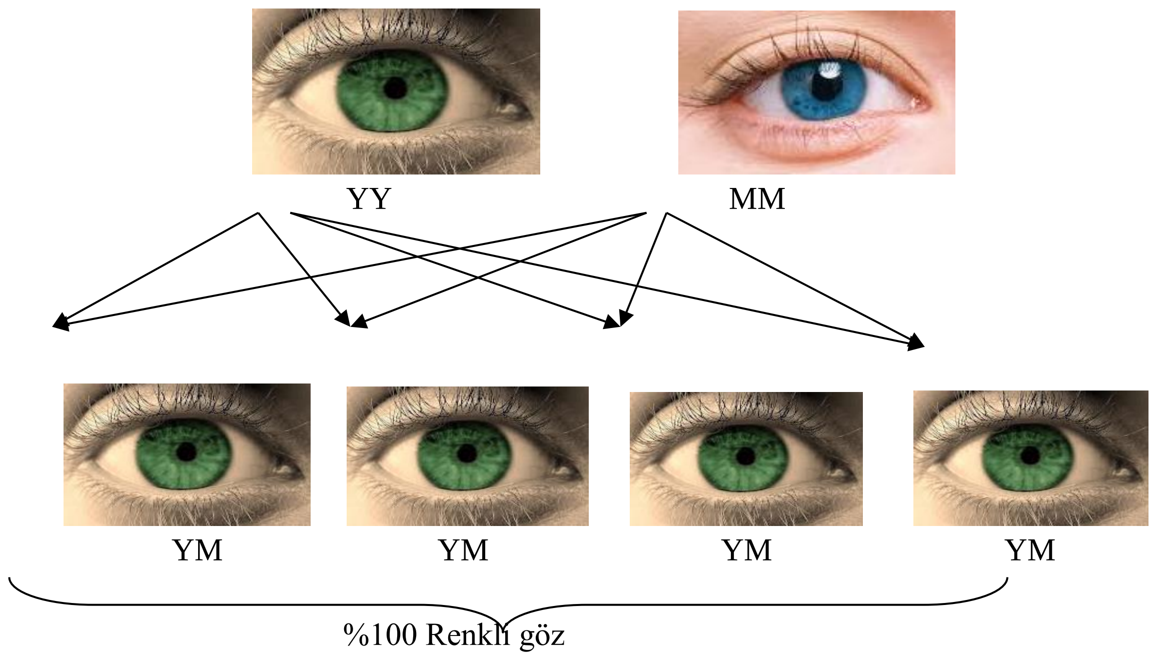
Örnek:



Örnek:



Örnek:



Veya hepsi de mavi olabilirdi, ya da 2 si yeşil ikisi mavi, 3 ü mavi biri yeşil, v.b. ama mutlaka renkli olacaktır.

12.4.3. Kan Grubu Deneyi

Temel kan grupları 0, A ve B'dir. "0" grubu kan çekinik, A ve B ise baskındırlar. Örneğin bir insanın kan grubu A ise, iki durum söz konusudur: ya saf A dır yani AA, ya da sıfırla karışıktır ama A baskın gen olduğu için A olarak ortaya çıkmıştır. Ancak bu melez durumlar sadece çaprazlama yaparken kullanılır, normalde birine sorsanız "kan grubum A0 demez, sadece A" der.

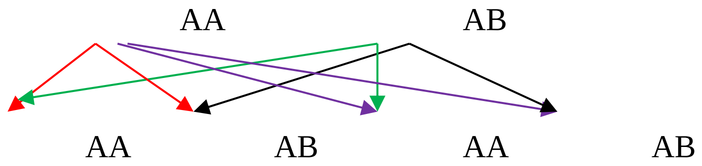
Örnek: Anne AB grubu, Baba B0 grubu ise çocukların muhtemel kan grupları neler olabilir?

AB B0

AB A0 BB B0

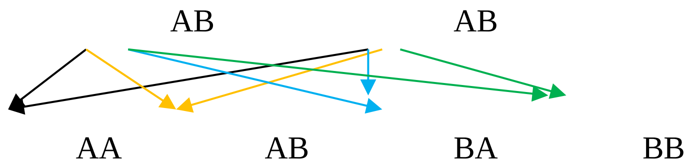
%25 : AB
%25: A
%50: B grubu olur.

Örnek: Anne AA grubu, Baba AB grubu ise çocukların muhtemel kan grupları neler olabilir?



%50: A
%50: AB

Örnek: Anne AB grubu, Baba AB grubu ise çocukların muhtemel kan grupları neler olabilir?



%25: A
%25: B
%50: AB

Bölüm Soruları

1. İki zar atılması deneyinde X rd. Üste gelen yüzlerin toplamı şeklinde tanımlı olsun. X rd'nin olasılık fonksiyonunu yazınız.

2. 1. Sorudaki X rd için E(X)=?

3. 1. Sorudaki X rd için V(X)=?

4. Her fonksiyon bir olasılık fonksiyonu mudur? Neden, açıklayınız.

5. X rd nin olasılık fonksiyonu aşağıdaki gibidir:

xi	-2	3	9
f(xi)=P(X=xi)	2/7	1/7	m

a. m =?

b. V(X)=?

6. Anne B0, baba BB kan gurubuna sahipse çocukların muhtemel kan gruplarının yüzdelerini bulunuz.

7. Anne KM (kahverengi gözlü ama çekinik olarak mavi gen var), baba KM (baba da aynı şekilde kahverengi gözlü ama çekinik mavi var) ise çocukların muhtemel göz rengi yüzdelerini bulunuz.

8.

xi	**……….**	**0**	**1**
f(xi)	**0.2**	**……………**	**0.5**

E(X)=0.3 olduğuna göre, V(X)=?

8. Bir ailenin 4. çocuğunun kız olma ihtimali nedir?

10.

xi	4	7	10
f(xi)	0.4	0.2	0.4

a. E(X) b. E(X-1) c. E(2X+4) d. V(X)

13.BİNOM DAĞILIMI VE POISSON DAĞILIMI

Bu Bölümde Neler Öğreneceğiz?

- Binom Dağılımı
- Poisson Dağılımı

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Binom Dağılımı ve Poisson Dağılımı	Kesikli rd lerin başlıca dağılım fonksiyonlarından Binom dağılımı ve Poisson dağılımı öğrenilecektir.	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kelimeler

- Binom dağılımı
- p
- q
- n
- npq
- poisson dağılımı
- λ

13.1. Bazı Özel Kesikli Dağılımlar

Bu bölümde kesikli rd lerin başlıca dağılım fonksiyonlarından Binom ve Poisson dağılımlarına yer vereceğiz.

13.1.1. Binom Dağılımı

Binom dağılımı, temeli Bernoulli denemelerine dayanır. Bernoulli denemeleri sadece iki sonucu olan ve bir defa tekrarlanan deneylerdir. Örneğin, bir paranın bir defa atılması, bir sınavın sonucunda geçme veya kalma, bir ürünün defolu veya sağlam olması, v.b. sadece 1 kez tekrar eden ve iki sonucu olan denemelerdir. Binom rd. ise, mümkün iki sonucu olan Bernoulli denemelerinin 1'den fazla sayıda tekrar edilmesiyle ortaya çıkmıştır. Yani, bir paranın 2 defa yada daha çok sayıda atılması, defolu olup olmama durumunun sadece 1 üründe değil 1'den fazla üründe incelenmesi, birden fazla sayıda dersten geçme veya kalma durumu,v.b.

Binom rd mümkün 2 sonucu olan bir rd dir demiştik. Bu sonuçların olasılıkları "p" ve "1-p" ile gösterilir. "1-p" ise kısaca "q" ile gösterilir. Bu durumda "p+q=1" dir. Şayet bir denemenin 2'den fazla sayıda olası sonucu bulunuyorsa, artık o rd'nin Binom dağıldığı söylenemez.

Binom Dağılımının Özellikleri

1. 2 sonuçlu n adet özdeş denemeden oluşur.

2. Denemeler birbirinden bağımsızdır. Örneğin bir paranın 2 kez atılması deneyinde, ikinci atışta Yazı gelmesi, birinci atışın sonucuna bağlı değildir. Her bir deneme ilk denem gibidir, birbirinden bağımsızdır. Deneylerin sonuçları bir sonraki denemenin sonucunu etkilemez.

3. İlgilenilen olasılık "p" ile ifade edilir, onun tümleyeni ise "1-p" dir, kısaca "q" ile ifade edilir. "p+q=1" dir.

Binom Dağılımının Olasılık Dağılım Fonksiyonu

$$P(X = x) = \binom{n}{x} p^x (q)^{n-x}$$

Binom dağılımının olasılık fonksiyonunda yer alan $\binom{n}{x}$, C_x^n, yani n'in x'li kombinasyonu demektir ve $\binom{n}{x} = \dfrac{n!}{x!(n-x)!}$ 'dir.

Binom dağılan X rd'nin beklenen değeri; E(X)=np ,Varyansı; V(X)=npq'dur.

Bu durum özet bir şekilde şöyle ifade edilir: X~Bin(np, $\sqrt{npq}$)
Sözel olarak açıklarsak, böyle bir gösterimle karşılaştığımızda, *"X rd; np ortalama ve* $\sqrt{npq}$ *sapma ile Binom dağılmaktadır*" deriz.

Örnek: Bir ütü fabrikasında üretilen ütülerin %15'i defoludur.
a) Tesadüfen seçilen 5 ütünün 3'ünün bozuk olması ihtimali nedir?
b) Tesadüfen seçilen 4 ütünün 2'sinin bozuk olması ihtimali nedir?

Çözüm:
a) Bu sorunun bir binom sorusu olduğunu nasıl anlıyoruz? Öncelikle deneyin mümkün sonuçlarına bakıyoruz. Defolu veya sağlam, yani sadece 2 sonucu var. Peki, deneme 12'den çok defa mı tekrarlanıyor? Evet, 5 ütü seçildiğine göre demek ki n deneme sayısı 5'tir. O halde koşullar binoma uyuyor demektir. Bize sorulan ise $P(X = 3)$ 'tür.

$P(X = x) = \binom{n}{x} p^x (q)^{n-x}$ 'tür.

b) $P(X = 2) = \binom{4}{2} 0.15^2 (0.85)^2$

Örnek: Bir fabrikadaki ürünlerin %10'unun defolu olduğu biliniyor. Tesadüfen seçilen 5 ürünün;
a) 3'ünün defolu olması ihtimali nedir?
b) 1'inin sağlam olması?

Çözüm:

a) $P(X=3)=\binom{5}{3}0.10^3(0.90)^2=0.0081$

b) 1'inin sağlam olması demek, 4'ünün defolu olması demektir.

$$P(X=4)=\binom{5}{4}0.10^4(0.90)^1=0.00045$$

Veya öyle de hesaplanabilirdi;

$$P(X=1)=\binom{5}{1}0.90^1(0.10)^4=0.00045$$

Örnek: Bir satın alma uzmanı mal alırken, satın alacağı bir seri malda %5'ten fazla defolu ürün olması halinde satın almaktan vazgeçmektedir. Bir seride 200 adet ürün olduğu biliniyorken, tesadüfen seçilen 10 ürünün kaçının defolu olduğu tespit edilmek isteniyor.

a) Seçilen 10 üründen 6'sının defolu olma ihtimali nedir?
b) Seçilen 10 üründen 2 veya daha fazlasının defolu olma ihtimali nedir?
c) E(X)=?
d) V(X)=?

Çözüm:

a) $P(X=6)=\binom{10}{6}0.05^6(0.95)^4=0.00000267$

b) $P(X\geq 2)=1-P(X<2)=P(X=0)+P(X=1)$

$$P(X=0)=\binom{10}{0}0.05^0(0.95)^{10}=0.59$$

$$P(X=1)=\binom{10}{1}0.05^1(0.95)^9=0.315$$

$P(X \geq 2) = 1\text{-}(0.059+0.315)=0.095$

c) E(X)=np=10*0.05=0.5

d) V(X)=npq=10*0.05*0.95=0.475

Örnek: X~Binom(1, $\sqrt{0.8}$) ise;

a) P(X=1)=?

b) P(X=2)=?

Çözüm:

np=1

npq= $\sqrt{0.8}$ olduğuna göre; q=0.8'dir. bu durumda p=1-0.8=0.2'dir. np=1 olduğuna göre, n=5'tir.

a) $P(X=1) = \binom{5}{1} 0.2^1 0.8^4$

b) $P(X=2) = \binom{5}{2} 0.2^2 0.8^3$

Örnek: n=25, p=0.5 ise V(X)=?

Çözüm:

E(X)=np=12.5

V(X)=npq=6.25

13.1.2. Poisson Dağılımı

Fransız matematikçi ***Siméon Denis Poisson*** tarafından 1837'de bulunmuştur. Kesikli bir dağılım olan Poisson dağılımı, sürekli bir uzayda (bir zaman dilimi, bir mekan) meydana gelen olayların sayılarını tanımlamakta kullanılır.

Örneğin,

- İstanbul'da bir hafta içinde meydana gelen trafik kazalarını sayısı
- Bir telefon santraline 5 dakika boyunca gelen aramaların sayısı
- Bir markette kasiyerin yarım saatlik bir zaman diliminde hizmet verdiği müşteri sayısı
- Bir kitabın bir sayfasındaki yazım hatalarının sayısı
- Bir limanda bir gün içinde yük boşaltmak için gelen gemilerin sayısı

Poisson dağılımının olasılık fonksiyonu şöyledir:

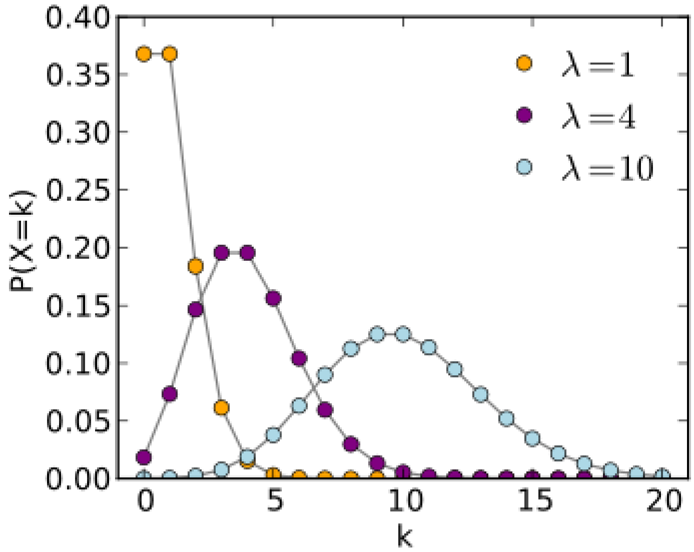
$$p(X = x) = \frac{e^{-\lambda} \times \lambda^{x}}{x!}$$

$$E(X) = V(X) = \lambda$$

$$e \cong 2,718$$

Görüldüğü üzere, dağılımın beklenen değeri ve varyansı birbirine eşittir ve λ dır. λ büyüdükçe, Poisson dağılımı normal dağılıma yaklaşır. Aşağıda Poisson dağılımının olasılık dağılım fonksiyonunun grafiği görülmektedir.

Şekil.13.1. Poisson Olasılık Dağılım Fonksiyonu[22]



Örnek: Bir markette bir kasiyerin bir saat içinde ortalama 15 müşteriye hizmet verdiği bilinmektedir. Bu bilgiye dayanarak;

a. Bir saat boyunca hiç müşteri gelmemesi?

b. En fazla 1 müşteri gelmesi?

c) Birden fazla müşteri gelmesi? Olasılıklarını hesaplayınız.

Çözüm:

a. λ =ortalama=15

bize sorulan P(X=0)'dır.

$$p(X=0)=\frac{e^{-2}\times 2^{0}}{0!}$$

b. P(X≤1)=P(X=0)+P(X=1)= $\frac{e^{-15}\times 15^{0}}{0!}+\frac{e^{-15}\times 15^{1}}{1!}$

c. P(X>1)=1-P(X=0)=1- $\frac{e^{-2}\times 2^{0}}{0!}+\frac{e^{-2}\times 2^{1}}{1!}$

Örnek: Bir kitapta yapılan yazım hatalarının sayısı sayfa başına ortalama 2'dir.

---

[22] http://upload.wikimedia.org/wikipedia/commons/1/16/Poisson_pmf.svg

a. Bir sayfada 5 yazım hatası olma olasılığı nedir?
b. En çok 1 yazım hatası olması?
c. Hiç yazım hatası olmaması?

Çözüm:

a. $p(X = 5) = \dfrac{e^{-2} \times 2^5}{5!}$

b. P(X≤1)=P(X=0)+P(X=1)= $\dfrac{e^{-2} \times 2^0}{0!} + \dfrac{e^{-2} \times 2^1}{1!}$

c. $p(X = 0) = \dfrac{e^{-2} \times 2^0}{0!}$

Bölüm Soruları

1. Poisson dağılımı ve binom dağılımı arasındaki farkları X rastlantı değişkenini baz alarak belirtiniz.

2. Bir kitapta yapılan yazım hatalarının sayısı sayfa başına ortalama 3'dir. Bir sayfada 4 yazım hatası olma olasılığı nedir?

3. Yukarıdaki soruda hiç hata ile karşılaşılmama olasılığını hesaplayınız

4. Bir fabrikadaki ürünlerin %15'unun defolu olduğu biliniyor. Tesadüfen seçilen 6 üründen 3'ünün defolu olması ihtimali nedir?

5. 4. Soruda ürünlerin hepsinin defolu olması olasılığı nedir?

14. UYGULAMA

Bu Bölümde Neler Öğreneceğiz?

- Final öncesi dönem tekrarı ve uygulaması yapılacaktır.

Bölümde Hedeflenen Kazanımlar ve Kazanım Yöntemleri

Konu	Kazanım	Kazanımın nasıl elde edileceği veya geliştirileceği
Uygulama, dönem tekrarı	Uygulama, dönem tekrarı yapılacaktır.	Metinler, çözümlü problemler, alıştırmalar ve interaktif materyaller ile konuların daha kolay anlaşılması sağlanacaktır.

Anahtar Kelimeler

- Uygulama

14.1. .Uygulama Soruları

Soru 1: Aşağıdaki ifadeleri boşluklara uygun şekilde yerleştiriniz.

"betimsel-çıkarımsal-çıkarımsal-çıkarımsal-devlet-ilerleme-istatistik-kollektif-olasılık-örneklem-rastlantı-rastlantısal-sayım-tipik"

*İstatistik konuları bakımından ikiye ayrılır; bunlardan biri ……………………… istatistik, diğeri ………………... istatistiktir.

*Logaritma cetveli, telefon rehberi, milli piyango sonuçları birer …………………… değildirler.

*Sonucu önceden kesin olarak bilinemeyen gözlem sürecine …………………………. süreç, ilgilenilen özelliğe ise …………………….. değişkeni denir.

*Hep aynı şekilde gerçekleşen olaylara ………………… olay, rastsal olarak ortaya çıkan ve değişik sonuçlar alabilen olaylara ise ……………………… olay denir.

*Örnekleme ve örnekleme dağılımı kavramlarına dayanan ……………………… istatistik, bir kitleden tesadüfi olarak seçilen bir ……………………. yardımıyla kitle parametrelerini tahmin etme, parametrelere ilişkin hipotezleri test etme ve geleceğe dönük tahminler yapma konularıyla ilgilenir.

*17., 18. ve 19. yy'lar boyunca Bernoulli, Gauss, Laplace ve Moivre gibi ünlü matematikçilerin ………………….. teorisini geliştirmeleriyle …………………….. istatistiğin temelleri atılmıştır.

*Bir ülkede devlet otoritesi zayıflayınca …………………. işleri de gevşemektedir. Bu bakımdan, bir ülkede istatistiğin ……………………. derecesi, devlet gücünün bir göstergesi sayılabilir.

* "İstatistik" kelimesi köken olarak "status" kelimesinden gelmektedir ve ………………… demektir.

Yanıt:

*İstatistik konuları bakımından ikiye ayrılır; bunlardan biri …………Betimsel……… istatistik, diğeri ………Çıkarımsal…... istatistiktir.

*Logaritma cetveli, telefon rehberi, milli piyango sonuçları birer ……istatistik……… değildirler.

*Sonucu önceden kesin olarak bilinemeyen gözlem sürecine ………………rastlantısal……. süreç, ilgilenilen özelliğe ise …rastlantı………….. değişkeni denir.

*Hep aynı şekilde gerçekleşen olaylara …………tipik……… olay, rastsal olarak ortaya çıkan ve değişik sonuçlar alabilen olaylara ise ……kollektir…………… olay denir.

*Örnekleme ve örnekleme dağılımı kavramlarına dayanan …………çıkarımsal…… istatistik, bir kitleden tesadüfi olarak seçilen bir ………örneklem……… yardımıyla kitle parametrelerini tahmin etme, parametrelere ilişkin hipotezleri test etme ve geleceğe dönük tahminler yapma konularıyla ilgilenir.

*17., 18. ve 19. yy'lar boyunca Bernoulli, Gauss, Laplace ve Moivre gibi ünlü matematikçilerin ………olasılık.. teorisini geliştirmeleriyle …çıkarımsal…………….. istatistiğin temelleri atılmıştır.

*Bir ülkede devlet otoritesi zayıflayınca …sayım……………… işleri de gevşemektedir. Bu bakımdan, bir ülkede istatistiğin ………ilerleme……. derecesi, devlet gücünün bir göstergesi sayılabilir.

* "İstatistik" kelimesi köken olarak "status" kelimesinden gelmektedir ve ………devlet…… demektir.

Soru 2: Riskten kaçınan bir yatırımcı düşünün. Elinde iki tane hisse senedi olsun. Söz gelimi, A hisse senedinin ortalama getirisi 50 tl ve standart sapması 20 tl olsun. B hisse senedinin ortalama getirisi 200 tl ve standart sapması 40 tl olsun. Sizce bu yatırımcı hangi hisse senedine yatırım yapmalıdır, neden?

$$DK = \frac{\sigma}{\overline{X}} * 100$$

Değişim Katsayısı A hisse senedi= (20/50)*100=40

Değişim Katsayısı B hisse senedi: (40/200)*100=20

Değişim katsayısının küçük olması riskin az olması anlamına gelir, bu anlamda B'yi seçmek rasyonel bir tercih olacaktır.

Soru 3: Aşağıdaki değişkenlerin veri türlerini belirtiniz.

Mezun olunan okullar – Bir kabın ağırlığı – Cinsiyet – Bir ailedeki çocuk sayısı – Göz rengi.

Yanıt:

Ordinal, sürekli nicel, nominal, kesikli nicel, nominal

Soru 4: Aşağıdaki serinin mod, medyan ve varyansını (σ^2) hesaplayınız.

Sınıflar	f_i
5-15	3
15-25	7
25-35	21
35-45	11
45-55	1

Yanıt:

mi	fi	mifi	fimi^2	küm. Frekans
10	3	30	300	3
20	7	140	2800	10
30	21	610	18900	31
40	11	440	17600	42
50	1	50	2500	43
150	43	1270	42100	

Ortalama=1270/43=29.54

$$\mathrm{mod} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} * c$$

Öncelikle mod sınıfı belirlenir, 25-35 sınıfı mod sınıfıdır.

$$\mathrm{mod} = 25 + \frac{(21-7)}{(21-7)+(21-11)} 10 = 30.83$$

Medyan için birikimli frekans serisine bakılarak, medyan sınıfı belirlenir.

$$med = Q_2 = L + \frac{\frac{n}{2} - F}{f_{med}} * c$$

$$medyan = 25 + \frac{22-10}{21} 10 = 30,71$$

Kareli Ortalama=979.06

Varyans= KO^2 -AO^2 = 979.06 – (29.54)^2 =106.45

Soru 5: Bir öğrencinin Fizik, Kimya ve Biyoloji derslerine ilişkin z skorları şöyledir: $z_{fizik}=2$, $z_{kimya}=0$, $z_{biyoloji}=-1$

Bu skorlar ne anlama gelmektedir? Öğrencinin en başarılı olduğu ders hangisidir, neden?

Yanıt:

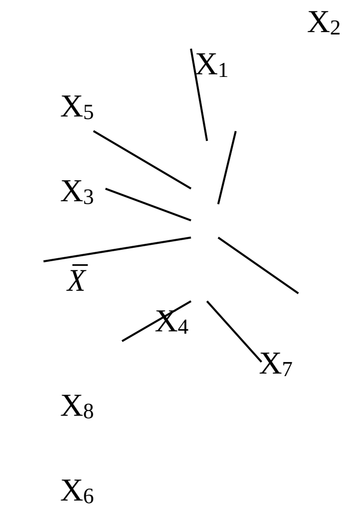
Z skorları öğrencinin söz konusu derslerdeki başarısının değerlendirilmesi için fikir verir, her ders için sınıf ortalamasının ne kadar uzağında yer aldığını görürürüz.

Kimya dersi için başarısı sınıf ortalaması düzeyinde iken, Fizik dersinde ortalama başarının üzerinde not almıştır. Biyoloji de ise sınıf ortalamasının altında not almıştır.

Öğrencinin en başarılı olduğu ders Fiziktir.

Soru 6:

$$\sigma^2 = \frac{\sum_{i=}^{N}(X_i - \mu)^2}{N}$$



Kareli Ortalama

Yukarıdaki formülü, şekli ve kareli ortalamayı ilişkilendiriniz.

Bu formül ile neyi hesaplıyoruz?

Şekilde anlatılmak istenen nedir?

Veri kümesindeki X_i gözlem değerlerinin homojen olması mı yoksa heterojen olması mı varyansı büyütür? Neden?

Yanıt:

Formül varyansa aittir. Şekilde her bir gözlemin ortalamadan ne kadar uzak olduğu görülmektedir. Kareli ortalamanın karesi ile aritmetik ortalamanın karesi arasındaki fark varyansı verir. Varyans, serideki her bir gözlemin ortalamadan sapmalarının karelerinin toplamı ile ilişklilidir. Gözlem değerlerinin homojen yapıda olması varyansı küçültürken, heterojen yapıda olması varyansı büyütür.

Soru 7:

İki hisse senedinin aylık ortalama getirileri ve standart sapmaları şöyledir:

$\mu_1 = 50$

$\sigma_1 = 4$

$\mu_2 = 100$

$\sigma_2 = 16$

Sizce, riskten hoşlanmayan bir yatırımcı hangi hisse senedini seçmelidir? Neden? (5 Puan)

$$DK = \frac{\sigma}{\overline{X}} * 100$$

Yanıt:

Birinci hisse senedi Değişim Katsayısı= (4/50)*100=8
İkinci hisse senedi Değişim Katsayısı= (16/100)*100=16
Birinci hisse senedinin seçilmesi yatırımcıyı riskten korur.

Soru 8:

Aşağıdaki veri kümesinin modunu ve medyanını hesaplayınız.

100 12 16 105 12 120 12 200 80

Yanıt:

Mod en sık tekrarlayan terim, bu yüzden 12

Medyan ise ortadaki terim. Öncelikle seri sıraya dizilmeli.

12 12 12 16 80 100 105 120 200

Toplam 9 gözlem var, (n+1)/2 =5. Medyan değeri 5. Gözleme yani 80'e karşılık gelir.

Soru 9:

Aşağıda Gini katsayıları verilen ülkeleri, gelir dağılımı en adil olandan, gelir dağılımı adil olmayana doğru sıralayınız.

Ülke	Gini katsayısı(%)
Paraguay	56,8
Danimarka	23,2
Sierra Leone	62,9
Türkiye	42,0
Macaristan	24,2

Yanıt:

Hatırlanacağı üzere Gini katsayısının 0'a yaklaşması gelir dağılımının adil olduğuna, 1'e yaklaşması ise adil olmadığına işaretti. Buna göre sıralama şöyle olur:

1- Danimarka
2- Macaristan
3- Türkiye
4- Paraguay
5- Sierra Leone

Soru 10:

İskambil destesinden bir kart çekiliyor. Çekilen bu kartın;

a) P(As veya Kupa)

b) P(Maça veya Sinek)

c) P(Karo/yedi)=?

Yanıt:

a) (13/52) +(13/52)
b) (13/52)+(4/52)-(1/52)
c) (1/52)/(4/52)

Soru 11:

x	-1	0	1
P(X=x)	0,2	0,3	0,5

E(X)=?
V(X)=?

Yanıt:

E(X)= (-1)* (0.2) +(0)*(0.3)+ (1)*(0.5)= 0.3

E(X^2)= $(-1)^2$* (0.2) +(0)*(0.3)+ $(1)^2$*(0.5)=0.7

V(X)= E(X^2)- E(X)=0.7-0.09=0.61

Soru 12:

Bir lisede öğrencilerin IQ ortalaması 105, standart sapması 9'dur. Öğrencilerin en az % kaçı 85.2 ile 124.8 arasında IQ'ya sahiptir?

Yanıt:

CHEBYSHEV TEOREMİ ile çözülmeli. Ortalaması μ, standart sapması σ olan bir verideki gözlemlerin $\mu \pm k\sigma$ aralığına düşenlerin oranı en az $1-\frac{1}{k^2}$ 'dir.

$$\mu - k\sigma = 85,2$$
$$\mu + k\sigma = 124,8$$

k=2.2

$$1-\frac{1}{(2,2)^2} = 0,79$$

Soru 13:

Olasılık dağılımı verilen X rd'nin E(X)=? V(X)=? Hesaplayınız.

x	-2	0	3
P(X=x)	0,1	0,2	0,7

Yanıt:

E(X)= (-2)* (0.1) +(0)*(0.2)+ (3)*(0.7)= 1.9

$E(X^2)$= $(-2)^2$* (0.1) +(0)*(0.2)+ $(3)^2$*(0.7)=6.7

V(X)= $E(X^2)$- E(X)=6.7 -3.61=0.19=3,09

Soru 14:

Aşağıdaki serinin çarpıklığını pearson katsayısı ile değerlendiriniz

X_i: 6 6 12 1 8

Yanıt:

Aritmetik ortalama= (4 +6+12+1+10+3)/ 6=6
Mod = 6

Medyan için sıraya dizilir, 1 6 6 8 12
Medyan =6

Seride ortalama=mod=medyan olduğu için seri simetriktir

Soru 15:

P(X) bir olasılık fonksiyonudur. X rd'nin beklenen değeri 2.8 olduğuna göre, X rd'nin varyansını bulunuz.

x	1	…….	5
P(X=x)	0.4	0.2	…….

Yanıt:

P(x) bir olasılık fonksiyonu ise, olasılıklar toplamı 1 olacağından tablo şöyle olur.

x	1	……	5
P(X=x)	0.4	0.2	0.4

E(X)= (1*0.4)+(x*0.2)+(5*0.4) =2.8
(x*0.2)=0.4
X=2
E(X^2)= (1^2*0.4)+(2^2*0.2)+(5^2*0.4)=11,2
V(X)= 11.2 – 2.8 ^{2}= 3,36

Soru 16:

X rd.~Binom(4, 0.5) ise;
a) E(X)=?
b) V(X)= ?

Yanıt:

a) E(X)= np = 4x0.5x0.5
b) V(x)= npq = 4x0.5x0.5

Soru 17:

Bir fabrikada üretilen ürünlerin %3'ü defoludur. 100 birimlik üretim bandından rastgele seçilen 4 ürünün defolu olması olasılığı nedir?

Yanıt: Bu bir Binom sorusudur. Çünkü iki sonucu var, ya sağlam ya defolu. Ayrıca deneme sayıs yani n=200>1 olduğu için Binomla çözeriz.

P(X=3) = $\binom{100}{4}(0.03)^4(0.97)^{96}$

Bu soruda rakamlar çok yüksek, p olasılığı da oldukça düşük. Böyle durumlarda Binom'un Poisson'a yaklaşmasını kullanarak çözüme gidebiliriz. Şartımız şu; eğer np<5 ise Binom dağılımı λ = np olan Poisson'a yakınsar. O halde sorumuzun cevabını şöyle bulacağız:

np = 100*0,03 = 3=λ

P(X=4)= $\frac{e^{-3}3^4}{4!}$ = 0.168

Soru 18: Bir ameliyatın başarı olasılığı %80'dir. Ameliyat edilen 10 hastadan

a. 6'sının iyileşme olasılığı?
b. En az 9'unun iyileşme olasılığı?
c. En fazla 7'sinin iyileşme olasılığı?
d. Ameliyatı başarılı geçen hastaların beklenen değeri ve varyansı nedir?

Yanıt:

Burada da yine 2 sonuç var ve deneme sayısı n=10>1, yani Binom sorusu.
n=10   p=0.8   q=0.2

a. $P(X=6)= \binom{10}{6}(0.8)^6(0.2)^4$
b. $P(X\geq 9) = P(9)+P(10) = \binom{10}{9}(0.8)^9(0.2)^1 + \binom{10}{10}(0.8)^{10}(0.2)^0$
c. $P(X\leq 7) = 1- [P(8)+P(9)+P(10)]$
d. $E(X) = np= 10*0.8=8$
$V(X)= npq= 10*0.8*.02 = 1.6$

Soru 19:

Bir şehirde ender rastlanan bir hastalıktan bir ay içinde ortalama ölen kişi sayısı 4'tür. Herhangi bir ayda bu hastalıktan;

a. Hiç kimsenin ölmemesi?
b. En az 2 kişinin ölmesi?

Yanıt: Bu bir Poisson sorusudur. Uzay sürekli, olaylar kesikli ve anlık.

$\lambda=4$

a. $P(X=0) = \frac{e^{-4}4^0}{0!} = 0.018$

b. $P(X\geq 2) = 1- [P(0)+P(1)] = 1- [\frac{e^{-4}4^0}{0!} + \frac{e^{-4}4^1}{1!}] = 1- (0.018+0.073) = 0.909$

KAYNAKÇA

Büyük Larousse, Cilt: 10

D.G.Rees, Essential Statistics, Third Edition, Chapman&Hall, 1995

Enis Sınıksaran, İstatistik, Filiz Kitabevi, İstanbul Üniversitesi.

Fikri Akdeniz, Olasılık ve İstatistik, 18. Baskı, Çukurova Üniversitesi, 2013.

http://www.google.com.tr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CBoQFjAA&url=http%3A%2F%2Fberkayvaz.files.wordpress.com%2F2013%2F10%2Folasilik-ve-istatistik-i_4-hafta.pptx&ei=87bHU7SgI8e6ygPSr4KQDg&usg=AFQjCNEIvmpReOsRZr_GidFPtudI0VAv7g&bvm=bv.71198958,d.bGQ (Öğr. Gör. Berk Ayvaz, İTİCÜ Mühendislik ve Tasarım Fakültesi Endüstri Mühendisliği Bölümü)

İstatistik, Murat Açıköğretim Yayınları, 2. Sınıf İktisat – İşletme

Meydan Larousse, Cilt: 6

Necmi Gürsakal, Betimsel İstatistik, Bursa, 2012.

Onur Özsoy, İktisatçılar ve İşletmeciler İçin İstatistik Excel Uygulamalı, Ankara,2010.

Özer Serper, Uygulamalı İstatistik 1, Uludağ Üniversitesi

www.wikipedia.org