Marmara Üniversitesi

İstatistik Bölümü

Örnekleme Teorisi (1)

Doç. Dr. Atıf Evren

"You don't have to eat the whole ox to know that it is tough."

Samuel Johnson

Giriş

Veri toplama (derleme) işlemi

i) Tamsayım
ii) örnekleme

yolları ile gerçekleştirilebilir.

- Hakkında bilgi edinilmek istenen topluluğu (ana kütleyi) oluşturan bütün birimlerin gözlenmesine **tamsayım** denir. Örneğin ülkemizde beş yılda bir yapılan genel nüfus sayımları tamsayım niteliğindedir.
- **Örnekleme**, bir ana kütleden seçilen ve daha az sayıda birimden oluşan bir örneklemi

incelemek sureti ile ana kütle hakkında genel yargılara varma işlemidir.

Belli bir ana kütleden alınan örneklem yardımıyla, bu ana kütlenin uyduğu bölünme (dağılım) şeklinin bir veya birkaç parametresinin değerini araştırma işlemi olan **kestirim** (estimation) ile **örnekleme** birbirinin ayrılmaz birer parçasıdır. Burada sözü edilen **parametre** bir ana kütlenin aynı yapıdaki diğer kütlelerden ayırt edilebilmesini sağlayan **sabit** bir değerdir. Sözgelimi normal dağılımın iki parametresi μ ve σ^2, bu dağılımın ortalaması ve varyansıdır. Poisson dağılımının tek parametresi olan λ, dağılımın ortalamasına ve varyansına eşittir. Yine Bernoulli dağılımının tek parametresi olan π, bu dağılımın başarı oranını (olasılığını) vermektedir.

"İstatistik" ise örneklem bilgisinden hareketle; değeri bilinmeyen bir parametrenin kestirilmiş (tahmin edilmiş) bir değeridir. Aşağıdaki tablo bazı parametreleri ve onların örneklemdeki karşılığı olan istatistikleri vermektedir:

Parametre	**İstatistik**
Ana kütle ortalaması (μ)	Örnek ortalaması ($\overline{X}$)
Ana kütle oranı (π)	Örnek oranı ($\hat{p}$)
Ana kütle varyansı (σ^2)	Örnek varyansı (S^2)
İki ana kütle ortalaması farkı ($\mu_1 - \mu_2$)	İki örnek ortalaması farkı ($\overline{X}_1 - \overline{X}_2$)
İki ana kütle oranı farkı ($\pi_1 - \pi_2$)	İki örnek oranı farkı ($\hat{p}_1 - \hat{p}_2$)

Tablo 1:Bazı parametre ve istatistikler

Örneklemeye başvurmak neden gereklidir?

Tamsayım kütle hakkında tam ve kesin bir fikir verebilirse de, bazı durumlarda uygulanması mümkün olmayabilir. Örneğin, genel nüfus sayımında erkek ve kadın sayısını, medeni hal bölünüşünü vb. kesinlikle belirlemek olası olduğu halde, bir ampulün maksimum ömrünü belirleyemeyiz. Çünkü bu takdirde bütün üretilmiş ampulleri denemeye tabi tutmak zorunda kalırız ki bu da bütün ampullerden yararlanma olanağını ortadan kaldırır. Yine herhangi bir deniz veya göldeki bütün balıklardan oluşan bir ana kütle ele alındığında da tamsayım yapmanın olanaksızlığı ortaya çıkacaktır. Bununla birlikte ana kütlede incelenecek birim sayısı az olduğunda örneklemeye gitmek gerekmez ve tamsayım uygulanır.

Örneklemenin üstün tarafları nelerdir?

a)Tamsayım, ana kütlenin bütün birimlerini dikkate aldığından örneklemeye göre gerek para gerekse işgücü ve zaman açısından daha büyük olanakları gerektirir. Örnekleme ise bunlardan **tasarruf** etmemizi sağlar.

b)Çok iyi elemanlar sağlanıp **eğitildiğinde**, onların ana kütlenin küçük bir parçasından elde ettikleri verilerin doğruluk derecesi de yüksek olur. Bu verilere dayanarak yapılacak ana kütle parametre kestirimlerinin ise daha güvenilir olacağı açıktır.

c) Bazı araştırmalarda ve özellikle kalite kontrol çalışmalarında tamsayım hiçbir zaman **uygulanamaz**. Dolayısıyla örneklemeye başvurulması gerekir. Belirsiz kütleler için de aynı şekilde davranılır.

Örneklemenin uygulanma alanları nelerdir?

Örnekleme daha ucuz, daha kolay ve daha çabuk olarak bilgi edinme, başka türlü elde edilmesi olanaksız durumlara çözüm bulma gibi yararları nedeniyle, geniş uygulama alanına sahiptir. Nitekim

i)sanayi işletmelerinde kalite kontrolünde,

ii)gelir dağılımı, eğitim vb. konularda bilgi toplamada,

iii)kamuoyu araştırmalarında,

iv)tüketim yapısının belirlenmesinde,

v)pazarlama araştırmalarında,

vi)tarım alanında üretim tahminleri yapmada,

vii)biyolojik, psikolojik, tıbbi vb. birçok araştırmada örneklemeden büyük ölçüde yararlanılmaktadır.

Ana kütle ve örneklem ne demektir?

Üzerinde araştırma yapılan herhangi bir canlılar ya da cansızlar topluluğuna **"ana kütle",** bunun içinden çekilen birimlerden oluşan alt topluluğa **"örneklem"** adı verilir. Diğer bir deyişle, ana kütle bir araştırmacının ilgilendiği ve ortak özelliklere sahip birimlerden oluşan topluluğun tamamı, örneklem ise bu ana kütlenin özelliklerini yansıtan bir **parçasıdır.**

Örneklemlerin seçildikleri ana kütleler, belirli sayıda birimlerden oluşabileceği gibi, sonsuz derecede büyük de olabilirler. İlk durumda **sonlu ana kütle**, ikinci durumda **sonsuz ana kütle** sözkonusu olur.

Ana kütleye dahil birimlerin sayısı **ana kütle hacmi**, örneklemin kapsadığı birimlerin sayısı **örneklem hacmi** olarak adlandırılır. **Ana kütle hacmi N**, **örneklem hacmi n** harfi ile sembolize edilir. **n/N oranına örneklem oranı** denir. Bunun tersi olan **N/n oranına büyütme oranı** denir.

Örneklenen ana kütle, hedef ana kütle ne demektir?

İçerisinden örneklem alınan ana kütleye **"örneklenen ana kütle"** hakkında bilgi toplanmak İstenen ana kütleye **"hedef ana kütle"** denir. Örneklenen ve hedef ana kütleler bazen aynı, bazen farklı olabilirler.

Örnekleme birimi ve gözlem birimi nedir?

Örnekleme biriminin ana kütlenin doğal ve fiziksel bir parçası olarak tanımlanması zorunlu değildir. Örneğin bir yerin nüfusuyla ilgili bazı istatistiklerin elde edilmesi söz konusu olduğunda, nüfus için bireyin birim olması gerekmez. Birey örnekleme birimi olarak seçilir veya seçilmez. Nitekim batı ülkelerinde, uzun yıllardan beri işsizlerin sayılarını tahmin için yapılan aylık örneklemelerde, konut veya belli sayıda konut kümesi, örnekleme birimi olarak tanımlanmaktadır.

"Gözlem birimi" hakkında ayrı ayrı bilgi toplanan ana kütlenin en küçük parçasıdır. Örnekleme birimi ile gözlem birimi ayrı olabileceği gibi farklı da olabilir. Genellikle örnekleme birimi birden fazla gözlem birimini kapsayacak şekilde tanımlanır. Örneğin, bireylerin çeşitli özelliklerine ilişkin bazı istatistikler elde etmek amacıyla yapılan bir örneklemede, aile **örnekleme birimi** olarak seçilirse, birey de **gözlem birimi** olur.

Çerçeve nedir?

"Çerçeve", ana kütleyi kapsayan ve birimlerin sınırlandırılmasına olanak sağlayan bir araçtır. Çerçeve bir adres listesi, harita, telefon rehberi, fiş dosyası vb. bir araç olabilir. Bir çerçeve olmadan ne örnekleme ne de tamsayım yapılabilir.

Örneklemede İzlenecek Aşamalar

1)Örnekleme amaçlarının belirlenmesi.

2)Ana kütlenin belirlenmesi.

3)Toplanacak bilgilerin belirlenmesi.

4)Bilgi toplama yönteminin seçimi (gözlem, anket, görüşme vb.).

5)Araştırmanın güven düzeyinin belirlenmesi.

6)Örneklem hacminin belirlenmesi.

7)Örnekleme tekniğinin seçimi.

8)Ön test uygulanması.

9)Alan çalışmalarının organizasyonu.

10)Uygulama, analiz ve sonraki araştırmalara hazırlık.

Örneklemede ortaya çıkabilecek hatalar nelerdir?

Örneklem istatistiğinin $(\hat{\theta})$ hesaplanmasında ana kütledeki birimlerin sadece n tanesinden yararlanıldığı için, bu tahmin bütün birimlerin dikkate alınmasıyla hesaplanacak ana kütle parametresine (θ) tam olarak eşit çıkması genellikle beklenemez. Örneklem istatistiğinin ana kütle parametresinden gösterdiği farka $(\hat{\theta} - \theta)$ **"toplam hata"** adı verilir.

Örneklemede önem taşıyan iki hata vardır**: Sistematik ve rassal hatalar.** Toplam hata ile sistematik ve rassal hatalar arasında aşağıdaki grafikte gösterilen bir bağlantı vardır:

Grafikten anlaşılacağı gibi, toplam hata şu şekilde tanımlanmaktadır:

$$Toplam\ Hata = \sqrt{(Sistematik\ Hata)^2 + (Rassal\ Hata)^2}$$

Örneklem istatistiği ana kütle parametresinden tek yönde sapma eğilimi gösterdiğinde **“sistematik hata”** söz konusudur. Bu tip hatanın varlığı halinde, örneklem istatistikleri, ya ana kütle parametresinden daima daha küçük ya da daima daha büyük olarak elde edilirler. Birimler rassal olarak seçildiklerinden sistematik hataya rastlanmaması gerekirse de bu hatalar;

a)Örneklemin seçiminde yararlanılan yaklaşımdan,

b)Örneklem istatistiğinden hareketle ana kütle parametresi tahmin edilirken kullanılan yaklaşımdan,

c)Örnekleme birimi ile gözlem biriminin birçok hallerde farklı olmasından,

d)Ana kütle iyi tanımlanmadığı için çerçevenin eksik olmasından veya aynı birimin birkaç defa kaydedilmiş bulunmasından,

e)Yanlış anlamaya yol açan sorunların varlığından,

f)Örnekleme dahil olan birimlerin bir kısmı hakkında bilgi sağlanamamasından dolayı ortaya çıkabilirler.

Bir ana kütleden çok sayıda n hacimlik örneklem çektiğimiz ve bunların sözgelimi aritmetik ortalamalarını hesapladığımız zaman ortalamaların birbirlerinden ve ana kütle ortalamasından farklı olduğunu görmekteyiz. Bu farklar veya hatalara **“rassal hata (örnekleme hatası)”** adı verilir. Rassal hatalar, her iki yönde de etkili olduklarından, birbirlerinin etkilerini yok edebilirler. Bu özelliğin bir sonucu olarak, bir örneklem istatistiği birbirinden farklı olsa bile, örnekleme dağılımının ortalaması ana kütle parametresine eşit olur. Rassal hatalar örneklem hacmi büyüdükçe **“Büyük Sayılar Kanunu”** gereğince küçülür.

Örnekleme teknikleri

Örnekleme teknikleri biri **olasılıklı olmayan,** diğeri de **olasılıklı (rassal)** teknikler olmak üzere ikiye ayrılır.

Olasılıklı Olmayan Örnekleme Teknikleri

1)Kolayda Örnekleme

Kolayda örnekleme tekniğinde örneklemede yer alan birimler kolay ve ucuz bir şekilde seçilir. Zaman ve maliyet içermediği için veri toplamada hızlı ve ucuz bir teknik olarak ele alınan bu teknikte, yapılan çalışmada yer alacak birimlere araştırmacı karar verir. Örneğin, bir konuda veri toplama için sokakta durup gelip geçenlerden bilgi almak, bu tür örnekleme tekniğidir.

2)Kota Örneklemesi

Bu teknikte ana kütle, araştırmanın amacına uygun olarak, araştırmacı tarafından belirlenen Değişkenlere (örneğin, yaş, eğitim durumu, nüfus grubu, cinsiyet vb.) göre sınıflandırılır. Dolayısıyla, ana kütle seçilen değişkenler açısından homojen alt gruplara ayrılır ve bu grupların ana kütle içindeki önemleri ile orantılı sayıda birim seçilir.

3)Amaçlı Örnekleme

Bu teknik, araştırmacıların açıkça tanımlanmış bir örneklem ile çalışmak istediğinde kullanılır. Örnekleme kimlerin dahil edileceği veya örneklemenin nerdede yapılacağı konusunda araştırmacı kendi yargılarını dikkate alır. Örneğin bir organizasyonda bulunan her bir kademeden kişilerin örnekleme alınması veya örneklemeye sadece 20-30 yaş arası üniversite mezunu ve evli olanların dahil edilmesi gibi.

4) Kartopu Örneklemesi

Bu teknikte öncelikle belirlenen özelliklere uygun olan katılımcı ile görüşme yapılır, görüşme yapılan katılımcının yardımı ile diğer potansiyel katılımcılara ulaşılır.

5) Tek Birimli Örnekleme

"Monografi tekniği" olarak da bilinen bu teknikte ana kütlenin belirlenen alt gruplarından, ana kütleyi temsil edeceği gerekçesi ile birer temsilci alınır. Bu temsilcilerin ait oldukları grupların tüm ortalama özelliklerini taşıması gerektiği için, ana kütle homojen olmalıdır. Olasılıklı olmayan örnekleme teknikleri ile elde edilen verilerde "taraflılık" söz konusu olduğu için, analizlerin sonuçları ana kütle için genellenmemelidir.

Olasılıklı (rassal) örnekleme teknikleri nelerdir?

Bu teknikler

i)basit rassal örnekleme,

ii)sistematik örnekleme,

iii)tabakalı örnekleme,

iv)katlı örnekleme

olarak sıralanabilir.

Olasılıklı örnekleme tekniklerinde, ana kütledeki her bir birime eşit seçilme şansı tanınmaktadır. Öte yandan, örneklemden elde edilen sonuçlar belli bir güven düzeyi ile ana kütle için genelleştirilir.

Örneklemede Çok Karşılaşılan Olasılık Dağılımları

Bernoulli Dağılımı

Bir rassal deney, başarılı olma/olmama; erkek olma /kadın olma; sigara içme/içmeme durumlarında olduğu gibi birbirini **kategorik olarak dışlayan iki farklı şekilde** sonuçlansın. Deney sonucunda karşılaşılan bu iki durumu, genel olarak, 'başarı' ve 'başarısızlık' durumu olarak ifade edelim. Yine bir denemenin 'başarı'ile sonuçlanma olasılığı p; 'başarısızlık' ile sonuçlanma olasılığı da 1-p olsun. X rastlantı değişkeni de bir denemedeki başarı sayısı olsun. Bu durumda Bernoulli olasılık fonksiyonu

$P(X=x)=\begin{cases} p & x=1 \\ 1-p & x=0 \end{cases}$ ile verilir. Bu fonksiyonu şu şekilde de ifade etmek mümkündür:

$f(x)=P(X=x)=p^{x}(1-p)^{1-x} \quad x=0,1$

Şimdi de bu dağılımın beklenen değerini ve varyansını aşağıdaki tablodan yola çıkarak bulalım:

x	**f(x)**	**x.f(x)**	$x^2.f(x)$
0	**1-p**	**0**	**0**
1	**p**	**p**	**P**

$E(X)=\sum x.f(x)=p$

$E(X^2)=\sum x^2.f(x)=p$

$Var(X)=E(X^2)-(E(X))^2$

$Var(X)=p(1-p)$

Binom Dağılımı (Bernoulli Dağılımının Genellenmesi)

Bir denemenin iki olası sonucu olsun. Her bir sonucun gerçekleşme ve gerçekleşmeme olasılıkları sırası ile p ve 1-p olsun. Deney **n bağımsız** Bernoulli denemesinden oluşsun. Yine her denemedeki başarı olasılığı önceki denemelerin sonuçlarından **bağımsız** varsayılsın.

X rastlantı değişkeni de gözlenen **başarı sayısı** olsun. Bu durumda X'in olasılık dağılımı binom dağılımı olarak adlandırılır ve

$f(x)=\binom{n}{x}p^x(1-p)^{n-x} \quad x=0,1,\ldots,n$ şeklinde verilir.

Örnek: Hilesiz bir para 10 kere atılsın. a) 3 kere tura gelme, b) 3 ile 5 arasında tura gelme, c)en az 3 kere tura gelme olasılığını hesaplayınız.

Çözüm: X rastlantı değişkeni n=10 ve p=0,5 parametreli bir binom dağılımına uyar.

a) P(X=3)= $\binom{10}{3}(1/2)^{10}=0,117$

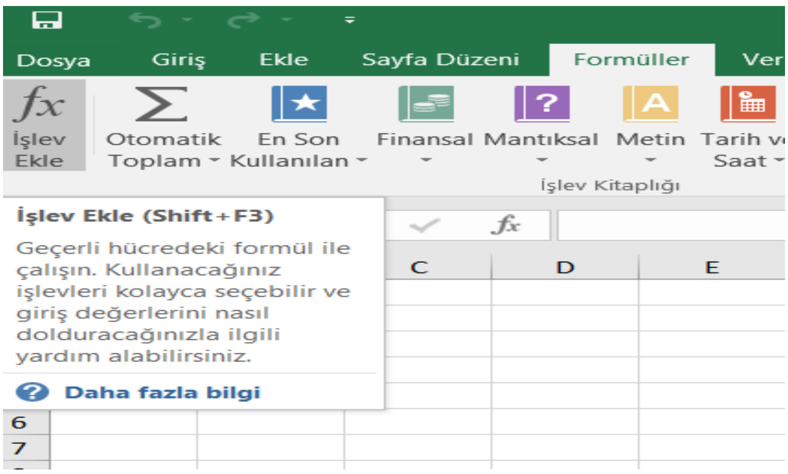
b) P(3≤X≤5)= $\sum_{x=3}^{5}\binom{10}{x}(1/2)^{10}\cong 0,57$

c) $P(X\geq 3)=\sum_{x=3}^{10}\binom{10}{x}(1/2)^{10}=1-P(X<3)=\sum_{x=0}^{2}\binom{10}{x}(1/2)^{10}\cong 0,95$ bulunur.

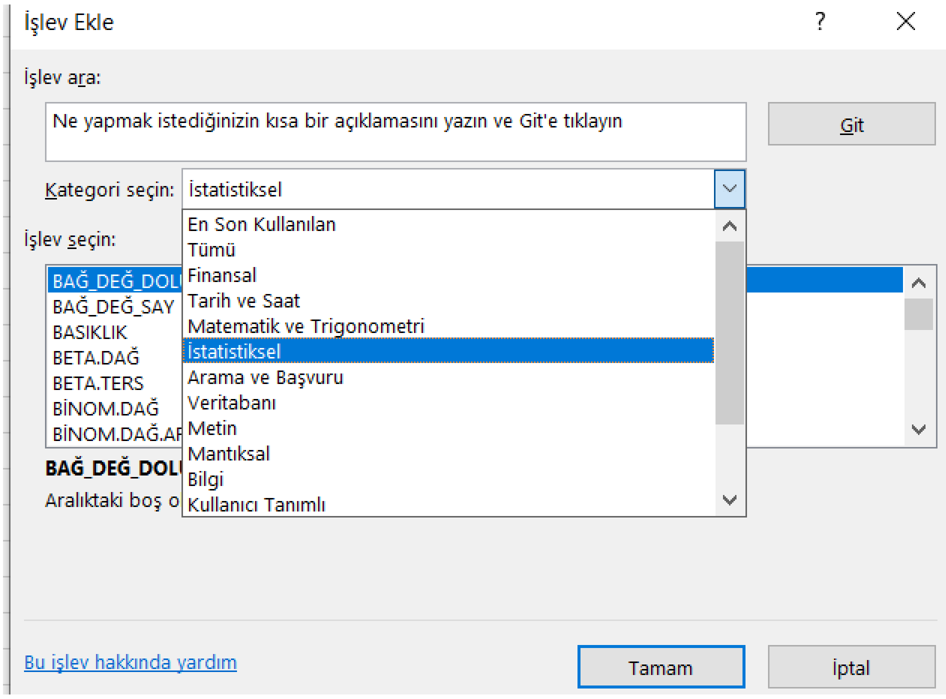
Örnek: Hilesiz bir para 10 kere atılsın. a) 3 kere tura gelme, b) 3 ile 5 arasında tura gelme, c)en az 3 kere tura gelme olasılığını EXCEL fonksiyonları yardımı ile hesaplayınız.

Çözüm:

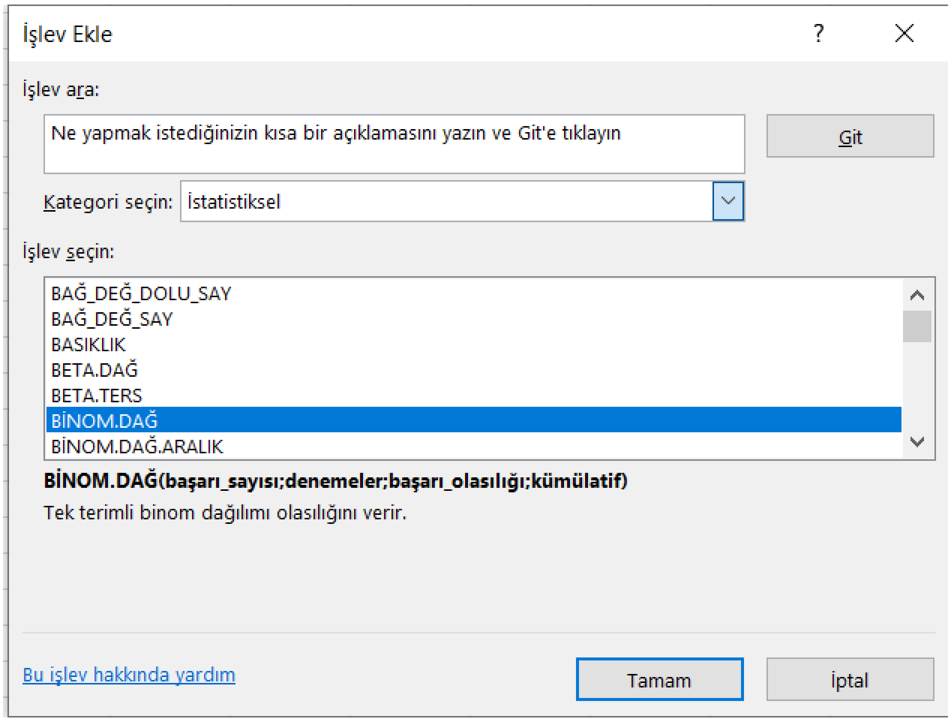
a) Excel dosyasında **"Formüller"** menüsünden **"İşlev ekle"** seçeneği işaretlenir:



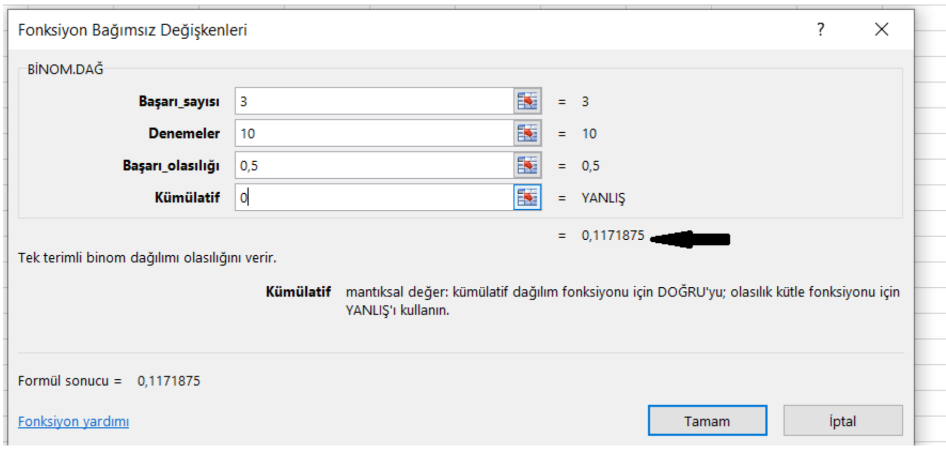
Daha sonra **“İstatistiksel”** fonksiyonlar seçilir:



Buradan da **“Binomdağ”** seçeneği işaretlenir:



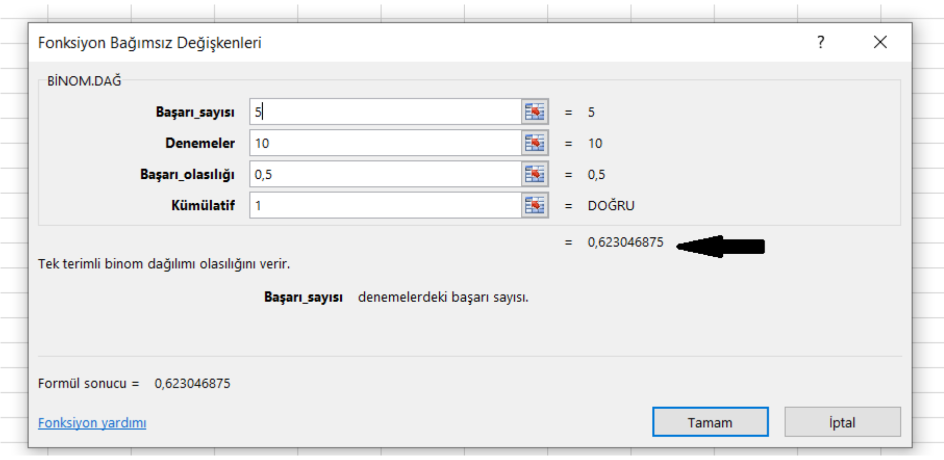
Açılan pencere aşağıdaki gibi doldurulur:



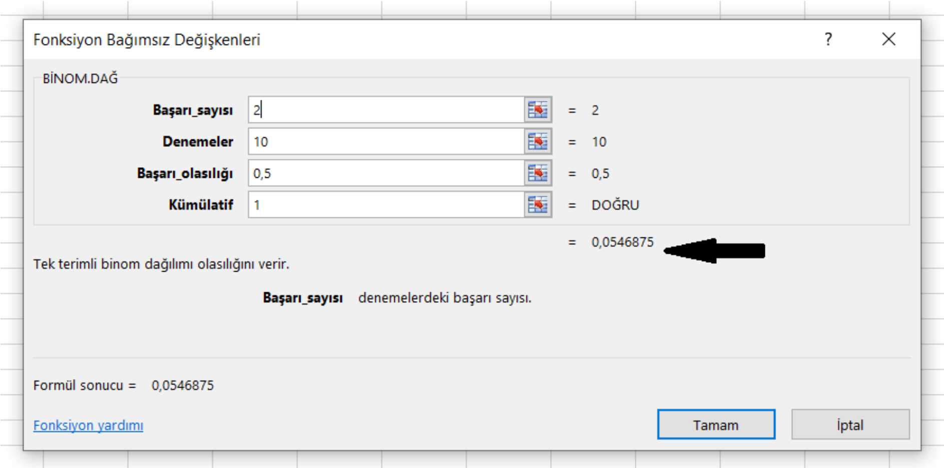
Sonuç yukarıdaki şekilde ok ile belirtildiği gibi 0,117 olarak bulunur. Dikkat edilecek olursa bu sonuç bir önceki sorunun "a" şıkkındaki cevap ile aynıdır.

b) P(3<=X<=5)=P(X<=5)-P(X<=2) Bu olasılığı hesaplamak için EXCEL'de tek bir fonksiyon yoktur ama P(X<=5) ve P(X<=2) olasılıkları ayrı ayrı hesaplanarak fark alınır:

P(X<=5) olasılığı aşağıdaki gibi bulunur:



P(X<=2) olasılığı da benzer şekilde



Dolayısı ile sorulan olasılık 0,623-0,05=0,573 olarak elde edilir.

c) P(X>=3)=1-P(X<3)=1-P(X<=2)=1-0,05=0,95 olarak bulunur.

<u>Binom Dağılımının Beklenen Değeri Ve Varyansı</u>

X rastlantı değişkeni n ve p parametreli bir binom dağılımına uysun.

E(X)=np

$Var(X) = np(1-p)$

<u>Hipergeometrik Dağılım</u>

Bir denemenin başarı ya da başarısızlık olmak üzere yine **iki sonucu** olsun. Rassal deney yine bu denemelerin toplamından oluşsun. Yalnız bu sefer denemeler **bağımsız olmasın**. Bu gibi durumlarda hipergeometrik olasılık dağılımından yararlanmak gerekmektedir. N birimlik bir toplulukta m tane "başarılı" birim olsun. Bu topluluktan n tanesi iadesiz yöntemle (ya da yerine koymaksızın) çekilsin**.** Binom dağılımındaki gibi rastlantı değişkeni bu örnekteki "başarı sayısı" olsun. Bu durumda X rastlantı değişkeninin hipergeometrik dağıldığını söylemek gerekecektir. X'in olasılık dağılımı ise şu şekildedir:

$$P(X=x) = f(x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}} \quad x = 0,1,2,\ldots,n$$

n, N, ve m de X'in olasılık dağılımının parametreleri olmaktadır.

Başka bir deyişle farklı n, N ve m değerleri için farklı olasılık dağılımları üretilebilir. N birimlik bir topluluktan n tanesi; iadesiz yöntemle $\binom{N}{n}$ kadar farklı şekilde seçilebilir. m başarı içerisinden x tanesi; $\binom{m}{x}$ kadar farklı şekilde ve N-m başarısızlık içerisinden n-x tanesi ; $\binom{N-m}{n-x}$ farklı şekilde seçilecektir. Dolayısıyla x tane başarı ve n-x tane başarısızlık birlikte $\binom{m}{x}.\binom{N-m}{n-x}$ farklı şekilde ortaya çıkacaktır. n birimlik bir örnekte x tane başarı gözleme olasılığı

$$P(X = x) = f(x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}} \qquad x = 0,1,2,\ldots,n$$ olur.

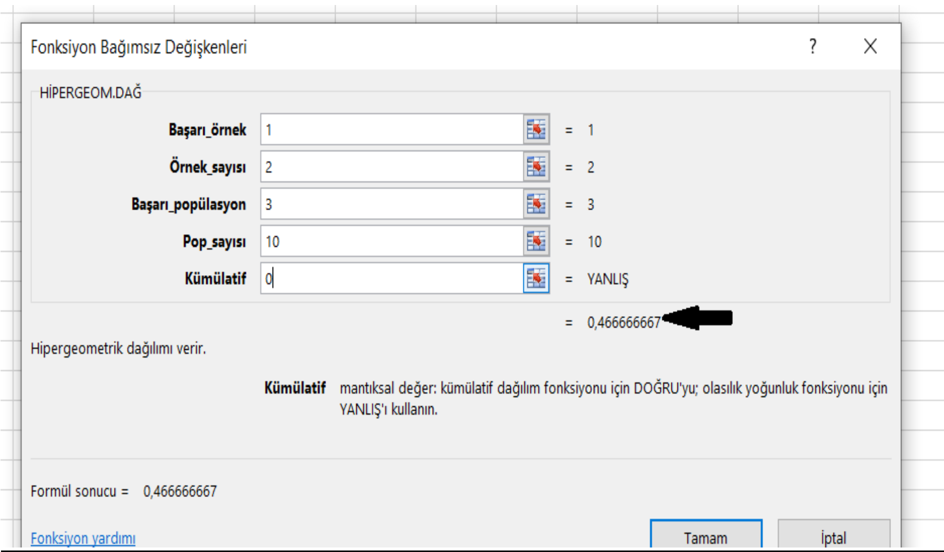
<u>Örnek:</u> 3 kırmızı ve 7 mavi topun bulunduğu bir torbadan iadesiz yöntemle 2 top seçilmektedir. Her iki renkten de birer adet gözleme olasılığı nedir?

<u>Çözüm:</u> X kırmızı top sayısı olsun.

,

P(X=1)= $\frac{\binom{3}{1}\binom{7}{1}}{\binom{10}{2}} = \frac{21}{45}$ bulunur.

<u>EXCEL Çözümü:</u>



Hipergeometrik Dağılımın Beklenen Değeri Ve Varyansı

E(X)= $\frac{mn}{N}$ $\qquad$ $Var(X) = \frac{(N-n)}{(N-1)}.np(1-p)$

olarak verilir. p değerinin de topluluk içindeki başarı oranına karşılık geldiğini de vurgulamak gerekiyor(p=m/N). Binom ve hipergeometrik dağılımın ortalama ve varyanslarını kıyaslamak için aşağıdaki tablo yararlı olacaktır:

Dağılım	**Beklenen Değer**	**Varyans**
Binom	E(X)=np	Var(X)=np(1-p)
Hipergeometrik	E(X)=nm/N =np	$Var(X) = \frac{(N-n)}{(N-1)} np(1-p)$

Tablo 1: Binom ve Hipergeometrik Olasılık Dağılımlarının Beklenen Değer ve Varyanslarının Kıyaslanması

Pratikte n/N≤0,05 olması halinde iki dağılımın birbirine yakın değerler aldığı bilinmektedir. $\frac{(N-n)}{(N-1)}$ oranı ise sonlu ana kütle düzeltme faktörü olarak bilinmektedir. Bazı kaynaklarda ise varyans yerine standart sapma kullanılırken bu ifadenin de karekökü alınacağı için $\sqrt{\frac{(N-n)}{(N-1)}}$ terimine sonlu ana kütle düzeltme faktörü denilmektedir. Literatürde bu konuda bir uyum bulunmamaktadır.

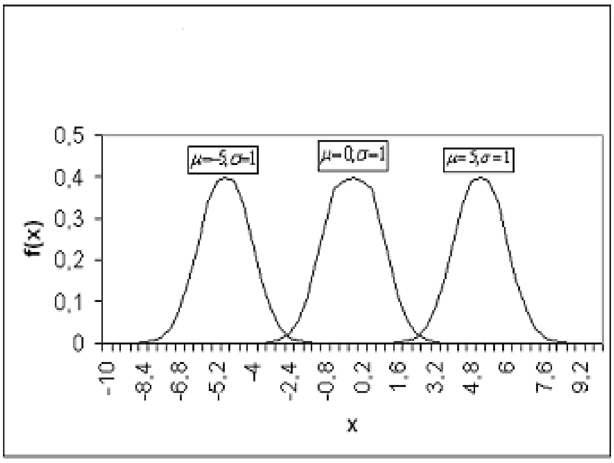
Normal Dağılım

Sürekli rastlantı değişkeni **X**'in aşağıdaki olasılık yoğunluk fonksiyonuna sahipse, normal dağıldığı söylenir:

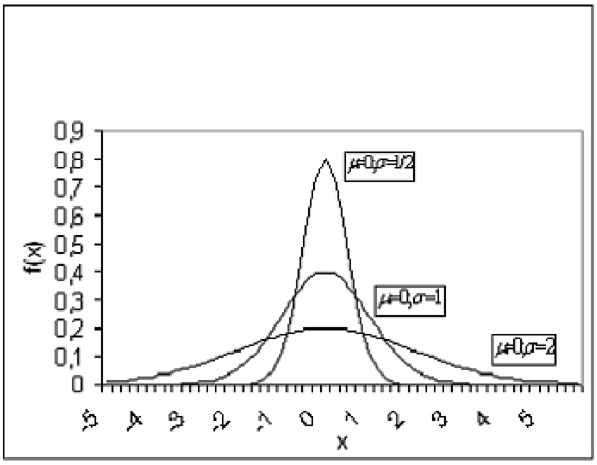
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

Burada μ ve σ dağılımın iki parametresidir. σ>0 olmalıdır. Ancak μ için böyle bir kısıtlama bulunmamaktadır.   μ=0 ve σ=1 için elde edilen normal dağılım standart normal dağılım   olarak adlandırılır.

Aşağıdaki grafikte farklı μ ve σ değerleri için elde edilen farklı normal dağılım eğrileri verilmektedir:



Şekil 12: Standart sapmaları aynı, ortalamaları farklı üç normal dağılım eğrisi



Şekil 13:Ortak Beklenen Değere ve Farklı Varyanslara Sahip Normal Dağılımlar

<u>Normal Dağılımın Beklenen Değeri</u>

μ ve σ parametreli bir normal dağılımın beklenen değeri

E(X)= μ

Aynı zamanda x=μ doğrusu normal dağılım eğrisinin simetri eksenine de eşit olacağı için dağılımın beklenen değeri ile medyanı çakışır.

Normal Dağılımın Varyansı

Var(X)= σ^2

Teorem:X rastlantı değişkeni μ, σ^2 parametreli bir normal dağılıma uysun. Bu durumda $Z = \dfrac{X-\mu}{\sigma}$ değişkeni standart normal dağılır. Başka bir deyişle E(Z)=0 ve Var(Z)=1 olur. Bu durum Z~Normal(0,1) biçiminde de gösterilir.

Örnek: Z rastlantı değişkeni standart normal dağılsın. Buna göre aşağıdaki olasılıkları aşağıdaki tablodan yararlanarak hesaplayınız.

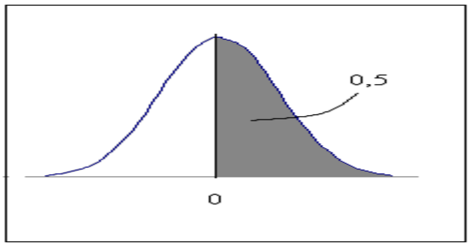
a)P(Z>0)=? b)P(Z>=0)=? c)P(Z<1)=? d)P(Z>1)=? e)P(-1<Z<1)=?

f)P(-2<Z<0)=? g)P(-1<Z<2)=? h)P(-2<Z<-1)=? i)P(-2<Z<1)=?

j) P(|Z|<1,96)=?

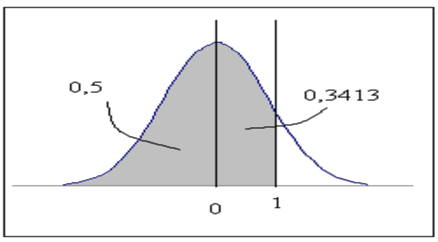
Çözüm:

a) Eğri z=0 ekseni etrafında simetrik olduğu için P(Z>0)=0,5 .



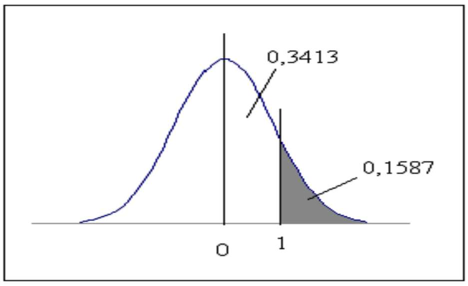
b) P(Z>=0)=0,5 olacaktır.

c)P(Z<1)=P(-∞<z<0)+P(0<=Z<1)=0,5+0,3413= 0,8513 .

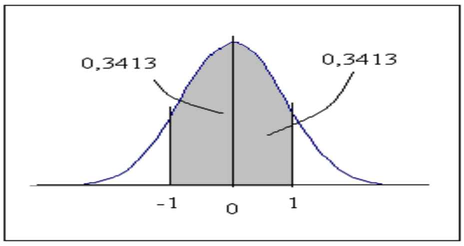


d) Tablodan hareketle P(0<Z<1) olasılığını 0,3413 olarak bulabiliriz.

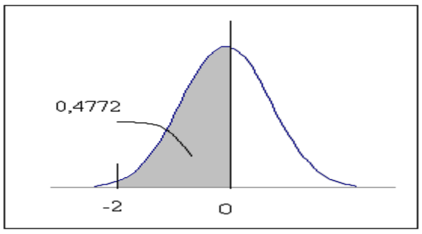
P(Z>1)=0,5-P(0<Z<1)=0,5-0,3413=0,1587 olarak bulunur.



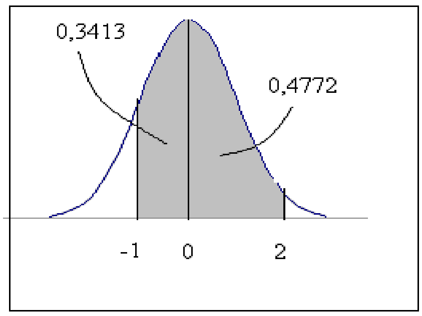
e) $P(-1<Z<1)=2*P(0<Z<1)=2*0,3413=0,6826$



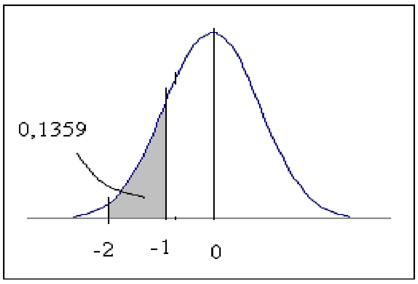
e) Simetri özelliğinden dolayı $P(-2<Z<0)=P(0<Z<2)$ olur. Bu olasılık ise tablodan yararlanılarak 0,4772 olarak bulunur.



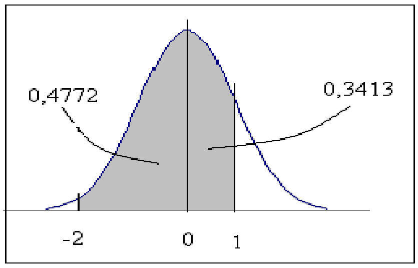
g) $P(-1<Z<2)=0,3413+0,4772=0,8185$ olur.



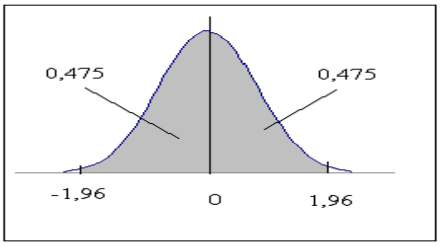
h) $P(-2<Z<-1)=0,4272-0,3413=0,1359$

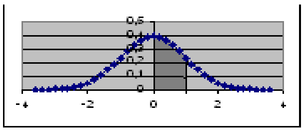


i)P(-2<Z<1)=0,4772+0,3413=0,8185



j)P(|Z|<1,96)=2*P(0<Z<1,96)=2*0,475=0,95 bulunur.



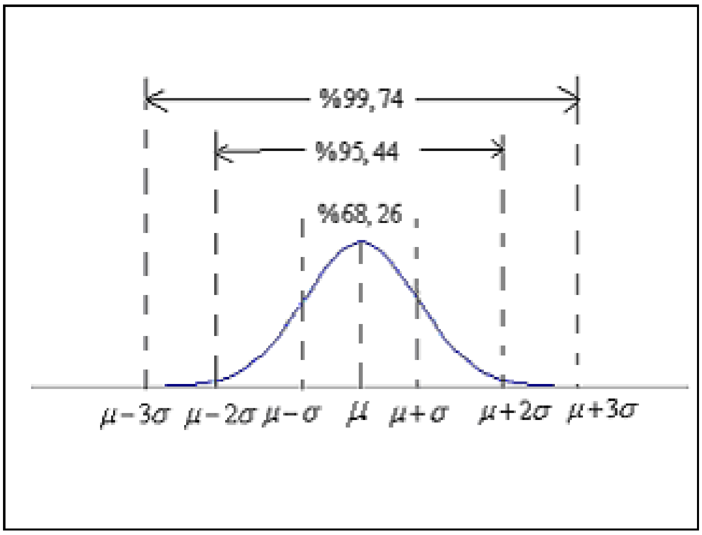


Aşağıdakı değerler standart normal dağılım eğrisi, Z=0 ve Z=z doğruları arasında kalan alanı vermektedir.

Z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0	0,004	0,008	0,012	0,016	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,056	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,091	0,095	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,133	0,1368	0,1406	0,1443	0,148	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,17	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,195	0,1985	0,2019	0,205	0,2088	0,2123	0,2157	0,219	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,239	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,258	0,2611	0,2642	0,2673	0,27	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,291	0,2939	0,2967	0,3	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,326	0,3289	0,3315	0,334	0,3365	0,3389
1	0,3413	0,3438	0,3461	0,3485	0,351	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,373	0,3749	0,377	0,379	0,381	0,383
1,2	0,3849	0,3869	0,3888	0,3907	0,393	0,3944	0,3962	0,398	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,41	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,425	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,437	0,438	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,45	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,459	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,467	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,474	0,4744	0,475	0,4756	0,4761	0,4767
2	0,4772	0,4778	0,4783	0,4788	0,479	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,483	0,4834	0,484	0,4842	0,4846	0,485	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,488	0,4878	0,4881	0,4884	0,4887	0,489
2,3	0,4893	0,4896	0,4898	0,4901	0,49	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,492	0,4922	0,4925	0,493	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,494	0,4941	0,4943	0,495	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,496	0,496	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,497	0,497	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,498	0,4978	0,4979	0,4979	0,498	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,498	0,4984	0,4985	0,4985	0,4986	0,4986
3	0,4987	0,4987	0,4987	0,4988	0,499	0,4989	0,4989	0,4989	0,499	0,499
3,1	0,499	0,4991	0,4991	0,4991	0,499	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,499	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,5	0,4996	0,4996	0,4996	0,4996	0,4997

Şekil 15: Standart Normal Dağılım için olasılık tablosu

Normal dağılımda beklenen değerin ve varyansın ne olduğunun herhangi bir önemi yoktur. Başka bir deyişle bütün normal dağılım eğrileri için $\mu \pm \sigma$; $\mu \pm 2\sigma$; $\mu \pm 3\sigma$;…; $\mu \pm k\sigma$ aralığına düşme olasılıkları hep sabittir ve ortalamadan ve standart sapmadan bağımsızdır.



Şekil 16: Ortalaması μ ve varyansı σ^2 olan bir normal dağılımda $\mu \pm \sigma$; $\mu \pm 2\sigma$; $\mu \pm 3\sigma$;…; $\mu \pm k\sigma$ aralığına düşme olasılıkları, ortalama ve varyansın ne olduğundan bağımsız olarak sabittir.

Örnek: Bir sınıfta yapılan bir sınavda öğrencilerin notları ortalaması 65 standart sapması da 15 olan bir normal dağılıma uysun. Bir öğrencinin geçer not alabilme olasılığını bulunuz. (Dersten başarılı olabilmek için 50 ve üzerinde not almak yeterlidir.)

Çözüm:

$$P(X \geq 50) = \int_{50}^{100} \frac{1}{\sqrt{2\pi}.15} e^{\frac{-(x-65)^2}{2.15}} dx \cong P\left(\frac{X-\mu}{\sigma} \geq \frac{50-65}{15}\right) = 0,8413$$

Binom Dağılımına Normal Dağılım Yaklaşımı

Genel kural olarak binom dağılımının deneme sayısı n çok büyük olduğunda ($n \to \infty$) binom olasılıkları normal dağılım yardımı ile hesaplanabilir. Bununla birlikte pratikte np>5 ve n(1-p)>5 olduğunda binom olasılıkları normal dağılım ile hesaplanabilir.

Örnek: Hilesiz bir para 500 kere atılsın. 75 ile 225 arasında (75 ve 225 de dahil olmak üzere) tura gelme olasılığını bulunuz.

Çözüm: Binom formülünden yararlanarak bu problemi

$P(75 \le X \le 225) = \sum_{x=75}^{225} \binom{500}{x} \left(\frac{1}{2}\right)^{500}$ biçiminde formüle etmek olasıdır. Bununla birlikte bu olasılığı hesaplamak çok pratik değildir. (Yine de Ms-Excel kullanarak bu olasılığın 0,01416 olduğunu söyleyebiliriz.) Öte yandan np=250>5 ve n(1-p)=250>5 koşulları gerçekleştiğine göre rahatlıkla normal dağılımdan yararlanabiliriz.

$E(X) = np = 250 \quad Var(X) = np(1-p) = 125$

$\sigma_x = \sqrt{Var(X)} = 11{,}18$

$$P(75 \le X \le 225) \cong P\left(\frac{75-250}{11{,}18} \le Z \le \frac{225-250}{11{,}18}\right)$$

$= P\left(-15{,}65 \le Z \le -2{,}23\right) = 0{,}0129$ olur.

Bu şekilde yapılan bir hesaplama diğerine göre daha pratik olmaktadır.

Gama Dağılımı

Olasılık teorisinde önemli rol oynayan dağılımlardan bir tanesi olan gama dağılımı, bir dizi olasılık dağılımının genel halini veren **zarf** niteliğinde bir dağılımdır. $\lambda > 0, \quad \alpha > 0$ olmak üzere **X** sürekli rastlantı değişkeni aşağıdaki olasılık yoğunluk fonksiyonuna sahip ise gama rastlantı değişkeni olarak adlandırılır:

$$f(x) = \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \quad x \ge 0$$

Burada $\Gamma(\alpha)$ gama fonksiyonudur ve

$$\Gamma(\alpha) = \int_0^{\infty} e^{-x} x^{\alpha-1} dx$$

şeklinde tanımlanmaktadır. α şekil parametresi; $1/\lambda$ de ölçü parametresi olarak adlandırılır. Tamsayı değerli α parametreli Gama dağılımına Erlang dağılımı da denmektedir.

$\alpha = n$ için $\Gamma(\alpha) = (n-1)!$ olduğu gösterilebilir.

Not1: α=1 için üstel dağılımın olasılık yoğunluk fonksiyonunun elde edileceğine dikkat ediniz. Başka bir deyişle λ parametreli bir üstel dağılım ile α=1 ve λ parametreli bir gama dağılımı birbirine özdeştir

Not2: Ki-kare dağılımı normal dağılımdan çekilen örnek varyansının modellenmesinde kullanılmaktadır. Ki-kare dağılımı gama dağılımının özel halidir. Gama dağılımında $\alpha = \frac{\nu}{2}$ ve $\lambda = \frac{1}{2}$ seçilecek olursa λ parametreli (ya da serbestlik dereceli) bir ki-kare dağılımı elde edilmektedir.

Örnek: X~Gama(α=2,λ=2) olsun. P(0<X<1) olasılığını bulunuz.

Çözüm: $\alpha = 2; \lambda = 2$ için $f(x) = 4xe^{-2x}$ olur.

$$P(0 < X < 1) = \int_0^1 4xe^{-2x}dx = 4\int_0^1 xe^{-2x}dx$$

$$u = x \Rightarrow du = dx; \quad dv = e^{-2x}dx \Rightarrow v = \frac{-1}{2}e^{-2x}$$

Kısmi integrasyon ile

$$= \int_0^1 xe^{-2x}dx = \frac{-xe^{-2x}}{2}\Big|_0^1 + \frac{1}{2}\int_0^1 e^{-2x}dx = \frac{-e^{-2}}{2} - \frac{1}{4}e^{-2x}\Big|_0^1 = \frac{-e^{-2}}{2} + \frac{1}{4}\left(1 - e^{-2}\right)$$

Buradan da $P(0 < X < 1) = \int_0^1 4xe^{-2x}dx$ =0,60

Gama Dağılımının Beklenen Değeri

$$E(X) = \frac{\alpha}{\lambda}$$

Gama Dağılımının Varyansı

$$Var(X) = \frac{\alpha}{\lambda^2}$$