İstatistik Nedir?

İstatistik doğru bir şekilde veri toplama ve verileri bilgiye dönüştürme bilimidir. İki türü vardır.

> **1– Tanımlayıcı İstatistik (Betimleyici – Tanımlayıcı İstatistik)** Elde edilen verilerin sınıflandırılması, ortalama ve yaygınlık ölçülerinin hesaplanması, tablo ve grafiklerle sunulmasını içerir.
>
> Tanımlayıcı istatistikler kategorisi içinde; ortalama, medyan ve mod gibi merkezi eğilim ölçüleri, standart sapma ve varyans gibi ortalamadan sapma ölçütleri çarpıklık ve basıklık gibi normalden sapma ölçütleri yer almaktadır.
>
> **2- Çıkarımsal İstatistik (Öngörüleyici – Tahmin İstatistiği)** Örneklemden elde edilen bulgular yardımıyla popülasyon (evren) hakkında tahminde bulunma, hipotezleri test etme ve karara varma gibi konuları içerir.

ÖNEMLİ KAVRAMLAR

1- Ölçme

Araştırma konusu ile ilgili sayısal değerleri elde etme işlemine ölçme denir.

2- Ölçek (Ölçü)

Ölçme faaliyetlerinin yerine getirilmesi için kullanılması gereken araçlar topluluğuna ölçek denir. Ölçek türleri <u>sayısal değişkenler</u> için eşit aralık ve eşit oran, <u>kategorik değişkenler</u> için nominal (sınıflama) ve ordinal (sıralama) olmak üzere 4'e ayrılır.

> **A- Eşit Aralıklı** ölçek türünde, ölçülen özelliğin belli bir başlangıç noktasına göre eşit aralıklarla ölçülmesi söz konusudur. Takvim, termometre, rakım, sınav başarısı, ilgi vb. İki temel özelliği vardır:
>
> **B- Eşit Oranlı** En hassas ölçme sonuçları bu ölçeklerle elde edilir. Metre, kronometre, tartı …
>
> **C- Nominal (Sınıflama) Ölçeği;** değişkenleri belirli bir özelliğe göre sınıflandırmak, gruplandırmak ve bu özelliğe göre de adlandırmak için yapılan ölçme işlemlerinde kullanılır. Nominal ölçeğe ait birimler sayı veya semboller ile ifade edilebilir ancak nominal değişkene hangi sayı veya değeri verirsek verelim bu kodlama sadece sınıflandırmayla ilgilidir. Örneğin, medeni durumu belirlemek için yapılan bir çalışmada bekar 0, evli 1, boşanmış 3 ile ifade edilebilir. Burada verilen bu sayılar, kişilerin medeni durumlarını birbirinden ayırt etmek için kullanılan sembollerdir.
>
> **D- Ordinal (Sıralama) Ölçeğinde;** değişkenlerin aldığı değerler önem derecesi ya da üstünlükleri baz alınarak sıralanır. Değişkenler büyükten küçüğe, küçükten büyüğe, kısadan uzuna gibi kriterlere göre mantıksal olarak sıralanır. Örneğin öğrenim durumu için bu ölçek kullanılabilir ve ilkokul (1), ortaokul(2), lise(3), lisans (4), lisansüstü (5) şeklinde kodlanılabilir.

3- Popülasyon (Evren)

İlgilenilen sonuçların, yanıtların, ölçümlerin veya sayımların tümünün oluşturduğu kümedir. İlgilendiğimiz büyük, ana kitledir. Mesela nüfus sayımı için kitle Türkiye'dir.

4- Örneklem

Popülasyonun içinden seçilen alt kümedir. Seçtiğimiz örneklem, popülasyonun özelliklerini taşımalı, tarafsız olmalı yani rastgele seçilmiş olmalı. Örneklem kullanarak hem bütünü incelemenin zorluğundan kurtuluruz hem de zamandan ve maliyetten tasarruf etmiş oluruz.

5- Örnekleme

Örneklem seçmek için kullanılan yöntemler topluluğuna örnekleme denir.

> **a. Olasılığa Bağlı Bazı Örnekleme Yöntemleri: Basit Örnekleme, Tabakalı Örnekleme, Sistematik Örnekleme, Küme Örneklemesi**
>
> **b. Olasılığa Bağlı Olmayan Bazı Örnekleme Yöntemleri: Kota Örneklemesi, Kartopu Örneklemesi, Uzman Örneklemesi**

6- Tam Sayım

Araştırma kapsamında, kitledeki tüm birimlere ulaşılarak istenen bilginin elde edilmesi işlemidir. Bazı durumlarda (nüfus sayımları gibi) kitle büyük olsa bile tam sayım yapılması zorunludur.

7- Gözlem Birimi (Gözlem – Denek)

Araştırmada incelediğimiz birimlerdir. Yani örneklemdeki her bir eleman gözlem birimidir. Diğer bir deyişle kitle ya da örneklemde yer alan her birime gözlem ya da denek denir.

> **Kitledeki Gözlem Sayısı : N Örneklemdeki Gözlem Sayısı: n**

8- Parametre

Popülasyonun ölçülebilen, gözlemlenebilen, sayısal olarak ifade edebildiğimiz özelliklerine parametre deriz.

Araştırma kitle yerine örneklem üzerinde uygulanıyorsa, parametre değerleri tahmin edilir bu durumda, örneklemden elde edilen sayısal değerlere istatistik denir.

9- Değişken

Bir durumdan diğerine, gözlemden gözleme farklılık gösterebilen, yeni değerler alabilen ya da sabit kalabilen nesnelere ya da özelliklere değişken adı verilir.

a. Nitel & Nicel Değişken

Kelimelerle ifade edilene nitel değişken denir. Örneğin; cinsiyet, eğitim durumu vb. Sayısal olarak ifade edilenlere nicel değişken denir. Örneğin; yaş, kilo vb.

b. Sürekli & Kesikli (Süreksiz) Değişken

İki birim arasında sonsuz bölünebilme şansı varsa bu ifade süreklidir. Metre, kilogram, yaş bunlara örnektir. Örneğin 3 ile 4 yaş arasında ay, hafta, gün ya da saat özelliğine varana kadar araya birçok değişik ifade girebilmektedir.

İki birim arasında bölünme şansı yoksa kesiklidir (süreksizdir). Ölçümler 0, 1, 2 gibi kesin değerler alır. Ara değerler söz konusu değildir. Nitel değişkenler genellikle kesikli değişkenlerdir. Cinsiyet, tam sayılar bunlara örnek teşkil eder.

VERİLERİN ÖZETLENMESİ

1- Veri

Veriler henüz anlamlı bir sonuç içermeyen, işlenmesi gereken, düzenlenmemiş, ham, organize olmayan sayılar kelimeler ya da nesnelerdir. Veriler organize olana kadar anlamsızdır.

Veriler toplandıktan sonra gruplanarak, sıralanarak ve özetlenerek, kişiler aracılığıyla ya da bilgisayarla işlenip bilgiye dönüştürüldüklerinde anlam kazanmaktalar.

a. Nicel Veri

Sayısal bir ölçekle ölçülerek elde edilmiş verilerdir. Boy uzunluğu, fiyat …

b. Nitel Veri

Kategoriler biçiminde sınıflandırılabilen verilerdir. Cinsiyet, göz rengi …

NOT: Sürekli veride alt sınır ve üst sınır arasında boşluk olmaz. Mesela (2,85 – 2,99) (2,99 – 3,13) (3,13 – 3,27) (3,27 – 3,41). Peki hoca bize 2,99 ya da 3,13 değerini nereye eklemeliyiz diye sorarsa ne yapacağız? (Yukardaki cümlede bahsi geçen örneği esas alarak cevap ver.)

Eğer 2,99'u üst sınırdaki 2,99'a eklediysek 3,14'ü de üst sınırdaki 3,14'e eklemeliyiz.

Eğer 2,99'u alt sınırdaki 2,99'a eklediysek 3,14'ü de alt sınırdaki 3,14'e eklemeliyiz.

2- Örnekleme

Örneklemenin amacı tahmin yapmaktır. Tahminde bulunma, örneklemeyle alınacak kısıtlı bilgiye dayanarak daha geniş çaptaki verilerin karakteristik özellikleri hakkında genelleme yapmaktır.

Örnekleme Sürecinin Aşamaları:

1- Ana kitlenin tanımlanması

2- Örnekleme çerçevesinin belirlenmesi

3- Örnek bireylerinin belirlenmesi

4- Örnekleme yönteminin tanımlanması

5- Örnek büyüklüğünün saptanması

6- Örnekleme planının belirlenmesi

7- Örneklerin seçimi

3- Dağılım

Kitlede ya da örneklemde yer alan her değişkene ilişkin veriler, araştırma konusuna ve araştırılan topluluğa özgü bir dağılım gösterirler. İstatistikte dağılım kavramı 3 başlık altında ele alınmaktadır.

a. Olasılık Dağılımları **b. Örneklem Dağılımları** **c. Sıklık (Frekans) Dağılımları**

Çeşitli yollarla toplanan veriler, özellikleri hakkında bilgi edinmek amacıyla, düzenlemeler yapılarak özet haline getirilir. Verilerin özetlenmesi, ilgilenilen olay ya da problem açısndan ilk yapılacak iştir.

Verilerin düzenlendiği çizelgelere sıklık (frekans) çizelgeleri, verilerin gösterildiği dağılıma sıklık (frekans) dağılımı denir.

Sıklık Çizelgesinin Elde Edilişi:

1- Verileri Sırala	Büyükten küçüğe veriler sıralanır.
2- Dağılım Sınırları	
3- Dağılım Genişliği (R)	Bir örneklemde en büyük değer ile en küçük değer arasındaki farka dağılım genişliği denir.
4- Sınıf (k)	Genelde 7 ile 20 arasında belirlenir. k = 1 + (3,3)*log(n), n: veri sayısı
5- Sınıfın Alt Sınırı (AS) ve Sınıfın Üst Sınırı (ÜS)	
6- Sınıf Aralığı (c)	İki sınıf arasındaki farka sınıf aralığı denir. c = $\frac{R+a}{k}$ R: Dağılım Genişliği k: Sınıf Sayısı a (+1): Sonucumuz tam sayı çıkarsa +1 eklemeye gerek yok ancak sonucumuz küsuratlı çıkarsa +1 ekleyip tekrar hesaplarız. ****Eğer böyle yaparsan sınav kağıdına böyle böyle yaptım diye belirt.****
7- Sınıf (Orta) Değeri (m_i ya da x_i ile gösterilir)	Bir sınıfın alt ve üst sınır değerlerinin ortalamasıdır.
8- Sıklık (Frekans) (f_i)	Bir sınıfa düşen veri sayısı.
9- Göreli Sıklık ya da Frekans (p_i)	Frekansın yüzdelik olarak gösterilme şeklidir.

NOT: Verilerin sınıf aralıklarının geniş tutulması halinde bilgi kaybı olmaktadır. Sebebi ise herhangi bir sınıfın aralığı içine düşen farklı değerler sadece o aralığın orta değeri ile temsil edilirler.

Sınıf Ara Değerleri (SA)

Sınıflar arasındaki değerlerdir. Birinci sınıfın üst sınırı ile ikinci sınıfın alt sınırlarının ortalaması, birinci sınıf ile ikinci sınıf arasındaki sınıf ara değerini verir.

İlk sınıfın ara değeri birinci sınıftan önce bir sınıf varmış gibi kabul edilerek, hesaplanan birinci ve ikinci sınıflar arasındaki değerinden sınıf aralığı (c) çıkarılarak bulunur.

Son sınıf ara değeri ise, sanki son sınıftan sonra bir sınıf daha varmış gibi kabul edilerek, hesaplanan son sınıf ara değerine sınıf aralığı (c) eklenerek bulunur. Sınıf ara değerlerinin sayısı (k+1)'dir.

Birikimli Sıklık (Frekans)

a. Den Daha Az Birikimli Sıklık

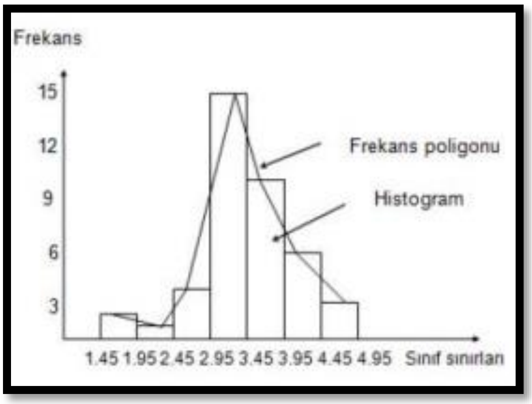
Sınıf ara değerinden daha az değeri olan sınıf sıklıklarının birinci sınıftan başlayarak eklenmesi ile elde edilir.

b. Den Daha Çok Birikimli Sıklık

Sınıf ara değerinden daha çok değeri olan sınıf sıklıklarının birinci sınıftan başlayarak eklenmesi ile elde edilir. Bir sınıf ara değerine karşı gelen den daha az ve den daha çok birikimli sıklıklar toplamı denek sayısına eşittir. Birikimli sıklıklar denek sayısına oranlanırsa den daha az birikimli sıklık yüzdeleri ve den daha çok birikimli sıklık yüzdeleri elde edilir. Birikimli sıklık yüzdeleri sınıf ara değerinden daha az ya da daha çok büyük değerli denek değerlerinin yüzdesini verir.

Histogram

X eksenine sınıf sınırları yani sınıf aralıkları, Y eksenine frekans değerleri yazılarak her bir sınıf için oluşturulan dikdörtgenlerin meydana getirdiği şekle histogram denir.



Şekildeki dikdörtgenlerin orta noktaların birleştirilmesiyle frekans poligonu elde edilir. Poligondaki kırık çizgilerin kaldırılmasıyla verilerin dağılım biçimi kabaca belirlenebilir.

Dağılımın simetrik değil de herhangi bir yöne çarpık olması;

-Örneklemenin hatalı yapıldığı

-Ölçüm skalasının yanlış seçildiği

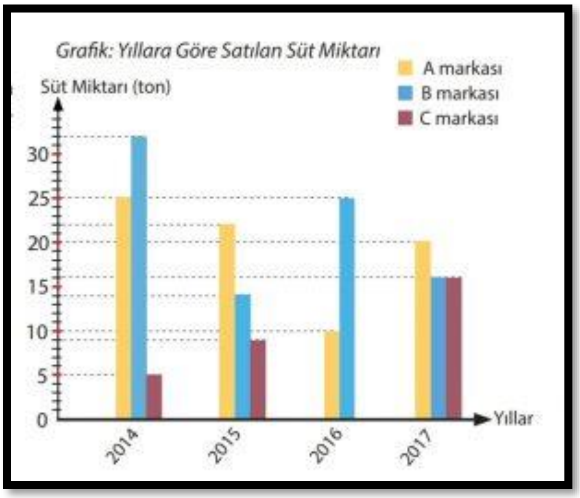
-Örnek büyüklüğünün yetersiz olduğu anlamına gelebilir.

Ayrıca çizilen histogramın veya frekans poligonunun tepe değerinin birden fazla olması;

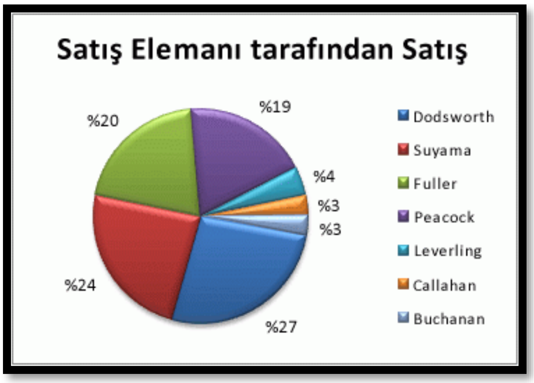
-Ana kütlenin homojen olmadığını ve örnek büyüklüğünün artırılması gerektiğini gösterir.

Sütun Grafiği

Miktarlar arasındaki ilişkiyi göstermek için X eksenine sınıf veya alt sınırlar, Y eksenine mutlak veya nispi miktarlar yerleştirilerek çizilen grafiklerdir.



Pasta Grafiği



MERKEZİ KONUM (EĞİLİM) ÖLÇÜLERİ

1- Aritmetik Ortalama ($\overline{x}$) (µ)

Sınıflandırılmamış verilerde:

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \ldots\ldots X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Sınıflandırılmış verilerde; her sınıfın kendi frekansıyla sınıf orta değeri çarpılıp sonra veri sayısına bölünür.

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

Aritmetik ortalamanın iki önemli özelliği vardır:

A- Verilerin, aritmetik ortalamadan sapmalarının toplamı sıfıra eşittir.

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

B- Verilerin, aritmetik ortalamadan sapmalarının kareleri toplamı minimumdur. Herhangi bir değeri a ile gösterirsek ifade aşağıdaki gibi yazılabilir.

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 < \sum_{i=1}^{n}(X_i - a)^2$$

2- Harmonik Ortalama

Verilerin, terslerinin ortalamasının, tersi bize harmonik ortalamayı verir.

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

3- Ağırlıklı Ortalama

Farklı krediye sahip dersler almış olan bir öğrencinin başarı ortalaması hesaplanırken aritmetik ortalama ile anlamsız sonuçlar bulunur. Böylesi durumlarda verilerin gözlenen frekansları ile çarpılıp toplanması ve elde edilen değerin toplam frekansa bölünmesi ile bulunan değer kullanılır. Bu işleme ağırlıklı ortalama denir.

X_i gözlenen i. veri,
W_i i. verinin gözlenen ağırlığı (frekansı) olmak üzere ağırlıklı ortalama:

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

4- Geometrik Ortalama

$$GO = \sqrt[n]{X_1 * X_2 * X_3 * \ldots\ldots * X_n}$$

NOT: Üç ortalama arasında AO >= GO >= HO ilişkisi vardır.

5- Medyan (Ortanca) ($\overline{x}'$)

Medyan, bir sayısal veri serisi sıralandığında ortada kalan sayıdır.
Veri sayısı tek ise, ortada kalan verimiz medyandır.

Ancak frekansımız (veri sayımız) çift ise; ortadaki iki verinin ortalamasını alacağız. Mesela frekansımız 120 olsun ortalaması 60 gelir ama veri sayımız çift olduğu için 60 ve 61. veriler hangi sınıfta ise o bizim medyan sınıfımız olacak. Eğer frekansımız tek ise direk 2'ye böleceğiz çıkan sonuç hangi sınıfta ise o bizim medyan sınıfımız olacaktır.
Sınıflandırılmış verilerde (frekans tablosunda) medyan hesabı:

$$\text{Medyan} = L + \frac{\frac{N}{2} - F_b}{F_{med}} * c$$

L: Medyan sınıfının alt sınırı
N: Toplam gözlem sayısı
c: Sınıf aralığı (genişliği)
F_{med}: Medyan sınıfının frekansı
F_b: Medyan sınıfından önceki sınıfların frekans toplamı

6- Mod (Tepe Değeri) ($\hat{x}$)

Mod, en çok tekrar eden sayıdır. Sınıflandırılmış verilerde (frekans tablosunda) mod hesabı:

$$\text{Mod} = L + \frac{d_1}{d_1 + d_2} * c$$

L: Mod sınıfının alt sınırı
d1: Mod sınıfı frekansı ile bir önceki sınıf frekansı arasındaki fark
d2: Mod sınıfı frekansı ile bir sonraki sınıfın frekansı arasındaki fark
c: Sınıf aralığı (genişliği)

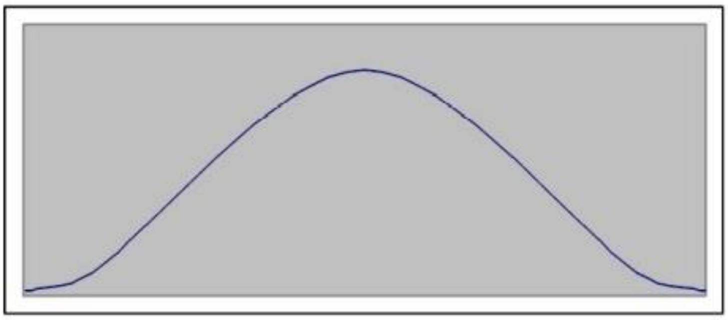
Frekansı en yüksek olan sınıf, mod sınıfı olarak adlandırılır.

Mod sınıfı eğer 1. veya sonuncu sınıf çıkarsa hesaplanmaz.

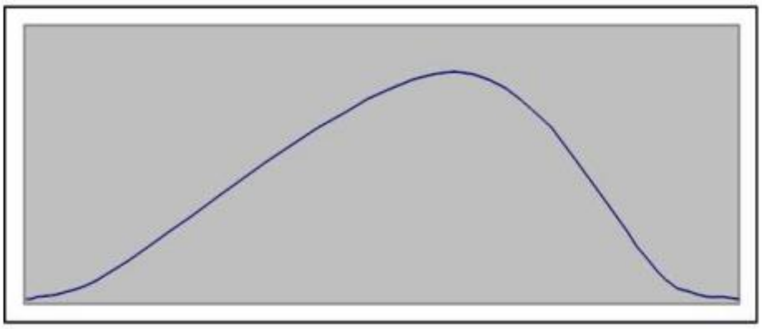
Bazen veri grubunun birden fazla modu olabilir. Bu durum, ilgili ana kütlenin, birkaç alt gruptan meydana geldiğini gösterir.

Ortalama ($\overline{x}$), Medyan ($\overline{x}'$), Mod ($\hat{x}$)Arasındaki Bağlantı

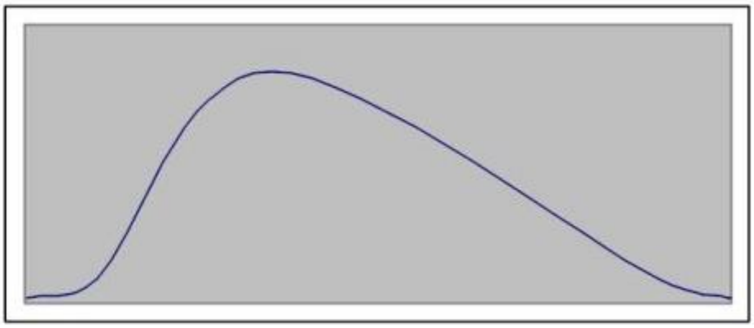
$\overline{x} = \overline{x}' = \hat{x}$ ise sıklık dağılımı simetrikdir.



$\overline{x} < \overline{x}' < \hat{x}$ ise sıklık dağılımı negatif yöne eğilimli ya da sola doğru çarpıktır.



$\overline{x} > \overline{x}' > \hat{x}$ ise sıklık dağılımı pozitif yöne eğilimli ya da sağa doğru çarpıktır.



DAĞILIM ÖLÇÜLERİ

1- Değişim Genişliği

Değişim Genişliği = En Büyük Değer – En Küçük Değer

2- Varyans (σ^2) (S^2)

Varyans, standart sapmanın karesidir. Varyansın büyüklüğü veri grubundaki değişkenliğin fazlalığını ve veri grubunun dağılımının yayvanlığını gösterir. Her bir veriyi ortalamadan çıkartıp karesini alıp hepsini topladıktan sonra veri sayısının bir eksiğine böldük.

Anakütle varyansı	Örnek varyansı
$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$	$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$

Varyans = p_i*(1 - p_i)

Sınıflandırılmış veride varyans:

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

f_i : i. Sınıfın frekansı
m_i: i. Sınıfın sınıf değeri
n: örnek sayısı

3- Standart Sapma (σ) (S)

Standart sapma, verilerin ne kadarının ortalamaya yakın olduğu bilgisini verir. Standart sapma küçükse veriler ortalamaya yakın, büyükse ortalamadan uzakta dağılım gösterirler.

NOT: Sınıflandırılmış veriyi bulmak için sınıflandırılmış varyansın formülünü kullanmalıyız. Sınıflandırılmamış veri için de sınıflandırılmamış varyansın formülünü kullanmalıyız. Ardından karekökünü alarak sonuca erişmiş oluruz.

NOT: Popülasyon varyansı, popülasyon büyüklüğü olan N ile hesaplanır. Ortalama tahminindeki hatayı telafi etmek için n-1 bölen olarak kullanılır.

5- Varyasyon Katsayısı (Değişim Katsayısı)

Ortalamaları birbirinden farklı olan ana kütlelerin, değişkenliklerinin karşılaştırılmasında değişim katsayısı ölçüsü kullanılır. Değişim katsayısının büyüklüğü arttıkça istenilen değerden uzaklaşılır. Değişim katsayısını, standart sapmayı ortalamaya böldükten sonra 100 ile çarparak bulabiliriz.

Anakütle değişim katsayısı	Örnek değişim katsayısı
$DK = \frac{\sigma}{\mu} * 100$	$DK = \frac{S}{\bar{X}} * 100$

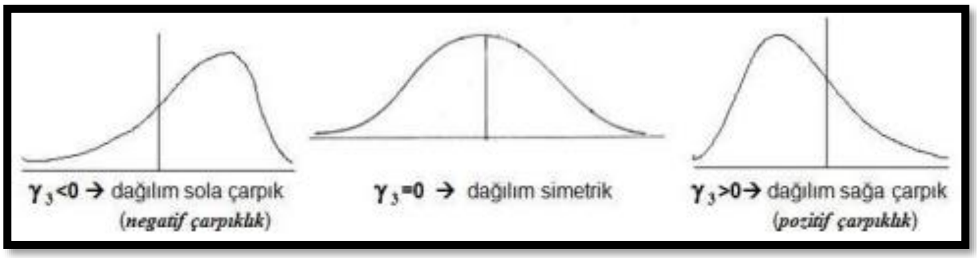
4- Serbestlik Derecesi

Örneklem varyansı n-1 miktarı ile hesaplanır. Bu niceliğe serbestlik derecesi denir.

6- Çarpıklık

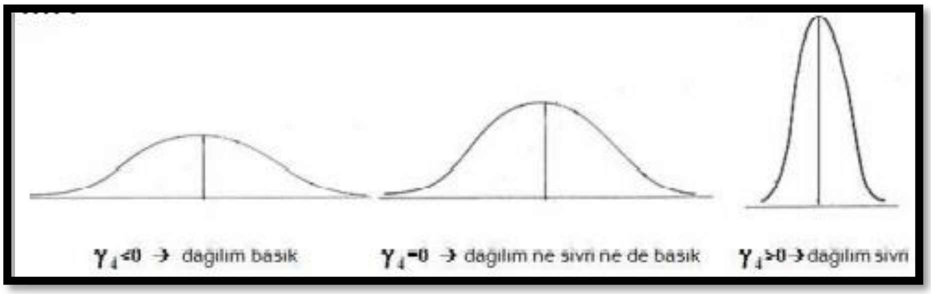
Çarpıklık, bir değişkenin dağılımının simetrik olamayışıdır. Standart sapmaya bölünür.

$$\gamma_3 = \frac{\dfrac{\sum_{i=1}^{n}(X_i-\mu)^3}{n}}{\sigma^3} = \frac{m^3}{\sigma^3}$$



7- Basıklık

Basıklık, dağılımın basıklığını/sivriliğini gösterir. Standart sapmaya bölünür.



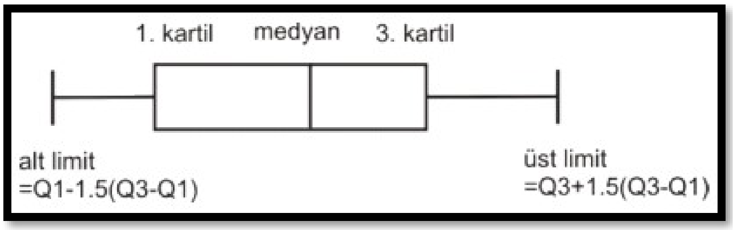
$$\gamma_4 = \frac{\dfrac{\sum_{i=1}^{n}(X_i-\mu)^4}{n}}{\sigma^4} - 3 = \frac{m^4}{\sigma^4} - 3$$

8- Yüzdelikler

Küçükten büyüğe sıralanmış verilerin belli yüzdesini altında bırakan noktadaki gözlenen değerdir. Örneğin; 40. yüzdelik kendisinden önce deneklerin %40'ını kendisinden sonra deneklerin %60'ının olduğu değerdir.

9- Box ve Whisker Grafiği (Kutu Grafiği)

Kutu grafiği; verilerin merkezi, yayılım şeklini ve üç değerlerini gösterir. Minimum değeri, 1. kartili, medyanı, 3. kartili, maksimum değeri gösterir.

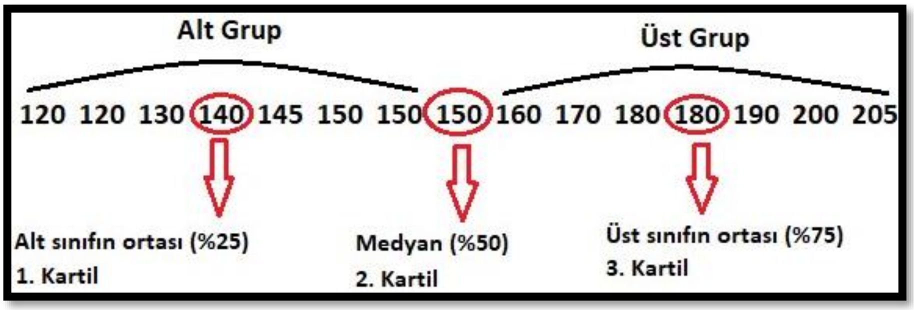


10- Kartiller

Kartiller küçükten büyüğe sıralanmış bir seriyi dört eşit parçaya bölen değerlerdir.

Sınıflandırılmamış verilerin hesabında:

$$Q_1 = \begin{cases} X_j & , j = \dfrac{n+1}{4} \text{ , n tek ise} \\ \dfrac{X_j + X_{j+1}}{2} & , j = \dfrac{n}{4} \text{ , n çift ise} \end{cases} \qquad Q_3 = \begin{cases} X_j & , j = \dfrac{3(n+1)}{4} \text{ , n tek ise} \\ \dfrac{X_j + X_{j+1}}{2} & , j = \dfrac{3n}{4} \text{ , n çift ise} \end{cases}$$



Yukarıdaki 1. kartilde 7 değerin ortancası 1. kartili verdi aynı şekilde 3. kartilde de ortadaki değer 3. kartili verdi, eğer tek değilde (7) çift sayı olsaydı ortadaki ikisini toplayıp ortalamasını aldığımızda yine 1. kartil değerini elde etmiş olacaktık, aynı şekilde bu 3. kartil içinde geçerli.

NOT: Q2 (2. kartil) değerini hesaplamak için Q3-Q1 yaparız bu aynı zamanda medyana eşittir.

NOT: n ya da n+1 dördün katı değilse, çeyrek değerler araya katma yolu bulunurlar.

Sınıflandırılmış verilerin hesabında:

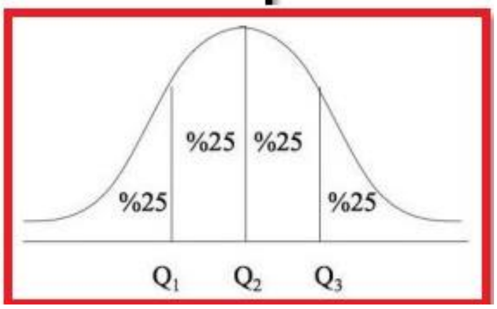
$$Q_1 = L_{Q1} + \frac{\frac{\sum f}{4} - \sum f_i}{f_{Q1}} \times c \qquad Q_2 = L_{Q2} + \frac{\frac{\sum f}{2} - \sum f_i}{f_{Q2}} \times c$$

$$Q_3 = L_{Q3} + \frac{\frac{3\sum f}{4} - \sum f_i}{f_{Q3}} \times c$$

$\sum f_i$: $Q1$'den önceki sınıfa kadar frekansların toplamı

f_{Q1}: $Q1$ sınıf frekansı

c: sınıf genişliği



Toplam f_i her kartil için o kartil sınıfından önceki sınıfların frekans toplamını ifade ediyor. Yani Q2 için Q2 sınıfından önceki sınıfların frekans toplamı, Q3 için de Q3 sınıfından önceki sınıfların frekans toplamı olacak.

NOT: L_{Q1}, ilk %25'lik verinin hangi sınıfa tekabül ettiğini bulur ve o sınıfın alt sınırını baz alınır. L_{Q2}, L_{Q3} ve L_{Q4}'de de mantık aynı L_{Q2} için %50'lik veri ve L_{Q3} için %75'lik veri.

OLASILIK

1- Olasılık Teorisi

Matematiğin belirsizlik taşıyan olaylarla ilgilenen bir dalıdır ve rastgele değişkenleri inceler.

2- Rastgele Değişken

Gelecekteki bir gözlemde, alacağı değer önceden kesinlikle bilinemeyen bir değişkendir. Örneğin; bir zar atışında gelecek sayının önceden bilinememesi gibi.

3- Belirsizliğin Kaynağı

- Daha önceden tahmin edilemeyen çok sayıda etkene bağlı olunması.

- Doğal olaylardaki mevcut değişkenliklerin olması.

Bu tür olaylarda değişkenler, deterministik (rastgele olmayan) bir yaklaşımla incelenemez.

Değişkenin alacağı değeri önceden kesinlikle belirleyen yasalar elde edilemez.

Bunun yerine probabilistik (olasılığa dayalı) yaklaşım gerekir.

Belirsizliklerden hareketle elde edilen verilerden bazı sonuçlar çıkarmak ve tahmin yapabilmek istatistiğin konusudur. Mesela "bu ameliyatın başarı düzeyi %95'tir". Elde edilen sonuçlar kesin olmamakla birlikte belirli bir güven (doğruluk payı) taşımaktadır.

4- Rastgele Olay

Bir rastgele değişkenin bir gözlem sırasında belli bir değeri almasına rastgele olay denir. Rastgele olayın görüleceği önceden kesinlikle bilinemez ancak rastgele olayın görülme ihtimalini belirlemek mümkündür. Örneğin bir zar atışında seçilen bir sayının görülmesi bir rastgele olay olup bunun ihtimali hesaplanabilir.

5- Örnek Uzayı

İlgilenen rastgele olayın alabileceği tüm değerleri içeren uzaydır. Örneğin bir zar atışında gelebilecek sayıların tümü, bir deneyde gözlemlenecek değerlerin tümü.

<u>Kümenin adı büyük harfle, elemanları bu harfe karşılık gelen küçük harf ile gösterilir.</u>

Örneğin; Türkçe'deki sesli harfler kümesi S= {a, e, ı, i, o, ö, u, ü}

6- Küme Kavramı

Bir kümenin bütün elemanları diğer bir kümenin de elemanları ise ilk küme ikinci kümenin alt kümesidir.

Bir elemanın bir kümeye ait olduğu $x \in X$ şeklinde gösterilir.

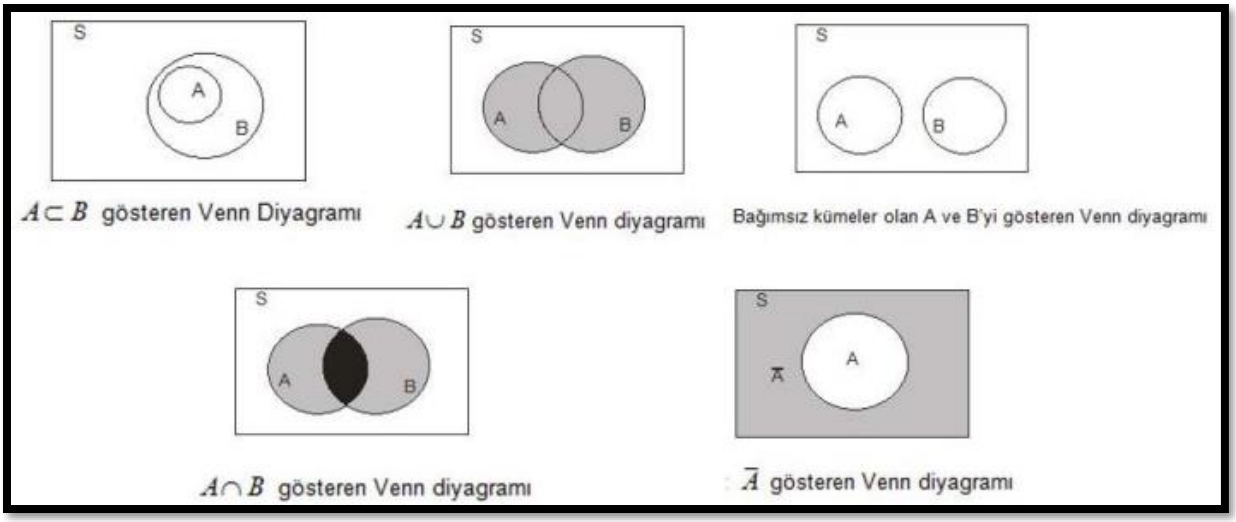
Bir elemanın bir kümeye ait olmadığı $x \notin X$ şeklinde gösterilir.

Hiçbir elemanı bulunmayan bir küme boş küme ($\emptyset$) olarak adlandırılır.

Herhangi A ve B kümeleri için A'nın tüm elemanları B kümesinde ise A, B'nin alt kümesi veya B, A'yı kapsar denir. $A \subset B$ şeklinde ifade edilir.

7- Venn Diyagramı

Ana küme ve alt kümeler arasındaki ilişkileri grafiksel gösterim kullanarak gösterme biçimidir.



Bir A kümesi ile B kümesinin ortak elemanları yok ise "A ∩ B = ∅ " yani bir olayın gerçekleşmesi diğer bir olayın oluşmasını önlüyorsa, bu iki olay "karşılıklı dışlamalı" ya da "birbirini engelleyen olaylar" (mutually exclusive) olarak adlandırılır. Diğer bir deyişle bir olayın olması durumunda diğer başka bir olayın gerçekleşme ihtimalinin sıfır olmasıdır.

Mutually Exclusive (Mesela bir parayı attığınızda ya yazı ya tura gelir. İkisi aynı anda gelemez.): Bir sepette 4 elma, 6 armut olsa elimizi daldırdığımızda elma veya armut alma olasılığımız kaç olur? Elma alma olasılığı 4/10, armut alma olasılığı 6/10. Birisini aldığımızda ötekini alamayız. Dolayısıyla bunlar mutually exclusive'dir. Yani ikisinden birini alma olasılığı 4/10 + 6/10 = 10/10 = 1 olur.

8- Olasılık Kavramı

Bir deneme farklı N sonucu ortaya koyuyor ve bunlardan n tanesinde A olayı meydana geliyorsa, A olayının ortaya çıkma olasılığı, P(A) = $\frac{n}{N}$

Rastgele değişkeni büyük harfle (X), rastgele değişkenin bir gözlem sırasında aldığı değeri küçük harfle (x) gösterirsek X = x_i rastgele olayının olasılığı: P(X = x_i) = p_i

9- Olasılık Aksiyomları

Aksiyom 1:

Herhangi bir E rasgele olayının ihtimali 0 ≤ P(E) ≤ 1

P(E): E rastgele olayının ihtimalini gösterir.

Olasılığın 0 olması söz konusu olayın hiçbir zaman meydana gelmeyeceğini gösterir.

Aksiyom 2:

Eğer örnek uzayı S ise P(S) = 1 yani örnek uzayındaki olayların olasılıkları toplamı 1'e eşittir.

1 olması söz konusu olayın kesinlikle her gözlemde meydana geleceğini gösterir.

Aksiyom 3:

Eğer $E_1, E_2, E_3, \ldots\ldots, E_n$ birbirlerini engelleyen (mutually exclusive) olaylar ise

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$$

Bu aksiyomdan hareketle aşağıdaki özellikler belirlenebilir:

$$P(\bar{E}) = 1 - P(E)$$

$$P(\varnothing) = 0$$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

10- Farklı – Bağımsız Olaylar

İstatistikte olayların bağımsızlığı, bir olay hakkındaki bilgi başka bir olaya bağlı değilse bu olay istatistiksel olarak bağımsızdır (independent).

Bağımsız olaylar, asla birbirlerini engelleyen olaylar (mutually exculsive) olmazlar.

Örnek-1: **Bir para üç kez atılsın. En az 2 kere tura gelmesi olasılığını bulunuz.**

YYY – YYT – YTY – TTT – YTT – TYY – TYT – TTY

P(X) = 4/8 = 1/2   X: En az 2 kere tura gelmesi

Örnek-2: **Bir torbada 5 kırmızı, 7 siyah ve 3 beyaz bilye bulunmaktadır. Bu torbadan rastgele çekilecek bir bilyenin kırmızı gelme olasılığı nedir?**

$$P(A) = \frac{n}{N} = \frac{5}{5+7+3} = 1/3$$

11- Olasılık Kuralları

Olasılık olayları, birbirini tamamıyla engelleyen ve birlikte meydana gelebilen olaylar olmak üzere ikiye ayrılır.

a- Toplama Kuralı

Karşılıklı olarak birbirini engelleyen olaylardan (mutually exclusive) birinin veya diğerinin ortaya çıkma olasılığı, bu olayların ayrı ayrı ortaya çıkma olasılıklarının toplamına eşittir.

A ve B gibi birbirini engelleyen (ayrık) iki olaydan herhangi birisinin meydana gelme olasılığı:

P(A ∪ B) = P(A) + P(B) zira P(A ∩ B) = 0

Mesela bir zar atıldığında 2 veya 3 gelme olasılığı nedir sorusuna cevap olarak: Bu olay <u>birbirini engelleyen</u> özellikte olup, herhangi bir anda sadece tek yüz ile karşılaşılacağından toplama kuralı kullanmalıdır. (1/6) + (1/6) = 1/3

b- Çarpma Kuralı

Birbirinden bağımsız ve aynı zamanda meydana gelebilen olayların olasılığı, bu olayların ayrı ayrı ortaya çıkma olasılıklarının çarpımına eşittir.

Mesela bir zar ve madeni para birlikte atıldığında, paranın yazı ve zarın 5 gelme olasılığı sorulduğunda cevap olarak: Bu olaylar birlikte meydana gelebilen özellikte olup, <u>birbirini engellemez</u>, bu nedenle çarpma kuralı kullanılmalıdır. (1/2) * (1/6) = 1/12

NOT: Bazı olaylarda ise hem birlikte çıkma hem de birbirlerini engelleme söz konusu olabilir. Bu gibi olaylarda çarpma ve toplama kuralı birlikte uygulanır.

Çarpma ve toplama kuralının birlikte kullanıldığı olay sayısı 2 ise yani:

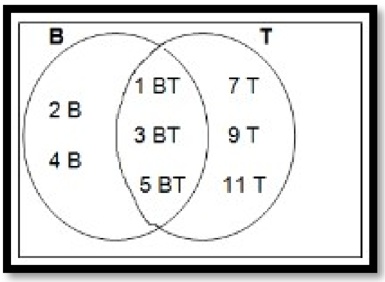
$$P(A \text{ veya } B) = P(A) + P(B) - P(A \text{ ve } B)$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Çarpma ve toplama kuralının birlikte kullanıldığı olay sayısı 3 ise:

$$P(A \text{ veya } B \text{ veya } C) = P(A) + P(B) + P(C) - P(A \text{ ve } B) - P(A \text{ ve } C) - P(B \text{ ve } C) + P(A \text{ ve } B \text{ ve } C)$$
$$P(A \cup B \cup C) = P(A) + P(B) + P(B) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Örnek-1: Bir torbada 1'den 5'e kadar numaralanmış 5 beyaz, 6'dan 12'ye kadar numaralanmış 7 tane siyah bilye vardır. Bu torbadan yapılacak bir çekilişte çıkacak bilyenin beyaz veya tek numaralı olması olasılığını hesaplayınız.

$$P(B \text{ veya } T) = P(B) + P(T) - P(B \text{ ve } T) = \frac{5}{12} + \frac{6}{12} - \frac{3}{12} = \frac{2}{3}$$



12- Koşullu Olasılık

Bir olayın ortaya çıkma olasılığı, başka bir olaya göre değişiyorsa sözü edilen olaylar arasında bağımlılık vardır ve koşullu olasılık uygulanır.

A olayının meydana gelmesi koşulu ile B olayının ortaya çıkma olasılığı P(B/A) şeklinde gösterilir.

$$P(B / A) = \frac{P(A \text{ ve } B)}{P(A)} = \frac{P(A \cap B)}{P(A)} \qquad P(A) \neq 0$$

Yukarıdaki ifade düzenlenirse: $P(A \cap B) = P(A).P(B|A)$

$$P(A| B) = P(A \cap B) / P(B)$$

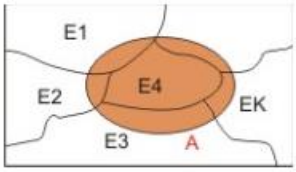
13- Bayes Teoremi

Koşullu olasılıkların hesabında kullanılır. Çeşitli nedenlerin aynı sonucu verebildiği durumlarda bazen sonuç bilindiği halde bunun hangi nedenden meydana gelmiş olduğu bilinemeyebilir. Söz konusu sonucun hangi olasılıkla hangi nedenden ortaya çıktığı araştırılmak istendiğinde Bayes Teoremi'nden yararlanılır. Diğer bir deyişle Bayes Teoremi sonuç belli iken geriye doğru analiz yapma imkânı sağlar.

P(A ∩ B) = P(A) . P(B|A) denkleminden yola çıkarak P(A) . P(B|A) = P(B) . P(A|B) şeklinde eşitlik elde edebiliriz ve P(A|B) = $\frac{P(A)\,.\,P(B|A)}{P(B)}$ **şekline çevirebiliriz.**

Bayes Teoremi:

$E_1, E_2, \ldots, E_n$ ayrık olaylar olsun ve hep birlikte şekilde verildiği gibi S örnek uzayını oluştursun. A olayı bu örnek uzayında bir olay ise



$$P(E_k|A) = \frac{P(E_k \cap A)}{P(E_1 \cap A) + P(E_1 \cap A) + \cdots + P(E_n \cap A)}$$

$$P(E_k \cap A) = P(E_k)P(A|E_k)$$

$$P(E_k|A) = \frac{P(E_k)P(A|E_k)}{P(E_1)P(A|E_1) + P(E_2)P(A|E_2) + \cdots + P(E_n)P(A|E_n)} = \frac{P(E_k)P(A|E_k)}{\sum_{i=1}^{n} P(E_1)P(A|E_i)}$$

Bayes Teoreminin uygulandığı durumlar:

- **Örnek uzayının E1, E2, …,** E_n **şeklinde ayrık olaylara bölündüğü**
- **Bu örnek uzayında, P(B) > 0 şartını sağlayan bir B rastgele olayın varlığında**
- **P(**E_k **∩ B) olasılığının hesaplanması istendiğinde**

Aşağıda tanımlanan ihtimallerden en az birinin bilindiği durumlarda:

- **Tüm** E_k **için P(**E_k **∩ B)**
- **Tüm** E_k **için P(**E_k**) ve P(B |** E_k**)**

Örnek-1: **Bir torbada 3 mavi, 4 beyaz ve 7 kırmızı bilye bulunmaktadır. Üst üste yapılacak iki çekilişten birincisinde mavi, ikincisinde beyaz bilye gelme olasılığını hesaplayınız.**

M: mavi bilye B: beyaz bilye

İadeli çekiliş ---> P(M ve B) = P(M) * P(B) ----> P(M ∩ B) = P(M) * P(B) = (3/14) * (4/14) = 2/196

İadesiz çekiliş --> P(M ve B) P(M) * P(B/M) ----> P(M ∩ B) = P(M) * P(B/M) =(3/14) * (4/13) = 12/182

NOT: **Yukarıdaki problemdeki gibi iki olay arasında bağımlılık varsa P(M) * P(B) ≠ P(M) * P(B/M) ilişkisi yazılabilir.**

Örnek-2: Kusursuz bir tavla zarı atıldığında sonucun çift bir sayı olduğu biliniyor. Bu sayının 4 çıkma olasılığını hesaplayınız.

S = {1, 2, 3, 4, 5, 6} X_1 = {2, 4, 6} X_2 = {4}

Bayes yaklaşımının en basit halini kullanarak çözelim: $P(X_2 / X_1) = \frac{P(X_1 \cap X_2)}{P(X_1)} = \frac{1/6}{3/6} = 1/3$

Örnek-3: Bir bölgede seçmenlerin %40 'ı A partisine %60 ise B partisine oy vermişlerdir. Bir kamuoyu yoklamasında A partisine oy verenlerin %30'u ile B partisine oy verenlerin %70'i Avrupa Birliğine girmeyi desteklemektedirler. Bu bölgeden rastgele seçilen birinin Avrupa Birliğini desteklediği bilindiğine göre B partisinin olma ihtimali kaçtır?

A: A partisine oy vermek   B: B partisine oy vermek   EU: Avrupa Birliğini desteklemek

P(A) = %40 = 0,4   P(B) = 0,6   P(EU/A) = 0,3   P(EU/B) = 0,7

$P(B/EU) = \frac{P(EU/B).P(B)}{P(EU)}$

P(EU) = P(A) . P(EU/A) + P(B) . P(EU/B) = (0,4 . 0,3) + (0,6 . 0,7) = 0,54

(0,7*0,6) / 0,54 = 7/9

Örnek-4: Elektrik ampulü üreten bir fabrikanın üretiminin %20'si A tipi, %80'ide B tipi ampullerden oluşmaktadır. Hatalı üretim oranı A tipi ampullerde %36, B tipi ampullerde ise %18'dir. Rastgele seçilen bir ampulün hatalı olduğu bilindiğine göre bu ampulün A tipi olma olasılığı nedir?

C: Hatalı üretim oranı   Bizden istenilen olasılık = P(A/C)

$P(A/C) = \frac{P(A \cap C)}{P(C)}$ ancak bu formülün payındaki ifade değeri bilinmiyor.

P(C) = P(A) * P(C/A) + P(B) * P(C/B) = (0,2 * 0,36) + (0,8 * 0,18) = 0,216 ifadesi ve

$P(C/A) = \frac{P(A \cap C)}{P(C)}$ -----> $P(A \cap C)$ = P(A) * P(C/A) ifadesi yerine koyuluyorsa

$P(A/C) = \frac{P(A \cap C)}{P(C)} = \frac{P(A) \, . \, P(C/A)}{P(C)} = \frac{0,2 \, . \, 0,36}{0,216}$ = 0,333 olarak bulunur.

Örnek-5: **Üşümesi, hafif baş ağrısı ve ateşi olan ancak burun akıntısı olmayan bir kişinin Bayesçi Karar'a göre soğuk algınlığı var mıdır?**

üşüme	burun akıntısı	baş ağrısı	ateş	soğuk algınlığı
E	H	Hafif	E	**H**
E	E	Yok	H	**E**
E	H	Şiddetli	E	**E**
H	E	Hafif	E	**E**
H	H	Yok	H	**H**
H	E	Şiddetli	E	**E**
H	E	Şiddetli	H	**H**
E	E	Hafif	E	**E**

SA = Soğuk Algınlığı Ü = Üşüme BAK = Burun Akıntısı BA = Baş Ağrısı A = Ateş

P(SA=E) = 5/8 = 0.625	P(SA=H) = 3/8 = 0.375
P(Ü=E|SA=E) = 0.6	P(Ü=E|SA=H) = 0.333
P(Ü=H|SA=E) = 0.4	P(Ü=H|SA=H) = 0.666
P(BAK=E|SA=E) = 0.8	P(BAK=E|SA=H) = 0.333
P(BAK=H|SA=E) = 0.2	P(BAK=H|SA=H) = 0.666
P(BA=Hafif|SA=E) = 0.4	P(BA=Hafif|SA=H) = 0.333
P(BA=Yok|SA=E) = 0.2	P(BA=Yok|SA=H) = 0.333
P(BA=Şiddetli|SA=E) = 0.4	P(BA=Şiddetli|SA=H) = 0.333
P(A=E|SA=E) = 0.8	P(A=E|SA=H) = 0.333
P(A=H|SA=E) = 0.2	P(A=H|SA=H) = 0.666

P(SA=E) * P(Ü=E|SA=E) * P(BAK=H|SA=E) * P(BA=Hafif|SA=E) * P(A=E|SA=E) = 0.018

P(SA=H) * P(Ü=E|SA=H) * P(BAK=H|SA=H) * P(BA=Hafif|SA=H) * P(A=E|SA=H) = 0.009

14- Permütasyon (Sıralama)

Olasılık hesaplarının yapılmasındaki en önemli nokta; olayın meydana gelbileceği yolların sayısı (N) ile istenen olayın meydana gelebileceği yolların sayısını (n) belirlemektir.

n farklı elemanın r tanesi sıra gözetilmek kaydıyla dizilecek olursa elde edilecek değişik düzenlerin sayısına permütasyon denir. Formülü ise P(n, r) = $\frac{n!}{(n-r)!}$

Örnek-1: 20 kişilik genel kurul toplantısında başkan, başkan yardımcısı ve sekreter olmak üzere 3 kişilik idare heyeti seçilecektir. Buna göre,

a) İdare heyeti için kaç farklı heyet oluşturulabilir?

b) Bilinen 3 kişiden A'nın başkan, B'nin başkan yardımcısı ve C'nin de sekreter seçilmesi olasılığı nedir?

3 pozisyon için yapılacak seçimde sıra gözetileceğinden (yani oluşturulan bir ABC heyetinde A başkan, B yardımcı, C sekreter iken, BAC heyetinde B başkan, A yardımcı ve C sekreterdir) permütasyon formülü kullanılır.

a) P(20,3) = $\frac{20!}{(20-3)!}$ **= 6840 heyet oluşturulabilir.**

b) İstenilen sonuç 6840 içinden sadece biridir. P(A, B, C) = $\frac{1}{6840}$ **= 0,00015**

15- Kombinasyon (Birleşim)

n farklı eleman arasından sıra farkı gözetmeksizin r elemanlı farklı grupların oluşturulması işlemine kombinasyon (birleşim) denir. Formülü ise C(n, r) = $\frac{n!}{(n-r)!\,.r!}$

Örnek-1: 10 profesörün bulunduğu bir gruptan seçilecek 3 kişilik jürinin istenen şahıslardan meydana gelme olasılığı nedir?

3 pozisyon için yapılacak seçimde sıra gözetilmeyeceğinden (yani oluşturulan bir XYZ heyetinde X jüri üyesi, Y jüri üyesi, Z jüri üyesi iken YXZ heyetinde de Y jüri üyesi, X jüri üyesi, Z jüri üyesidir) kombinasyon formülü kullanılır.

10 C_3 **=** $\frac{10!}{(10-3)!\,.\,3!}$ **= 120 farklı jüri oluşturulabilir.**

İstenilen sonuç 120 farklı jüriden sadece biridir. -----> P(X, Y, Z) = 1/120

Örnek-2: 4 tarih, 3 felsefe, 3 matematik kitabı olmak üzere toplam 10 kitap rafta kaç değişik şekilde sıralanabilir?

$\frac{10!}{4!\,.\,3!\,.\,3!}$ **= 4200 farklı sıralama yapılabilir.**

Karışık Soru:

1- **1, 2, 3, 4 sayılarıyla yazılabilen 5 basamaklı sayıların ne kadarında 1 ve 2 sayıları bulunur?**

|E| = 4^5 A1 = {1 bulunmayan} A2 = {2 bulunmayan} A1 ∩ A2 = {1 ve 2 bulunduran}

|A1 ∩ A2| = |E| - |(A1 ∩ A2)'| = |E| - |A1' U A2'| = |E| - |A1'| + |A2'| - |A1' ∩ A2'|

4^5 - 3^5 + 3^5 - 2^5 = 570

2- **1, 2, 3, 4, 5, 6 sayıları ile kullanılan sayı tekrar kullanılmamak üzere 6 basamaklı sayılar yazılıyor. Bunların kaçında 1,2,3 veya 5,6 sayıları verilen sırayla yan yana bulunmaktadır?**

A1 = {1, 2, 3 yan yana bulunduran} A2 = {5, 6 yan yana bulunduran}

|A1 U A2| = |A1| + |A2| - |A1 ∩ A2|

= 4! + 5! - 3! = 138

3- **3 evli çift eşler yan yana gelmeyecek şekilde bir sıraya kaç şekilde dizilebilir?**

|E| = 6! A = {1. çiftin yan yana oturması} B = {2. çiftin yan yana oturması}
C= {3. çiftin yan yana oturması}

|A| = |B| = |C| = 5! * 2!

|A' ∩ B' ∩ C'| = |E| - |A U B U C|

= |E| - (|A| + |B| + |C| - |A ∩ B| - |B ∩ C| - |A ∩ C| + |A ∩ B ∩ C|

|A ∩ B ∩ C| = 3! * 2! * 2! * 2!

|B ∩ C| = |A ∩ B| = |A ∩ C| = 4! * 2! * 2!

|E| - (|A| + |B| + |C| - |A ∩ B| - |B ∩ C| - |A ∩ C| + |A ∩ B ∩ C|

6! – (6 * 5! – 12 * 4! + 8 * 3!) = 240

RASTGELE DEĞİŞKENLER

Rastgele değişken, örnek uzaydaki her elemana verilmiş olan sayısal bir değerdir. Rastgele değişkenleri diğer değişkenlerden ayıran özellik, almış olduğu değerleri belli bir olasılıkla almasıdır.

Olasılık kuramında rastgele değişkenler genellikle X, Y, Z, … gibi büyük harflerle, değişkenin aldığı değerler ise x, y, z, … gibi küçük harflerle belirtilir.

Rastgele değişken, aldığı sayısal değerlere göre iki farklı şekilde olabilir:

- **Kesikli (Süreksiz) Rastgele Değişken**
- **Sürekli Rastgele Değişken**

Süreksiz (Kesikli) Rastgele Değişkenler

Bir x rastgele değişkeninin alacağı değerlerinin sayısı sonlu veya sayılabilir ise X'e kesikli (süreksiz) rastgele değişken denir.

Örneğin; zar atışı, paranın yazı tura atışı, fabrikanın haftalık üretimindeki kusurlu mal sayısı.

Sürekli (Sayılamayan) Rastgele Değişkenler

Bir x rastgele değişkeni, belli bir aralıkta bütün değerleri alabiliyorsa yani sayı çizgisinde bir aralığı kaplıyorsa X'e sürekli (sayılamayan) rastgele değişken adı verilir. Ölçümler doğruluk ve hassasiyetlerine bağlı olarak herhangi bir değer alabilirler.

Örneğin; bir sınıftaki öğrencilerin boy uzunluğu, aracın belirli bir zamanda aldığı yol, bir koşuya katılanların bitirme süreleri.

Rastgele değişkenlerin kesikli ve sürekli ayrımı basit olarak; tanımlanan rastgele değişkenin tam sayı olma durumunda kesikli, diğer durumlarda sürekli şekliyle olabilir.

Olasılık Matematiksel Modelleri (Fonksiyonları)

Rastgele değişkene ait matematiksel modeller (fonksiyonlar):

Kesikli rastgele değişkenler için olasılık fonksiyonu

Sürekli rastgele değişkenler için olasılık yoğunluk fonksiyonu olarak tanımlanır.

Olasılık Fonksiyonu

Kesikli bir rastgele değişken olduğunda, $P\{x_i\} = P\{X = x_i\}$ fonksiyonuna rastgele değişkeninin olasılık fonksiyonu denir. X kesikli rastgele değişkenin değer kümesi A = $\{X_i \mid = 1, 2, 3, \ldots\}$

$X_i \in A$ ----> $P\{X = x_i\} = P(x_i)$ **P(X): X'in**
$X_i \notin A$ ----> $P\{x_i\} = 0$ **olasılık fonksiyonu**

$$f(x_i) \geq 0$$
$$\sum_{i=1}^{n} f(x_i) = 1$$
$$f(x_i) = P(X = x_i)$$

Olasılık fonksiyonu olabilmesi için aşağıdaki nitelikler olmalı:

1- $X_i \in A$ için $0 \leq P\{x_i\} \leq 1$

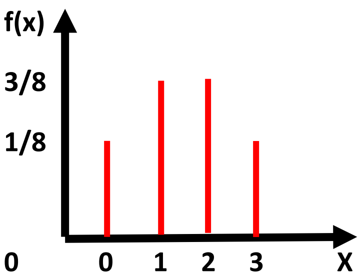
2- $\sum_A P\{x_i\} = 1$

"Olasılık kütle fonksiyonu (1'den küçüktür) olarak da geçer."

Örnek-1: **Bir adet hilesiz bozuk paranın üç kez atılması denemesinde, X rastgele değişkeni gelen yazı sayısı olduğuna göre, X rastgele değişkeninin aldığı değerleri ve bu değerleri alması olasılıklarını belirtiniz.**

Örneklem uzayı; S = {(Y,Y,Y) (Y,Y,T) (Y,T,Y) (T,Y,Y) (Y,T,T) (T,Y,T) (T,T,Y), (T,T,T)} olacaktır. Görüldüğü gibi X rastgele değişkeninin alabileceği değerler 0, 1, 2 ve 3 tür.

P(X=0) = f(0) = 1/8 P(X=1) = 3/8 P(X=2) = 3/8 P(X=3) = 1/8



Örnek-2: İki hilesiz zarın atılması olayını dikkate alalım. X; atılan iki zardaki sayıların toplamını gösteren rastgele değişkeni olsun.

X = {2, 3, … , 12}

P(x=2) = P[(1,1)] = 1/36 P(x=3) = P[(1,2), (2,1)] = 1/8 P(x=4) = P[(1,3), (2,2), (3,1)] = 1/12

Aşağıdaki olasılık değerlerini, olasılık fonksiyon kullanarak elde ediniz.

a) X'in olasılık fonksiyonunu bulunuz.

X_i	$P(x_i)$
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

b) Toplamın 7 veya 11 olması

A = {x=7 olması}

B = {x=11 olması}

P(A U B) = P(A) + P(B)
= P(x=7) + P(x=11)

= 6/36 + 2/36 = 8/36

c) Toplamın 8'den büyük olması

P(x>8) = P(9)+P(10)+P(11)+P(12)

$$= \frac{4+3+2+1}{36} = \frac{5}{18}$$

d) Toplamın 3'den büyük fakat 9'dan küçük olması

$$P(3 < x < 9) = \frac{3+4+5+6+5}{36} = \frac{23}{36}$$

Kümülatif Dağılım Fonksiyonu

X kesikli rastsal değişkeninin kümülatif dağılım fonksiyonu F(x) ile gösterilir.

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

X kesikli rastsal değişkeninin, kümülatif dağılım fonksiyonu aşağıdaki özellikleri gösterir:

1- $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$
2- $0 \leq F(x) \leq 1$
3- Eğer $x \leq y$ ise o zaman; $F(x) \leq F(y)$ olur

Örnek-1: Aşağıdaki kümülatif dağılım fonksiyonunu kullanarak, X rastsal değişkeninin olasılık kütle fonksiyonunu oluşturunuz:

$$F(x) = \begin{cases} 0 & x < 2 \\ 0,2 & -2 \leq c < 0 \\ 0,7 & 0 \leq x < 2 \\ 1 & 2 \leq x \end{cases}$$

f(-2) = 0,2 – 0 = 0,2

f(0) = 0,7 – 0,2 = 0,5

f(2) = 1 – 0,7 = 0,3

Beklenen Değer

Bir rastsal değişkenin herhangi bir olasılık fonksiyonundan aldığı tüm değerlerin ortalaması o rastsal değişkenin beklenen değeridir.

$$\mu = E(X) = \sum_x x f(x)$$

Örnek-1: 2 parayı havaya atalım, x = tura ise x rassal değişkeninin beklenen değeri nedir?

E(x) = (0 * 0,25) + (1 * 0,5) + (2 * 0,25) = 1,0

Varyans ve Standart Sapma

$$\sigma^2 = V(X) = E(X - \mu)^2 = \sum_x (X - \mu)^2 \,.\, f(x) = \sum_x x^2 \,.\, f(x) - \mu^2$$

Diğer bir deyişle $E(x^2)$: X rastsal değişkeninin karesinin beklenen değeri iken:

$E(x^2) = \sum_x x^2 f(x)$

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

Standart Sapma:

$\sigma = \sqrt{\sigma^2}$

Örnek-1: Bir bilgisayar ağında bir saatte gönderilen mesajların dağılımı aşağıdaki gibidir:

x = mesaj sayısı	10	11	12	13	14	15
f(x)	0,08	0,15	0,3	0,2	0,2	0,07

Saatlik mesaj sayısının ortalama ve standart sapmasını bulunuz.

E(x) = (10*0,08) + (11*0,15) + (12*0,3) + (13*0,2) + (14*0,2) + (15*0,07) = 12,5

$E(x^2) = (10^2*0,08) + (11^2*0,15) + (12^2*0,3) + (13^2*0,2) + (14^2*0,2) + (15^2*0,07) = 158,1$

$V(x) = 158,1 - (12,5)^2 = 1,85$

$\sigma = 1,36 = \sqrt{V(x)}$

Örnek-2: Bir otomobil galerisinin günlük satışlarının dağılımı aşağıdaki gibidir:

X	0	1	2	3	4	5	6	7	8
P(X)	0,02	0,08	0,15	0,19	0,24	0,17	0,1	0,04	0,01

Bu dağılışa göre galerinin

a) 5'ten fazla otomobil satma ihtimali nedir? 0,1 + 0,04 + 0,01 = 0,15

b) Satışların beklenen değerini hesaplayınız. E(x) = (0*0,02) + (1*0,08) + … + (8*0,01) = 3,72

c) Satışların varyansını bulunuz.

$E(x^2) = (0^2*0,02) + (1^2*0,08) + … + (8^2*0,01) = 16,68$

$V(x) = 16,68 - (3,72)^2 = 2,84$

$\sigma = 1,68$

SÜREKSİZ OLASILIK DAĞILIMLARI

1- Düzgün (Üniform) Dağılım

X süreksiz rastgele değişkeninin tümü, eşit olasılıklı N sonuca sahipse düzgün dağılıma sahiptir.

$$f(x) = P(X = x) = \frac{1}{N}, \; x = x_1, x_2, \ldots, x_N$$

Ortalama ve varyans $x_1 = a$, $x_2 = a + 1, \ldots, x_N = b$ **ve** $a \leq b$ **olduğu zaman**

$$\mu = E(X) = \frac{b + a}{2} \qquad \sigma^2 = \frac{(b - a + 1)^2 - 1}{12}$$

2- Bernoulli Dağılımı

Bir X rastgele değişkeni için yalnızca iki sonuç varsa X'e Bernoulli rastgele değişkeni denir. Örneğin, paranın tek atışında Y ya da T gelmesi, tek bir oyunda kazanma ya da kaybetme vb.

X rastgele değişkeni 0 ve 1 değerlerini alsın. X'in olasılık fonksiyonu:

$P(X=1) = p \qquad P(X=0) = 1\text{-}p$ **veya** $f(x) = P(X=x) = p^x * (1-p)^{1-x}$, **x=0,1'dir.**

Bernoulli Dağılımının ortalaması ve varyansı sırasıyla:

$$\mu = E(X) = p \qquad \sigma^2 = E(x^2) - [E(x)]^2 = pq = p(1-q)$$

3- Binom Dağılımı

Bernoulli deneyi gibi iki sonuçlu olayların olasılığının hesaplanmasında kullanılır.

Binom dağılımına uyması için aşağıdaki şartların sağlaması gerekir:
- **Deneme belirli sayıda (n) tekrarlanır.**
- **Her deneyin başarılı ve başarısız olmak üzere iki sonucu vardır.**
- **Deneydeki her bir deneme diğer denemelerden bağımsızdır.**
- **Başarı olasılığı (p) ve başarısız olma olasılığı q = 1 – p'dir.**
- **n deneyde elde edilen başarılı sonuçlar x değişkenine atanır.**

Binom dağılımının olasılık fonksiyonu:

$$P(n, X, p) = \left(\frac{n!}{(n-X)!\,X!}\right) p^X q^{n-X}$$

n: Deneyin tekrarlanma sayısı
X: İstenen sonuç sayısı
p: İstenen başarılı sonucun olasılığı
q: Başarısızlık olasılığı

Binom dağılımının ortalaması ve varyansı aşağıdaki formüllerle hesaplanır:

$$E(X) = \mu = n * p \qquad V(X) = \sigma^2 = n * p * q$$

Örnek-1: 10 yazı/tura atmada 4 yazı gelme olasılığını hesaplayınız.

P(10, 4, (0,5)) = $\frac{10!}{6!\,.\,4!}$ * $0,5^4$ * $0,5^6$ = 0,205

Bir zarın 20 kez atılması durumunda tam 12 kez altı gelme olasılığını hesaplayınız.

P(20, 12, $\frac{1}{6}$) = $\frac{20!}{8!\,.\,12!}$ * $(\frac{1}{6})^{12}$ * $(\frac{5}{6})^{8}$ = 0,0000135

4- Poisson Dağılımı

İlgilenilen zaman aralığı, uzunluk veya hacimde nadir rastlanan olayların olasılık dağılımları Poisson ile ifade edilir. Örneğin; belirli bir trafik noktasında meydana gelen trafik kazası sayısı, 1 m^2 kumaştaki kusur sayısı, 1 cm^3 kandaki anormal hücre sayısı vb. sayılabilir.

Poisson dağılımının olasılık fonksiyonu aşağıda verilmiştir.

P(X) = $\frac{\mu^x\,.\,e^{-\mu}}{X!}$ $\quad\quad$ $\mu = \sigma^2 = n * p$

olarak ifade edildiğinden dağılımın tek parametresi olduğu söylenebilir.

Örnek-1: Bir sınıftaki öğrenciler üzerine yapılan bir araştırmada dersi dinlemeyen öğrenci sayısının ortalama olarak 3 kişi olduğu belirlenmiştir. Herhangi bir derste;

a) En az bir kişinin dersi dinlememesi olasılığını hesaplayınız.

P(x ≥ 1) = 1 – P(x = 0) = 1 - $\frac{e^{-3}\,.\,\mu^0}{0!}$ = 1 - e^{-3} = 0,95

b) En fazla iki kişinin dersi dinlememesi olasılığını hesaplayınız.

P(x ≤ 2) = P(x=0) + P(x=1) + P(x=2)

$$= \frac{e^{-3}.\,3^0}{0!} + \frac{e^{-3}.\,3^1}{1!} + \frac{e^{-3}.\,3^2}{2!} = 8,5 * e^{-3} = 0,423$$

5- Geometrik Dağılım

Bir Bernoulli deneyi ilk başarılı (p) sonuç elde edilmesine kadar tekrarlansın ve başarıya ulaşana kadar yapılan deneylerin sayısı da X rastsal değişkeni olsun.

Tek bir denemede başarısızlık ihtimali q = 1 - p olduğunda X'in olasılık fonksiyonu:

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

Ortalama değer ve varyans sırasıyla

$\mu = E(X) = \frac{1}{p}$ $\quad\quad$ $\sigma^2 = E(x^2) - [E(x)]^2 = \frac{q}{p^2} = \frac{1-p}{p^2}$

Örnek-1: 1 elde edilinceye kadar bir zar atalım.

a) Bağımsız atışlar dizisinde, ilk 1'in elde edilmesi için gereken atışların sayısının olasılık fonksiyonu nedir?

p = $\frac{1}{6}$ q = $\frac{5}{6}$ F(x) = $(\frac{5}{6})^{x-1} \cdot \frac{1}{6}$

b) 3. atışta 1 bulmanın olasılığı nedir?

F(3) = $\frac{5^2}{6^3}$ = 25/216

Örnek-2: Bir hastalığa yakalanma olasılığı 0,05'tir. Bu hastalığa yakalanan kişiler tespit edilmek istenmiştir. Seçilen 5. kişinin, bu hastalığa yakalanan ilk kişi çıkma olasılığı nedir? p=0,05

X: Hastalığa yakalanan ilk kişi bulunana kadar seçilen kişi sayısı

P(X=5) = p * $(1-p)^{x-1}$ = (0,05) * $(1-0,05)^{5-1}$ = 0,041

Sürekli Rastgele Değişkenler

Değerleri ölçümle ya da tartımla elde edilen, bir başka ifadeyle sayımla elde edilemeyen bir rassal değişkene sürekli rassal değişken adı verilir.

Örneğin; bir sınıftaki öğrencilerin boy uzunluğu, aracın belirli bir zamanda aldığı yol, bir koşuya katılanların bitirme süreleri.

Rastgele değişkenlerin kesikli ve sürekli ayrımı basit olarak şu şekilde ifade edilebilir; tanımlanan rastgele değişkenin tamsayı olma durumunda kesikli, diğer durumlarda sürekli şeklindedir.

X rastgele değişkeninin R 'deki değer kümesi olan A sayılamaz bir küme ise X'e sürekli rastgele değişken denir. Değer kümesi A = {X | a ≤ X ≤ b} şeklinde yazılır.

Olasılık Yoğunluk Fonksiyonu

X sürekli rastgele değişkenin değer kümesi A = {X | a ≤ X ≤ b}

$x_i \notin$ A ---> f(x_i) = 0

A_i = (a, b] ⊂ A için P(X ∈ A_i) = $\int_a^b f(x)dx$ ----> f(x): X'in olasılık yoğunluk fonksiyonu

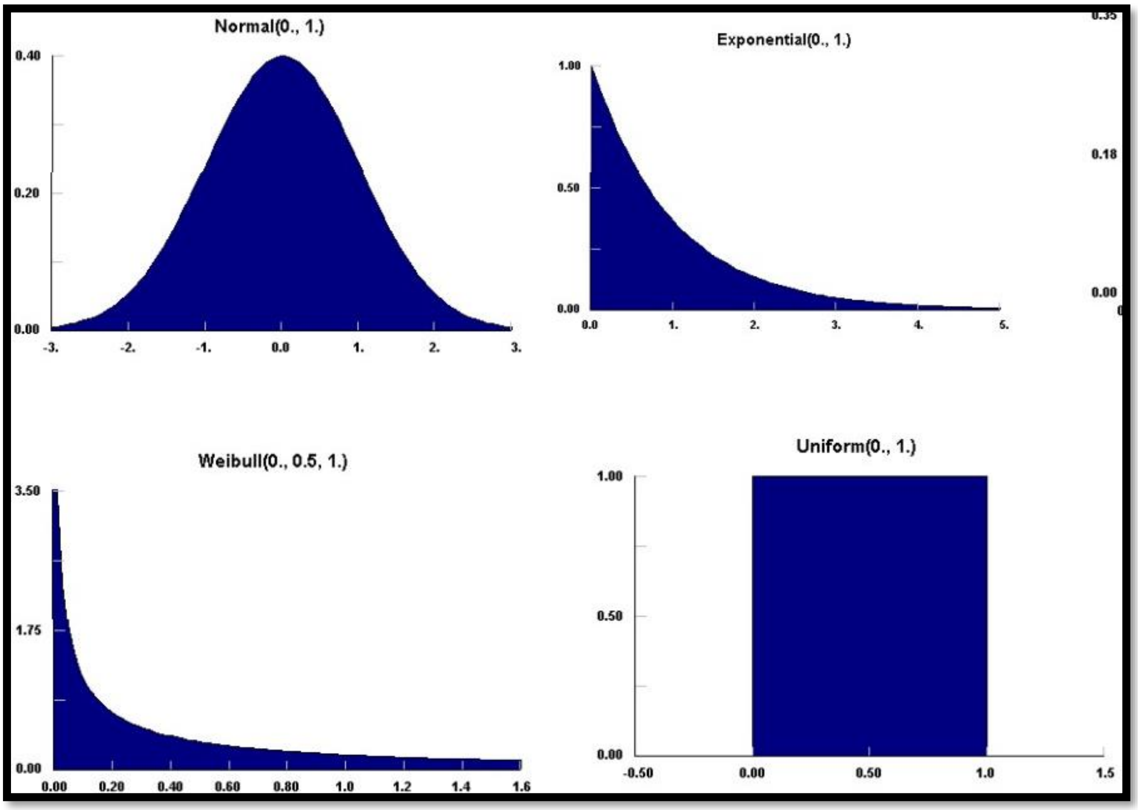
Olasılık yoğunluk fonksiyonunun özellikleri:

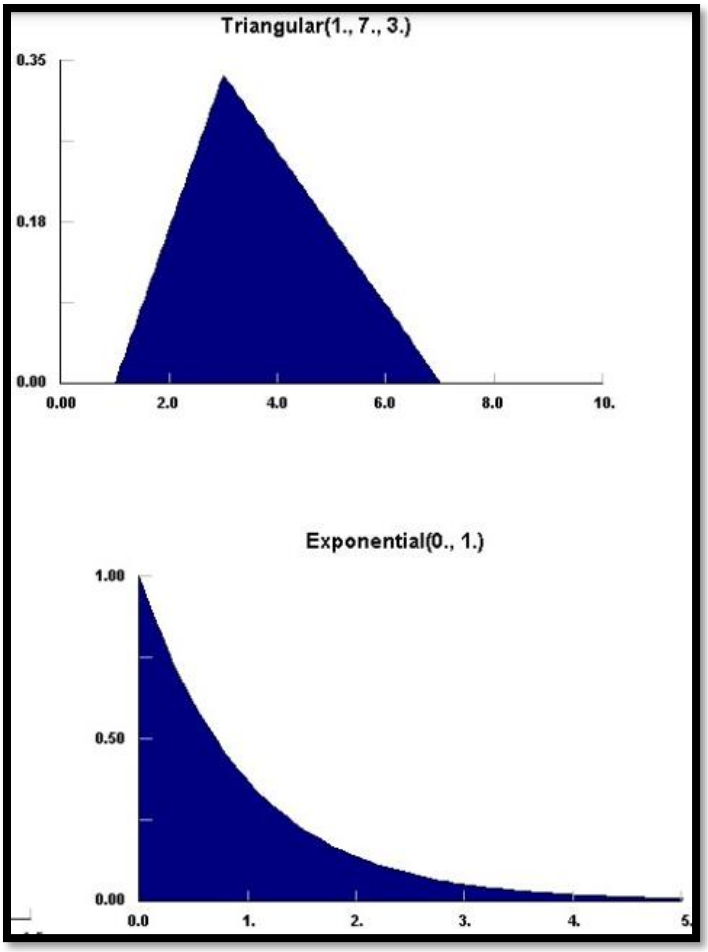
P(X = x) = 0 $-\infty < x < \infty$

P(X < x) = P(X ≤ x) = $\int_{-\infty}^{x} f(t)dt$

P(a ≤ X ≤ b) = P(a ≤ X < b) = P(a < X < b) = P(a < X ≤ b) = $\int_a^b f(t)dt$ a < b için

İlgilenilen rastgele olayın olasılıkları grafikten kolaylıkla gözlemlenebilir:





Örnek-1: X, aşağıda verilen olasılık fonksiyonuna sahip bir sürekli rastgele değişken ise

$f(x) = 2x \qquad 0 < x < 1$

$\qquad 0 \qquad$ diğer

a) P(X ≤ 0,4) olasılığını hesaplayınız.

Olasılık yoğunluk fonksiyonu kullanılarak bir sürekli rastgele değişkene ait olasılık hesaplaması aşağıdaki ifade ile gerçekleştirilir.

$P(X \le x) = \int_{-\infty}^{x} f(t)dt$

Dolayısıyla P(X ≤ 0,4) değerini hesaplamak için aşağıdaki işlemin hesaplanması gerekir.

$P(X \le 0,4) = \int_{-\infty}^{0,4} f(x)dx$ ancak olasılık yoğunluk fonksiyonunun tanımlı olduğu aralık dikkate alınarak

$P(X \le 0,4) = \int_{-\infty}^{0,4} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{0,4} f(x)dx$

$P(X \le 0,4) = \int_{-\infty}^{0} 0\, dx + \int_{0}^{0,4} 2x\, dx = 0,16$

b) P(X > 0,2) olasılığını hesaplayınız.

$P(X > 0,2) = P(0,2 < X < \infty) = \int_{0,2}^{\infty} f(x)dx$

$P(X > 0,2) = \int_{0,2}^{\infty} 2x\, dx = 0,96$

c) P(0,2 < X < 0,4) olasılığını hesaplayınız.

$P(0,2 < X < 0,4) = \int_{0,2}^{0,4} f(x)dx = \int_{0,2}^{0,4} 2x\, dx = 0,12$

d) P(X = 0,5) olasılığını hesaplayınız.

$P(X = 0,5) = 0$

X rastgele değişkenin verilen bir değere eşit veya küçük çıkma olasılığı ile ilgilenilen olaylar olasılık hesaplamalarında önemli bir yer tutar. Örneğin; makine elemanına gelen kuvvetlerin, emniyetli kuvvet değerinden küçük olma ihtimali vb. Bu tür hesaplamalar için dağılım fonksiyonu kullanılır.

$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt \qquad -\infty < x < \infty$

Bu fonksiyona eklenik dağılım fonksiyonu (kümülatif ya da birikimli dağılım fonksiyonu) denir.

Örnek-2: **Yüksek bir yapının rüzgâr hesabında kullanılacak olan maksimum rüzgâr hızı için aşağıda verilen olasılık yoğunluk fonksiyonu kabul edilmiştir.**

f(x) = 0,033.$e^{-0,033x}$ x ≥ 0

0   diğer

a) Dağılım fonksiyonunu elde ediniz.

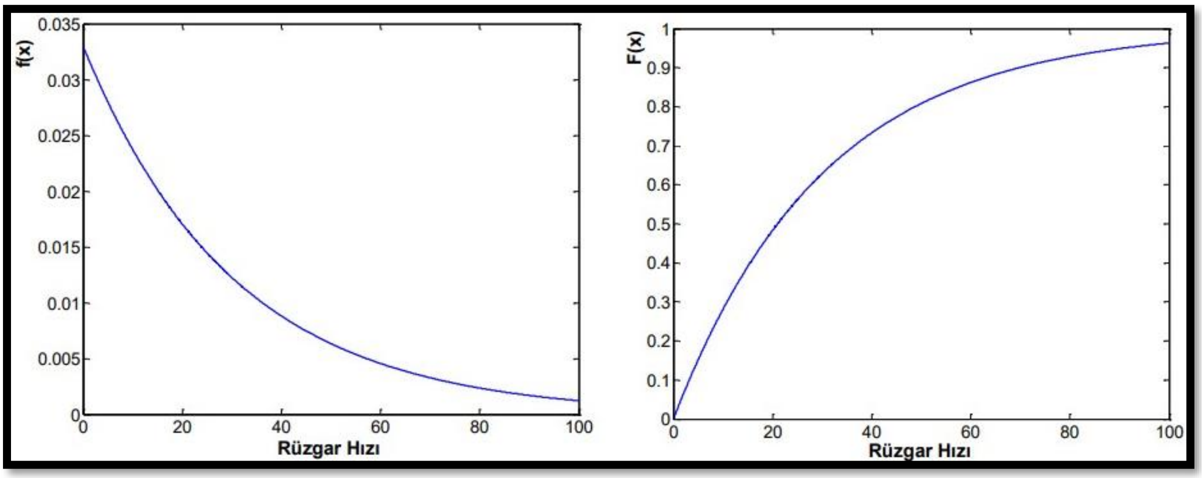
Dağılım fonksiyonu, olasılık yoğunluk fonksiyonun belirli bir aralıkta integre edilmesi ile elde edilebilir.

F(x) = P(X ≤ x) = $\int_{-\infty}^{x} f(t)dt$ $-\infty$ < x < ∞

F(x) = $\int_{0}^{x} 0,033.e^{-0,033x}\, dx$ = 1 - $e^{-0,033x}$

b) Olasılık yoğunluk fonksiyonu ve dağılım fonksiyonunu çiziniz.

Dağılım fonksiyonu, olasılık yoğunluk fonksiyonun belirli bir aralıkta integre edilmesiyle elde edilebilir.



c) Rüzgâr hızının 80 km/h aşma olasılığını hesaplayınız. "P(X > 80) = ?"

Rüzgâr hızının 80km/h aşma olasılığı dağılım fonksiyonu veya dağılım fonksiyonu grafiği yardımıyla elde edilebilir. Grafikten okunacak olasılık değeri

F(x) = P(X ≤ 80) = 1 - $e^{-0,033x}$ = 1 - $e^{-0,033.80}$ = 0,92

İstenilen olasılık ise, tüm olasılık 1'e eşit olduğundan:

P(X > 80) = 1 – 0,92 = 0,08

Sürekli Rastgele Değişkenler İçin Beklenen Değer, Varyans ve Standart Sapma

Beklenen Değer: **μ = E(X) = $\int_{-\infty}^{\infty} xf(x)dx$**

Varyans: **σ^2 = V(X) = $\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$ = $\int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$**

Standart Sapma: **σ = $\sqrt{\sigma^2}$**

SÜREKLİ OLASILIK DAĞILIMLARI

1- Düzgün (Üniform) Dağılım

X sürekli rastgele değişken belirli bir aralıktaki her değerinin meydana gelme olasılığı eşit ise bu rastgele değişkenin dağılımı düzgün (üniform) dağılımdır.

Üniform dağılıma ait olasılık fonksiyonu:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & \text{diğer} \end{cases} \qquad \mu = E(X) = \frac{a+b}{2} \quad \text{ve} \quad \sigma^2 = V(X) = \frac{(b-a)^2}{12}$$

Örnek-1: **Süper marketteki kasaya 30 dakikalık periyotta bir müşteri gelmiştir. Bu müşterinin son 5 dakikada gelmiş olma ihtimalini hesaplayınız.**

5/30 = 1/6 ya da

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & \text{diğer} \end{cases} \qquad f(x) = \begin{cases} \dfrac{1}{30-0} & a \leq x \leq b \\ 0 & \text{diğer} \end{cases}$$

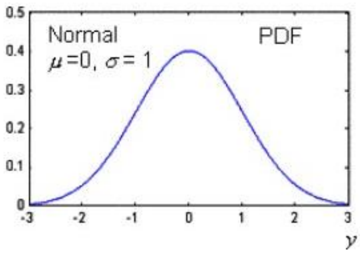
$$P(25 \leq X \leq 30) = \int_{25}^{30} f(x)\, dx = \frac{30-25}{30} = \frac{1}{6}$$

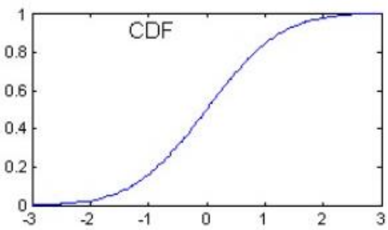
2- Normal Dağılım

Sürekli olasılık dağılımlarının en önemlisi ve en çok kullanılanı normal dağılımdır. Normal dağılıma, Gauss-Laplace dağılımı ya da çan eğrisi de denilmektedir.

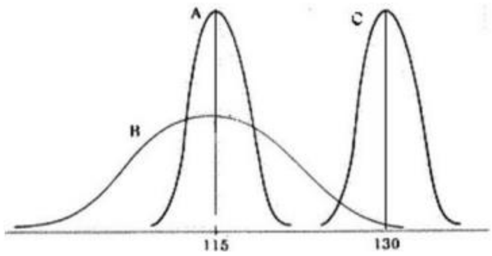
Normal dağılımın olasılık yoğunluk fonksiyonu $f(x) = \dfrac{1}{\sqrt{2\prod\sigma^2}} \,.\, e^{\frac{1(x-\mu)^2}{2\,.\,\sigma^2}}$

Normal yoğunluk fonksiyonu iki parametreye sahiptir; ortalama (μ) ve standart sapma (σ). Normal dağılım fonksiyonu ve kümülatif dağılım fonksiyonu grafiksel olarak aşağıda verilmiştir.





Ortalama ve standart sapma değerlerine bağlı olarak normal dağılımın yeri ve biçimi değişmektedir. Örneğin: Aşağıda şekilleri verilen A, B ve C normal dağılmış rastgele değişkenler arasında: $\mu_A = \mu_B < \mu_C$ $\qquad \sigma_A^2 = \sigma_C^2 < \sigma_B^2$



Normal dağılımın süreklilik özelliğinden dolayı X rastgele değişkeninin sadece belirli bir aralıkta değer alması söz konusudur.

İlgilenilen aralıkta değer alma olasılığı, olasılık yoğunluk fonksiyonunun entegrali ile elde edilir.

Örneğin; P(a ≤ X ≤ b) olasılığını hesaplamak için

P(a ≤ X ≤ b) = $\int_a^b f(x)\,dx = \int_a^b (\frac{1}{\sqrt{2\prod\sigma^2}} \cdot e^{\frac{1(x-\mu)^2}{2 \cdot \sigma^2}})$ dx işlemi yapılmalıdır.

Görüleceği üzere oldukça fazla işlem yükü gelmektedir. İşlem yükünü azaltmak için bu dağılım yerine geliştirilen standart normal dağılım kullanılmaktadır. X rastgele değişkeni normal dağılıyorsa aşağıdaki şekilde gösterilir: X ~ N(μ; σ^2)

Standart Normal Dağılım

X normal değişkeni sonsuz değer alabileceğinden nümerik olarak çözüm elde edilebilmesi için normal dağılmış rastgele fonksiyon standart normal dağılmış rastgele değişkene dönüştürülür: Z ~ N(0; 1)

Standart normal dağılım: Ortalaması 0 ve varyansı 1 olacak şekilde dönüşüm yapılır: Z = $\frac{X-\mu}{\sigma}$

Bu ifade normal rastgele değişkenin olasılık yoğunluk fonksiyonuna yazılırsa standart normal değişkene ait olasılık yoğunluk fonksiyonu elde edilir: f(x) = $\frac{1}{\sqrt{2\prod}} \cdot e^{\frac{-z^2}{2}}$ belirli entegraline eşit olur.

Dağılımın genel özellikleri dikkate alınarak standart normal değişken (Z) için

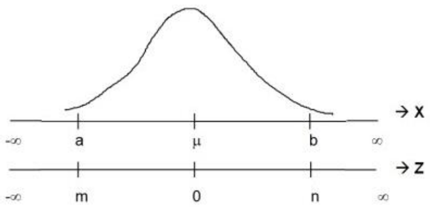
P(0 ≤ Z ≤ z_0) = $\int_a^{z_0} f(z)\,dz = \int_k^{z_0} (\frac{1}{\sqrt{2\prod}} \cdot e^{\frac{-z^2}{2}})$dz

P(Z ≤ z_0) = $\int_{-\infty}^{z_0} f(z)\,dz = \int_{-\infty}^{z_0} (\frac{1}{\sqrt{2\prod}} \cdot e^{\frac{-z^2}{2}})$ dz
integralleri hesaplanarak standart normal dağılımla ilgili tablolar hazırlanmıştır.

- **Z tablosu olarak adlandırılan bu tablolar farklı şekillerde düzenlenmektedir.**
- **Bu ders kapsamında kullanılacak olan tablo P(Z > z_0) olasılığını vermektedir.**
- **Verilen tablo yardımıyla normal dağılıma ait her türlü olasılık hesaplanabilmektedir.**
- **Ayrıca, dağılım simetrik olup dağılımın tepe noktasının yatay ekseni kestiği noktanın koordinatı sıfırdır (dağılımın ortalamasıdır) ve eğri altında kalan alanın değeri 1'e eşittir.**
- **Dağılım simetrik olduğu için P(Z > 0) = P(Z < 0)= 0.5 dir.**
- **Bu nedenle, ortalamanın sağında kalan kısmı tablolarda verilmekte, diğer yarısının aynı olduğu bilinmektedir.**

İstenen X rastgele değişkeninin belirli aralıkta değer alma olasılığını hesaplamak için izlenecek yaklaşımlar şöyle özetlenebilir:

1. Verilen $a < X < b$ aralığı $m < Z < n$ aralığına dönüştürülür. Yani,



Bu amaçla

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

dönüşümü kullanılır.

2. Karşı gelen P(m<Z<n) değeri tablo yardımıyla belirlenir. Öyle ise P(A<X>b):

$$P\left(\underbrace{\frac{a-\mu}{\sigma}}_{m} \leq \underbrace{\frac{x-\mu}{\sigma}}_{x} \leq \underbrace{\frac{b-\mu}{\sigma}}_{n}\right) \Rightarrow P(m \leq Z \leq n)$$

hesaplanır.

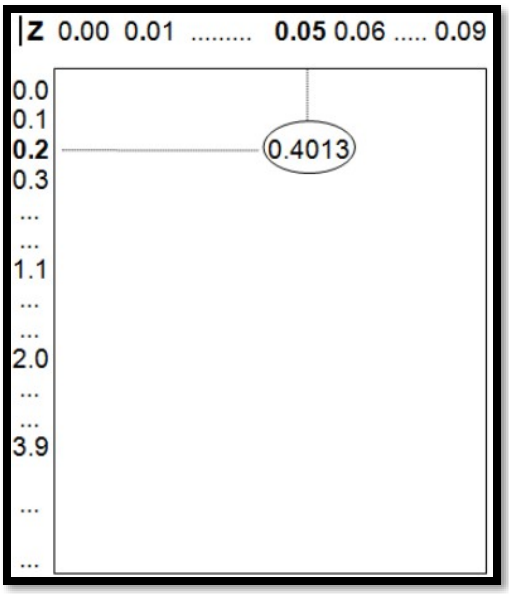
Z tablosundan istenilen olasılık değeri bulunulurken verilen değer; tamsayı kısmı ile birinci ondalık kısmı, ikinci ondalık kısmı olmak üzere iki parçaya ayrılır

Z tablosundan bir olasılık değeri okumak için aşağıdaki adımlar takip edilir.

1. Tamsayı kısmı ile birinci ondalık kısmı düşey eksende işaretlenir.

2. ikinci ondalık kısmı için yatay eksende eksende işaretlenir.

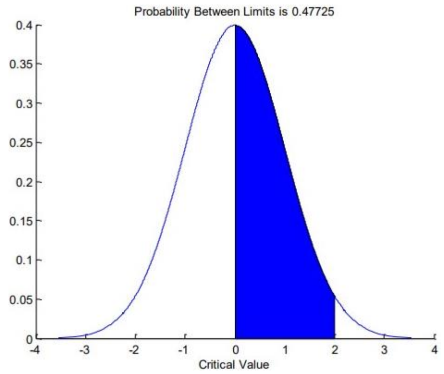
3. Bu değerlere yatay ve düşey eksende karşı gelen değerlerin kesiştiği hücredeki değer aranan olasılık değeridir.



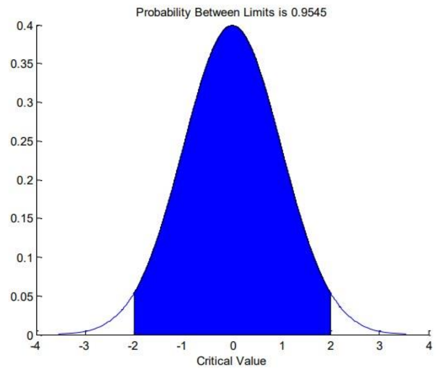
	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Örnek-1: Eğer Z standart normal dağılmış bir rastgele değişkense aşağıdaki olasılıkları hesaplayınız.

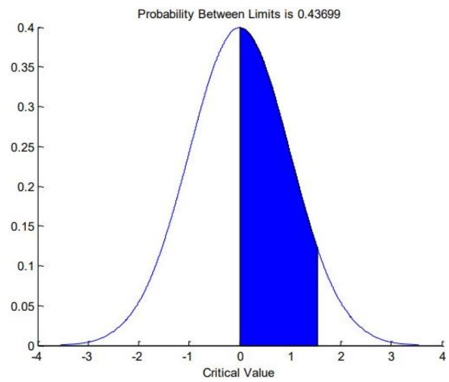
a) P(0 ≤ Z ≤ 2) = 0,4772 --> Satırdan 2,0'ı, sütundan 9'ı bulduğumuzda 0,4772'yi elde ederiz.



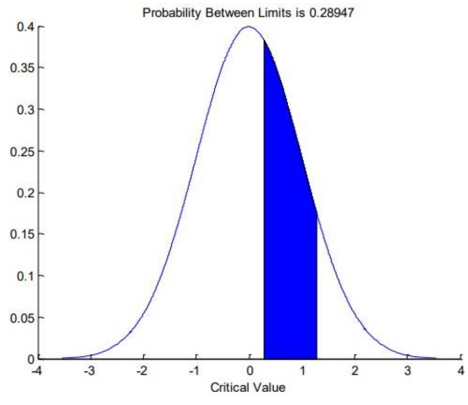
b) P(-2 ≤ Z ≤ 2) = 0,9544



c) P(0 ≤ Z ≤ 1,53) = 0,4370



d) P(0,28 < Z < 1,28) = 0,3997 – 0,1103 = 0,2894



Örnek-2: P(Z > z1) = 0,025 ise z1 kaçtır?

Önceki problemlerde eksen değerlerinden hareketle olasılık değeri bulunurken bu problemde olasılık değerinden hareketle eksen değerleri bulunmaktadır. Yani tabloya bakış yönteminde değişiklik var. Tablodan 0.025 olasılık değerine karşı gelen z değeri araştırılırsa bunun 1.96 (yani z1=1.96) olduğu görülür. *** HOCANIN ÇÖZÜMÜ ------> 0,5 – 0,025 = 0,475 ***

Örnek-3: P(-z1 < Z < z1) = 0,90 ise z1 kaçtır?

Değer çift taraflı olduğundan (her iki kuyruğu kapsadığından) her parçanın olasılığı $\frac{1-0,90}{2}$ = 0,05'tir. Tablodan 0,05 olasılık değerine karşı gelen z değeri araştırılırsa bunun 1.64 (yani z1=1.64) olduğu görülür.

Örnek-4: Bir imalathanede üretilen millerin çaplarının ortalaması 3.0005 inç ve standart sapmalarının ise 0.001 inç olan normal dağılıma uyduğu tespit edilmiştir. Üretilen miller eğer 3.000±0.002 inç aralığının dışında iseler bu miller hatalı üretim kabul edilmektedir. Buna göre toplam üretimdeki hatalı ürün miktarını bulunuz.

İstenilen olasılık ifadesi: Bu olasılık değerini hesaplamak için X sürekli normal değişkeni standart normal hale dönüştürülür.

Z = $\frac{X-\mu}{\sigma}$ -----> $z_1 = \frac{2,998-3,0005}{0,001}$ = -2,5 $z_2 = \frac{3,002-3,0005}{0,001}$ = 1,5

P(hatalı ürün) = 1 – P(-2,5 ≤ Z ≤ 1,5) = 1 – 0,927 = 0,073

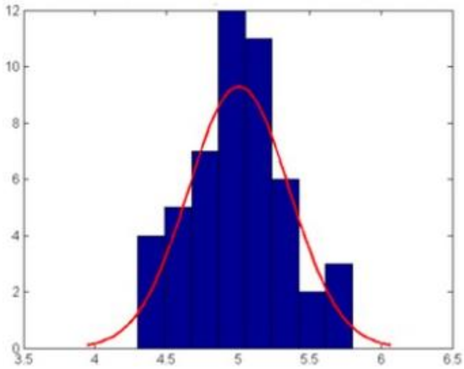
Dağılım Tipinin Belirlenmesi

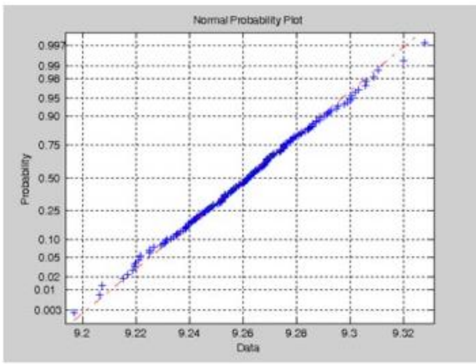
Ham olarak elde edilen rasgele değişkene ait verilerin dağılım tipini (Normal, Exponensiyal, Log-nomal vb.) belirlemek rasgele değişken kullanılarak yapılacak analizler için çok önemlidir. Bu işlemlerde rasgele değişkenin nasıl bir dağılım davranışı gösterdiği ve bu dağılımın parametreleri kullanılmaktadır. Ham olarak elde edilen bu verilere bir dağılım uydurmak (distirbution fitting) için aşağıda verilen adımlar takip edilir:

Dağılım tipini grafiksel olarak belirlemek.
Belirlenen bu dağılım tipine ait parametreleri tahmin etmek.
Belirlenen bu dağılım tipinin uygunluğunu test etmek.

Dağılım Tipini Grafiksel Olarak Belirleme

Ham olarak elde edilmiş rasgele değişkene ait verilerin hangi dağılım tipine uygun olduğunu belirlemede genellikle bu dataların grafiksel olarak gösterimi ile birlikte uygunluk testi (goodness-of-fit) uygulanarak elde edilir.





Dağılım Parametrelerinin Tahmini

Belirlenen dağılıma ait parametrelerin (ortalama, standart sapma, çarpıklık, basıklık gibi) için başlıca iki metot kullanılır:

- Momentler metodu (method of moments)
- Maksimum olabilirlik metodu (method of maximum likelihood)

Bu metotlar vasıtasıyla edilen parametreler daha sonra gerçekleştirilecek analizlerde rasgele değişkenlerin kullanılmasını sağlar.

Seçilen Dağılım Fonksiyonun Uygunluk Testi

Son adım olarak, rasgele değişkenlere ait belirlenmiş dağılım tipinin uygunluk testi yapılarak istatistiksel olarak ne kadar uygun olduğu tespit edilir. Bu adımda kullanılan belli başlı uygunluk testi yöntemleri:

- Ki-kare uygunluk testi (Chi Square test)
- Kolmogorov Smirnov test
- Anderson Darling test

ÖRNEKLEME TEORİSİ

İstatiksel metotlar, popülasyon hakkında sonuç çıkarmak için kullanılır. Sonuç çıkarmaya istatiksel çıkarsama da denir. Bu teknikler sonuç çıkarsamada, örneklemdeki bilgiyi kullanırlar.

Örneklem dağılımı, bir popülasyondan seçilen belirli boyuttaki bir örnek için örnek istatistiğinin tüm olası değerlerinin dağılımıdır.
Örneğin üniversitemizde, ortalamayla ilgili olarak 50 öğrenciyi örneklediğimizi varsayalım. Eğer farklı 50 kişilik numuneler aldıysanız, her numune için farklı bir ortalama hesaplayacaksınız. İşte biz de herhangi bir 50 öğrenci örneği için hesaplayabileceğimiz potansiyel ortalamaların dağılımı ile ilgileniyoruz.

Aynı popülasyondan aynı boyuttaki farklı numuneler, farklı örnek ortalamaları sağlayacaktır. Örnekten örneğe göre değişen ortalamadaki bu değişime, ortalamanın standart hatası denir.

X'in örneklem dağılımı için Z değeri:

- $$Z = \frac{(\overline{X} - \mu_{\overline{X}})}{\sigma_{\overline{X}}} = \frac{(\overline{X} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

$\overline{X}$ = örnek ortalaması
μ = popülasyon ortalaması
σ = popülasyon standart sapması
n = örnek büyüklüğü

Merkezi Limit Teorisi

Ortalaması μ ve varyansı σ^2 olan herhangi bir anakütleden rastgele çekilen n birimlik örneklerin ortalamalarının dağılımı normal, ortalaması μ ve varyansı $\frac{\sigma^2}{n}$'dir.

Bu gibi durumlarda kullanılacak Z eşitliği yandaki biçimde olacaktır.

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

Bu teoremin bir sonucu olarak; örnekteki birim sayısı yeterince büyük olduğunda X ~ N (μ ; $\frac{\sigma^2}{n}$) ilişkisi anakütlenin dağılımına bakılmaksızın yazılabilmektedir.

Popülasyon istatistikleri	Örneklem istatistikleri
μ = ortalama (aritmetik)	$\bar{X}$ = ortalama
σ^2 = varyans	s^2 = varyans
σ = standart sapma	s = standart sapma

Örnek-1: **Bir bölgedeki telefon görüşmeleri üzerine yapılan incelemede ortalama görüşme süresinin 8 dakika ve varyansının 4 olduğu belirlenmiştir. Rastgele seçilen 49 telefon görüşmesinde ortalama görüşme süresinin 7,8 ile 8,4 dakika arasında çıkma olasılığı nedir?**

Anakütlenin dağılımı bilinmese de örnek ortalamasının dağılımının merkezi limit teoremine göre normal dağılım olacağından normal dağılım yardımıyla istenen olasılık değeri hesaplanabilir.

ralık 2021 Salı 15:44

$n = 49 \quad \mu_{\bar{x}} = \mu \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$\mu = 8 \quad \sigma^2 = 4$

$$P(7.8 < \bar{x} < 8.4) = P\left(\frac{7.8-8}{\sqrt{\frac{4}{49}}} < Z < \frac{8.4-8}{\sqrt{\frac{4}{49}}}\right)$$

$$= P\left(\frac{-0.2}{2/7} < Z < \frac{0.4}{2/7}\right)$$

$$= P(-0.69 < Z < 1.38)$$

$$= 0.2549 + 0.4162 = \boxed{0.6711}$$

Örnek-2: **Kablo üreticisi iki firmanın ürettikleri kabloların kopma ortalamasının, sırasıyla 200 kg/cm2 ve 180 kg/cm2 , standart sapmalarının 13,5 kg/cm2 ve 9 kg/cm2 olduğu belirtilmiştir. Bu iddianın doğru olup olmadığını test etmek isteyen tüketici bir firma ilk firmanın üretiminden rastgele 100 parça kablo, ikinci firmanın üretiminden rastgele 50 parça kablo almıştır. Üretici firmaların beyanatlarının doğru olduğu kabul edilirse; birinci ve ikinci firmanın kablolarının kopma mukavementleri ortalamaları arasındaki farkın;**

a) En fazla 17 kg/cm2 çıkma olasılığı nedir?

b) En az 15 kg/cm2 çıkma olasılığı nedir?

7 Aralık 2021 Salı 15:52

$\mu_1 = 200\ kg/cm^2 \quad \mu_2 = 180\ kg/cm^2$

$\sigma_1 = 13.5\ kg/cm^2 \quad \sigma_2 = 9\ kg/cm^2$

$n_1 = 100 \quad n_2 = 50$

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$a)\ P(\bar{X}_1 - \bar{X}_2 < 17) = P\left(Z < \frac{17-(200-180)}{\underbrace{\sqrt{\frac{13.5^2}{100} + \frac{9^2}{50}}}_{1.86}}\right) = P(Z < -1.61)$$

$= 0.5 - 0.4463 = 0.0537$

-1.61

$$b)\ P(\bar{X}_1 - \bar{X}_2 \geq 15) = P\left(Z > \frac{15-20}{1.86}\right) = P(Z > -2.69)$$

$= 0.5 + 0.4964 = 0.9964$

0.5

-2.69

Örnek-3: **Bir marketten 50 TL veya daha fazla bedelli ürün alan müşterilerin ortalama olarak %30'unun kredi kartı kullandığı belirlenmiştir. 50 TL veya daha fazla bedelli ürün alan müşteriler arasından rastgele seçilen 100 müşteriden ödemesini kredi kartı ile yapanların oranının %20 ile %25 arasında çıkması olasılığı nedir?**

Aralık 2021 Salı 16:46

$\mu = \frac{\cancel{n}.p}{\cancel{n}} = p \qquad \sigma = \frac{1}{n^{\cancel{2}}}\ \cancel{n}.p.q = \frac{p.q}{n}$

$\mu = 0.3 \qquad \sigma = \frac{0.3 \times 0.7}{100}$

$$P(0.2 < \hat{p} < 0.25) = P\left(\frac{0.2-0.3}{\sqrt{\frac{0.3\ 0.7}{100}}} < Z < \frac{0.25-0.3}{\sqrt{\frac{0.3 \cdot 0.7}{100}}}\right)$$

$= P(-2.18 < Z < -1.09)$

$= 0.4854 - 0.3621$

$= 0.1233$

$-2.19 \quad -1.09$

Örnek-4: Pil üreten iki fabrikanın ürettiği pillerin dayanma süreleri aşağıda açıklanmıştır.

Birinci Fabrika: Pillerimizin %80'i 200 saatin üzerinde dayanır.
İkinci Fabrika: Pillerimizin %73'ü 200 saatin üzerinde dayanır.

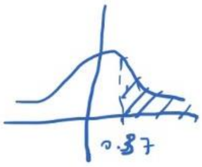
Bunu test etmek isteyen bir tüketici örgütü birinci fabrikanın üretiminden rastgele 50 adet pil, ikinci fabrikanın üretiminden rastgele 60 pil almıştır.
Birinci ve ikinci fabrikada üretilen pillerin dayanma oranları arasındaki farkın en az %10 olma olasılığı nedir?

k 2021 Salı 16:55

$$P_1 = 0.8 \quad P_2 = 0.73 \qquad P(\hat{P}_1 - \hat{P}_2 \geq 0.10) = ?$$
$$n_1 = 50 \quad n_2 = 60$$

$$P\left(Z \geq \frac{0.1 - (0.8 - 0.73)}{\sqrt{\frac{0.8 \times 0.2}{50} + \frac{0.73 \times 0.27}{60}}}\right)$$

$$= P(Z \geq 0.37) = 0.5 - 0.1443$$
$$= 0.3557$$



Güven Aralığı

- Anakütle varyansı (σ^2) biliniyorsa **Z** dağılımı
- Anakütle varyansı (σ^2) bilinmiyorsa
  - **n≥30** ise **Z** dağılımı
  - **n<30** ise **t** dağılımı

Örneklem = n ile ifade edilmektedir.

Örnek-1: Bir tezgahta üretilen parçaların dış çaplarının standart sapması σ=2,4 cm'dir. Tezgahın üretiminden rastgele seçilen 16 parçanın dış çap ortalaması 3,2 cm olarak bulunmuştur. %5 hata (%95 güven) seviyesinde anakütle ortalamasının güven aralığını tahmin ediniz.

Salı 17:05

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} \qquad = 0.475$$

$$\sigma = 2.4\,cm \quad n = 16 \quad \bar{X} = 3.2\,cm \qquad 0.5 - 0.025$$

$$1 - \alpha = 0.95 \quad \alpha = 0.05 \quad \alpha/2 = 0.025 \quad Z_{0.025} = 1.96$$

$$\bar{X} - Z_{\alpha/2}\,\sigma_{\bar{x}} \leq \mu \leq \bar{X} + Z_{\alpha/2}\,\sigma_{\bar{x}}$$

$$3.2 - 1.96\sqrt{\frac{2.4^2}{n}} \leq \mu \leq 3.2 + 1.96\sqrt{\frac{2.4^2}{16}}$$

$$2.024 \leq \mu \leq 4.376$$

Örnek-2: **Bir tezgahta üretilen parçaların dış çaplarının standart sapması σ=2,4 cm'dir. Tezgahın üretiminden rastgele seçilen 16 parçanın dış çap ortalaması 3,2 cm olarak bulunmuştur. %5 hata (%95 güven) seviyesinde örnek ortalaması (tahmin edilen değer) ile anakütle ortalaması (gerçek değer) arasındaki farkın (yani hatanın) 1 cm veya daha az olması için alınması gereken örnek hacmi ne olmalıdır?**

14 Aralık 2021 Salı 16:07

$\alpha = 0,05$

$d = \bar{X} - \mu = 1\,cm \quad \sigma = 2.4\,cm \quad z_{0.025} = 1.96$

$$n = \left(\frac{z_{\alpha/2}\,\sigma}{d}\right)^2 = \left(\frac{1.96 \times 2.4}{1}\right)^2 \cong 22.13 \cong \underline{\underline{23}}$$

Ortalamanın Güven Aralığı

Örnek-3: **Bir işyerinde çalışan işçilerin boylarına göre tezgah yüksekliklerinin ayarlanması amacıyla bir araştırma yürütülmüştür. Farklı bölümlerden rasgele 25 işçi seçilmiş ve boyları ölçülmüştür. İşçilerin boyları ortalaması 1,72 m ve varyansı 0,18 olarak belirlendiğine göre %99 güven (%1 hata) seviyesinde anakütle ortalamasının güven sınırlarını tahmin ediniz.**

Aralık 2021 Salı 16:15

$t_{\alpha/2,n-1} = t_{0.005,24} = 2.797$

$SD = 25-1 = 24 \quad \bar{X} = 1.72 \quad S^2 = 0.18 \quad 1-\alpha = 0.99 \quad \alpha = 0.01$

$$\bar{X} - t_{\alpha/2,n-1} S_{\bar{X}} \le \mu \le \bar{X} + t_{\alpha/2,n-1} S_{\bar{X}} \qquad S_{\bar{X}} = \sqrt{\frac{S^2}{n}}$$

$$1.72 - 2.797\sqrt{\frac{0.18}{25}} < \mu < 1.72 + 2.797\sqrt{\frac{0.18}{25}}$$

$$1.48 < \mu < 1.96$$

İki Ortalama Farkının Güven Aralığı

Örnek-4:

İçinde kusurlu ürün bulunduğu bilinen 8 koli ile kusurlu ürün bulunmadığı bilinen 9 kolinin ortalama ağırlıkları kg olarak aşağıda verilmiştir:

Kusurlu koli	125	120	119	123	126	116	118	119	
Kusursuz koli	130	130	128	126	125	120	132	127	128

%95 güven düzeyinde ortalamalar arasındaki farkın güven aralığını oluşturunuz.

1 Aralık 2021 Salı 16:31

$\bar{X}_1 \to$ kusursuz $\quad n_1 = 9 \qquad \bar{X}_2 \to$ kusurlu $\quad n_2 = 8$

$\bar{X}_1 = 127.33 \quad S_1^2 = 12.25 \qquad \bar{X}_2 = 120.75 \quad S_2^2 = 12.5$

$$s^2 = \frac{(9-1)\,12.25 + (8-1)\,12.5}{9+8-2} = 12.4 \qquad S_{\bar{x}_1-\bar{x}_2} = \sqrt{12.4\left(\frac{1}{9}+\frac{1}{8}\right)} = 1.71$$

$$s^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \qquad S_{\bar{x}_1-\bar{x}_2} = \sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$SD = n_1+n_2-2 = 15 \quad 1-\alpha = 0.95 \quad \alpha = 0.05 \quad \alpha/2 = 0.025 \quad t_{0.025,15} = 2.131$

Aralık 2021 Salı 16:37

$$(\bar{X}_1-\bar{X}_2) - t_{\alpha/2, n_1+n_2-2}\, S_{\bar{x}_1-\bar{x}_2} < \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2) + t_{\alpha/2, n_1+n_2-2}\, S_{\bar{x}_1-\bar{x}_2}$$

$$(127.33-120.75) - 2.131 \cdot 1.71 < \mu_1-\mu_2 < (127.33-120.75) + 2.131 \cdot 1.71$$

$$6.58 - 3.64401 < \mu_1-\mu_2 < 6.58 + 3.64401$$

$$2.94 < \mu_1-\mu_2 < 10.22$$

Bir Oranın Güven Aralığı

Örnek-5: Rassal olarak 75 adet otomobil cıvatası incelenmiş ve bunlardan 12 tanesinin yüzey düzgünlüğünün verilen spesifikasyonlar dışında olduğu belirlenmiştir. Buna göre yüzey düzgünlüğü belirlenen spesifikasyonlar dışında olan cıtaların anakütle oranının %99 güven seviyesinde güven aralığını bulunuz.

Aralık 2021 Salı 15:11

$n = 75 \quad 1-\alpha = 0.99 \quad \sigma = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{0.16 \cdot 0.84}{75}}$

$\hat{p} = \frac{12}{75} = 0.16 \quad \hat{q} = 0.84 \quad \alpha = 0.01 \quad \alpha/2 = 0.005 \quad z_{0.005} = 2.58$

$$\hat{p} - z_{0.005}\,\sigma < p < \hat{p} + z_{0.005}\,\sigma$$

$0.5 - 0.005 = 0.495$

$$0.16 - 2.58 \times 0.042 < p < 0.16 + 2.58 \times 0.042$$

$$0.05 < p < 0.27$$

İki Oran Farkının Güven Aralığı

Örnek-6: 75 adet otomobil mili incelenmiş ve bunlardan 12 tanesinin yüzey düzgünlüğünün verilen spesifikasyonlar dışında olduğu belirlenmiştir. Mühendisler yüzey düzgünlüğünü sağlamak için iyileştirme çalışmaları yapmış ve tekrar süreçten 85 adet örnek almıştır. Alınan örneklerden 10 tanesinin yüzey düzgünlüğü hatalı çıkmıştır. Buna göre yapılan iyileştirmenin yüzey pürüzlülüğünü gidermek için faydalı olup olmadığını %95 güven seviyesinde yorumlayınız.

21 Aralık 2021 Salı 15:28

$\hat{p}_1 = 0.16 \quad \hat{q}_1 = 0.84 \qquad 10/85 \quad \hat{p}_2 = 0.12 \quad \hat{q}_2 = 0.88 \qquad \alpha = 0.05 \quad z_{0.025} = 1.96$

$n_1 = 75 \qquad n_2 = 85 \qquad \alpha/2 = 0.025$

$0.5 - 0.025 = 0.475$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} = \sqrt{\frac{0.16 \cdot 0.84}{75} + \frac{0.12 \times 0.88}{85}} =$$

$$= \sqrt{0.0018 + 0.0012} \approx 0.055$$

$$(\hat{p}_1 - \hat{p}_2) - z_{0.025}\, 0.055 < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{0.025}\, 0.55$$

$$+0.04 - 1.96 \cdot 0.056 < p_1 - p_2 < 0.04 + 1.96 \times 0.55$$

$$-0.07 < \hat{p}_1 - p_2 < 0.15$$

Varyansın Güven Aralığı – Ki Kare Dağılımı

Örnek-7: Bir fabrikanın üretiminden rasgele alınan 20 birimlik örneğin varyansı 35 olarak belirlendiğine göre %99 güven düzeyinde fabrikanın üretimine (yani anakütleye) ait varyansın güven aralığını oluşturunuz.

lık 2021 Salı 15:46

$s^2 = 35 \qquad SD = n-1 = 20-1 = 19 \qquad 1-\alpha = 0.99 \qquad \alpha = 0.01 \qquad \alpha/2 = 0.005$

$1 - \alpha/2 = 0.995$

$$\chi^2_A = \chi_{0.995,19} = 6.844 \qquad \chi^2_B = \chi^2_{0.005,19} = 38.582$$

$$\frac{(n-1)s^2}{\chi^2_A} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_B}$$

$$\frac{19 \cdot 35}{6.844} < \sigma^2 < \frac{19 \cdot 35}{38.582}$$

$$17.237 < \sigma^2 < 97.222$$

Bir spor malzemesi üreticisi, ürettiği pilates bantlarının dayanma mukavemeti ortalamasının **15 kg/cm²**, standart sapmasının **0.5 kg/cm²** olduğunu açıklamıştır.

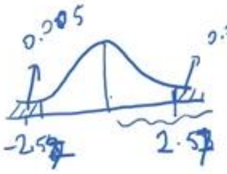
Firmanın bu iddiasını test etmek isteyen bir tüketici örgütü firmanın üretiminden rastgele **50** pilates bantı almış ve ortalamasını **14.8 kg/cm²** olarak belirlemiştir.

%1 hata seviyesinde pilates bantları mukavemetinin **15 kg/cm²** ye eşit olup olmadığını test ediniz.

21 Aralık 2021 Salı 16:41

$0.5 - 0.005 = 0.495$

1. $H_0: \mu = 15 \text{ kg/cm}^2$
   $H_1: \mu \neq 15 \text{ kg/cm}^2$

2. 

$\alpha = 0.01$

$\alpha/2 = 0.005$

$z_{0.005} = 2.57$

3. $z_h = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \dfrac{14.8 - 15}{\sqrt{0.5^2/50}} = \dfrac{-0.2}{0.0707} = -2.829$

4. $|z_h| = |-2.829| > |z_{0.005}| = |\mp 2.57| \rightarrow H_0$ reddedilir

5. %1 hata seviyesinde pilates bantları mukavemeti ortalaması 15 kg/cm²'den farklıdır diyebiliriz.

Varyansları 56 ve 65 olan iki anakütleden sırasıyla 25 ve 30 birimlik örnekler alınmış ve birinci örneğin ortalaması 92, ikinci örneğin ortalaması da 88 olarak hesaplanmıştır.

Anakütle ortalamalarının farklı olup olmadığını %5 hata seviyesinde test ediniz.

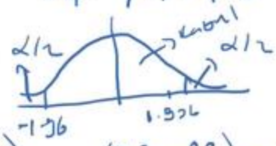
21 Aralık 2021 Salı 17:11

$\sigma_1^2 = 56$ $\quad \sigma_2^2 = 65$

$n_1 = 25$ $\quad n_2 = 30$

$\bar{x}_1 = 92$ $\quad \bar{x}_2 = 88$

1) $H_0: \mu_1 = \mu_2$
   $H_1: \mu_1 \neq \mu_2$



2) $\alpha = 0.05$
   $\alpha/2 = 0.025$ $\quad 0.5 - 0.025 = 0.475$
   $z_{0.025} = \mp 1.96$

3) $z_h = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{(92-88) - 0}{\sqrt{\dfrac{56}{25} + \dfrac{65}{30}}} = \dfrac{4}{\sqrt{2.24 + 2.17}} = \dfrac{4}{2.099} = 1.906$

4) $|z_h| = 1.906 < |z_{0.025}| = 1.96 \rightarrow H_0$ kabul

5) %5 hata ile iki ana kütle ortalamasının aynı olduğu söylenebilir.

Bir bölümde okuyan öğrencilerin not ortalamasının 70 olduğu bölüm yöneticileri tarafından iddia edilmektedir. Bu iddianın doğru olup olmadığını test etmek üzere Bölüm öğrencileri arasından rasgele seçilen 26 öğrencinin notları tespit edilmiş ve ortalaması 67, varyansı 9 olarak hesaplanmıştır.

%99 güven (%1 hata) seviyesinde Bölüm yöneticilerinin iddiasının doğru olduğu söylenebilir mi?

(Başka bir ifadeyle, sözü edilen bölümdeki öğrencilerin not ortalamalarının 70 den az olduğu söylenebilir mi?)

21 Aralık 2021 Salı 17:28

$\bar{x} = 67$ $n = 26$

$S^2_{\bar{x}} = 9$ $\mu = 70$

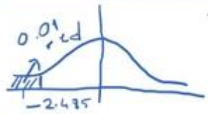
1) $H_0: \mu = 70$

$H_1: \mu < 70$

2) $1-\alpha = 0.99$ SD = n-1 = 25

$\alpha = 0.01$

$t_{\alpha,n-1} = t_{0.01,25} = 2.485$



3) $t_h = \frac{\bar{x}-\mu}{S_{\bar{x}}} = \frac{67-70}{\sqrt{9/26}} = \frac{-3}{0.588} = -5.099$

4) $|t_h| = 5.099 > |t_{0.01,25}| = 2.485 \rightarrow H_0$ ret

5) %1 hata seviyesinde öğrenci ortalamalarının 70'ten küçük olduğu söylenebilir.

Ki-kare (χ^2) Dağılımı ile Yapılan Testler

UYGUNLUK (UYUM) TESTİ

Örnek:

Bir zar 120 defa atılmış ve gelen değerler aşağıda verilmiştir.

X_i	1	2	3	4	5	6	Toplam	
m_i	17	18	24	26	21	14	120	→ *Verilen*
b_i	20	20	20	20	20	20	120	→ *Hesaplanan*

Bu veriler esas alınarak sözü edilen zarın hilesiz olduğu söylenebilir mi? (α=0.05)

28 Aralık 2021 Salı 15:41

1. $H_0: p = 1/6$ veya $b_i = 20$

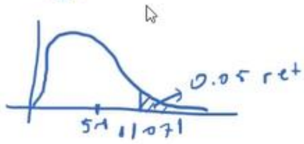
$H_1: p \neq 1/6$ veya $b_i \neq 20$

2. $\alpha = 0.05$ SD = k-1 = 6-1 = 5

$\chi^2_{\alpha,k-1} = \chi^2_{0.05,5} = 11.071$

3. $\chi_h^2 = \sum \frac{(m_i-b_i)^2}{b_i} = \frac{(17-20)^2}{20} + \frac{(18-20)^2}{20} + \dots + \frac{(14-20)^2}{20} = 5.1$

4. $\chi_h^2 = 5.1 < \chi^2_{0.05,5} = 11.071 \rightarrow H_0$ kabul



5. 0.05 hata ile zarın hilesiz olduğu söylenebilir.

Ki-kare (χ^2) Dağılımı ile Yapılan Testler

VARYANSLA İLGİLİ TESTLER

Örnek:

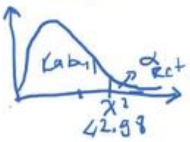
Bir çimento fabrikasında üretilen çimentodan yapılan betonların dayanımına ilişkin standart sapmanın 1.24 kg/mm^2 olduğu iddia edilmektedir. Bu iddiayı test etmek üzere fabrikanın üretiminden rasgele alınan 25 birimlik örneğin ortalama dayanma miktarı ölçülmüştür. Daha sonra ölçülen verilerin ortalaması 25 kg/mm^2 ve varyansı da 2.4 olarak hesaplanmıştır. Bu verilere göre fabrika yöneticilerinin iddiasının doğru olduğu söylenebilir mi? (α=0.01)

28 Aralık 2021 Salı 14:54

$\sigma^2 = (1.24)^2$ $\quad n = 25$ $\quad \bar{X} = 25$ kg/mm^2 $\quad S_x^2 = 2.4$ $\quad \alpha = 0.01$

1. $H_0: \sigma^2 = (1.24)^2$

   $H_1: \sigma^2 > (1.24)^2$

2. 

   $SD = n-1 = 25-1 = 24$

   $\chi^2_{0.01,24} = 42.98$

3. $\chi_h^2 = \dfrac{(n-1)S_x^2}{\sigma^2} = \dfrac{24 \cdot 2.4}{(1.24)^2} = \dfrac{24 \cdot 2.4}{1.5376} = 37.46$

4. $\chi_h^2 = 37.46 < \chi_{0.01,24} = 42.98 \rightarrow H_0$ reddedilemez

5. 0.01 hata ile yöneticilerin iddiası doğrudur yani $\sigma^2 = (1.24)^2$ olduğu söylenebilir.

Bir atölyede aynı parçayı üreten 3 tezgah vardır. Her tezgah farklı bir operatör tarafından çalıştırılmaktadır. Altı saatlik bir üretim devresinde her bir saatte üretilen kusurlu birim sayıları belirlenmiş ve aşağıda verilmiştir. Bir saat içinde üretilen kusurlu parçaların normal dağıldığı varsayılmaktadır. %95 güven düzeyinde (α=%5 hata seviyesinde) tezgahların kusurlu üretim ortalamalarının aynı olup olmadığına karar veriniz.

	Tezgah Çeşidi			
	1.Tezgah	2.Tezgah	3.Tezgah	
Gözlemler (Çıktılar)	3	2	4	
	4	3	5	
	4	4	6	
	5	5	6	
	7	--	7	Genel toplam
	--	--	8	
Örnek Hacmi	5	4	6	15
Örnek Toplamı	23	14	36	73

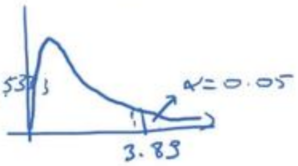
28 Aralık 2021 Salı 16:58

1. $H_0: \mu_1 = \mu_2 = \mu_2$

H_1: En az biri farklı

2. $\alpha = 0.05$ $\quad v_1 = k-1 = 3-1 = 2$ $\quad v_2 = N-k = 15-3 = 12$

$F_{0.05,(2;12)} = 3.89$



3. $SS_B = \sum_{i=1}^{k} \frac{T_i^2}{n_i} - \frac{T^2}{n} = \left(\frac{23^2}{5} + \frac{14^2}{4} + \frac{36^2}{6}\right) - \frac{73^2}{15} = \frac{932}{60} = 15.5333$

$SS_W = \sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij}^2 - \sum_{i=1}^{k} \frac{T_i^2}{n_i} = 3^2 + 4^2 + \ldots + 7^2 + 8^2 - (\ ) = \frac{1452}{60} = 24.2$

$SS_T = SS_B + SS_W = 15.5333 + 24.2 = 39.7333$

Korelasyon Analizi

Örnek:

Yeni doğan bebeklerin ağırlık (*kg*) ve beden (göğüs) genişliği (*cm*) arasındaki ilişkiyi belirlemek için bir araştırma yapılmış ve aşağıdaki sonuçlar elde edilmiştir.

Beden Genişliği (X)	29.5	26.3	32.2	36.5	27.2	27.7	28.3	30.3	28.7
Ağırlık (Y)	2.75	2.15	4.41	5.42	3.21	4.32	2.31	4.30	3.71

Bu sonuçlara göre;

a) Korelasyon katsayısını hesaplayarak yorumlayınız.

β)α=0.01 hata seviyesinde hesaplanan korelasyon katsayısının anlamlı olup olmadığını test ediniz.

χ)α=0.01 hata seviyesinde hesaplanan korelasyon katsayısının 0.80'e eşit olup olmadığını test ediniz

Korelasyon Analizi

$$r = \frac{\hat{\Sigma} xy - n\bar{x}\bar{y}}{\sqrt{(\Sigma x^2 - n\bar{x}^2)(\Sigma y^2 - n\bar{y}^2)}}$$

Örnek Çözüm:

a) $\overline{X} = \frac{266.7}{9} = 29.6333$ $\qquad \overline{Y} = \frac{32.58}{9} = 3.62$

$$r = \frac{986.62 - 9 \cdot 29.6333 \cdot 3.62}{\sqrt{\left(7980.83 - 9 \cdot (29.6333)^2\right) \cdot \left(127.57 - 9 \cdot (3.62)^2\right)}} = 0.774$$

	Beden Genişliği (X)	Ağırlık (Y)	X^2	Y^2	XY
	29.5	2.75	870.25	7.5625	81.125
	26.3	2.15	691.69	4.6225	56.545
	32.2	4.41	1036.84	19.4481	142.002
	36.5	5.42	1332.25	29.3764	197.830
	27.2	3.21	739.84	10.3041	87.312
	27.7	4.32	767.29	18.6624	119.664
	28.3	2.31	800.89	5.3361	65.373
	30.3	4.30	918.09	18.49	130.29
	28.7	3.71	823.69	13.7641	106.477
Toplam	**266.7**	**32.58**	**7980.83**	**127.57**	**986.62**

Bebeklerin ağırlıkları ile beden genişlikleri arasında aynı yönlü ve iyi bir ilişki vardır.

HOCA, SADECE İKİ DEĞİŞKENLİ KORELASYON ANALİZİNİ İŞLEDİ.

4 Ocak 2022 Salı 15:42

$r=0.774 \quad \alpha=0.01$

i) $H_0: \rho=0$

$H_1: \rho>0$

ii) $SD=n-2=9-2=7$

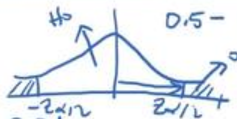
$t_{0.01,7}=2.998$

iii) $t_h = r\sqrt{\dfrac{n-2}{1-r^2}} = 0.774\sqrt{\dfrac{7}{1-0.3589}} = 3.234$

iv) $|t_h| = 3.234 > t_{0.01,7}=2.998 \rightarrow H_0$ ret

v) r anlamlıdır.

4 Ocak 2022 Salı 16:09



$0.5-0.005=0.495$

$r=0.774$

i) $H_0: \rho=0.8$

$H_1: \rho\neq 0.8$

ii) $\alpha=0.01$

$\alpha/2=0.005 \quad z_{0.005}=\mp 2.57$

iii) $Z_h = \sqrt{\dfrac{n-3}{2}}\ \ln\left[\dfrac{(1+r)(1-\rho_0)}{(1-r)(1+\rho_0)}\right] = \sqrt{\dfrac{9-3}{2}}\ \ln\left[\dfrac{(1+0.774)(1-0.8)}{(1-0.774)(1+0.8)}\right]$

$= -0.168$

iv) $|Z_h| = 0.168 < |z_{0.005}| = 2.57 \rightarrow H_0$ kabul

v) 0.01 hata seviyesinde $\rho=0.8$ e eşit olduğu söylenebilir.

Regresyon Analizi

Örnek:

Belirli bir ağırlık taşıyan bir plastik malzemenin verilen sıcaklıklar altında uğradığı şekil değişimleri Tabloda verilmiştir.

a) En küçük kareler yöntemini kullanarak regresyon denklemini oluşturunuz.
b) Regresyon katsayısının anlamlı olup olmadığını test ediniz (α=0.01).
c) 80 ^{0}C sıcaklıkta şekil değişimini belirleyiniz
d) Korelasyon katsayısını hesaplayarak yorumlayınız.

Tek değişkenli doğrusal regresyon denkleminin genel yazılışı:

Anakütle için	Örnek için
$\hat{Y}_i = \alpha + \beta * X_i + \cancel{\varepsilon_i}$	$\hat{y}_i = a + b * X_i + \cancel{e_i}$
α: *Doğrunun Y eksenini kestiği noktanın koordinatı (sabit değer)* **β**: *Doğrunun eğimi* ~~ε_i: *i. tahminin rassal hatası*~~	**a**: *Doğrunun Y eksenini kestiği noktanın koordinatı (sabit değer)* **b**: *Doğrunun eğimi* ~~e_i: *i. tahminin rassal hatası*~~