İstatistiksel Kavramların Gözden Geçirilmesi

Anlamlı Basamaklar Konusu ve Olasılık

Ekonometri 1 – Konu 1

Sürüm 2,0 (Ekim 2011)

UADMK Açık Lisans Bilgisi

A. Talha Yalta
TOBB Ekonomi ve Teknoloji Üniversitesi
Ekim 2011 BY-NC-SA

ULUSAL AÇIK DERS MALZEMELERİ KONSORSİYUMU

Ders Planı

ULUSAL AÇIK DERS MALZEMELERİ KONSORSİYUMU

Anlamlı Basamaklar

Anlamlı Basamaklar

Ondalık bir sayının "anlamlı basamakları" (significant digits), o sayının kesinlik ve doğruluğuna katkıda bulunan tüm basamaklarını gösterir.

- Veri ve ölçümleri elde etmek için çeşitli süreç ve işlemler kullanılabilmektedir.
- Eğer eldeki ölçüme ait bazı rakamlar, o ölçümü elde etmek için kullanılan sürecin doğruluk sınırı dışındaysa, bunları kullanmanın anlamı yoktur.
- Örnek olarak, kol saatimize bakıp "saat 10:18:37:3" demek anlamlı değildir. Saat 10:18'dir.

Anlamlı Basamakları Belirleme Kuralları

1. Sıfır olmayan tüm basamaklar anlamlıdır.
   Örnek: 123456 sayısının anlamlı basamak sayısı altıdır.
2. İki sıfır-dışı basamak arasındaki tüm sıfırlar anlamlıdır.
   Örnek: 103,406 sayısının anlamlı basamak sayısı altıdır.
3. Baştaki sıfırlar anlamsızdır.
   Örnek: 000012 ve 0,012 için anlamlı basamak sayısı ikidir.
4. Ondalık ayraç içeren sayılarda sondaki sıfırlar anlamlıdır.
   Örnek: 1,20300 için anlamlılık düzeyi altı basamaktır.
5. Tam sayılarda sondaki sıfırlar anlamlı ya da anlamsız olabilir.
   Örnek: ($10\bar{0}00$), ($10\underline{0}00$), (1230000) ve (100,) sayıları için anlamlılık düzeyi üçtür. Sonuncu örnekte ondalık ayraçının anlamlılık düzeyini vurgulamak için kullanılmış olduğuna dikkat ediniz.

Bilimsel Gösterim

- "Bilimsel gösterim" (scientific notation), baştaki ve sondaki anlamlı olmayan sıfırları kullanmayarak anlamlı basamak sayısındaki olası bir karışıklığı önlemeyi hedefler.
- Kısaca bilimsel gösterimde tüm basamaklar anlamlıdır.
- "Üstel gösterim" (exponential notation) adı da verilen bilimsel gösterimde tüm sayılar $a \times 10^b$ biçiminde yazılır.
- Burada b bir tam sayıdır. a ise $1 \leq |a| < 10$ olan bir "oranlı sayı" (rational number) biçimindedir.
  Örnek: 0,00123 bilimsel gösterimi $1{,}23 \times 10^{-3}$'tür.
  Örnek: 0,0012300 bilimsel gösterimi $1{,}2300 \times 10^{-3}$'tür.
  Örnek: 1230000 eğer dört basamağa kadar anlamlı ise $1{,}230 \times 10^6$ diye gösterilir.
  Örnek: Üç basamağa kadar anlamlıysa da $1{,}23 \times 10^6$ olur.
- Dikkat: Bilimsel gösterimde, baştaki oranlı sayının her zaman 1 ile 10 arasında olduğuna dikkat ediniz.

Yuvarlama Kuralları

"Yuvarlama" (rounding) kavramı anlamlı basamak kavramı ile yakından ilişkilidir.

Çeşitli hesaplamalarda sıradan yuvarlama yerine "istatistikçi yuvarlaması" (statistician's rounding) yöntemini kullanmak, sonuçların yukarı "yanlı" (biased) olmasını önlemede gereklidir:

1. Tutulacak son basamak seçilir. Bir sonra gelen basamak eğer < 5 ise tutulacak basamak değişmez.
   Örnek: 1,2345 sayısı üç basamağa yuvarlanırsa 1,23 olur.
   Örnek: 1230000 iki basamağa yuvarlanırsa 1200000 olur.
2. Bir sonraki basamak > 5 ise tutulacak basamak bir artırılır.
   Örnek: 0,126 sayısı iki basamağa yuvarlanırsa 0,13 olur.
3. Bir sonra gelen basamak $= 5$ ise; tutulacak basamak tek sayıysa bir artırılır, çift sayıysa değiştirilmez.
   Örnek: 13500 sayısı iki basamağa yuvarlanırsa 14000 olur.
   Örnek: 0,125 sayısı iki basamağa yuvarlanırsa 0,12 olur.

Anlamlı Basamaklar ve Aritmetik

Anlamlı basamaklar ile ilgili olarak, veri ve ölçümler arası aritmetik işlemlerinde aşağıdaki kurallar uygulanır:

1. Öncelikle, örnek olarak $0,12$ gibi bir değerin gerçekte $0,115$ ile $0,125$ arasında olduğu unutulmamalıdır.
2. Toplama ve çıkarma işlemlerinde sonuç, girdiler içinde en az ondalık basamak içeren sayı ile aynı ondalık basamak sayısında olacak şekilde yuvarlanmalıdır.
   Örnek: $0,12 + 0,1277$ yanıtı 0,2477 değil 0,25 olmalıdır.
3. Çarpma ve bölme işlemlerinde sonuç, girdiler içindeki en az anlamlı basamak içeren sayı ile aynı anlamlılık düzeyinde olmalıdır.
   Örnek: $0,12 \times 1234$ yanıtı 148,08 değil 150 olmalıdır.
4. Ancak ara işlemlerde izleyici basamakları elde tutmak gereklidir. Böylece yuvarlama hataları azaltılmış olur.

Örneklem Uzayı ve Örneklem Noktası

Örneklem Uzayı ve Örneklem Noktası

"Rastsal" (random) bir deneyin olabilecek tüm sonuçlarına "örneklem uzayı" (sample space), bu örneklem uzayının her bir üyesine de "örneklem noktası" (sample point) denir.

- Örnek: İki madeni para ile yazı-tura atma deneyinin 4 örneklem noktalı bir örneklem uzayı vardır:

$$Y = \{\text{YY, YT, TY, TT}\}$$

Rastsal Olay ve Karşılıklı Dışlamalı Olay

Rastsal Olay

Rastsal bir deneye ait örneklem uzayının olası her bir alt kümesine "rastsal olay" (random event) denir.

- Örnek: Bir yazı ve bir tura gelmesi olayı: $\{YT, TY\}$

Karşılıklı Dışlamalı Olay

Bir olayın gerçekleşmesi diğer bir olayın oluşmasını önlüyorsa, bu iki olay "karşılıklı dışlamalı" (mutually exclusive) olaylardır.

- Örnek: $\{YY, YT, TY\}$ ve $\{TT\}$ karşılıklı dışlamalıdır.

Rastsal Değişken

Rastsal Değişken

Değerleri rastsal bir deney sonucu belirlenen değişkene "rastsal değişken" (random variable) ya da kısaca "rd" (rv) denir.

- Rastsal değişkenler genellikle X, Y, Z gibi büyük harflerle ve aldıkları değerler de x, y, z gibi küçük harflerle gösterilir.
- Rastsal bir değişken ya "kesikli" (discrete) ya da "sürekli" (continuous) olur.
- Kesikli bir rd ancak sonlu sayıda farklı değerler alabilir. Örnek: Zar.
- Sürekli bir rd ise belli bir aralıkta her sayısal değeri alabilir. Örnek: Rastsal olarak seçilmiş bir kişinin boyu.

Olasılık

Olasılık

A, örneklem uzayındaki bir olay olsun. Rastsal deney sürekli yinelendiğinde, A olayının gerçekleşme sıklık oranına A olayına ait "olasılık" (probability) denir, $P(A)$ ya da $Prob(A)$ ile gösterilir.

- $P(A)$ aynı zamanda "göreli sıklık" (relative frequency) olarak da adlandırılır.

Olasılığın Özellikleri

$P(A)$ gerçek değerli bir "işlev" (function) olup, şu özellikleri taşır:

1. Her A için $0 \leq P(A) \leq 1$'dir. ($1 = \%100$)
2. $A, B, C, \ldots$ örneklem uzayını oluşturuyorsa şu geçerlidir:
$$P(A + B + C + \ldots) = 1$$
3. A, B ve C karşılıklı dışlamalı olaylar ise şu geçerlidir:
$$P(A + B + C) = P(A) + P(B) + P(C)$$

Örnek: Altı yüzlü bir zarı atma deneyi düşünelim:
Bu deneyde örneklem uzayı= $\{1, 2, 3, 4, 5, 6\}$ biçimindedir ve
$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$'dır.
Ayrıca, $P(1) + P(2) + P(3) + P(4) + P(5) + P(6) = 1$ olur.

Kesikli Olasılık Yoğunluk İşlevi

Kesikli Bir Değişkenin Olasılık Yoğunluk İşlevi

X değişkeni $x_1, x_2, x_3, \ldots$ gibi ayrık değerler alan bir rd olsun.

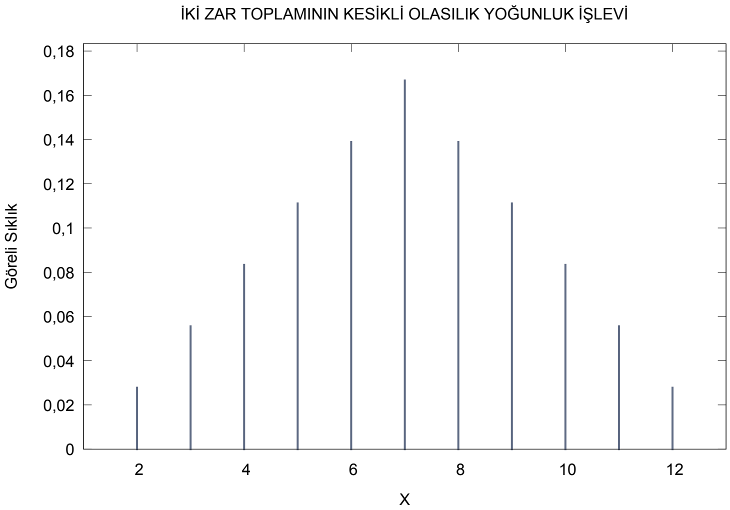
$$\begin{aligned} f(x) &= P(X = x_i) \quad && i = 1, 2, \ldots, n \text{ için} \\ &= 0 && X \neq x_i \text{ için} \end{aligned}$$

işlevine X'e ait "kesikli olasılık yoğunluk işlevi" (discrete probability density function) denir.

- Örnek: İki zar atıldığında zarların toplam değerini gösteren kesikli rastsal değişken X, 11 farklı değer alabilir:

$$x = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$
$$f(x) = \{\tfrac{1}{36}, \tfrac{2}{36}, \tfrac{3}{36}, \tfrac{4}{36}, \tfrac{5}{36}, \tfrac{6}{36}, \tfrac{5}{36}, \tfrac{4}{36}, \tfrac{3}{36}, \tfrac{2}{36}, \tfrac{1}{36}\}$$

Kesikli Olasılık Yoğunluk İşlevi Örnek



Sürekli Olasılık Yoğunluk İşlevi

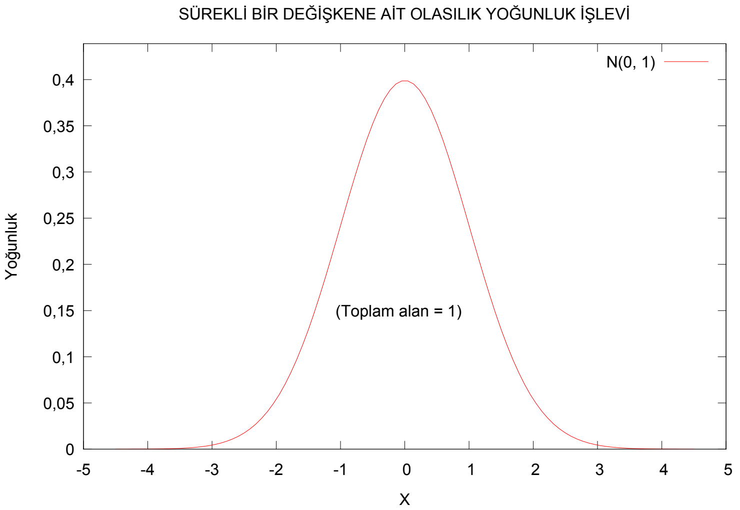
Sürekli Bir Değişkenin Olasılık Yoğunluk İşlevi

X sürekli bir rd olsun.

$$\begin{aligned} f(x) &\geq 0, \\ \int_{-\infty}^{\infty} f(x)dx &= 1, \\ \int_{a}^{b} f(x)dx &= P(a \leq x \leq b) \end{aligned}$$

Eğer yukarıdaki koşullar sağlanırsa, $f(x)$'e X'in "sürekli olasılık yoğunluk işlevi" (continuous probability density function) denir.

Sürekli Olasılık Yoğunluk İşlevi Örnek



Birleşik Olasılık Yoğunluk İşlevi

Birleşik Olasılık Yoğunluk İşlevi

X ve Y iki kesikli rd olsun.

$$\begin{aligned} f(x,y) &= P(X = x_i \wedge Y = y_j), \\ &= 0 \qquad\qquad\qquad\qquad X \neq x_i \wedge Y \neq y_j \text{ için} \end{aligned}$$

işlevi, "kesikli birleşik olasılık yoğunluk işlevi" (discrete joint probability density function) adını alır.

- Birleşik OYİ, X'in x_i değerini ve Y'nin de y_j değerini aynı anda almasının birleşik olasılığını gösterir.

Birleşik Olasılık Yoğunluk İşlevi Örnek

- Aşağıdaki çizelgede X ve Y kesikli değişkenlerine ait bir birleşik OYİ gösterilmektedir:

		X		
		1	**2**	**3**
Y	**0**	0,2	0,3	0,1
	1	0,1	0,1	0,2

- Buna göre $X = 2$ değerini aldığında $Y = 0$ olma olasılığı $f(2, 0) = 0{,}3$ ya da diğer bir deyişle %30'dur.
- Tüm olasılıklar toplamının 1 olduğuna dikkat ediniz.

Marjinal Olasılık Yoğunluk İşlevi

Marjinal Olasılık Yoğunluk İşlevi

$f(x, y)$ birleşik OYİ'sine ilişkin olarak $f(x)$ ve $f(y)$ işlevlerine "marjinal olasılık yoğunluk işlevi" (marginal probability density function) adı verilir:

$$f(x) = \sum_y f(x, y) \quad X\text{'in marjinal OYİ'si}$$
$$f(y) = \sum_x f(x, y) \quad Y\text{'nin marjinal OYİ'si}$$

Marjinal Olasılık Yoğunluk İşlevi Örnek

- Önceki örnekteki verileri ele alalım. X'in marjinal OYİ'si:

$$\begin{array}{rclclcl} f(x=1) & = & \sum_y f(x=1,y) & = & 0{,}2+0{,}1 & = & 0{,}3 \\ f(x=2) & = & \sum_y f(x=2,y) & = & 0{,}3+0{,}1 & = & 0{,}4 \\ f(x=3) & = & \sum_y f(x=3,y) & = & 0{,}1+0{,}2 & = & 0{,}3 \\ & & & & & + & \\ \hline & & & & & & 1{,}0 \end{array}$$

- Aynı şekilde Y'nin marjinal OYİ'si de aşağıdaki gibidir:

$$\begin{array}{rclclcl} f(y=0) & = & \sum_x f(y=0,x) & = & 0{,}2+0{,}3+0{,}1 & = & 0{,}6 \\ f(y=1) & = & \sum_x f(y=1,x) & = & 0{,}1+0{,}1+0{,}2 & = & 0{,}4 \\ & & & & & + & \\ \hline & & & & & & 1{,}0 \end{array}$$

İstatistiksel Bağımsızlık

İstatistiksel Bağımsızlık

X ve Y rastsal değişkenlerinin ancak ve ancak

$$f(x, y) = f(x) \cdot f(y)$$

çarpımı olarak yazılabilmeleri durumunda bunlara "istatistiksel bağımsız" (statistically independent) değişkenler denir.

İstatistiksel Bağımsızlık Örnek

- Örnek olarak bir torbada üzerlerinde 1, 2, 3 yazılı üç top olduğunu düşünelim. Torbadan iki top (X ve Y) yerine koyularak çekilirse, X ve Y'nin birleşik OYİ'si şöyle olur:

		X: 1	2	3
Y	1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
	2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
	3	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

- Burada $f(x=1, y=1) = \frac{1}{9}$'dur.
- $f(x=1) = \sum_y f(x=1, y) = \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{1}{3}$
- $f(y=1) = \sum_x f(x, y=1) = \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{1}{3}$

- Bu örnekte $f(x, y) = f(x) \cdot f(y)$ olduğuna göre, bu iki değişken istatistiksel olarak bağımsızdır diyebiliriz.

Olasılık Dağılımlarının Beklemleri

- Matematikte, bir noktalar kümesinin nasıl bir şekil gösterdiğini anlatan sayısal ölçüye "beklem" (moment) denir.
- Dolayısıyla, bir olasılık dağılımı o dağılıma ait bir dizi beklem ile özetlenebilir.
- Beklemler, "merkezi beklem" (central moment) ve "ham beklem" (raw moment) olarak ikiye ayrılır.
- En yaygın kullanılan iki beklem ise "ortalama" (mean) (μ) ve "varyans" (variance) (σ^2) olarak karşımıza çıkar.
- Ortalama, aynı zamanda "beklenen değer" (expected value) olarak da adlandırılır.

Beklenen Değer

Beklenen Değer

Kesikli bir rd olan X'e ait ortalama ya da beklenen değer $E(X)$ şöyle tanımlanır:

$$E(X) = \sum_x xf(x)$$

- Örnek olarak, iki zarın toplamını gösteren kesikli rd X'in olasılık dağılımını ele alalım:

$$E(X) = \sum_x x\, f(x) = 2\tfrac{1}{36} + 3\tfrac{2}{36} + 4\tfrac{3}{36} + \cdots + 11\tfrac{2}{36} + 12\tfrac{1}{36} = 7$$

- Demek ki iki zar atıldığında gözlenecek sayıların beklenen değeri 7'dir.

Beklenen Değerin Özellikleri

Beklenen değer kavramına ilişkin bazı özellikler şunlardır:

1. Sabit bir sayının beklenen değeri kendisidir.
   Örnek: Eğer $b = 2$ ise $E(b) = 2$'dir.
2. Eğer a ve b birer sabitse, $E(aX + b) = aE(X) + b$'dir.
3. Eğer X ve Y bağımsız rd ise, $E(XY) = E(X)E(Y)$'dir.
4. X, $f(X)$ olasılık yoğunluk işlevli bir rd ve $g(X)$ de X'in herhangi bir işleviyse, şu kural geçerlidir:

$$E[g(X)] = \begin{cases} \sum_x g(X)f(x) & X \text{ kesikli ise,} \\ \int_{-\infty}^{\infty} g(X)f(x)dx & X \text{ sürekli ise.} \end{cases}$$

Buna göre eğer $g(X) = X^2$ ise:

$$E(X^2) = \begin{cases} \sum_x x^2 f(X) & X \text{ kesikli ise,} \\ \int_{-\infty}^{\infty} x^2 f(X)dx & X \text{ sürekli ise.} \end{cases}$$

Beklenen Değer Örnek

- Örnek olarak, aşağıdaki OYİ'yi ele alalım:

$$\begin{array}{rcllll} x & = & \{\text{-}2, & 1, & 2\} \\ f(x) & = & \{\frac{5}{8}, & \frac{1}{8}, & \frac{2}{8}\} \end{array}$$

- Buna göre X'in beklenen değeri şudur:

$$\begin{aligned} E(X) = \sum_x xf(x) &= -2\tfrac{5}{8} + 1\tfrac{1}{8} + 2\tfrac{2}{8} \\ &= -\tfrac{5}{8} \end{aligned}$$

- Ayrıca X^2'nin beklenen değeri ise şudur:

$$\begin{aligned} E(X^2) = \sum_x x^2 f(x) &= 4\tfrac{5}{8} + 1\tfrac{1}{8} + 4\tfrac{2}{8} \\ &= \tfrac{29}{8} \end{aligned}$$

Varyans (Değişirlik)

Varyans (Değişirlik)

X bir rd ve $E(X) = \mu$ ise, X değerlerinin beklenen değerleri etrafındaki yayılımı "varyans" (variance) ile ölçülür:

$$\begin{aligned} \text{var}(X) = \sigma_X^2 &= \textstyle\sum_x \ (X-\mu)^2 f(x) && X \text{ kesikli ise,} \\ &= \textstyle\int_{-\infty}^{\infty} (X-\mu)^2 f(x)dx && X \text{ sürekli ise.} \end{aligned}$$

- σ_X^2'nin artı değerli kare kökü σ_X, X'e ait "ölçünlü sapma" (standard deviation) olarak adlandırılır.
- Varyans ve ölçünlü sapma, her bir rastsal x değerinin X'in ortalaması etrafında ne genişlikte bir alana yayıldığının göstergesidir.

Varyansın Özellikleri

Varyans kavramına ilişkin bazı özellikler şunlardır:

1. Sabit bir sayının varyansı sıfırdır.
2. Eğer a ve b birer sabitse, $\text{var}(aX + b) = a^2\text{var}(X)$'dir.
3. Eğer X ve Y bağımsız birer rd ise şu yazılabilir:
$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$$
$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y)$$
4. Eğer X ve Y bağımsız birer rd ve a, b, c de birer sabit ise, aşağıdaki kural geçerlidir:
$$\text{var}(aX + bY + c) = a^2\text{var}(X) + b^2\text{var}(Y)$$

Varyans Örnek

- Hesaplama kolaylığı bakımından varyans formülü şöyle de yazılabilir:

$$\begin{aligned}
\text{var}(X) = \sigma_X^2 &= (1/n)\sum\left((X_i - E(X))^2\right) \\
&= (1/n)\sum\left(X_i^2 - 2X_iE(X) + E(X)^2\right) \\
&= \sum(X_i^2)/n - \sum 2X_iE(X)/n + \sum E(X)^2/n \\
&= E(X^2) - 2E(X)E(X) + E(X)^2 \\
&= E(X^2) - E(X)^2
\end{aligned}$$

- Buna göre önceki örnekteki rastsal değişkenin varyansı şudur:

$$\text{var}(X) = \frac{29}{8} - \left(-\frac{5}{8}\right)^2 = \frac{207}{64}$$

Kovaryans (Eşdeğişirlik)

Kovaryans (Eşdeğişirlik)

X ve Y rd'lerinin ortalamaları sırasıyla $E(X)$ ve $E(Y)$ olsun. Bu iki değişkenin birlikte değişirlikleri "kovaryans" (covariance) ile ölçülür:

$$\begin{aligned} \text{cov}(X,Y) &= \sum_y \sum_x XY\, f(x,y) - E(X)E(Y) && \text{kesikliyse,} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} XY\, f(x,y)\, dxdy - E(X)E(Y) && \text{sürekliyse.} \end{aligned}$$

- Kovaryans formülü şöyle de gösterilebilir: $\text{cov}(X,Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$
- Görüldüğü gibi bir değişkenin varyansı aynı zamanda kendisiyle olan kovaryansıdır.

Kovaryansın Özellikleri

Kovaryans kavramına ilişkin birkaç önemli özellik şunlardır:

1. Eğer X ve Y bağımsız rd'ler ise kovaryansları 0 olur:
$$\begin{aligned}\text{cov}(X, Y) &= E(XY) - E(X)E(Y)\\ &= E(X)E(Y) - E(X)E(Y) = 0\end{aligned}$$
2. Eğer a, b, c, d birer sabitse şu kural geçerlidir:
$$\text{cov}(a + bX, c + dY) = bd\,\text{cov}(X, Y)$$
3. Bağımsız olmayan X ve Y rd'lerinin bileşimlerinin varyanslarını hesaplarken kovaryans bilgisi de gereklidir:
$$\text{var}(aX + bY) = a^2\text{var}(X) + b^2\text{var}(Y) + 2ab\text{cov}(X,Y)$$

İlinti Katsayısı

İlinti Katsayısı

"İlinti katsayısı" (correlation coefficient) iki rd arasındaki doğrusal ilişkinin bir ölçüsüdür ve $[-1, 1]$ değerleri arasında yer alır:

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y}.$$

- Yukarıdaki formülden şu görülebilir: $\text{cov}(X, Y) = \rho\sigma_x\sigma_y$

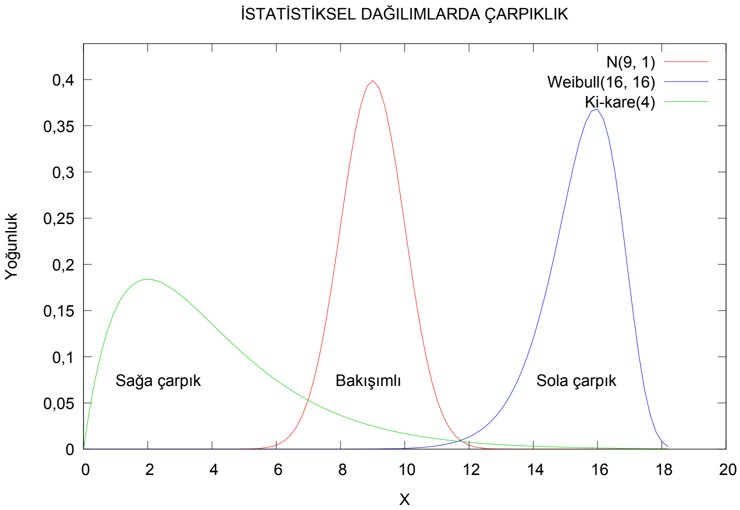
Diğer Merkezi Beklemler

- Genel olarak, $f(x)$ tek değişkenli OYİ'sinin kendi ortalaması dolayındaki merkezi beklemleri şöyle tanımlanır:

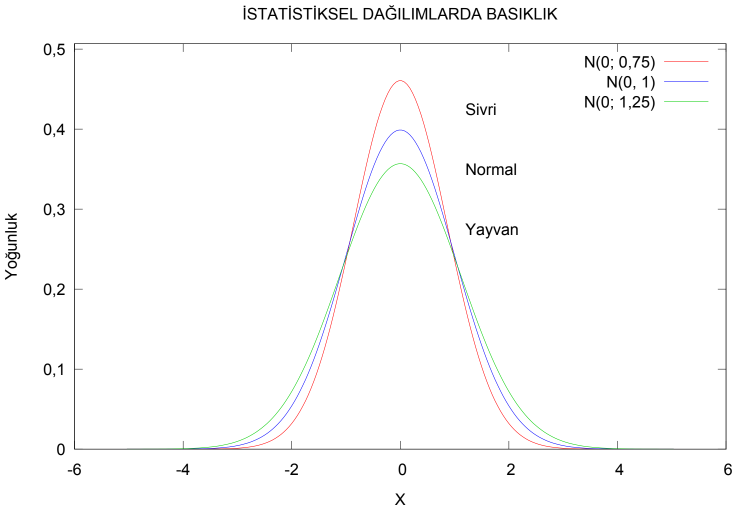
Beklem	Tanım	Açıklama
1	$E(X-\mu)$	0
2	$E(X-\mu)^2$	varyans
3	$E(X-\mu)^3$	çarpıklık
4	$E(X-\mu)^4$	basıklık
$\vdots$	$\vdots$	$\vdots$
n	$E(X-\mu)^n$	n. derece

- "Çarpıklık" (skewness), bakışımdan uzaklığı ölçer.
- "Basıklık" (kurtosis), yayvanlığı incelemek için kullanılır.
- Bir rastsal değişkenin normal dağılıma uyup uymadığını anlamak için çarpıklık ve basıklık değerlerine bakılabilir.

İstatistiksel Dağılımlarda Çarpıklık



İstatistiksel Dağılımlarda Basıklık



Önümüzdeki Dersin Konusu

Önümüzdeki ders

Bazı kuramsal olasılık dağılımları