

Bilişimde İstatistiksel Analiz

Ders 4

Ali Mertcan KOSE Msc.

`amertcankose@ticaret.edu.tr`

İstanbul Ticaret Üniversitesi



İSTANBUL TİCARET
ÜNİVERSİTESİ

*Meydana gelmesi arzu edilen olay sayısının, olayın içinde bulunduğu yığın mevcuduna olan oranına **olasılık** denir.*

Eğer bir olay kesinlikle meydana gelecekse, buna kesin olay(veya belirli bir durum) denir ve olayın meydana gelme olasılığının sayısal değeri **1** dir.

Eğer bir olayın belli koşullar altında meydana gelmesi olanaksız ise, olayın ortaya çıkması olasılığı **0** dır, bu durum önceden bilinmektedir, kesindir. Başka bir anlamda, istenen şekilde meydana gelmeyeceği kesin olan bir olayın istenen şekilde meydana gelmesi olasılığı sıfır olacaktır.

Yapılan açıklamalarda anlaşılabacağı üzere, herhangi bir olasılık ölçüsü **0 ile 1** sayıları arasında bir değere sahip olacaktır. Elverişli şekil olasılığı çok kere **p**, elverişsiz şekil olasılığı da **q** simgesiyle gösterilir; bir olayın istenen veya istenmeyen şekillerde meydana çıkması olasılıkları **1**'e eşittir.

Bir E olayının, aynı meydana gelme “*şansına*” sahip n tane olanaklı durum içerisinde r kere meydana gelebileceğini varsayalım; r elverişli durum olarak adlandırılır ve E olayının meydana gelme olasılığı aşağıdaki gibi gösterilir.

$$P(E) = p = \frac{r}{n}$$

E olayının elverişsiz şekilde meydana gelme olasılığı ise

$$P(\bar{E}) = q = \frac{n-r}{n} = 1 - \frac{r}{n} = 1 - p$$

Örneklem Uzayı: Bir olayın olası tüm sonuçlarının bulunduğu kümeye denir.

Bir sonucun tümleyeni: Kendisi dışındaki diğer sonuçların tümüne, ilgili sonucun tümleyeni denir.

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

$$P(A) + P(\bar{A}) = 1$$

Birleşim: İki ya da daha fazla olayın herhangi birinin ortaya çıkması olayına denir.

$$A \cup B$$

Kesişim: İki ya da daha fazla ayrık olmayan olayın bir arada ortaya çıkması olayına denir.

$$A \cap B$$

Koşullu Olasılık, Bağımsız ve Bağımlı Olaylar

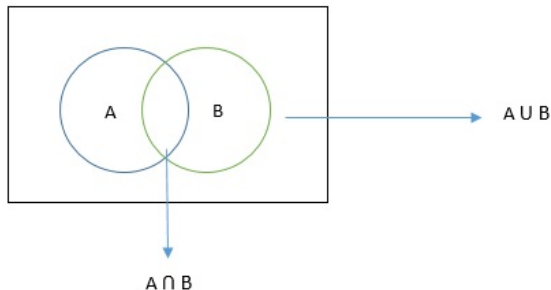


Figure 1: Birleşim ve Kesişim Kümeleri.

Olasılığın Oransal Sıklık olarak Tanımlanması

Olasılıkla ilgili klasik tanımda yer verilen “*aynı meydana gelme şansı*”, “*aynı meydana gelme olasılığın*”dan başka bir şey değildir. Üzerinde durulan istatistiksel olasılık tanımına göre bu olayla ilgili deneysel olasılık, gözlem sayısının çok büyük olması durumunda olayın meydana çıkışıyla ilgili oransal sıklıktır. Olasılık, gözlemlerin sonsuz büyük sayıda arttırılması durumunda oransal sıklığın limiti olarak kabul edilmiştir.

Olasılığın Oransal Sıklık olarak Tanımlanması

Örnek 1:

Family History of Mood Disorders			
	Early =18(E)	Later>18(L)	Total
Negative(A)	28	35	63
Bipolar disorder(B)	19	38	57
Unipolar (C)	41	44	85
Unipolar and bipolar(D)	53	60	113
Total	141	177	318

$$P(E) = \frac{141}{318} = 0.4434$$

Koşullu Olsailık, Bağımsız ve Bağımlı Olaylar

Koşullu olasılık: Bütün gruptaki alt kümeden meydana gelen bütün mümkün sonuçlardır. Diğer bir ifadeyle, koşullu olasılık ilgilendiğiniz grubun genişliğinden dolayı koşulların bütün gruplara uygulanamamasından dolayı indirgenmesidir. Olasılıklar payda olarak bütün grupların alt kümeleri ile hesaplanıldığında, bu sonuçlar *koşullu olasılık* olarak nitelendirilir. Kısacası bir olayın sonucu biliniyorken başka bir olayın ortaya çıkması olasılığıdır.

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

Örnek 1'deki değerler incelendiğinde $P(A/E) = ?$

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

$$= \frac{28}{141} = 0.1986$$

Koşullu Olasılık, Bağımsız ve Bağımlı Olaylar

Bağımsız olaylar: Birinin ortaya çıkma olasılığının diğerine bağlı olmadığı olaylara *bağımsız olaylar* denir

$$P(A \cap B) = P(A) \times P(B)$$

Bağımlı olaylar: Birinin ortaya çıkma olasılığının diğerine bağlı olduğu olaylara *bağımlı olaylar* denir.

$$P(A \cap B) \neq P(A) \times P(B)$$

Ayrık olaylar: Kesişimi boş küme olan olaylara *ayrık olay* denir.

Note

Ayrık olmayan olaylarda: $A \cup B = P(A) + P(B) - P(A \cap B)$

Ayrık olan olaylarda: $A \cup B = P(A) + P(B)$

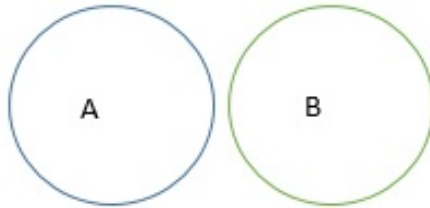


Figure 2: Ayrık Küme.

Permütasyon(Sıra Düzeni): İnsanlar uzun yıllar boyunca, bir takım nesnenin değişik düzenlerde sıralanma sayısının bulunmasını gerektiren sorunlarla ilgilenmişlerdir: On üç kişi, bir sıraya kaç ayrımlı düzende oturabilir; yirmi pul, yirmi numaralı yere kaç değişik düzeyde yerleştirilebilir. Değişik sıra durum sayılma koşuluyla N tane farklı şeyden alınabilecek n tane şeyin çeşitli sıralanışlardan her birine **sıradüzen** denir. Elde edilecek olan sıra düzen sayısı $P_{N,n}$, $P(N,n)$ veya ${}_NP_n$ simgeleriyle gösterilir. aşağıdaki formül yardımıyla hesaplanır.

Permütasyon ve Kombinasyon

$${}_N P_n = \frac{N!}{(N-n)!}$$

Eğer N tane farklı şeyin tamamı alınıp bir sıralama yapılacak olursa, elde edilecek sıradüzen sayısı aşağıdaki gibi olur.

$${}_N P_n = N(N-1)(N-2)\dots (1)=N!$$

Birbirinin benzeri olan N_1, N_2, \dots, N_k birimlik gruplardan meydana gelmiş N tane şeyden oluşturulacak sıradüzen sayısı aşağıda gösterilmiştir.

$${}_N P_n = \frac{N!}{(N-n)!} \quad (N=N_1, N_2, \dots, N_k)$$

Örnek:

“Biyostatistik” sözcüğünün harflerinin sıradüzen sayısını hesaplayınız

Permütasyon ve Kombinasyon

Kombinasyon(Birleşim): Aynı şeyleri değişik bir sıralamayla kapsayan grupları aynı durumda saymak koşuluyla N tane farklı şeyden n tane alarak meydana getirecek gruplara birleşim(kombinasyon) denir. N tane farklı şeyden n tane almak koşuluyla oluşturulacak birleşim sayısı aşağıda gösterilmiştir.

$${}_NC_n = \frac{N(N-1)\dots(N-n+1)!}{(N-n)!n!}$$

veya

$${}_NC_n = \frac{{}_NC_n}{n!}$$

Diğer taraftan ${}_NC_n = {}_NC_{N-n}$ olduğu bilinmelidir.

Örnek

x,y,z harflerinden 2 tanesini seçerek oluşturulabilecek birleşim sayısını belirleyiniz

Önsel olasılıklar(önceden bilinen) yardımı ile herhangi bir olaya ilişkin olasılığın bulunmasında kullanılan eşitliktir.

Tıpta karar verme sürecinde; herhangi tanı yöntemi ile verilen kararın doğru olma olasılığını bulmada yararlanılan bir eşitliktir. *Koşullu olasılığın* bir düzenlenmesidir.

$$P(B_i/A) = \frac{P(B_i)P(A/B_i)}{\sum_{x=a}^b P(B_i)P(A/B_i)} , i=1,2,\dots,n \text{ için}$$

Örnek 3

Alzheimer Hastalığı üzerine yapılan çalışmada hastalara ilişkin sonuçlar aşağıda yer almaktadır. 65 yaş ve üzeri ABD popülasyonun %11,3'nün alzheimer hastası bulunduğu göre, rasgele seçilen 450 alzheimer hastası ve semptom göstermeyen 500 hastanın alzheimer olma olasılığı nedir?

Test Result	Yes(D)	No(D')	Total
Positive(T)	436	5	441
Negative(T')	14	495	509
Total	450	500	950

Çözüm:

$$P(D/T) = \frac{P(D) \times P(T/D)}{P(D) \times P(T/D) + P(D') \times P(T/D')}$$

$$P(D) = 0.113$$

$$P(D') = 1 - 0.113$$

$$P(T/A) = 436/450$$

$$P(T/D') = 5/500$$

$$P(D/T) = \frac{(0.9689)(0.113)}{(0.9689)(0.113) + (0.01)(1 - 0.113)} = 0.93$$

7 farklı hasta arasından seçilen 3 hasta bir boş oda içerisinde yer alan 3 boş yatağa kaç farklı şekilde yerleştirilebilir?

A ve B ayrık olaylar ve $P(B) > 0$ olduğuna göre $P(A/B)$ olasılığı kaçtır?

Bir Hastanede yer alan Röntgen Aletinin hatalı sonuç verme olasılığı $p=(0.03)$ ise 5. hasta geldiğinde Aletin hatalı sonuç vermesi olasılığı kaçtır?

HIV virüsü testi sonucunun doğru olma olasılığı %95 olarak ifade edilirken, yaklaşık olarak ABD'de 308 milyonda 1 kişinin HIV virüsüne yakalandığı bilinirken. HIV virus testinin doğru olup, HIV virüsüne yakalanma olasılığı nedir(H/P)?