IST108
OLASILIK VE İSTATİSTİK

ÖRNEKLEME İSTATİSTİKLERİNİN DAĞILIMI

İçerik

İçerik

Örnek Ortalaması

Yetişkinlerden oluşan bir topluluk düşünelim.
- Yığın

Her biri bir sayısal değerle (rastgele değişken) eşleştirilmiş olabilir.
- Yıllık gelir
- Boy uzunluğu
- Yaş

Bu rastgele değişkene ait beklenti μ
- Yığın ortalaması

Bu rastgele değişkene ait varyans ise σ^2
- Yığın varyansı

Örnek Ortalaması

$X_1, X_2, \dots, X_n$ bu yığından alınan n adet örneği göstersin. Bu durumda örnek ortalaması

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Örnek ortalaması örnekteki rastgele değişkenlerin toplamının n'ye bölünmesi ile belirlendiğinden, $\bar{X}$ de bir rastgele değişken olarak düşünülebilir. Bu durumda bu rastgele değişkene ait beklenti ve varyans da hesaplanabilir.

Örnek Ortalamasının Beklentisi

Örnek ortalamasının beklentisi aşağıdaki şekilde hesaplanabilir.

$$E[\bar{X}] = E\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right]$$

$$E[\bar{X}] = \frac{1}{n} E[X_1 + X_2 + \cdots + X_n]$$

$$E[\bar{X}] = \frac{1}{n}(E[X_1] + E[X_2] + \cdots + E[X_n])$$

$$E[\bar{X}] = \frac{1}{n}(n\mu)$$

$$E[\bar{X}] = \mu$$

Örnek Ortalamasının Varyansı

Örnek ortalamasının varyansı aşağıdaki şekilde hesaplanabilir.

$$Var(\bar{X}) = Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$$

$$Var(\bar{X}) = \frac{1}{n^2} Var(X_1 + X_2 + \cdots + X_n)$$

$$Var(\bar{X}) = \frac{1}{n^2}\big(Var(X_1) + Var(X_2) + \cdots + Var(X_n)\big)$$

$$Var(\bar{X}) = \frac{1}{n^2}(n\sigma^2)$$

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

Merkezi Limit Teoremi

Merkezi Limit Teoremi: Genel olarak, çok sayıda sayıda birbirlerinden bağımsız rastgele değişkenlerin toplamı yaklaşık olarak normal dağılım şeklinde davranır.

$X_1, X_2, \dots, X_n$, birbirinden bağımsız ve her biri beklentisi μ ve varyansı ise σ^2 olan aynı dağılımlara sahip rastgele değişkenler olsun. Eğer n çok büyük ise, $X_1 + X_2 + \dots + X_n$ rastgele değişkenine ait dağılım, beklentisi $n\mu$ ve varyansı ise $n\sigma^2$ olan bir normal dağılıma yakınsanabilir.

$$X = X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2)$$

$E[X] = n\mu$ ve $Var(X) = n\sigma^2$

Merkezi Limit Teoremi

$P(X_1 + X_2 + \cdots + X_n < a)$ olasılığını merkezi limit teoremini kullanarak aşağıdaki gibi bulabiliriz.

Toplamın normal dağılım olduğu kabul edilirse olasılık standart normal dağılıma geçiş yapılarak bulunabilir. Burumda

$$P(X_1 + X_2 + \cdots + X_n < a) = P\left(\frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}} < \frac{a-n\mu}{\sigma\sqrt{n}}\right)$$

$$\cong P\left(Z < \frac{a-n\mu}{\sigma\sqrt{n}}\right)$$

Örnek 1

Bir sigorta şirketi 25000 araç poliçesi yaptıran müşteriye sahiptir. Poliçe sahiplerinin yıllık sigorta talepleri, ortalaması 320 ve standart sapması 540 olan bir rastgele değişken ile ifade edilebiliyorsa, yıllık sigorta taleplerinin 8,3 milyonu aşması olasılığını hesaplayınız.

Örnek 1

Bir sigorta şirketi 25000 araç poliçesi yaptıran müşteriye sahiptir. Poliçe sahiplerinin yıllık sigorta talepleri, ortalaması 320 ve standart sapması 540 olan bir rastgele değişken ile ifade edilebiliyorsa, yıllık sigorta taleplerinin 8,3 milyonu aşması olasılığını hesaplayınız.

X, yıllık talebi göstersin. Poliçe sahiplerini numaralandıralım ve X_i, i. poliçe sahibinin yıllık sigorta talebi olsun. $n = 25000$ ile merkezi limit teoremine göre $X = \sum_{i=1}^{n} X_i$, beklentisi $25000 \times 320 = 8\ milyon$ olan ve standart sapması $540 \times \sqrt{25000} = 85831$ olan bir normal dağılıma yakınsanabilir.

Örnek 1

Bir sigorta şirketi 25.000 araç poliçesi yaptıran müşteriye sahiptir. Poliçe sahiplerinin yıllık sigorta talepleri, ortalaması 320 ve standart sapması 540 olan bir rastgele değişken ile ifade edilebiliyorsa, yıllık sigorta taleplerinin 8,3 milyonu aşması olasılığını hesaplayınız.

Ortalaması: $n\mu = 25000 \times 320 = 8 \times 10^6$olan ve

Standart sapması: $\sigma\sqrt{n} = 540 \times \sqrt{25000} = 85831$

$$P(X > 8{,}3 \times 10^6) \approx P\left(Z > \frac{8{,}3\times10^6 - 8\times10^6}{85831}\right)$$

$$\approx P(Z > 3{,}51) \approx 0{,}00022$$

Örnek 2

İnşaat mühendislerinin hesabına göre, yapısal bir zarar oluşturmadan bir köprünün taşıyabileceği yük miktarı W (ton olarak) beklentisi 400 ve standart sapması 40 olan bir normal dağılımla dağıtılmıştır. Bir arabanın ağırlığı (ton olarak) beklentisi 3 ve standart sapması 0,3 olan bir normal dağılımla dağıtılmış olsun. Köprüde kaç araç olduğunda yapısal zarar olasılığı 0,1'i geçer?

Örnek 2

İnşaat mühendislerinin hesabına göre, yapısal bir zarar oluşturmadan bir köprünün taşıyabileceği yük miktarı W (ton olarak) beklentisi 400 ve standart sapması 40 olan bir normal dağılımla dağıtılmıştır. Bir arabanın ağırlığı (ton olarak) beklentisi 3 ve standart sapması 0,3 olan bir normal dağılımla dağıtılmış olsun. Köprüde kaç araç olduğunda yapısal zarar olasılığı 0,1'i geçer?

P_n, köprü üzerinde n tane araç olduğunda yapısal zarar ihtimali olsun. Bu durumda soruda sorulan $P_n > 0{,}1$ olduğunda n kaç olur?

Örnek 2

P_n, köprü üzerinde n tane araç olduğunda yapısal zarar ihtimali olsun. Bu durumda soruda sorulan $P_n > 0{,}1$ olduğunda n kaç olur?

X_i, köprü üzerindeki i. aracın ağırlığı olsun. Köprü üzerindeki araçların toplam ağırlığı köprünün taşıyabileceğinden fazla olursa zarar verecek.

$P_n = P(X_1 + X_2 + \cdots + X_n \geq W)$

$P_n = P(X_1 + X_2 + \cdots + X_n - W \geq 0)$

$X = X_1 + X_2 + \cdots + X_n$ olsun.

$P_n = P(X - W \geq 0)$

Örnek 2

P_n, köprü üzerinde n tane araç olduğunda yapısal zarar ihtimali olsun. Bu durumda soruda sorulan $P_n > 0{,}1$ olduğunda n kaç olur?

Merkezi Limit Teoremine göre araçların toplam ağırlığı $X = X_1 + X_2 + \cdots + X_n$ yaklaşık olarak ortalaması $3n$ ve varyansı $0{,}09n$ olan bir normal dağılım ile modellenebilir.

$E[X] = \mu n = 3n$ ve $Var(X) = \sigma^2 n = 0{,}3^2 n = 0{,}09n$

X ve W birbirlerinden bağımsızdır.

$E[X - W] = E[X] - E[W] = 3n - 400$

$Var(X - W) = Var(X) + Var(W) = 0{,}09n + 1600$

Örnek 2

P_n, köprü üzerinde n tane araç olduğunda yapısal zarar ihtimali olsun. Bu durumda soruda sorulan $P_n > 0{,}1$ olduğunda n kaç olur?

$U = X - W$

$P_n = P(X - W \geq 0) = P(U \geq 0) > 0{,}1$

$E[U] = 3n - 400$

$Var(U) = 0{,}09n + 1600$

$P(U \geq 0) = P\left(\frac{U-(3n-400)}{\sqrt{0{,}09n+1600}} \geq \frac{0-(3n-400)}{\sqrt{0{,}09n+1600}}\right) > 0{,}1$

$P(U \geq 0) = P\left(Z \geq \frac{-(3n-400)}{\sqrt{0{,}09n+1600}}\right) > 0{,}1$

Örnek 2

P_n, köprü üzerinde n tane araç olduğunda yapısal zarar ihtimali olsun. Bu durumda soruda sorulan $P_n > 0{,}1$ olduğunda n kaç olur?

$P(Z \geq 1{,}28) \approx 0{,}1$ olduğundan

$\frac{-(3n-400)}{\sqrt{0{,}09n+1600}} \leq 1{,}28 \ \rightarrow n \geq 117$

Merkezi Limit Teoremi ve Binom Rastgele Değişkeni

Merkezi limit teoremi, Binom rastgele değişkene uygulanabilir.

X, (n, p) parametreli bir Binom rastgele değişken olsun.

X, her birinin başarılı olma olasılığı p, başarısız olma olasılığı $1 - p$ olan n adet bağımsız deneyden başarılı olanların sayısını ifade eder.

Buradaki her bir deney ise $(X_1, X_2, X_3, \ldots, X_n)$, bir Bernoulli rastgele değişken ile ifade edilebilir.

$X = X_1 + X_2 + X_3 + \cdots + X_n$

$X_i = \begin{cases} 1 & i.\, deney\ başarılı\ ise \\ 0 & diğer\ durumlarda \end{cases}$

Merkezi Limit Teoremi ve Binom Rastgele Değişkeni

$E[X_i] = p \qquad Var(X_i) = p(1 - p)$

$E[X] = np \qquad Var(X) = np(1 - p)$

$\frac{X-np}{\sqrt{np(1-p)}} \approx N(0{,}1)$

Örnek 3

Bir banka şubesi geçmiş tecrübelerinden, sıramatikten sıra numarası alan müşterilerden %80'inin işlem yaptırdığını hesaplamıştır. Banka şubesinde günlük olarak 480 müşterinin işlemin yapılabilmektedir. Bundan dolayı banka şubesindeki sıramatik 600 müşteriye sıra numarası vermektedir. Bir günde 480'den fazla müşterinin işlem yaptırma olasılığını hesaplayınız.

Örnek 3

X_B, işlem yaptıran müşteri sayısını gösteren rastgele değişken olsun.

Her müşterinin işlem yaptırma olasılığı birbirinden bağımsızdır.

X_B'ye, bu durumda $(n = 600, p = 0{,}8)$ parametreli bir Binom rastgele değişken diyebiliriz.

Binom kesikli bir rastgele değişken, Normal dağılım ise sürekli rastgele değişken olduğundan, normal yakınsama uygulanırken $P(X = i)$ olasılığını hesaplamak için $P(i - 0{,}5 < X < i + 0{,}5)$ olasılığını hesaplamak gerekir. Buna **süreklilik düzeltmesi** denir.

Örnek 3 – Süreklilik Düzeltmesi

X_B, işlem yaptıran müşteri sayısını gösteren Binom rastgele değişken olsun.

Soruda $P(X_B > 480)$ olasılığı soruluyor.

X_B kesikli rastgele değişken olduğu için sadece tamsayı değerler alabilir. Dolayısıyla $P(X_B > 480) = P(X_B \geq 481)$'dir.

Süreklilik düzeltmesinde sınır değer, 480'i dışarıda bırakıp 481'i kapsayacak şekilde seçilmelidir.

Örnek 3 – Süreklilik Düzeltmesi

X_N rastgele değişkeni, X_B Binom rastgele değişkeninin Normal dağılım olarak modellenmiş hali olsun.

Bu durumda $P(X_B > 480) \approx P(X_N > 480{,}5)$ ile yaklaşık olarak hesaplanabilir.

Soruda $P(X_B \geq 480)$ istenseydi, süreklilik düzeltmesi ile $P(X_N > 479{,}5)$ normal dağılım yakınsaması ile bulunabilirdi.

Örnek 3

$P(X_B > 480) \approx P(X_N > 480{,}5)$

$$P(X_N > 480{,}5) = P\left(\frac{X-(600)(0{,}8)}{\sqrt{600(0{,}8)(0{,}2)}} > \frac{480{,}5-(600)(0{,}8)}{\sqrt{600(0{,}8)(0{,}2)}}\right)$$

$$= P\left(Z > \frac{0{,}5}{9{,}8}\right) = P(Z > 0{,}05) = 1 - P(Z < 0{,}05)$$

$$= 1 - \Phi(0{,}05) = 1 - 0{,}5199 = 0{,}4801$$

Merkezi Limit Teoremi ve Binom Rastgele Değişkeni

Şu ana kadar işlediğimiz konularda Binom rastgele değişkene ait iki yakınsama metodu öğrenildi: Poisson ve Merkezi limit teoremi.

Poisson: n çok büyük ve p çok küçük olduğunda iyi bir yakınsama sunar.

Merkezi Limit Teoremi: $np(1-p)$ büyük olduğunda (ör. ≥ 10) iyi bir yakınsana sunar.

Merkezi Limit Teoremi ve Örnek Ortalaması

Merkezi limit teoremi örnek ortalamasının dağılımını yakınsamak için de kullanılabilir.

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

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0,1)$$

Merkezi Limit Teoremi ve Örnek Ortalaması

Merkezi limit teoreminin geçerli olabilmesi için örnek boyutu n'in ne olması gerekir?

Eğer her bir değere ait dağılım normal dağılım ise, bu değerlerin toplamına ait dağılım, n değerine bağlı olmaksızın normal olacaktır.

Genel bir kural olarak eğer n en az 30 ise normal dağılıma yaklaştırmak uygun olacaktır.

Ama bir çok durumda daha düşük boyutlar için bile merkezi limit teoremi çok iyi bir yakınsama sunar.

Örnek 4

Bir işçi popülasyonunda işçilerin ağırlıklarına ait beklenti 167 ve standart sapma 27'dir.

A) 36 tane işçi bu popülasyondan rastgele seçilirse, örnek ortalamalarının 163 ile 171 arasında olmasını yaklaşık olarak hesaplayınız.

B) 144 işçi seçimi için cevabınız ne olurdu?

Örnek 4

A) 36 tane işçi bu popülasyondan rastgele seçilirse, örnek ortalamalarının 163 ile 171 arasında olmasını yaklaşık olarak hesaplayınız.

Merkezi Limit Teoremine göre örnek ortalaması, beklentisi 167 olan ve standart sapması $\frac{27}{6} = 4{,}5$ olan bir normal dağılıma yakınsanabilir.

$$P(163 < \bar{X} < 171) = P\left(\frac{163-167}{4{,}5} < \frac{\bar{X}-167}{4{,}5} < \frac{171-167}{4{,}5}\right)$$

$$= P(-0{,}8889 < Z < 0{,}8889) = 2P(Z < 0{,}8889) - 1 = 0{,}6259$$

Örnek 4

B) 144 işçi seçimi için cevabınız ne olurdu?

Bu durumda beklenti aynı kalırken standart sapma $\frac{27}{12} = 2{,}25$ olur.

$$P(163 < \bar{X} < 171) = P\left(\frac{163-167}{2{,}25} < \frac{\bar{X}-167}{2{,}25} < \frac{171-167}{2{,}25}\right)$$

$$= P(-1{,}7778 < Z < 1{,}7778) = 2P(Z < 1{,}7778) - 1 = 0{,}9246$$

Örnek Varyansı ve Standart Sapması

Örnek varyansı ve standart sapması da aynı örnek ortalaması gibi bir rastgele değişken olarak düşünülebilir.

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

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

Sınırlı Yığından Örnekleme

N elemandan oluşan bir yığında, yığın, p oranında belirli bir karakteristiği gösteriyor olsun.

Yani N elemanlı grubun Np tanesi belirli bir karakteristiğe sahiptir. $N(1-p)$ tane eleman bu karakteristiği göstermemektedir.

Bu yığından az sayıda n elemanlı bir örnek seçelim. Eğer yığının n boyutundaki tüm alt kümelerinin örnek olma ihtimali eşitse bu her bir örneğe **rastgele örnek** denir.

Örneğin, $\{a, b, c\}$ elemanlarından oluşan bir popülasyon için, 2 elemanlı her bir alt küme $\{a, b\}$, $\{a, c\}$ ve $\{b, c\}$ aynı olasılıkla örnek oluşturabilirlerse bir rastgele örnek seçilebilir.

Sınırlı Yığından Örnekleme

n elemanlı bir rastgele örneğin N elemanlı bir yığından seçildiği varsayalım ve aşağıdaki tanımlamayı yapalım.

$$X_i = \begin{cases} 1 & i.\text{üye belirli karakteristiğe sahip ise} \\ 0 & \text{diğer durumlarda} \end{cases}$$

Bu durumda belirli bir karakteristiği gösteren elemanların sayısı bu rastgele değişkenlerin toplamı olacaktır.

$$X_1 + X_2 + \cdots + X_n$$

Bu değerlerin ortalaması ise örnek içinde belirli bir karakteristiğe sahip elemanların oranı olacaktır.

Sınırlı Yığından Örnekleme

N eleman içindeki her bir elemanın örnekteki i. eleman olma olasılığı eşittir.

$$P(X_i = 1) = \frac{Np}{N} = p$$

$$P(X_i = 0) = 1 - P(X_i = 1) = 1 - p$$

Sınırlı Yığından Örnekleme

$X_1, X_2, \dots, X_n$ değerleri birbirinden bağımsız değildir. Neden?

Örnek için seçilen **ikinci** eleman, N elemandan herhangi biri olabileceğinden ikinci elemanın olasılıkları aşağıdaki gibi olmalı diye düşünebiliriz.

$$P(X_2 = 1) = \frac{Np}{N} = p$$

$$P(X_2 = 0) = 1 - P(X_2 = 1) = 1 - p$$

Ancak, birinci seçilen elemanın karakteristiğe sahip olup olmadığı verildiğinde durum değişir.

Sınırlı Yığından Örnekleme

İlk seçilen eleman karakteristiğe sahipse tüm yığın içinden seçilen ikinci elemanın ilgilenilen karakteristiğe sahip olması:

- $P(X_2 = 1|X_1 = 1) = \frac{Karakteristiğe\ sahip\ geriye\ kalan\ eleman\ sayısı}{Geriye\ kalan\ eleman\ sayısı} = \frac{Np-1}{N-1}$

İlk seçilen eleman karakteristiğe sahip değilse tüm yığın içinden seçilen ikinci elemanın ilgilenilen karakteristiğe sahip olması:

- $P(X_2 = 1|X_1 = 0) = \frac{Karakteristiğe\ sahip\ geriye\ kalan\ eleman\ sayısı}{Geriye\ kalan\ eleman\ sayısı} = \frac{Np}{N-1}$

Olasılıklar p olmadığı için bağımlıdırlar.

Sınırlı Yığından Örnekleme

Ancak N çok büyükse, bu etki çok küçük olacaktır. Örneğin, $N = 1000$ ve $p = 0,4$ için bu koşullu olasılık, koşulsuz olasılığa çok yakın olur.

$P(X_2 = 1|X_1 = 1) = \frac{399}{999} = 0,3994$

$P(X_2 = 1|X_1 = 0) = \frac{400}{999} = 0,4004$

$P(X_2 = 1) = \frac{1000 \times 0,4}{1000} = 0,4$

N eğer n'ye göre çok büyük ise $X_1, X_2, \ldots, X_n$ değerleri yaklaşık olarak birbirinden bağımsızdır ve X toplamı (n, p) parametreli bir Binom rastgele değişken gibi düşünülebilir.

Örnek 5

Nüfusun gelecek seçimlerde %45 oranında belirli bir partiyi desteklediğini varsayın. Eğer nüfustan rastgele 200 kişi seçilirse,

A) Bu partiyi destekleyeceklerin sayısına ait beklenti ve standart sapma nedir?

B) Yarısından fazlasının bu partiyi desteklemesi ihtimali nedir?

Örnek 5

Nüfusun gelecek seçimlerde %45 oranında belirli bir partiyi desteklediğini varsayın. Eğer nüfustan rastgele 200 kişi seçilirse,

A) Bu partiyi destekleyeceklerin sayısına ait beklenti ve standart sapma nedir?

$E[X] = 200 \times 0,45 = 90$

$\sqrt{Var(X)} = \sqrt{200 \times 0,45 \times 0,55} = 7,0356$

Örnek 5

Nüfusun gelecek seçimlerde %45 oranında belirli bir partiyi destekledigini varsayın. Eğer nüfustan rastgele 200 kişi seçilirse,

B) Yarısından fazlasının bu partiyi desteklemesi ihtimali nedir?

Bir program vasıtası ve Binom ile çözülürse $P(X \geq 101) = 0{,}0681$

Normal yakınsama ile aşağıdaki gibi çözülebilir.

$$P(X \geq 101) = P(X \geq 100{,}5) = P\left(\frac{X-90}{7{,}0356} \geq \frac{100{,}5-90}{7{,}0359}\right)$$
$$= P(Z \geq 1{,}4924) = 0{,}0678$$