

## İSTATİSTİK DAĞILIMLAR

Üzerinde çalışılan kesikli ve sürekli değişkenlere ait verilerin elde edildiği örnekler genellikle dağılım fonksiyonları bilinen populasyonları temsil ederler. Bunlardan en çok rastlanan üç tanesi bu kitabın kapsamında incelenecektir. **Binom**, **Poisson** ve **Normal** dağılımlar fonksiyonları bilinen ve en çok rastlananlardır.

### 3.1. Binom Dağılımı

Adından da anlaşılacağı gibi binom dağılmış populasyonda değişkenlerin iki hali söz konusudur. Bunlardan biri üzerinde durulan (araştırılan, istenen) hal, diğeri de üzerinde durulmayan (istenmeyen) haldir. Eğer ele alınan özelliğın ikiden fazla hali söz konusu ise multinom dağılım söz konusudur. Ancak böyle durumlar da binom dağılımı olarak incelenebilir. Bu ikiden fazla halden sadece bir tanesi araştırmacı için önemli ise bunu birinci hal olarak alır, diğerkleri de ikinci hali oluşturur. Binom dağılımında populasyonda üzerinde durulan halin olasılığı (ihtimali)  $\pi$  ile gösterilir. Ele alınan populasyonda  $\pi$ 'nin değişmediği, yani populasyondaki her bireyin üzerinde durulan hali taşıma olasılığının sabit olduğu varsayılır. Üzerinde durulan halin olasılığı  $\pi$  olduğuna göre ( $\pi < 1$ ), üzerinde durulmayan halin olasılığı da  $(1-\pi)$ 'dir.  $\pi$  bilinmediği zaman, diğerk parametrelerde olduğu gibi, örnekten tahmin edilir. Bu durumda üzerinde durulan halin olasılığı  $p$ , diğerk halin olasılığı  $(1-p)$  veya kısaca  $q$  ile gösterilir.

Herhangi bir toplumda doğan bebeklerin cinsiyetlerinin erkek veya kız olması gibi sadece iki hal söz konusudur. Biyolojik kurallardan da çocuğın cinsiyetinin kız veya erkek olma olasılığının  $1/2$  olduğu bilinmektedir.

Üzerinde durulan hal kız çocuk ise, bunun olma olasılığı  $\pi=0.5$ 'dir. Yani parametre bilinmektedir. İki çocuklu aileler ele alındığında, aşağıdaki durumlar söz konusudur.

SÖZ KONUSU HALLER	OLASILIĞI
a. İlk ve ikinci çocuk erkektir.	$\frac{1}{4}$
b. İlk çocuk erkek ikinci çocuk kızdır.	$\frac{1}{4}$
c. İlk çocuk kız ikinci çocuk erkektir.	$\frac{1}{4}$
d. İlk ve ikinci çocuk kızdır.	$\frac{1}{4}$

Kız çocuk sayısı ele alınan değişken ( $X$ ) ise bunun yukarıda aldığı değerler (0, 1, 1, 2) şeklindedir. Birinci veya ikinci çocuğın kız olması önemli olmadığından  $b$  ve  $c$  halleri birlikte ele alınabilir. Bu durumda  $X$  değişkeninin aldığı değerler (0,1 ve 2)'dir. Bunların olasılıkları sırasıyla ( $1/4$ ,  $2/4$  ve  $1/4$ )'dür. Bu durum da şöyle de yazılabilir:

$[(1)(\frac{1}{4}), (2)(\frac{1}{4}), (1)(\frac{1}{4})]$ . Bunlar her halin olasılıklarıdır. Sırasıyla 0.25, 0.5 ve 0.25'tir ve toplamları (1.0)'dir. Bu  $[\pi+(1-\pi)]^2$  şeklinde de gösterilebilir. Bu açılım  $(1)(\pi^2)+(2)(\pi)(1-\pi)+(1)(1-\pi)^2$ 'dir. Bilindiği şekilde buna binom açılımı da denir. Burada 1,2,1 katsayıları

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

n olayda (burada iki), üzerinde durulan halin 2, 1 ve 0 kere görülme sayısıdır, yani kombinasyon sayısıdır.  $nCr$  veya  $\binom{n}{r}$  şeklinde gösterilir. Bu da:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \text{ 'dir.}$$

Bu bilgilere göre binom dağılımında üzerinde durulan halin r kere olma olasılığı  $P(r)$  aşağıdaki gibi gösterilir.

$$\begin{aligned} P(r) &= C(n,r)\pi^r(1-\pi)^{(n-r)} \\ &= \frac{n!}{(n-r)!r!} \pi^r(1-\pi)^{(n-r)} \quad \dots(3.1) \end{aligned}$$

3.1 numaralı eşitlik binom dağılımının olasılık fonksiyonudur. Bu fonksiyonda n deneme sayısı, r bu denemede istenen olayın görünme sayısı,  $P(r)$  istenen olayın n denemede r defa görünme olasılığıdır.  $nCr$  ise n tane farklı şeyden r tanesi alınarak sıra gözetmeksizin yapılabilecek dizilişlerin sayısıdır.

Faktöriyel ile ilgili aşağıdaki bilgilerin hatırlanması yararlıdır.

**\*\* 0!=1'dir.**

**\*\* 1!=1'dir.**

**\*\* 4!=4.3.2.1=24'tür.**

### 3.1.1. İstenen Olayın Oluş İhtimalinin Hesaplanması

Bütün binom popülasyonlarda n denemede istenen olayın r defa meydana gelme olasılığı yukarıda verilen olasılık fonksiyonu kullanılarak hesaplanabilir.

Yapılan bir araştırmada öğrencilerin %25'inin sigara içtiği saptanmıştır. Rastgele seçilecek dört öğrenciden;

a. 4'ünün de sigara içmeme olasılığı,

b. 4 öğrenciden 3'nün sigara içmeme olasılığı,

c. 4 öğrenciden 2'sinin sigara içmeme olasılığı,

d. 4 öğrenciden 1'inin sigara içmeme olasılığı,

e. 4 öğrenciden 4'ünün de sigara içme (veya hepsinin sigara içme) olasılığı yukarıda verilen olasılık fonksiyonu kullanılarak hesaplanabilir.

Bu olasılıkları hesaplamak için önce istenen ve istenmeyen tip olayların olasılıklarının bulunması gerekir. Sigara içmeme olayının olasılığı (istenen tip)  $\pi = \frac{3}{4}$  ve

istenmeyen tip (sigara içme) olayın olasılığı ise  $(1-\pi) = 1 - \frac{3}{4} = \frac{1}{4}$  şeklinde hesaplanır. Bu

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

örnekte deneme sayısı (n) 4'tür. Deneme sayısı 4 olduğu zaman mümkün olan ve olasılıkları istenen 5 durum söz konusudur.

$$\begin{aligned} \text{a. } P(4) &= C(4,4)\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^{4-4} = \frac{4!}{(4-4)!4!}\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^0 \\ &= \left(\frac{3}{4}\right)^4 = \frac{81}{256} = 0.316 \end{aligned}$$

Bu hesaplanan olasılık 4 öğrenciden 4'ünde sigara içmeme olasılığıdır.

$$\begin{aligned} \text{b. } P(3) &= C(4,3)\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^{4-3} = \frac{4!}{(4-3)!3!}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^1 \\ &= \frac{4!}{(4-3)!3!}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right) = 4\frac{27}{256} = \frac{108}{256} = 0.422 \end{aligned}$$

Bu hesaplanan ihtimal 4 öğrenciden 3'ünün sigara içmeme ihtimalidir.

$$\begin{aligned} \text{c. } P(2) &= C(4,2)\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^{4-2} = \frac{4!}{(4-2)!2!}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 \\ &= \frac{4!}{(4-2)!2!}\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2 = 6\frac{9}{256} = \frac{54}{256} = 0.211 \end{aligned}$$

Bu hesaplanan olasılık 4 öğrenciden 2'sinin sigara içmeme olasılığıdır.

$$\begin{aligned} \text{d. } P(1) &= C(4,1)\left(\frac{3}{4}\right)^1\left(\frac{1}{4}\right)^{4-1} = \frac{4!}{(4-1)!1!}\left(\frac{3}{4}\right)^1\left(\frac{1}{4}\right)^3 \\ &= \frac{4!}{(4-1)!1!}\left(\frac{3}{4}\right)^1\left(\frac{1}{4}\right)^3 = 4\frac{3}{256} = \frac{12}{256} = 0.047 \end{aligned}$$

Bu hesaplanan olasılık 4 öğrenciden 1'inin sigara içmeme olasılığıdır.

$$\begin{aligned} \text{e. } P(0) &= C(4,0)\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^{4-0} = \frac{4!}{(4-0)!0!}\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^4 \\ &= \frac{4!}{(4-0)!0!}\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^4 = \frac{1}{256} = \frac{1}{256} = 0.004 \end{aligned}$$

Bu hesaplanan ihtimal 4 öğrencinin hepsinin sigara içme ihtimalidir. Mümkün olan durumların oluş ihtimallerinin toplamı 1'e eşittir, yani;

$$P(4) + P(3) + P(2) + P(1) + P(0) = 1.0$$

$$\frac{81}{256} + \frac{108}{256} + \frac{54}{256} + \frac{12}{256} + \frac{1}{256} = 1.0$$

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Binom dağılımına ait olasılık fonksiyonu kullanılarak hesaplanan olasılıklar  $(\pi+(1-\pi))^n$  binomunun açılımındaki terimlere karşılık gelir. Yukarıda verilen örnek için binom açılımı kullanılacak olursa;

$$(\pi + (1 - \pi))^4 = \pi^4(1 - \pi)^0 + 4\pi^3(1 - \pi)^1 + 6\pi^2(1 - \pi)^2 + 4\pi^1(1 - \pi)^3 + \pi^0(1 - \pi)^4$$

$$\left(\frac{3}{4} + \frac{1}{4}\right)^4 = \underbrace{\left(\frac{3}{4}\right)^4\left(\frac{1}{4}\right)^0}_{a} + 4\underbrace{\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^1}_{b} + 6\underbrace{\left(\frac{3}{4}\right)^2\left(\frac{1}{4}\right)^2}_{c} + 4\underbrace{\left(\frac{3}{4}\right)^1\left(\frac{1}{4}\right)^3}_{d} + \underbrace{\left(\frac{3}{4}\right)^0\left(\frac{1}{4}\right)^4}_{e}$$

Mümkün olan haller için olasılıklar bulunduktan sonra araştırmacı istediği olasılıkları hesaplayabilir. Örneğin:

a. En az 2 öğrencinin sigara içmeme olasılığı;

**$P(r \geq 2) = P(2) + P(3) + P(4)$** 'e eşittir, yani en az iki öğrencinin sigara içmeme olasılığı 2, 3 ve 4 öğrencinin sigara içmeme olasılıklarının toplamıdır.

b. En fazla 2 öğrencinin sigara içmeme olasılığı;

**$P(r \leq 2) = P(0) + P(1) + P(2)$** 'ye eşittir, yani en fazla 2 öğrencinin sigara içmeme olasılığı 0, 1 ve 2 öğrencinin sigara içmeme olasılıklarının toplamıdır.

c. 2 veya 3 öğrencinin sigara içmeme olasılığı  $P(2)$  ve  $P(3)$ 'ün toplamına eşittir.

### 3.1.2. Binom Dağılımına Göre Beklenen Frekanslar

Verilen örnekte araştırmacı bir kere 4 öğrenciyi rastgele seçtiği zaman yukarıdaki olasılıklar hesaplanır. Fakat rastgele 4 öğrenci seçme işlemi 200 kere tekrarlanmış ve aşağıdaki gibi bir frekans dağılımı tablosu oluşturulmuş olabilir. 4' er öğrencilik 200 örnek için sigara içmeyen öğrenci sayısı bakımından dağılımı (gözlenen frekanslar) Tablo 3.1'de verilmiştir. Araştırmacı, bu örnek için binom dağılımına göre beklenen frekansları da hesaplayarak aşağıdaki frekans dağılımı tablosunda vermiştir. Binom dağılımına göre beklenen frekanslar şu şekilde hesaplanmaktadır: 4 öğrencilik bir örnekte hepsinin sigara içme olasılığı 0.004 (1/256) ise 200 örnekten  $200(0.004)=0.8$  tanesinde 4 öğrencinin 4'ünde de sigara içmesi beklenir ki bu 1. sınıfın binom dağılımına göre beklenen frekansıdır. Aynı şekilde 2. sınıf için beklenen frekans: 4 öğrencilik bir örnekte 4 öğrenciden 1'inin sigara içmeme olasılığı 0.047 ise 200 örnekten  $200(0.047)=9.4$  tanesinde 4 öğrenciden 1 tanesinin sigara içmemesi beklenir. Bu şekilde diğer sınıflar içinde beklenen frekanslar hesaplanarak Tablo 3.1'deki frekans dağılımı tablosunda verilmiştir. Frekans dağılımı tablosunda da görüldüğü gibi beklenen frekansların toplamı da 200'dür.

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

TABLO 3.1. 4'er öğrencilik 200 örnek için gözlenen ve  $\pi=0.75$  olan binom dağılımına göre beklenen frekanslar.

Sigara içmeyen öğrenci sayısı (r)	Gözlenen frekans	Binom dağılımına göre beklenen frekans
0	2	0.8
1	12	9.4
2	45	42.2
3	80	84.4
4	61	63.2
Toplam	200	200

### 3.1.3. Binom Dağılımının Parametreleri

Binom dağılımının parametreleri "n" ve " $\pi$ " dir. Dağılımın ortalaması,  $\mu=n\pi$  ve varyansı,  $\sigma^2=n\pi(1-\pi)$ 'dir. Yukarıda verilen örnek için ortalama,  $\mu=4(3/4)=3$ 'tür. Yani 200 kere rastgele 4 öğrencinin bulunduğu örnekler oluşturulsa sigara içmeyen öğrenci sayısı bakımından ortalama 3'tür. Bu ortalama Tablo 3.1'de verilen frekans dağılım tablosunda teorik frekanslardan aşağıdaki şekilde de hesaplanabilir.

$$\mu = \frac{0 \times 0.8 + 1 \times 9.4 + 2 \times 42.2 + 3 \times 84.4 + 4 \times 63.2}{200} = 2.99 \cong 3$$

Rastgele oluşturulan 4 öğrencilik 200 örnekte sigara içmeyen öğrenci sayısı bakımından değişimin ölçüsü de varyanstır ve örneğimizde aşağıdaki şekilde hesaplanır:

$$\sigma^2 = 4 \left( \frac{3}{4} \right) \left( 1 - \frac{3}{4} \right) = 0.75$$

Gözlenen frekanslardan hesaplanan ortalama ise istatistiktir ve aşağıdaki gibidir:

$$\bar{X} = \frac{(0)(2) + (1)(12) + (2)(45) + (3)(80) + (4)(61)}{200} = 2.93 \text{ tür.}$$

Örnekten hesaplanan varyans ise 0.859'tur.

### 3.1.4. Binom Dağılımının Şekli

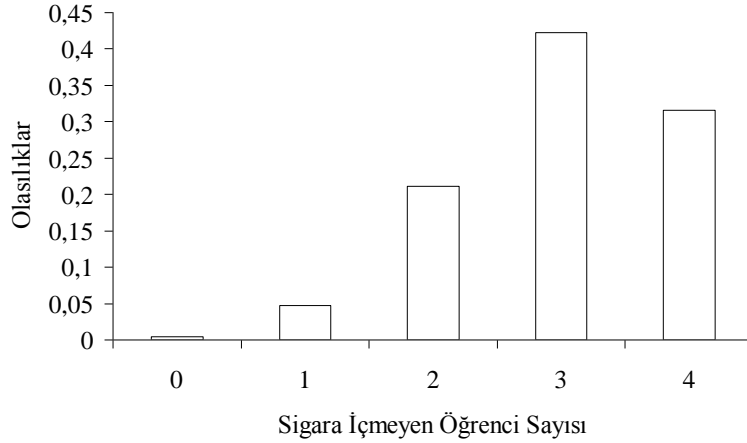
Binom dağılımının şekli "n" ve " $\pi$ "'ye bağlıdır. Deneme sayısı ne olursa olsun  $\pi=1/2$  için dağılımın şekli simetriktir. İstenen tip ve istenmeyen tip olayların oluş olasılığı farklı olduğu zaman dağılımın şekli değerlerden büyük olanın tarafına doğru yatıktır (çarpıktır). Eğer  $\pi > 1/2$  ise dağılım sol tarafa, fakat  $\pi < 1/2$  ise dağılım sağ tarafa yatıktır.  $\pi$  ve  $(1-\pi)$  değerlerinin eşit olmadığı durumlarda deneme sayısının (n'nin) artması dağılımın simetriye yaklaşmasını sağlar.

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Binom dağılımında sınıf sayısı deneme sayısının bir fazlasına (yani  $n+1$ 'e) eşittir, çünkü istenen olayın sıfır olması durumu da söz konusudur.

### ÖRNEK:

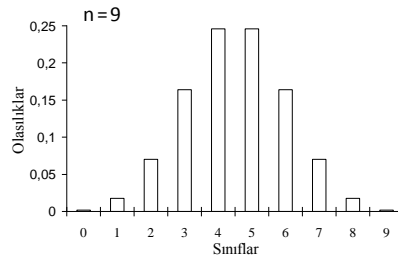
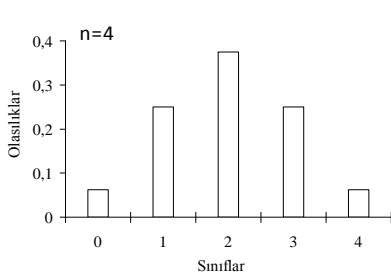
Yukarıda verilen ve sigara içen ve içmeyen öğrenciler ile ilgili binom dağılımının şekli Şekil 3.1'de verilmiştir.



ŞEKİL 3.1. Sigara içmeyen öğrencilerin sayılarına ait dağılım

Şekil 3.1'den de görüldüğü gibi  $\pi > 1/2$  olduğu zaman dağılım sola yatık (çarpık) olmaktadır.

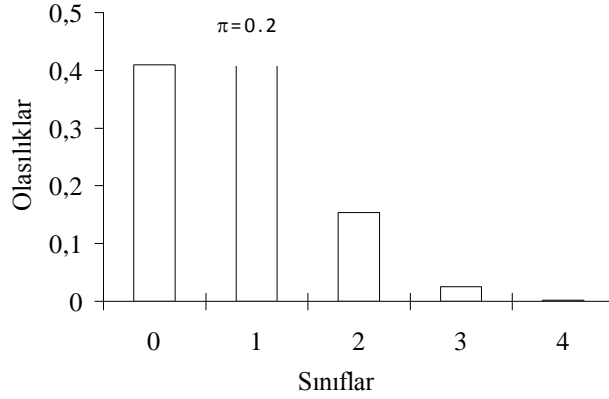
" $\pi$ " ve " $n$ "nin değişik değerlerine bağlı olarak binom dağılımın şekli aşağıdaki grafiklerde gösterilmiştir. Şekil 3.2,  $\pi=0.5$  ve çeşitli " $n$ " değerleri için binom dağılımlarının şeklini göstermektedir. Görüldüğü üzere deneme sayısına bağlı olmaksızın  $\pi=0.5$  olduğu zaman dağılım simetriktir. Şekil 3.3,  $\pi=0.2$  ( $n=4$ ) ve Şekil 3.4,  $\pi=0.8$  ( $n=4$ ) için binom dağılımlarının şeklini göstermektedir.



ŞEKİL 3.2.  $\pi=0.5$  ve deneme sayısı 4 ve 9 için binom dağılımının şekli

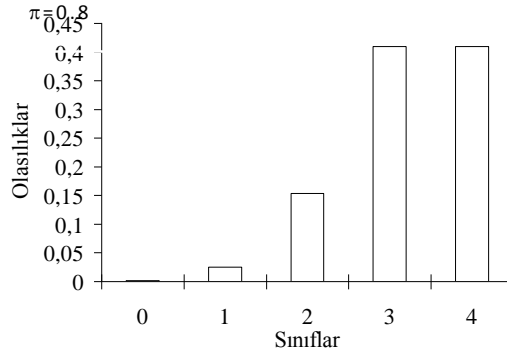
## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94



ŞEKİL 3.3.  $\pi=0.2$  ve  $n=4$  için binom dağılımının şekli

Şekil 3.3'de gösterildiği gibi  $\pi < 0.5$  olduğu zaman dağılım sağa yatık (çarpık) olmaktadır, yani istenen olayın  $n$  denmede  $r$  kere meydana gelme olasılığı  $r$  değeri arttıkça azalmaktadır



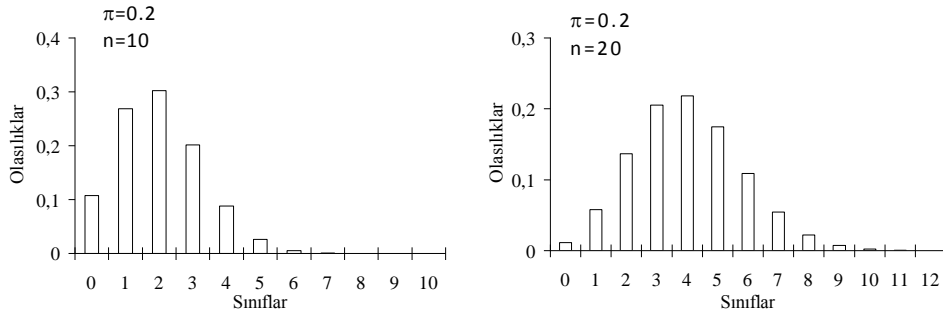
ŞEKİL 3.4.  $\pi=0.8$  ve  $n=4$  için binom dağılımının şekli

Şekil 3.4'te gösterildiği gibi  $\pi > 0.5$  olduğu zaman dağılım sola yatık (çarpık) olmaktadır, yani istenen olayın  $n$  denmede  $r$  kere meydana gelme olasılığı  $r$  değeri arttıkça artmaktadır.

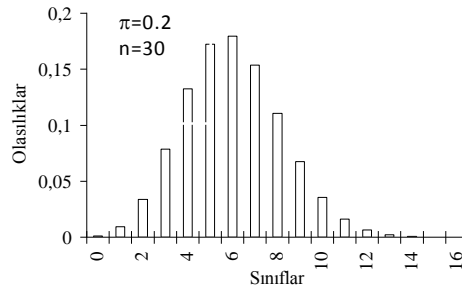
Şekil 3.5 ise istenen olayın herhangi bir değeri için deneme sayısının artması ile dağılımın şeklinin simetrikleştiğini göstermektedir. Şekil 2'de  $\pi=0.2$  ve  $n=4$  olduğu zaman dağılımın sağa yatık bir dağılım olduğu gösterilmişti. Aşağıdaki şekillerde ise  $\pi=0.2$  olduğu durumda deneme sayısının artması ile dağılımın şeklinin simetriye yaklaştığını göstermektedir.

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94



ŞEKİL 3.5.  $\pi=0.2$  için artan deneme sayısının binom dağılımının şekli üzerine etkisi



ŞEKİL 3.5. (devam).  $\pi=0.2$  için artan deneme sayısının binom dağılımının şekli üzerine etkisi

Şekil 3.5’de de görüldüğü üzere  $\pi=0.2$  olmasına rağmen deneme sayısı arttıkça dağılımın şekli simetrikleşmektedir. Bu  $n=10$ ,  $n=20$  ve  $n=30$  için gösterilmiştir. Burada gösterildiği gibi  $\pi$  değeri ne olursa olsun deneme sayısının artması ile dağılım simetriye yaklaşır.

**ÖRNEK:**

Çok sayıda kişinin katıldığı ve sınav sonuçlarının “BAŞARILI” ve “BAŞARISIZ” olarak açıklandığı bir sınavda tesadüfen seçilen 300 adaydan 180 adayın başarılı olduğu saptanmıştır. Bu sınava katılanlar arasından rastgele 6 kişi seçilse;

- Hepsinin başarılı olma olasılığı,
- 5 kişinin başarılı olma olasılığı,
- 4 kişinin başarılı olma olasılığı,
- 3 kişinin başarılı olma olasılığı,
- 2 kişinin başarılı olma olasılığı,
- 1 kişinin başarılı olma olasılığı,
- Hiçbirinin başarılı olma olasılığı nedir.



## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

300 kişinin katıldığı bir sınavda 180 kişi başarılı olduğuna göre başarılı olma (istenen olayın) olasılığı  $p = \frac{180}{300} = 0.6$  ve başarısız olma (istenmeyen olayın) olasılığı  $q=1-p=1-0.6=0.4$ 'tür. Bu durumda;

$$P(6) = C(6,6)(0.6)^6(0.4)^0 = 0.047$$

$$P(5) = C(6,5)(0.6)^5(0.4)^1 = 0.187$$

$$P(4) = C(6,4)(0.6)^4(0.4)^2 = 0.311$$

$$P(3) = C(6,3)(0.6)^3(0.4)^3 = 0.276$$

$$P(2) = C(6,2)(0.6)^2(0.4)^4 = 0.138$$

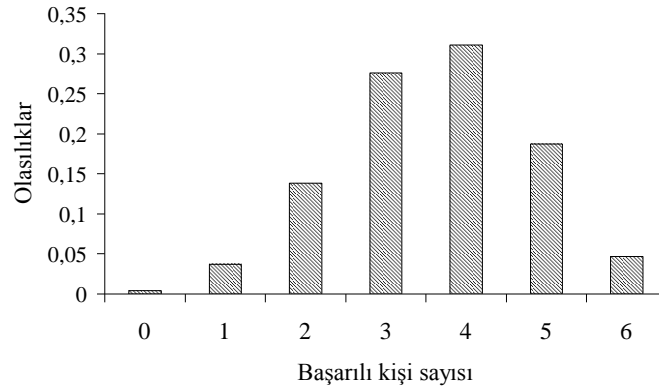
$$P(1) = C(6,1)(0.6)^1(0.4)^5 = 0.037$$

$$P(0) = C(6,0)(0.6)^0(0.4)^6 = 0.004$$

Deneme sayısı 6 olduğu zaman mümkün olan 7 hal vardır. Bu hallerin olasılıklarının toplamı 1'e eşittir;

$$P(0)+P(1)+P(2)+P(3)+P(4)+P(5)+P(6)=1 \text{ 'dir.}$$

Bu örnek için binom dağılımının şekli ise Şekil 3.6'da gösterilmiştir.



ŞEKİL 3.6. Başarılı kişilerin sayısına ait olasılıklar

Şekil 3.6'da verilen grafikten de görüldüğü üzere  $p=0.6$  olduğu için dağılımın şekli simetriye çok yakındır.

Sınava giren 300 kişi arasından geriye iadeli olarak tesadüfen seçilmiş 6'şar kişi bulunan 175 tane grup oluşturan bir araştırmacı başarılı kişi sayısı bakımından dağılımı ve binom dağılımına göre beklenen frekansları Tablo 3.2'deki frekans dağılımı tablosunu oluşturarak vermiştir.

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

TABLO 3.2. 6'şar bireylik 175 grupta gözlenen ve  $p=0.6$  olan binom dağılımına göre beklenen başarılı birey sayısı

Başarılı kişi sayısı (r)	Gözlenen frekans	Binom dağılımına göre beklenen frekans
0	1	0.700
1	9	6.475
2	25	24.150
3	47	48.000
4	51	54.425
5	35	32.725
6	7	8.225

Tablo 3.2'de verilen örnek için ortalama,  $\mu=6(0.6)=3.6$ 'dır. Rastgele seçilmiş 6'şar kişilik 175 grupta başarılı kişi sayısı bakımından ortalama 3.6'dır. Bu ortalama Tablo 3.2'de verilen frekans dağılım tablosundaki teorik frekanslar kullanılarak aşağıdaki şekilde de hesaplanabilir.

$$\mu = \frac{0 \times 0.70 + 1 \times 6.475 + 2 \times 24.15 + 3 \times 48.0 + 4 \times 54.425 + 5 \times 32.725 + 6 \times 8.225}{175} = \frac{629.45}{175} \cong 3.6$$

Rastgele oluşturulan 6 kişilik 175 örnekte başarılı kişi sayısı bakımından değişimin ölçüsü de varyanstır ve örneğimizde aşağıdaki şekilde hesaplanır:

$$\sigma^2 = 6(0.6)(1 - 0.6) = 1.44$$

### 3.2. Poisson Dağılımı

Poisson dağılımı da binom dağılımı gibi kesikli bir dağılımdır. Bu dağılım, çok az rastlanan fakat belirli (sabit) bir olasılıkla meydana gelen olayların dağılımıdır. Poisson dağılımı gösteren olaylara örnekler aşağıdaki gibi verilebilir:

- Belirli bir nüfus içinde 110 yaşına kadar yaşayan insanların sayısı,
- Belirli bir günde yanlış düşen telefon numaralarının sayısı,
- Bir günde başlayan savaş sayısı,
- Bir hayat sigortası tarafından sigortalanan belirli sayıda insandan belirli bir zaman aralığında ölenlerin sayısı,
- Çok az satılan bir maldan bir dükkânda bir günde satılanların sayısı,
- Bir kitaptaki yazım hatalarının sayısı,
- Çok sayıda birimden oluşan ambalajlanmış mallardaki bozuk olanların sayısı.

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Yukarıda örnek olarak verilen olayların meydana geliş sayıları Poisson dağılımı gösterirler. Bu olayların belirli bir sayıda meydana gelme olasılıkları Poisson dağılımı fonksiyonu kullanılarak hesaplanır.

Binom dağılımında “n” deneme sayısının çok büyümesi ve “π” istenen olayın oluş olasılığının çok küçük olması durumlarında da Poisson dağılım fonksiyonu kullanılır. Çünkü bu durumda istenen olayın belirli bir sayıda oluş olasılığının binom fonksiyonu kullanılarak hesaplanması zorlaşır.

### 3.2.1. Olayların Oluş Olasılıklarının Hesaplanması

Poisson dağılımının fonksiyonu Fransız matematikçisi Poisson tarafından bulunan aşağıdaki eşitliğe göre belirlenir:

$$P(r) = \frac{\mu^r}{r!} e^{-\mu} \quad \dots(3.2)$$

Formülde,  $\mu$  dağılımın ortalaması, r nadir olarak görülen olayın meydana geliş sayısı, e tabii logaritma tabanı olup  $e \approx 2.718$ 'dir. Araştırmacı istenen olayın herhangi bir sayıda görülme olasılığını hesapladıktan sonra aşağıda gösterildiği şekilde olayın birbirini izleyen sayıda görülme olasılıklarını hesaplayabilir.

$$P(0) = \frac{\mu^0}{0!} e^{-\mu} = e^{-\mu}$$

$$P(1) = \frac{\mu^1}{1!} e^{-\mu} = \mu P(0)$$

$$P(2) = \frac{\mu^2}{2!} e^{-\mu} = \frac{\mu}{2} \frac{\mu}{1} e^{-\mu} = \frac{\mu}{2} P(1)$$

$$P(3) = \frac{\mu^3}{3!} e^{-\mu} = \frac{\mu}{3} \frac{\mu}{2!} e^{-\mu} = \frac{\mu}{3} P(2)$$

$$P(n) = \frac{\mu^n}{n!} e^{-\mu} = \frac{\mu}{n} P(n-1)$$

#### ÖRNEK 1:

Bir hayat sigortası bir yılda 5000 kişiyi kazalara karşı sigortalamış olsun. Bir yılda sigortalanmış kişilerden herhangi birinin kazadan ölme olasılığı 0.001 olarak bilindiğine göre bir yıl boyunca 4 kişinin kazadan ölme olasılığı nedir.

Bu olasılık binom fonksiyonu kullanılarak hesaplanacak olursa bu  $P(r=4) = C(5000,4)\pi^4(1-\pi)^{5000-4}$ ’dür. Bunun hesaplanması zordur. Hâlbuki Poisson dağılımının fonksiyonu kullanılarak kısaca hesaplanabilir.

Bu örnek için ortalama,  $\mu = n\pi$ ,  $5000(0.001) = 5$ ’tir. 4 kişinin kazadan ölme olasılığı ise;

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

$$P(4) = \frac{\mu^4}{4!} e^{-5} = \frac{5^4 (2.718)^{-5}}{24} = \frac{625(0.0067)}{24} = 0.1745$$

**ÖRNEK 2.**

100 kişilik bir sınıfta 1 Ocak'ta doğanların sayısı Poisson dağılımı gösterir. Herhangi bir kişinin 1 Ocak'ta doğmuş olma olasılığı  $p=1/365$ 'tir. Bu dağılımın ortalaması,

$$\mu = np = \frac{100}{365} \cong 0.2740. \quad 3 \text{ kişinin 1 Ocak'ta doğmuş olma olasılığı;}$$

$$P(3) = \frac{0.2740^3}{3!} e^{-0.2740} = \frac{0.02057(2.718)^{-0.2740}}{6} = \frac{0.02057(0.76035)}{6} = 0.002606$$

**ÖRNEK 3:**

300 sayfalık bir kitabın hazırlanması sırasında 150 yazım hatası yapılmış ise sayfa başına ortalama hata  $\mu=150/300=1/2$ 'dir. Bir sayfada hiç hata yapılmamış olma ihtimali;

$$P(0) = \frac{0.5^0}{0!} 2.718^{-0.5} \cong \frac{1}{\sqrt{2.718}} = 0.6066$$

300 sayfalık bu kitabın kaç sayfasında hiç hata yapılmadığı hesaplanmak istenirse bunun için de hesaplanan olasılık ile sayfa sayısı çarpılır ve  $300(0.6066) \cong 182$  olarak bulunur.

**3.2.3. Poisson Dağılımının Parametresi ve Dağılım Şekli**

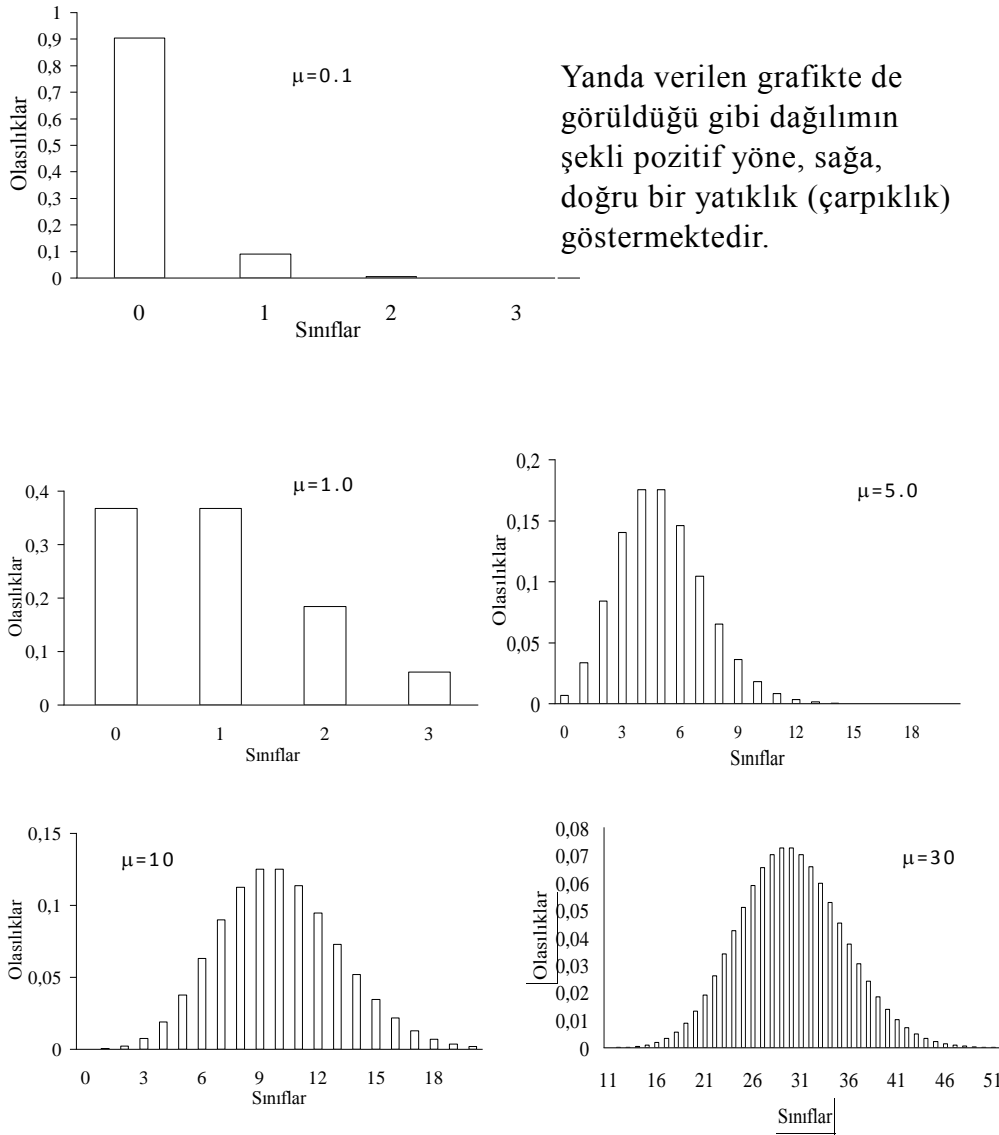
Poisson dağılımları 3.2 numaralı eşitlikte verilen fonksiyondan da görüldüğü gibi birbirlerinden ortalamaları ile ayrılır. Yani fonksiyonun tek parametresi dağılımın ortalaması ( $\mu$ )'dür. Poisson dağılımında ortalama ve varyans birbirine eşittir ve aşağıdaki şekilde hesaplanır:

$$\mu = \sigma^2 = n \cdot \pi$$

Eşitlikte,  $n$  birey sayısı,  $\pi$  ise olayın oluş olasılığıdır.

Poisson dağılımının şekli ortalamaya bağlıdır. Şekil 3.7 farklı ortalamalara sahip Poisson dağılımlarının şeklini göstermektedir.

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı, Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94



ŞEKİL 3.7. Farklı ortalamalara sahip Poisson dağılımları

Şekil 3.7’de verilen grafiklerde de görüldüğü gibi Poisson dağılımının şekli ortalaması büyüdükçe simetriye yaklaşmaktadır.

#### ÖRNEK 4.

Hastaneler için imal edilen ve içinde 300 tablet bulunan ambalajlardan 200 adet rastgele alınmış ve bunlardaki kırık tablet sayısı belirlenmiştir. Ambalajda bulunan tabletler ya kırıktır veya sağlamdır. Üzerinde durulan değişkenin iki hali söz konusudur. Kırık tablet oranı çok düşüktür. Kırık tabletlerin binom dağılımı gösterdiği düşünülebilir. Ancak kırık tablet oranı çok düşüktür, n ise 300 gibi büyük bir sayıdır. Bu durumda Poisson dağılımı uygulanabilir. Yukarıda belirtildiği gibi teorik olarak Poisson dağılımında ortalama varyansa eşittir. Örnekte eşit olmayabilir. Çekilen örnek sayısı sonsuz olduğu zaman eşit olur. Herhangi bir örnekte değişken kesikli, n sayısı çok ve

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

ortalama ile varyans birbirine yakın ise Poisson dağıldığı düşünülebilir. Aşağıdaki örnekte (Tablo 3.3) önce ortalama ve varyans hesaplanmıştır.

TABLO 3.3. İçinde 300 tablet bulunan 200 ambalajın kırık tablet bakımından dağılımı

Kırık tablet sayısı	$f_i$	$f_i X_i$	$f_i X_i^2$
0	56	0	0
1	77	77	77
2	40	80	160
3	20	60	180
4	6	24	96
5	1	5	25
Toplam	200	246	538

$$\bar{X} = \frac{246}{200} = 1.23$$

$$\sum d_x^2 = 538 - \frac{(246)^2}{200} = 235.42$$

$$S^2 = \frac{235.42}{199} = 1.183$$

Ortalama ve varyans birbirine yakındır. Bu örneğin ortalaması 1.23 olan bir Poisson dağılımı gösterdiği varsayılarak teorik frekanslar,  $f'_i$ , hesaplanabilir. Sonuçlar Tablo 3.4'de özetlenmiştir.

TABLO 3.4. 300 tablet bulunan 200 ambalaj için gözlenen ve Poisson dağılımına göre beklenen frekanslar

Sınıflar	Kırık tablet sayısı	$f_i$	$f'_i$
1	0	56	58.46
2	1	77	71.91
3	2	40	44.22
4	3	20	18.13
5	4	6	5.58
6	5	1	1.70

Birinci sınıfta yani herhangi bir ambalajda hiç kırık tablet bulunmama olasılığı;

$$P(0) = \frac{1.23^0}{0!} e^{-1.23} = 0.2923$$

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Birinci sınıfta beklenen teorik frekans  $f_1' = (200)(0.2923) = 58.46$  olarak bulunur. En son sınıfta 5 ve daha fazla kırık tableti kapsar. Bu sınıfın teorik frekansı bulunurken bundan önceki sınıfların teorik frekansları toplamı olan 198.30'un 200'den farkı alınır yani,  $f_6' = 200 - (58.46 + 71.91 + 44.22 + 18.13 + 5.58) = 1.70$  olarak bulunur.

Görüldüğü gibi teorik frekanslar ile gözlenen frekanslar arasında farklar vardır. Bunların tesadüften ileri gelip gelmediğinin testi ilerideki konularda görülecektir.

### 3.3. Normal Dağılım

Normal dağılım, binom ve Poisson dağılımlarının aksine sürekli bir dağılımdır. Aralarındaki farklılıklar tesadüften ileri gelen gözlem değerlerinin oluşturduğu populasyonların dağılımı normal dağılıma uygunluk gösterir. Örneğin, aynı koşullarda yetişen belirli yaş grubunda ve cinsiyetteki sağlıklı öğrencilerin ağırlıkları arasındaki farklılık her öğrenciye etkisi rastgele ve küçük olan çok sayıda etkenden (faktörden) ileri gelir. Bu gibi etkiler altında oluşan özellikler normal dağılım gösterir. Bu nedenle ele alınan öğrenciler ağırlık bakımından bir **normal populasyonu** temsil ederler.

Normal populasyonlarda dağılım:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \dots(3.3)$$

olasılık yoğunluk fonksiyonu ile ifade edilir. Normal dağılım fonksiyonunun iki parametresi vardır. Bunlar ortalama,  $\mu$ , ve standart sapma,  $\sigma$ 'dır. Fonksiyondaki  $e$  ve  $\pi$  sabit değerlerdir:  $e \approx 2.718$ , yani tabii logaritma tabanı ve  $\pi = 3.1416$ 'dır. Bir normal dağılımın ortalama ve standart sapması biliniyorsa yukarıda verilen olasılık fonksiyonu kullanılarak belirli değerler arasındaki olasılıklar hesaplanabilir. Örneğin, öğrencilerin ağırlıkları ortalaması,  $\mu = 60$  kg ve standart sapması  $\sigma = 5$  kg olan bir normal dağılım gösteriyorsa bu populasyonda öğrencilerin ağırlıklarının 65 ile 69 kg arasında olma olasılığı hesaplanmak istenirse normal dağılım olasılık fonksiyonunun bu aralıkta integralinin alınması gerekir:

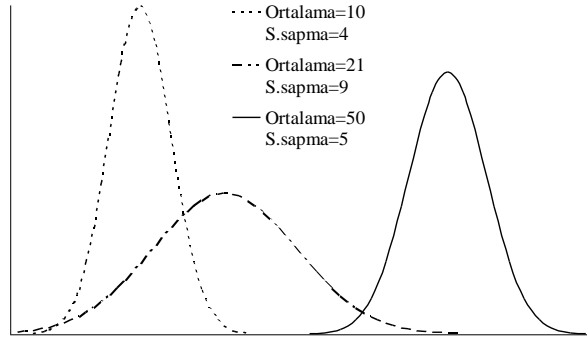
$$\int_{65}^{69} f(x) dx = \int_{65}^{69} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-60}{5}\right)^2} dx$$

Bu integralin alınması ile bulunacak değer öğrencilerin 65 ile 69 kg arasında olma olasılığını verecektir. Başka bir deyişle, ağırlıkları 65 ile 69 kg arasında olan öğrencilerin % miktarını verecektir.

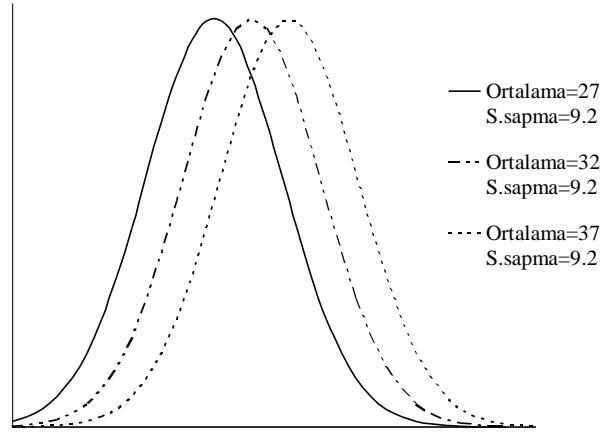
Normal dağılım fonksiyonundan da görüldüğü gibi normal dağılımlar ortalama ve standart sapmaları ile birbirlerinden ayrılırlar. Şekil 3.8 ortalama ve standart sapmaları birbirinden farklı normal dağılımları göstermektedir. Şekil 3.9 ise standart sapmaları aynı fakat ortalamaları farklı olan normal dağılımları göstermektedir. Şekil 3.10'da ortalamaları aynı fakat standart sapmaları farklı olan normal dağılımlar verilmiştir.

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

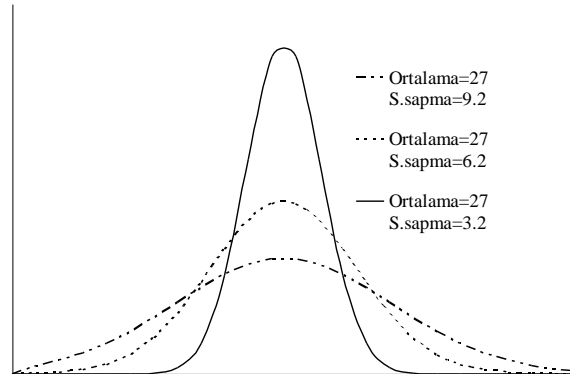


ŞEKİL 3.8. Parametreleri birbirinden farklı normal dağılımlar



ŞEKİL 3.9. Standart sapmaları aynı fakat ortalamaları birbirinden farklı normal dağılımlar

Şekil 3.9'da görüldüğü gibi normal dağılımlar aynı standart sapmaya, fakat farklı ortalamalara sahip olabilir. Bu durumda dağılımları oluşturan gözlem değerleri arasındaki varyasyon aynı fakat ortalamaları farklıdır.



ŞEKİL 3.10. Ortalamaları aynı fakat standart sapmaları birbirinden farklı normal dağılımlar



## 7. HAFTA DERS NOTLARI

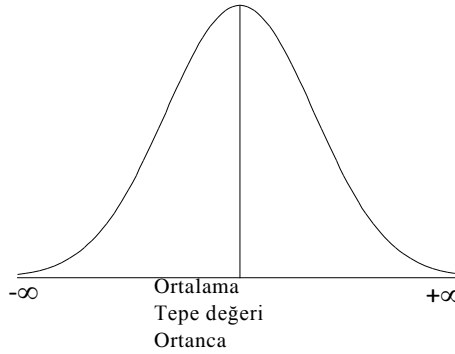
Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Şekil 3.10'da görüldüğü gibi normal dağılımlar aynı ortalamaya, fakat farklı standart sapmalara sahip olabilir. Bu durumda dağılımları oluşturan gözlem değerlerinin ortalamaları aynı olmasına karşın standart sapma arttıkça aralarındaki varyasyon artacaktır.

### 3.3.1. Normal Dağılımın Özellikleri

Normal dağılımın özellikleri aşağıdaki şekilde sıralanabilir:

1. Normal dağılım çan eğrisi şeklindedir (Şekil 3.11),
2. Normal dağılım ortalama etrafında simetriktir. Ortalama, tepe değeri ve ortanca değer aynıdır (Şekil 3.11).



ŞEKİL 3.11. Normal dağılım şekli

3. Normal dağılımın simetrik olması sebebi  $\mu$  ile  $\mu+c$  bulunan bireylerin %'si,  $\mu$  ile  $\mu-c$  arasında bulunan bireylerin %'sine eşittir.

4. Normal dağılım sürekli bir dağılım olduğu için  $-\infty$ 'dan  $+\infty$ 'a kadar bütün değerleri kapsar, yani  $x$ 'in  $-\infty$  ile  $+\infty$  arasında bulunma ihtimali 1'dir.

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx = 1.0$$

### 3.3.2. Standart Normal Dağılım

Ortalaması ve standart sapması bilinen bir normal dağılımda herhangi bir  $X$  değerinin belir sınırlar arasında olma olasılığı hesaplanmak istenirse olasılık fonksiyonun söz konusu sınırlar arasında integralinin alınması gerektiği yukarıda belirtilmişti. Bu olasılığın hesaplanması aşağıdaki integralin alınmasını gerektirir. Bu hem çok zaman gerektirir hem de istatistiği kullanmak durumunda olan herkes bu ölçüde matematik bilmek zorunda değildir.

$$\int_a^b f(x) dx = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx$$

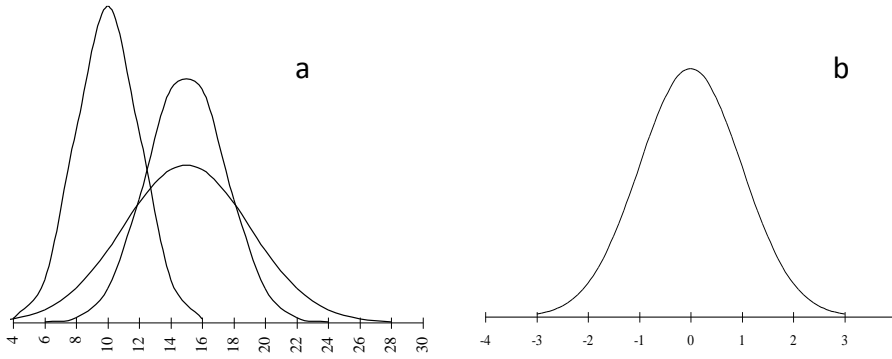
## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

Parametreleri farklı normal dağılımları oluşturan gözlem değerlerinden ortalamaları çıkarılıp standart sapmalarına bölünerek standardize edilebilir, standardize edilmiş değerler (Z değerleri) aşağıdaki şekilde hesaplanır:

$$Z_i = \frac{X_i - \mu}{\sigma}$$

Bu durumda ortalaması ve standart sapmaları farklı olan normal dağılımlar (Şekil 3.12 a), ortalamaları ve standart sapmaları aynı olan değerlere dönüştürülmüş olur (Şekil 3.12 b). Standardize edilmiş değerlerin dağılımına "**Standart Normal**" dağılımı denir.



ŞEKİL 3.12. Parametreleri farklı normal dağılımlar ve standart normal dağılım

Standart normal dağılımın olasılık yoğunluk fonksiyonu;

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

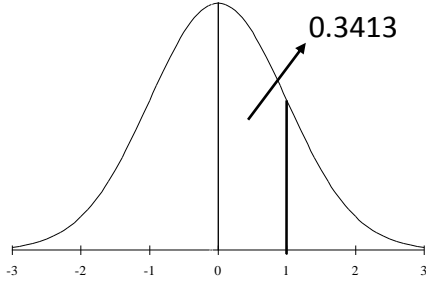
şeklindedir. Fonksiyonda görüldüğü gibi normal dağılımın olasılık yoğunluk fonksiyonunda olduğu gibi X değerleri değil, X değerinin ortalama ve standart sapma bakımından standartlaştırılmış değeri, Z, kullanılmıştır. Z-dağılımı da olarak isimlendirilen standart normal dağılımın ortalaması 0 ve standart sapması 1'dir. Bu dağılım da ortalaması etrafında simetriktir, yani ortalama ile +Z arasında olma olasılığı, ortalama ile -Z değeri arasında olma olasılığına eşittir. Ortalamadan küçük Z değerlerinin oranı %50 (yani 0.5) ve ortalamadan büyük Z değerlerinin oranı da %50 (yani 0.5)'dir. Bu dağılıma dahil olan bütün Z değerlerinin oluş olasılıkları toplamı 1'e eşittir. Parametresi bilinen ve normal dağılım gösteren X değerleri standardize edilerek Z değerlerine çevrilebilirler. Z dağılımında ortalama (0 ile) belirli Z değerleri arasında kalan alanlar, olasılık fonksiyonunun integrali alınarak hesaplanmış ve tablo halinde düzenlenmiştir (TABLO A). Bu tablo kullanılarak kolayca ortalama ile belirli Z değerleri arasındaki alan bulunabilir.

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

**ÖRNEK 1:**

Z değerlerinin 0 ile 1.0 değerleri arasında olanların oranı nedir?

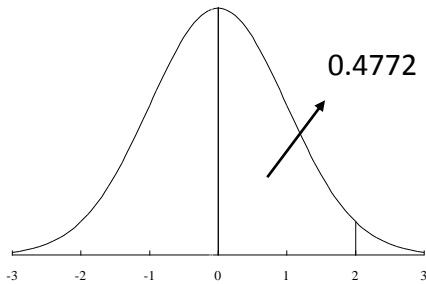


Tablo A'dan bakıldığı zaman 0 ile +1 arasındaki alan 0.3413'tür. Yani sıfır ile +1 arasındaki bütün Z değerleri, dağılımı oluşturan Z değerlerinin %34.13 kadardır. Başka bir deyişle; standart normal popülasyonda herhangi bir Z değerinin sıfır ile 1 arasında olma olasılığı %34.13'tür.

Normal popülasyonlarda gözlem değerlerinin %34.13'ü ortalama ile bundan 1 standart sapma uzaklıktaki nokta arasında olduğu söylenebilir.

**ÖRNEK 2:**

Z değerlerinin 0 ile 2.0 değerleri arasında olanların oranı nedir.



Tablo A'dan bakıldığı zaman 0 ile +2 arasındaki alan 0.4772'dir. Yani sıfır ile +2 arasındaki bütün Z değerleri, dağılımı oluşturan Z değerlerinin %47.72 kadardır. Normal popülasyonda gözlem değerlerinin %47.72'si ortalama ile bundan 2 standart sapma uzaklıktaki nokta arasında olduğu söylenebilir.

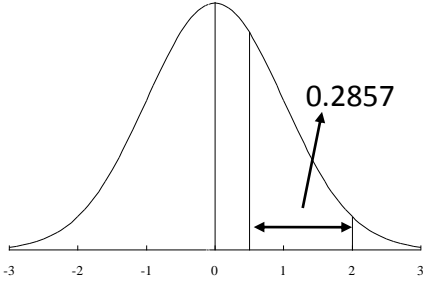
**ÖRNEK 3:**

Standart normal dağılımda tüm Z-değerlerinin % ne kadarı 0.5 ile 2 arasındadır. Bu örnekte istenen ihtimal:

$$P(0.5 < Z < 2.0) = P(0 < Z < 2.0) - P(0 < Z < 0.5)$$

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94



Bu olasılığı hesaplamak için Z değerlerinin 0 ile 2.0 arasında olma olasılığından Z değerlerinin 0 ile 0.5 arasında olma olasılığı çıkarılır. Tablo A'dan bakıldığı zaman  $P(0 < Z < 2.0) = 0.4772$  ve  $P(0 < Z < 0.5) = 0.1915$ 'dir. Bu durumda istenen olasılık:

$$0.4772 - 0.1915 = 0.2857 \text{ 'dir.}$$

**ÖRNEK 4:**

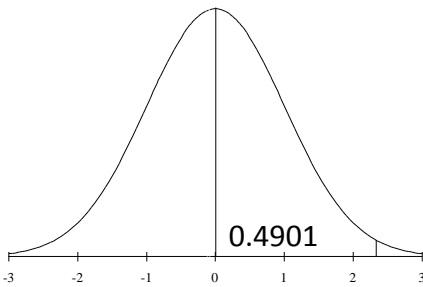
Bir doğum evinde doğan bebeklerin ağırlıklarının ortalamasının 3.2 kg ve standart sapmasının ise 0.3 kg olan bir normal dağılım gösterdiği bilinmektedir. Buna göre;

- Bu doğum evinde 3.2 kg ile 3.9 kg arasında doğan bebeklerin nisbi miktarı nedir?
- Bu doğum evinde 3.2 kg'dan daha hafif doğan bebeklerin nisbi miktarı nedir?
- Bu doğum evinde 2.8 kg ile 3.6 kg arasında doğan bebeklerin nisbi miktarı nedir?
- Bir günde ortalama 200 bebeğin doğduğu kabul edilirse bu bebeklerden kaç tanesinin ağırlığı 4 kg'dan daha fazladır?
- Bebeklerden en ağır %2.5'ini sınırlayan alt değer nedir?
- Bebeklerden en hafif %5'i hangi ağırlıktan daha düşüktür?

Yukarıda istenen olasılıkların hesaplanabilmesi için her X değerinin standardize edilerek bunlara karşılık gelen Z değerinin hesaplanması gerekir.

**a.** Burada istenen olasılık  $P(3.2 < X < 3.9)$ 'tur. Bunun için X değerleri standardize edilerek Z değerleri hesaplanır ve Tablo A kullanılarak söz konusu olasılık aşağıdaki şekilde bulunur.

$$\begin{aligned} P(3.2 < X < 3.9) &= P\left(\frac{3.2 - 3.2}{0.3} < Z < \frac{3.9 - 3.2}{0.3}\right) \\ &= P(0 < Z < 2.33) \text{ 'dır.} \end{aligned}$$



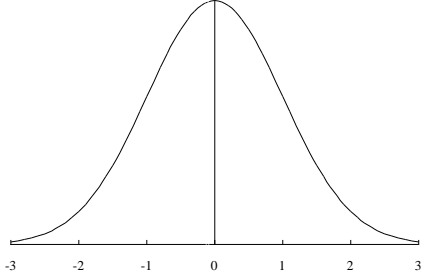
Tablo A'dan  $P(0 < Z < 2.33) = 0.4901$  olarak bulunur. Yani bu doğum evinde doğan bebeklerin %49.01'inin ağırlığı 3.2 kg ile 3.9 kg arasındadır.

**b.** Bu şıkta istenen doğan bebeklerin ağırlığının 3.2 kg'dan daha hafif olma olasılığı, yani  $P(X < 3.2)$ 'dir. Burada da yine 3.2 değerine karşılık gelen Z değeri

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

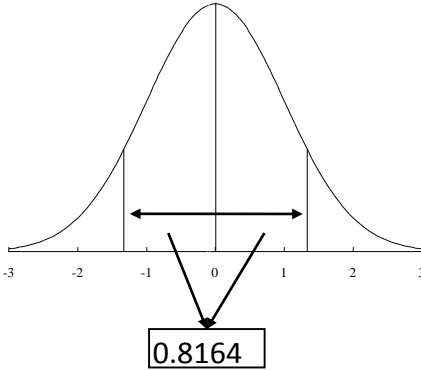
bulunmalıdır. Bu değer a şıkında  $Z=0$  olarak çıkmıştır. Burada istenen olasılık  $Z$  değerlerinin 0'dan küçük olma olasılığıdır.



$P(Z < 0) = 0.5$  yani standart normal dağılımda  $Z$  değerlerinin 0'dan küçük olma olasılığı %50'dir. Başka bir deyişle, doğan bebeklerin 3.2 kg'dan daha hafif olma olasılığı %50'dir.

**c.** Doğan bebeklerin 2.8 kg ile 3.6 kg arasında olanların nisbi miktarı  $P(2.8 < X < 3.6)$ 'dir. İlk olarak yapılması gereken bu değerlerine karşılık gelen  $Z$  değerlerinin hesaplanmasıdır.

$$\begin{aligned} P(2.8 < X < 3.6) &= P\left(\frac{2.8 - 3.2}{0.3} < Z < \frac{3.6 - 3.2}{0.3}\right) \\ &= P(-1.33 < Z < 1.33) \end{aligned}$$



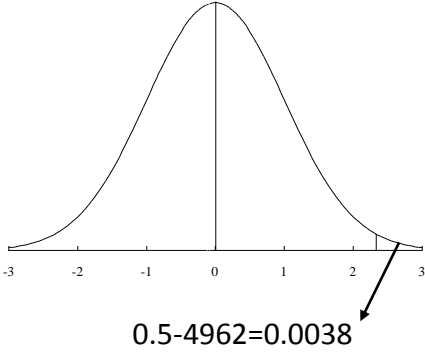
Burada istenen -1.33 ile 0 ve 0 ile 1.33 arasındaki  $Z$  değerlerinin nisbi miktarlarının toplamıdır. Başka bir deyişle  $Z$  değerlerinin -1.33 ile 0 arasında olma olasılığı ile  $Z$  değerlerinin 0 ile 1.33 arasında olma olasılıklarının toplamıdır. Tablo A'dan;  
 $P(-1.33 < Z < 0) = 0.4082$  ve  
 $P(0 < Z < 1.33) = 0.4082$ 'dir ve istenen olasılığı  $0.4082 + 0.4082 = 0.8164$ 'tür. Yani bu doğum evinde doğan bebeklerin %81.64'ünün doğum ağırlığı 2.8 kg ile 3.6 kg arasındadır.

**d.** Bu şıkta bir günde doğan 200 bebekten kaç tanesinin doğum ağırlığının 4 kg'dan daha fazla olduğu isteniyor. Bu durumda ilk olarak doğan bebeklerden % kaçının doğum ağırlığının 4 kg'dan daha fazla olduğunun hesaplanması gerekir.

$$P(X > 4) = P\left(Z > \frac{4 - 3.2}{0.3}\right) = P(Z > 2.67)$$

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

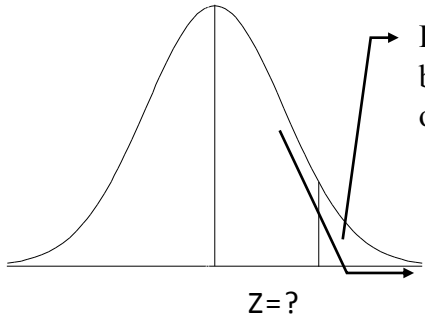


Burada istenen olasılık;

$$0.5-P(0 < Z < 2.67) = 0.5 - 0.4962 \\ = 0.0038 \text{ dir.}$$

Yani bir günde doğan 200 bebeğin %0.38'inin ağırlığı 4 kg'dan daha fazladır. Bu durumda doğum ağırlığı 4 kg'dan daha fazla olan bebek sayısı  $200 \cdot 0.0038 = 0.76$ , yaklaşık 1'dir.

**e.** Bu şıkta sorulan doğan bebeklerin en fazla ağırlığa sahip olan %2.5'uğunun hangi ağırlığın üstünde olduğudur. Bu ağırlığın hesaplanabilmesi için standart normal dağılım tablosunun tersten kullanılması gerekmektedir, yani 0 ile belirli bir Z değerinin arasındaki alan değil, 0.4750'lik alanın 0 ile hangi Z değeri arasında olduğuna bakılacaktır. Bu Z değeri bulunduktan sonra  $Z = \frac{X - \mu}{\sigma}$  eşitliğinde dağılım parametreleri ve Z değeri yerine konarak bu Z değerine karşılık gelen X değerinin hesaplanması gerekir.



En ağır doğum ağırlığına sahip bebeklerin %2.5'uğunun oluşturduğu alan

Doğum ağırlığı ortalama ile %2.5'lik en ağır doğum ağırlığına sahip bebekler arasındaki alan %47.5'tur. Bu durumda standart normal dağılım tablosundan 0 ile hangi Z değeri arasındaki alanının %47.5 olduğunun bulunması gerekir. Tablo A'ya

bakıldığı zaman 0 ile  $Z=1.96$  arasında kalan alanının 0.4750 olduğu görülüyor. Bu durumda %2.5'lik alanın başladığı ağırlığı bulmak için;  $Z = \frac{X - \mu}{\sigma}$  eşitliğinde Z değeri ve dağılımın parametreleri yerine konarak X değeri aşağıdaki şekilde hesaplanır:

$$1.96 = \frac{X - 3.2}{0.3} \text{ ise}$$

$$X = 1.96(0.3) + 3.2 = 3.788 \text{ kg}$$

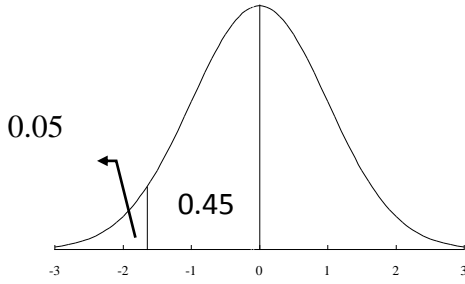
olarak bulunur.

**f.** Bu şıkta ise en az doğum ağırlığına sahip bebeklerin %5'nin hangi doğum ağırlığından başladığı sorulmaktadır. Burada da yine standart normal dağılım tablosu

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

tersinden kullanılmalıdır ve 0 ile hangi Z değeri arasındaki alanın %45 olduğu bulunmalıdır. Tablo A'ya bakıldığı zaman 0 ile  $Z=1.64$  arasındaki alanın 0.4495 ve 0 ile  $Z=1.65$  arasındaki alanın ise 0.4505 olduğu görülür. 0.4495 ile 0.4505'in ortalaması alınırsa 0.45 değeri elde edilir. 0.45 değerine karşılık gelen Z değerini bulmak için de 1.64 ile 1.65'in ortalaması alınırsa 1.645 bulunur yani 0 ile  $Z=1.645$  arasındaki alan 0.45'tir.



Burada en az ağırlığa sahip bebeklerin %5'nin hangi ağırlıktan başladığı sorulduğu için bizi ilgilendiren ortalamadan küçük değerlerin bulunduğu taraftır. Fakat yukarıda da belirtildiği gibi dağılım simetrik olduğu için 0 ile 1.645 arasındaki

alan ile -1.645 ile 0 arasındaki alan birbirine eşittir. Bu durumda %5.0'lik alanın başladığı ağırlığı bulmak için;  $Z = \frac{X - \mu}{\sigma}$  eşitliğinde  $Z = -1.645$  ve dağılımın parametreleri yerine konarak X değeri aşağıdaki şekilde hesaplanır.

$$-1.645 = \frac{X - 3.2}{0.3} \text{ ise}$$

$$X = -1.645(0.3) + 3.2 \cong 2.707 \text{ kg}$$

olarak bulunur.

### 3.3.3. Frekans Dağılım Tablosunda Normal Dağılıma Göre Olması Beklenen Frekansların Hesaplanması

Üzerinde çalışılan konu ile ilgili bilgilerin toplanması için seçilen örnekten veriler elde edildikten sonra araştırmacı için frekans dağılım tablosunu oluşturarak örneğin ortalaması ve standart sapmasını hesaplamış olsun. Frekans dağılım tablosunu oluşturan sınıflar için, ortalaması ve standart sapması örnekten hesaplanan gibi olan bir normal dağılıma göre her sınıfta beklenen frekansları hesaplanabilir.

Frekans dağılımı tablosunda normal dağılıma göre beklenen frekansların nasıl hesaplanacağı I. BÖLÜM'de 120 bebeğin doğum ağırlığı için düzenlenen frekans dağılımı tablosu kullanılarak açıklanacaktır. Bu örnek için ortalama 3.414 ve standart sapma da 0.2152 olarak hesaplanmıştı. Örnekten hesaplanan bu değerler parametre yerine kullanılarak, ortalaması 3.414 ve standart sapması 0.2152 olan normal dağılıma göre beklenen frekanslar şu şekilde hesaplanır: Tablo 3.5'deki frekans dağılımı tablosunda da görüleceği gibi **birinci sınıf 2.945** üst gerçek sınırında bitmektedir. Bu sınıf 2.945'ten daha küçük değere sahip gözlem değerlerini kapsar. Bu durumda söz konusu normal dağılımda gözlem değerlerinin % ne kadarının 2.945'ten daha küçük olduğunun hesaplanması gerekir. Bu olasılığı ( $P(X < 2.945)$ ) hesaplayabilmek içinde ilk olarak 2.945 değerine karşılık gelen Z-değeri bulunur.

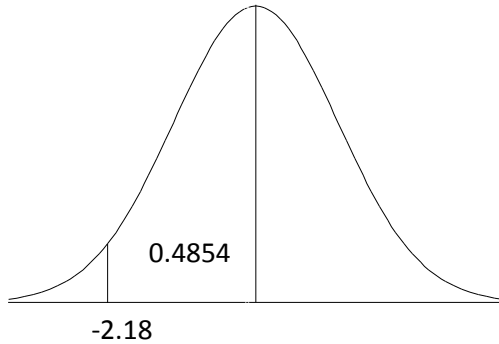
## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

TABLO 3.5. 120 bebeğin doğum ağırlığına ait gözlenen ve ortalaması 3.414 ve standart sapması 0.2152 olan normal dağılıma göre beklenen frekanslar

Gerçek üst sınır	Olasılık	Frekans	Beklenen frekans
2.945	0.0146	1	1.750
3.045	0.0290	2	3.480
3.145	0.0620	9	7.440
3.245	0.1092	15	13.104
3.345	0.1597	19	19.164
3.445	0.1812	24	21.744
3.545	0.1734	19	20.808
3.645	0.1286	14	15.432
3.745	0.0805	8	9.660
3.845	0.0390	5	4.680
3.945	0.0160	3	1.920
4.045	0.0068	1	0.816
Toplam		120	119.998

$$P(X < 2.945) = P\left(Z < \frac{(2.945 - 3.414)}{0.2152}\right) = P(Z < -2.18) \text{ bulunur.}$$



Bu normal dağılımda gözlem değerlerinin 2.945'ten veya standart normal dağılımda  $Z = -2.18$ 'den daha küçük değere sahip olma olasılığı;

$0.5 - 0.4854 = 0.0146$ 'dir. Yani tüm gözlemlerin (120 gözlem değerinin)

%1.46'sı bu sınıftadır. Bu durumda 1. sınıf için normal dağılıma göre beklenen frekans  $120(0.0146) \cong 1.75$ 'tir.

**İkinci sınıfta** ise doğum ağırlığı 2.945 ile 3.045 kg arasında olan bebekler vardır. Bu sınıf için söz konusu normal dağılıma göre olması beklenen frekansların hesaplanabilmesi için  $P(2.945 < X < 3.045)$  olasılığının bulunması gerekir.



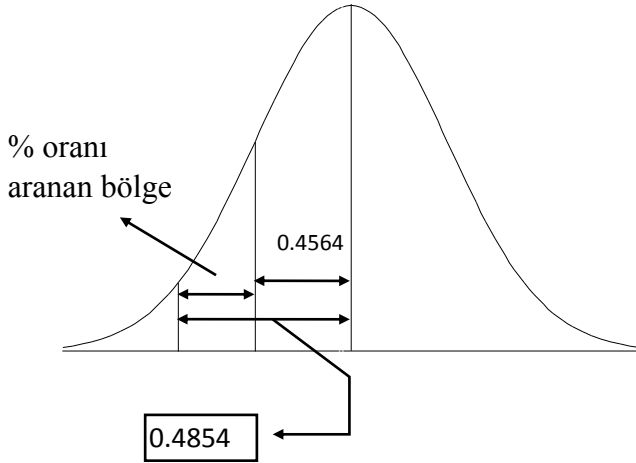
## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94

$$P(2.945 < X < 3.045) = P\left(\frac{2.945 - 3.414}{0.2152} < Z < \frac{3.045 - 3.414}{0.2152}\right)$$

$$= P(-2.18 < Z < -1.71)$$

Bu olasılığın hesaplanabilmesi için Z değerlerinin -2.18 ile 0 arasında olma olasılığından Z değerlerinin -1.71 ile 0 arasında olma olasılığının çıkarılması gerekir. Yani aşağıdaki grafik üzerinde gösterilen alan hesaplanacaktır.



Şöyle ki;  
Gözlem değerlerinin 2.945 ile 3.045 arasında olma olasılığı:  
 $0.4854 - 0.4564 = 0.029$ . Yani gözlem değerlerinin %2.9'unun bu sınıfta olması beklenir. Bu durumda bu sınıf için beklenen frekans  $120(0.029) = 3.48$  olarak bulunur.

**III.sınıf için:**  $P(3.045 < X < 3.145) = P(-1.71 < Z < -1.25)$  yani bebeklerin doğum ağırlıklarının 3.045 ile 3.145 kg arasında olma olasılığı  $0.4564 - 0.3944 = 0.062$ 'dir. Bu sınıfın beklenen frekansı ise  $120(0.062) \approx 7.44$ 'tür.

**IV.sınıf için:**  $P(3.145 < X < 3.245) = P(-1.25 < Z < -0.79)$  yani bebeklerin doğum ağırlıklarının 3.145 ile 3.245 kg arasında olma olasılığı  $0.3944 - 0.2852 = 0.1092$ 'dir. Bu sınıfın beklenen frekansı ise  $120(0.1092) \approx 13.104$ 'tür.

**V. sınıf için:** Bu sınıfta yer alan bebeklerin doğum ağırlığı 3.245 ile 3.345 kg arasındadır. Burada  $P(3.245 < X < 3.345)$  olasılığı hesaplanacaktır. Bu X değerlerine karşılık gelen Z değerleri hesaplandığı zaman  $P(-0.79 < Z < -0.32)$  olarak bulunur.

$$P(-0.79 < Z < -0.32) = 0.2852 - 0.1255 = 0.1597$$

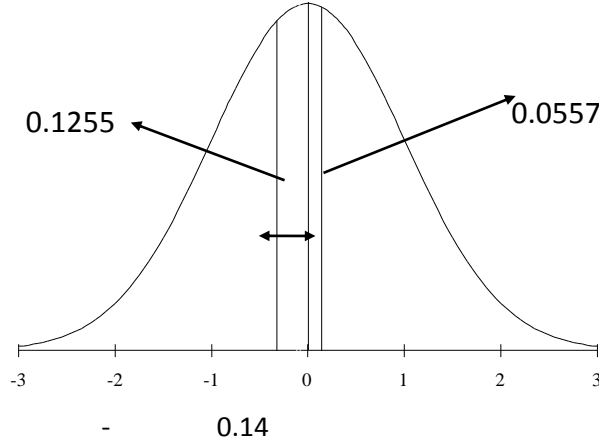
Bu sınıfta beklenen teorik frekans  $= 120(0.1597) = 19.164$ 'dür.

**VI.sınıf için:**  $P(3.345 < X < 3.445) = P(-0.32 < Z < 0.14)$  'tür. Dikkat edilecek olursa bu sınıfa ait Z değerlerinin biri pozitif diğeri ise negatiftir. Bunun anlamı aranan aralık ortalamasının iki tarafındadır.

$P(-0.32 < Z < 0.14) = P(-0.32 < Z < 0) + P(0 < Z < 0.14)$  şeklinde de ifade edilebilir. Bu ifade aşağıda daha açık olarak normal dağılım üzerinde gösterilmiştir.

## 7. HAFTA DERS NOTLARI

Kaynak Kitap: BİYOİSTATİSTİK (2007) Kesici T. ve Kocabaş Z., İkinci Baskı,  
Ankara Üniversitesi Eczacılık Fakültesi Yayın No:94



$$P(-0.32 < Z < 0) = 0.1255$$

$$P(0 < Z < 0.14) = 0.0557$$

$P(-0.32 < Z < 0.14) = 0.1255 + 0.0557 = 0.1812$  yani ele alınan 120 bebeğin doğum ağırlığının dahil olduğu normal dağılımda 3.345 ile 3.445 kg arasında olanların nisbi miktarı %18.12'dir. Bu sınıfta beklenen frekans  $120(0.1812) = 21.744$ 'dür.

Diğer sınıflar içinde benzer şekilde teorik frekanslar hesaplanarak tabloda verilmiştir. Burada dikkat edilmesi gereken son sınıfın teorik frekansının hesaplanmasıdır. Bu sınıf 3.945'ten büyük tüm verileri kapsar. Yani bunların olasılığı  $P(3.945 < X)$ 'dir. Standart normal dağılımda ise

$$P\left(\frac{3.945 - 3.414}{0.2152} < X\right) = P(2.47 < Z) \text{ dir.}$$

Bunun da sayısal değeri  $0.5000 - 0.4932 = 0.0068$ 'dir. Bu sınıfta beklenen teorik frekans  $= 120(0.0068) = 0.816$ 'dır.