ECZ158 BİYOİSTATİSTİK DERSİ

8. HAFTA DERS NOTLARI

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

ÖRNEKLEME DAĞILIMLARI

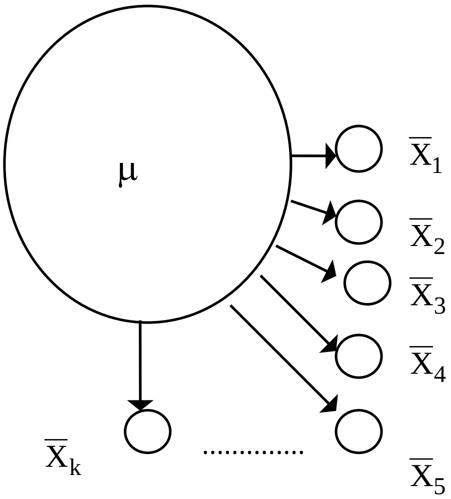
Bir populasyondan belirli bir örnek genişliğinde rastgele mümkün olan sayıda örnekler çekilse ve bu örneklerde herhangi bir istatistik hesaplansa, hesaplanan bu istatistiğin değeri örnekten örneğe değişecektir. Hesaplanan bu istatistikler bir dağılım gösterir. Bu dağılıma, hesaplanan istatistiğe ait **ÖRNEKLEME DAĞILIMI** denir.

Örnekleme dağılımları teorik dağılımlardır ve hipotez kontrolleri için gereklidir. Üzerinde çalışılan bir örneğin hipotezle belirtilen (hipotez ve kontrolleri hakkında bilgi için BÖLÜM VI'ya bakınız.) populasyondan rastgele çekilmiş örneklerden biri olma olasılığı, yani söz konusu istatistiğe ait örnekleme dağılımına dahil olama olasılığı hesaplanır. Bu hesaplanan olasılığa göre araştırıcının kurmuş olduğu hipotez ret veya kabul edilir.

Herhangi bir istatistiğe ait örnekleme dağılımının şekli ve parametreleri, örneklerin çekildiği populasyonun şekline, parametrelerine ve örnek genişliğine bağlıdır.

4.1. Ortalamaya ait Örnekleme Dağılımı

Ortalaması μ ve standart sapması σ olan bir normal populasyondan n örnek genişliğinde çok sayıda geri iadeli rastgele örnekler çekilse ve bu örneklerde ortalamalar hesaplansa bu ortalamaların her biri populasyon ortalamasının bir tahminidir. Örneklerden hesaplanan ortalamalar örnekten örneğe değişir ve bir dağılım gösterir. Bu dağılıma "**Ortalamaya ait Örnekleme Dağılımı**" denir. Bu dağılımın parametreleri $\mu_{\overline{X}}$ ve $\sigma_{\overline{X}}$ 'dir. Ortalamaya ait örnekleme dağılımının parametreleri yukarıda da değinildiği gibi rastgele çekilen örneklerin genişliğine, populasyonun parametrelerine ve şekline bağlıdır.



Parametreleri bilinen bu populasyondan "n" örnek genişliğine sahip çok sayıda (k tane) örnek çekilse ve ortalamaları hesaplansa, bu ortalamalar örnekten örneğe değişecek ve bir dağılım gösterecektir. Bu dağılım $k\rightarrow\infty$'a gittiği zaman "ortalamaya ait örnekleme" dağılımıdır.

Bu dağılımın parametrelerinden biri $\mu_{\overline{X}}$ yani ortalamadır ve populasyon ortalamasına, μ_x'e, eşittir. Diğer parametresi ise standart sapma, $\sigma_{\overline{X}}$ 'dir ve aşağıdaki gibidir:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \qquad \ldots(4.1)$$

Yani ortalamaya ait örnekleme dağılımının populasyondan hesaplanan standart sapmasıdır.

Fakat çoğu zaman populasyon varyansı bilinmediği için örnekten tahmin edilir. Bu durumda ortalamaya ait örnekleme dağılımının standart sapması;

$$S_{\bar{x}} = \frac{S_x}{\sqrt{n}} \qquad \ldots(4.2)$$

şeklinde tahmin edilir. Buna ortalamaya ait örnekleme dağılımının örnekten tahmin edilen standart sapması veya kısaca ortalamanın **standart hatası** denir.

Ortalamaya ait örnekleme dağılımının şekli örneklerin çekildiği populasyonun şekline ve tesadüfen çekilen örneklerin genişliğine bağlıdır. Eğer örneklerin çekildiği populasyon normal dağılım gösteren bir populasyon ise örnek genişliği ne olursa olsun ortalamaya ait örnekleme dağılımı normal dağılım gösterir. Fakat örneklerin çekildiği populasyon yatık (sağa veya sola çarpık) bir dağılım gösteriyorsa bu durumda örnekleme dağılımın şekli örneklerin genişliğine bağlıdır. Normal dağılım dışındaki dağılımlardan çekilen örneklerin genişliği arttıkça ortalamaya ait örnekleme dağılımı örnek genişliğine bağlı olarak normal dağılıma yaklaşır. Aşağıdaki örneklerde önce normal dağılımdan çekilen örneklerin genişliğinin dağılım tipini etkilemediği daha sonra normal olmayan dağılımlardan çekilen örneklerin genişliği arttıkça ortalamaya ait örnekleme dağılımının normal dağılıma yaklaştığı gösterilecektir.

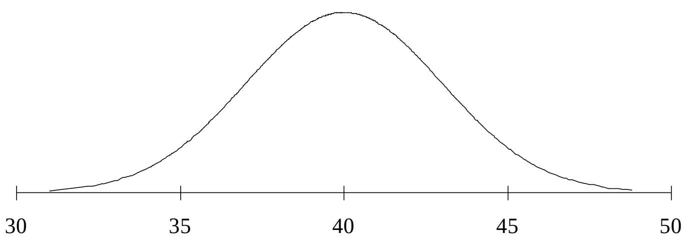
ÖRNEK 1:

Şekil 4.1.a'da MINITAB paket programında benzetim (simulasyon) yöntemi ile üretilmiş, ortalaması, μ=40 ve standart sapması, σ=3 olan bir normal dağılım görülmektedir. Bu dağılımdan n=4 olan mümkün olan sayıda örnekler çekildiği zaman hesaplanan ortalamaların her biri populasyon ortalamasının bir tahmini olup örnekten örneğe değişir ve Şekil 4.1.b'de görüldüğü gibi bir normal dağılım gösterir. Bu ortalamaya ait örnekleme dağılımıdır. Şekil 4.1.b'de görüldüğü gibi örnek genişliği 4 olmasına rağmen ortalamaya ait örnekleme dağılımı normal dağılım göstermiştir. Bu dağılımın parametreleri ise ortalama ve standart sapma olup aşağıdaki gibidir:

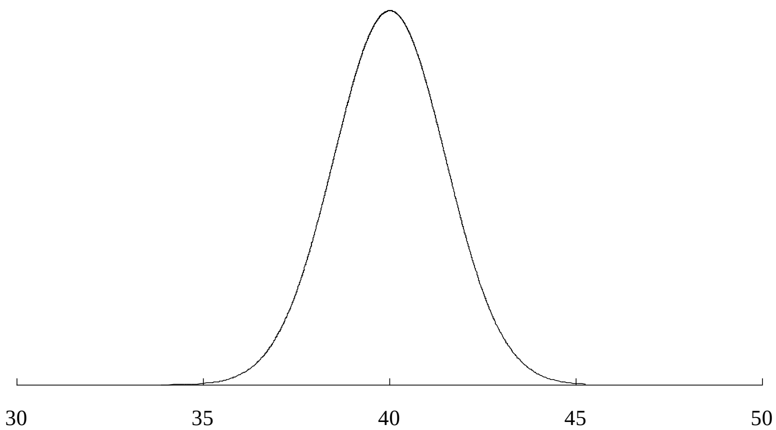
$$\mu_{\bar{x}} = 40$$

$$\sigma_{\bar{x}} = \frac{3}{\sqrt{4}} = 1.5$$

a.



b.



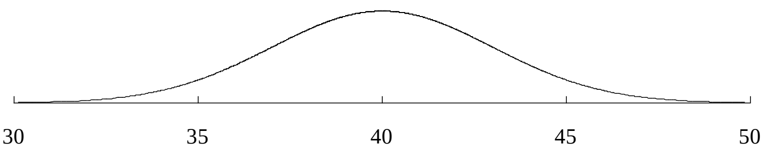
ŞEKİL 4.1. Benzetim (simulasyon) yöntemi ile üretilmiş, ortalaması, μ=40 ve standart sapması, σ=3 olan bir normal dağılım (a) ve n=4 bireylik örneklerden elde edilen ortalamaya ait örnekleme dağılımı (b)

ÖRNEK 2:

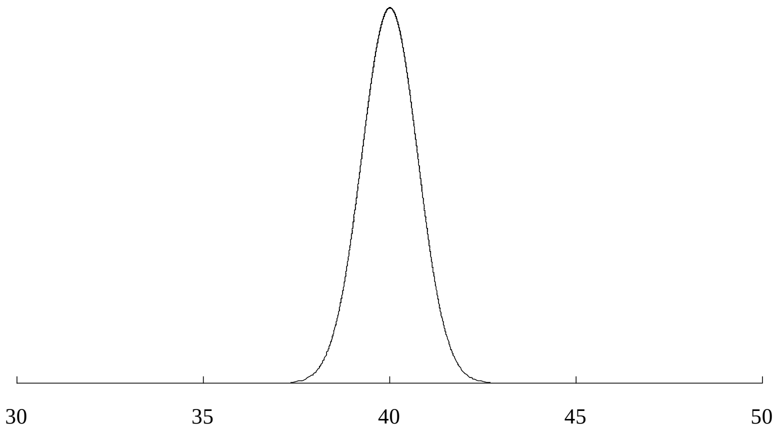
Ortalaması μ=40 ve σ=3 olan ve Şekil 4.1.a'da verilen normal dağılımdan bu sefer "n=16" örnek genişliğine sahip mümkün olan sayıda örnekler çekilse ve ortalamaları hesaplansa, ortalamaya ait örnekleme dağılımı yine normal dağılım gösterecektir (Şekil 4.2. a ve b). Şekil 4.1.b'de verilen grafikten görüldüğü ve Şekil 4.2.b ile karşılaştırıldığı zaman da anlaşılacağı gibi örnek genişliği arttığı zaman ortalamaya ait örnekleme dağılımının standart sapması küçülmektedir.

Bu dağılımın parametreleri ise; ortalaması, $\mu_{\overline{X}} = 40$, fakat örnek genişliğine bağlı olarak standart sapması, $\sigma_{\bar{x}} = \frac{3}{\sqrt{16}} = 0.75$'dir.

a.



b.



ŞEKİL 4.2. Benzetim (simulasyon) yöntemi ile üretilmiş, ortalaması, μ=40 ve standart sapması, σ=3 olan bir normal dağılım (a) ve n=16 bireylik örneklerden elde edilen ortalamaya ait örnekleme dağılımı (b).

ÖRNEK 3:

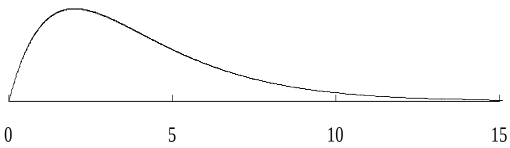
Örnekler yatık (çarpık) bir dağılım gösteren populasyondan çekiliyorsa ortalamalara ait örnekleme dağılımının şekli populasyondan rastgele çekilen örneklerin büyüklüğüne (genişliğine) bağlıdır. Örneğin 4 serbestlik dereceli ki-kare dağılımı gösteren bir populasyondan genişliği 4 olan örnekler çekilse populasyon ve ortalamanın örnekleme dağılımı Şekil 4.3.a ve b'de verildiği gibidir.

Şekil 4.3.a'da simulasyon yöntemi ile üretilmiş 4 serbestlik dereceli bir ki-kare dağılımı görülmektedir. Bu dağılımın ortalaması 4 (ki-kare dağılımında ortalama serbestlik derecesidir) ve varyansı 8'dir (ki-kare dağılımında varyans serbestlik derecesinin iki katıdır.). Şekil 4.3.b'de ise Şekil 4.3.a'da verilen dağılımı gösteren populasyondan çekilen ve genişliği 4 olan örneklerden hesaplanan ortalamalara ait dağılım görülmektedir. Bu dağılımın ortalaması yine 4, fakat varyansı 8/4'tür.

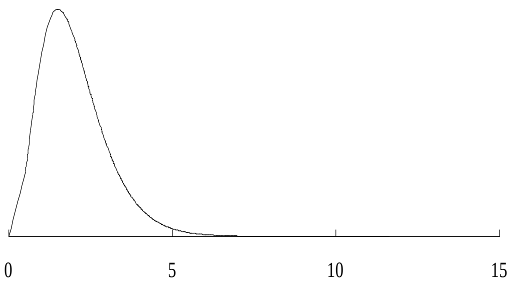
Şekil 4.3.c'de ise 4 serbestlik dereceli ki-kare populasyonundan çekilmiş 32 bireylik örneklerden hesaplanmış ortalamalara ait örnekleme dağılımı görülmektedir. Bu dağılımın da ortalaması 4, fakat varyansı 8/32'dir.

Şekil 4.3.b'den de görüldüğü gibi ki-kare dağılımından çekilen 4 bireylik örneklerden hesaplanan örneklerin ortalamalarına ait dağılımda yatıklık (çarpıklık), Şekil 4.3.a'da verilen dağılıma nazaran biraz azalmıştır. Eğer Şekil 4.3.a'da verilen dağılımı gösteren populasyondan 32 bireylik örnekler çekilirse ve bu örneklerde ortalamalar hesaplanırsa, hesaplanan ortalamalara ait dağılımın şekli ise Şekil 4.3.c'de verildiği gibi olacaktır. Şekil 4.3.c'de de görüldüğü gibi dağılım yatık (çarpık) bir dağılım olmasına rağmen, söz konusu populasyondan çekilen örneklerin genişliği arttıkça ortalamaya ait örnekleme dağılımının şekli de normal dağılıma yaklaşmaktadır.

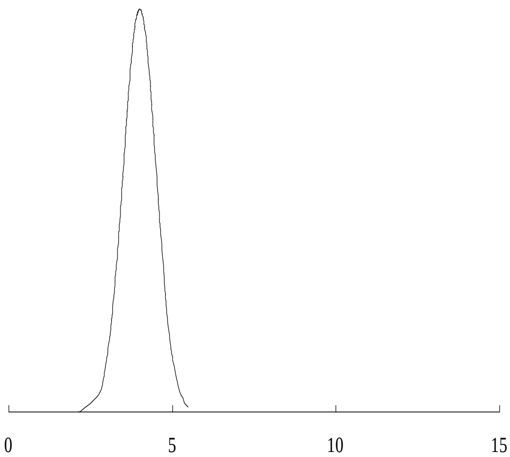
a.



b.

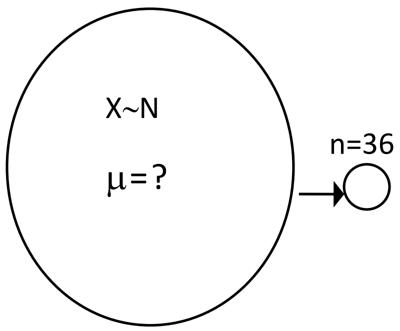


c.



ŞEKİL 4.3. Simulasyon yöntemi ile üretilmiş 4 serbestlik dereceli ki-kare dağılımı (a), n=4 bireylik örneklerden elde edilen ortalamaya ait örnekleme dağılımı (b), ve n=32 bireylik örneklerden elde edilen ortalamaya ait örnekleme dağılımı (c).

ÖRNEK 4:

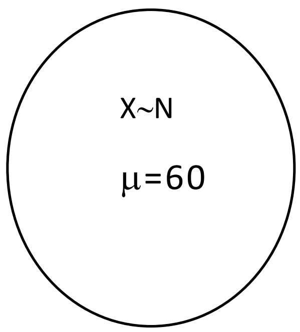


Ortalaması ve standart sapması bilinmeyen bir populasyondan 36 bireylik bir örnek çekilmiş ve bu örneğin ortalaması 56.4 ve standart sapması da 7.2 olarak bulunmuştur. Ortalamaya ait örnekleme dağılımının parametrelerinin örnekten tahminleri $\bar{X}$ ve $S_{\bar{x}}$'dir.

Çünkü populasyonun standart sapması örnekten tahmin edilmiştir. Bu dağılımın parametrelerinin tahminleri aşağıdaki şekilde bulunur:

$$\bar{X} = 56.4 \qquad \text{ve} \qquad S_{\bar{x}} = \frac{7.2}{\sqrt{36}} = 1.2$$

ÖRNEK 5:



Parametreleri yandaki şekilde verilen bir populasyon vardır. Elimizde de bu populasyondan çekilip çekilmediğini bilmediğimiz 9 bireylik bir örnek bulunmaktadır. Bu örnekte ortalama 56.4 olarak hesaplanmış ise bu örneğin, ortalamaya ait örnekleme dağılımına dahil olma olasılığı nedir?

Bu örnekte, ortalamaya ait örnekleme dağılımının parametreleri $\mu_{\bar{x}} = 60$ ve $\sigma_{\bar{x}} = \frac{6.3}{\sqrt{9}} = 2.1$'dir. 56.4 olarak hesaplanan ortalamanın, ortalamaya ait örnekleme dağılımına dahil olma olasılığını bulmak için bu değere karşılık gelen Z değerinin aşağıda verilen eşitlik kullanılarak hesaplanması gerekir. Aranan olasılık P($\bar{X}$ <56.4)'tür.

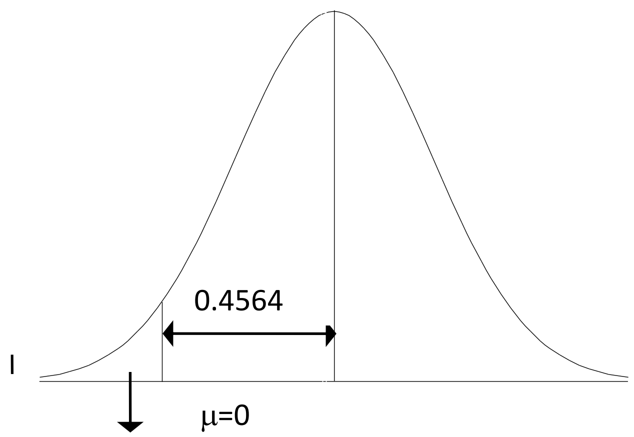
$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu_{\bar{x}}}{\frac{\sigma_{x}}{\sqrt{n}}}$$

Eşitlikte görüldüğü gibi standart sapma olarak ortalamaya ait örnekleme dağılımının standart sapması kullanılmaktadır. Çünkü hesaplanmak istenen, örnekten hesaplanan ortalamanın ortalamaya ait örnekleme dağılımına dahil olma olasılığıdır.

Yukarıda verilen ortalamaya karşılık gelen Z-değeri;

$$Z = \frac{56.4 - 60}{2.1} = -1.71$$

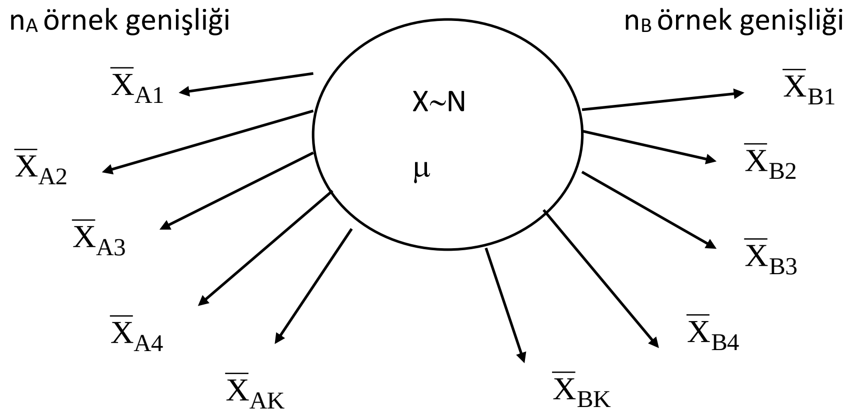
Söz konusu ortalamanın, ortalamaya ait örnekleme dağılımına dahil olma olasılığının bulunması ise aşağıda şekil üzerinde gösterilmiştir. Aşağıdaki şekilde gösterildiği gibi standart normal dağılımda Z değerlerinin -1.71 ile 0 arasında olma olasılığı 0.4564'tür.



Şekilde işaretlenmiş olan alan söz konusu ortalamanın, ortalamaya ait örnekleme dağılımına dahil olma olasılığını verir. Bu olasılık da 0.5-0.4564=0.0436' ya eşittir. Yani bu ortalamanın bu dağılıma dahil olma olasılığı %5'ten daha azdır.

4.2. Ortalamalar arası Farka ait Örnekleme Dağılımı

Ortalaması μ ve varyansı σ^2 olan bir normal populasyondan n_A örnek genişliğinde geri iadeli örnekler alınarak ortalamaları hesaplansa ve daha sonra aynı populasyondan n_B örnek genişliğinde yine geri iadeli örnekler çekilse ve yine ortalamaları hesaplansa. Hesaplanan bu ortalamalar tamamen tesadüfi olarak yan yana getirilerek ortalamalar arasındaki fark alınsa bu farklar bir dağılım gösterir. Bu dağılıma k→∞ için "**ortalamalar arası farka ait örnekleme**" dağılımı denir ve bu farklar da normal bir dağılım gösterir.



$$\bar{X}_{A1} - \bar{X}_{B50} = D_1$$
$$\bar{X}_{A5} - \bar{X}_{B27} = D_2$$
$$\bar{X}_{A12} - \bar{X}_{B60} = D_3$$
$$\bar{X}_{A29} - \bar{X}_{B72} = D_4$$
.
$$\bar{X}_{A99} - \bar{X}_{B121} = D_k$$

Her bir ortalama populasyon ortalamasının bir tahmini olduğu için hesaplanan farkların sıfır olması beklenir. Fakat farklar normal bir dağılım gösterir. Bu dağılım yukarıda da belirtildiği gibi ortalamalar arası farka ait örnekleme dağılımıdır.

Bu dağılımın ortalaması, $\mu_D=0$'dır. Çünkü $\mu_D = \mu_{\bar{X}_A} - \mu_{\bar{X}_B}$ ve $\mu_{\bar{X}_A} = \mu$, $\mu_{\bar{X}_B} = \mu$'dür.

Dağılımın varyansı $\sigma^2_D = \sigma^2_{\bar{X}_A} + \sigma^2_{\bar{X}_B}$'dir[1]. $\sigma^2_{\bar{X}_A} = \frac{\sigma^2_{X_A}}{n_A}$ ve $\sigma^2_{\bar{X}_B} = \frac{\sigma^2_{X_B}}{n_B}$ olduğuna ve $\sigma^2_{X_A} = \sigma^2_{X_B} = \sigma^2$ olduğuna göre dağılımın varyansı;

$$\sigma^2_D = \frac{\sigma^2}{n_A} + \frac{\sigma^2}{n_B} = \sigma^2 \frac{(n_A + n_B)}{n_A n_B} \qquad ...(4.3)$$

olarak bulunur. Dağılımın standart sapması ise;

$$\sigma_D = \sqrt{\sigma^2 \frac{(n_A + n_B)}{n_A n_B}} \qquad ...(4.4)$$

$n_A = n_B = n$ ise dağılımın varyansı;

---

[1] X ve Y gibi iki bağımsız değişkenin toplamlarının varyansı,

$$\sigma^2_{(X+Y)} = \sigma^2_X + \sigma^2_Y$$

X ve Y gibi iki bağımsız değişkenin farkının varyansı,

$$\sigma^2_{(X-Y)} = \sigma^2_X + \sigma^2_Y$$

X ve Y gibi iki bağımlı değişkenin toplamlarının varyansı,

$$\sigma^2_{(X+Y)} = \sigma^2_X + \sigma^2_Y + 2\mathrm{Cov}(X,Y)$$

X ve Y gibi iki bağımlı değişkenin farkının varyansı,

$$\sigma^2_{(X-Y)} = \sigma^2_X + \sigma^2_Y - 2\mathrm{Cov}(X,Y)$$

$$\sigma_D^2 = \frac{\sigma^2}{n} + \frac{\sigma^2}{n} = \frac{\sigma^2(n+n)}{n.n} = \frac{2n\sigma^2}{n^2} = \frac{2\sigma^2}{n}$$

ve standart sapması da;

$$\sigma_D = \sqrt{2\frac{\sigma^2}{n}} \qquad ...(4.5)$$

şeklinde bulunur.

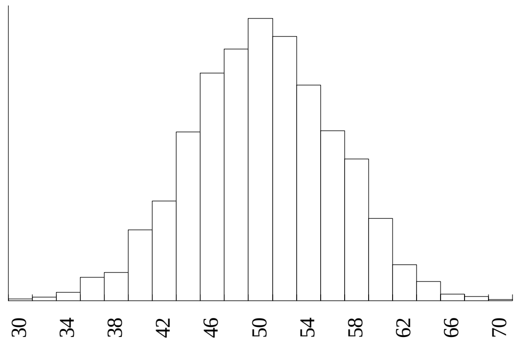
Ortalamalar arası Farka ait Örnekleme Dağılımının Şekli

Normal dağılımdan alınan örneklerin içerdiği birey sayısı ne olursa olsun elde edilecek ortalamalar arası farka ait örnekleme dağılımının şekli normal olacaktır.

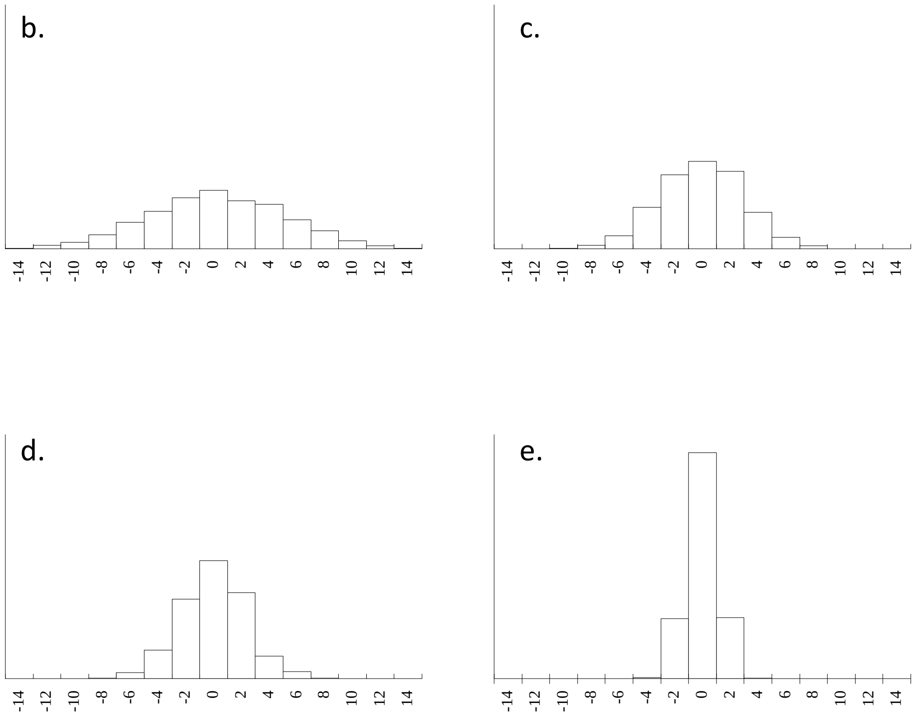
Simulasyon yöntemi ile ortalaması 50 ve standart sapması 6 olan normal dağılan değerler üretilmiştir. Bunların histogramı Şekil 4.4.a'da görülmektedir. Bu normal populasyondan 4, 9, 15 ve 60 birey içerecek şekilde simulasyon yöntemi ile çekilen örneklerden elde edilen ortalamalar arası farka ait histogramlar da Şekil 4.4.b, c, d ve e'de görülmektedir.

Şekil 4.4.b, c, d ve e'de görüldüğü gibi normal dağılımdan çekilen örnek ortalamaları arası farkların dağılımı örnek genişliği ne olursa olsun normaldir. Sadece örnek genişliği artıkça ortalamalar arası farka ait örnekleme dağılımının standart sapması azalmaktadır.

a.



ŞEKİL 4.4. Simulasyon yöntemi ile üretilmiş a. ortalaması 50 ve standart sapması 6 olan populasyon ve bu populasyondan çekilen ve b.)$n_A = n_B = 4$, c.)$n_A = n_B = 15$, d.)$n_A = n_B = 30$ ve e.)$n_A = n_B = 60$ olan örneklerden elde edilen ortalamalar arası farkların histogramları

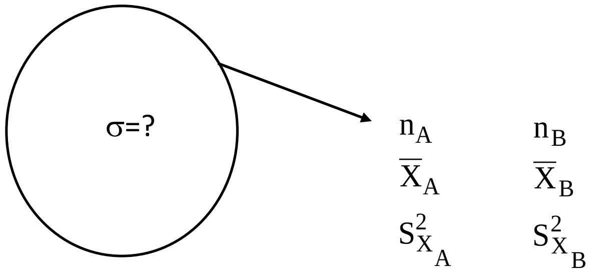


ŞEKİL 4.4. (devam). Simulasyon yöntemi ile üretilmiş a. ortalaması 50 ve standart sapması 6 olan populasyon ve bu populasyondan çekilen ve b.)$n_A = n_B = 4$, c.)$n_A = n_B = 15$, d.)$n_A = n_B = 30$ ve e.)$n_A = n_B = 60$ olan örneklerden elde edilen ortalamalar arası farkların histogramları

Şekil 4.4.b'de verilen dağılımın standart sapması, $\sigma_D = 6\sqrt{\frac{2}{4}} = 4.24$ iken, Şekil 4.4.c'de verilen dağılımın standart sapması 2.828, Şekil 4.4.d'de verilen dağılım standart sapması 2.191 ve Şekil 4.4.e'de verilen dağılımın standart sapması ise 1.549'tur. Örnek genişliğine bağlı olarak elde edilecek örnekleme dağılımında farklar arasındaki varyasyonun azaldığı verilen şekillerden de görülebilmektedir. Örnek genişliği 4 olduğu zaman deneyde elde ettiğimiz farklar -14 ile 14 arasında değişirken, örnek genişliği 60 olduğu zaman elde edilen farklar yaklaşık olarak -3.5 ile 3.5 arasında değişmektedir.

Toplanmış Varyans:

Çoğu zaman örneklerin çekildiği populasyonun varyansı bilinmez. Bu durumda ortalamalar arası farka ait örnekleme dağılımının varyansı (veya standart sapması) örnekten tahmin edilir.



Yukarıdaki şekilde gösterildiği gibi eğer örnek genişlikleri n_A ve n_B olan iki örnek çekilmiş ise bu durumda

$$\sigma_D^2 = \sqrt{\sigma^2 \frac{(n_A + n_B)}{n_A n_B}}$$

eşitliğinde σ^2 yerine konacak populasyon varyansının en iyi tahmini **toplanmış varyanstır**. Toplanmış varyans populasyondan çekilen örneklerin varyanslarının serbestlik derecesi ile tartılı ortalamasıdır ve populasyon varyansının en iyi tahminidir ve aşağıdaki şekilde hesaplanır:

$$S^2 = \frac{(n_A - 1)S^2_{X_A} + (n_B - 1)S^2_{X_B}}{(n_A - 1) + (n_B - 1)} \qquad \text{...(4.6)}$$

$\sum d^2_{X_A} = (n_A - 1)S^2_{X_A}$ ve $\sum d^2_{X_B} = (n_B - 1)S^2_{X_B}$ olduğu bilindiğine göre toplanmış varyans aşağıdaki şekilde de yazılabilir;

$$S^2 = \frac{\sum d^2_{X_A} + \sum d^2_{X_B}}{(n_A - 1) + (n_B - 1)}$$

Bu durumda ortalamalar arası farka ait örnekleme dağılımın örnekten tahmin edilen varyansı S_D^2 olarak gösterilir ve aşağıdaki şekilde hesaplanır:

$$S_D^2 = \frac{\sum d^2_{X_A} + \sum d^2_{X_B}}{(n_A - 1) + (n_B - 1)} \cdot \frac{(n_A + n_B)}{(n_A n_B)}$$

Ve dağılımın örnekten hesaplanan standart sapması da;

$$S_D = \sqrt{\frac{\sum d_{x_A}^2 + \sum d_{x_B}^2}{(n_A - 1) + (n_B - 1)} \cdot \frac{(n_A + n_B)}{(n_A n_B)}} \quad ...(4.7)$$

şeklinde hesaplanır. Bu örnek ortalamaları arası farkın standart hatasıdır.

4.3. Oranlara ait Örnekleme Dağılımı

Üzerinde çalışılan olayın oluş olasılığı π olan populasyondan örnekler çekilse söz konusu oran hesaplansa, bunların her biri π'nin birer tahminidir. Tahmin edilen bu olasılıklar örnekten örneğe değişecek ve bir dağılım gösterecektir. Söz konusu bu dağılıma "**oranlara ait örnekleme dağılımı**" denir. Bu dağılımın parametreleri olan ortalama ve varyans, μ_p ve σ_p^2'dir. Oranlara ait örnekleme dağılımının ortalaması, populasyonda üzerinde durulan olayın oluş olasılığına eşittir, yani $\mu_p=\pi$'dir. Dağılımın varyansı ise;

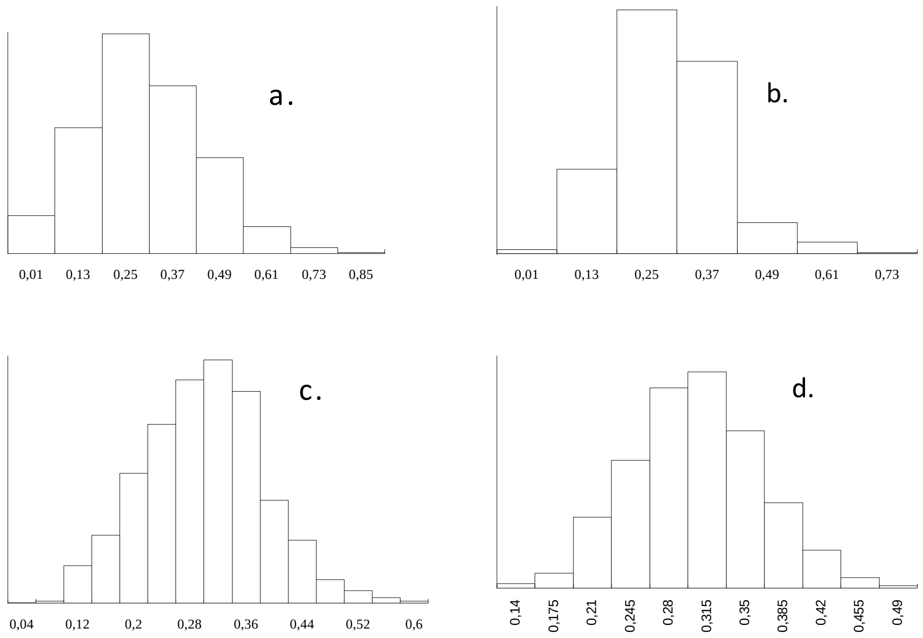
$$\sigma_p^2 = \frac{\pi(1-\pi)}{n} \quad ...(4.8)$$

Eğer π bilinmiyor ise yerine örnekten hesaplanan p değeri kullanılır. Bu durumda örnekleme dağılımına ait varyansın tahmini;

$$S_p^2 = \frac{p(1-p)}{n} \text{ dir.} \quad ...(4.9)$$

Oranlara ait örnekleme dağılımının şekli populasyonun π'sine ve çekilen örneklerin genişliğine bağlıdır. Eğer populasyonda üzerinde durulan olayın oluş olasılığı, $\pi \leq 1/2$ ise oranlara ait örnekleme dağılımının normal dağılıma yaklaşabilmesi için $n.\pi \geq 10$ olmalıdır. $\pi > 1/2$ olduğu durumlarda ise oranlara ait örnekleme dağılımının normal dağılıma yaklaşabilmesi için ise $n.(1-\pi) \geq 10$ şartının yerine getirilmiş olması gerekir. Aslında hiçbir zaman tam olarak normal dağılım olmayıp bu şartlar sağlandığı zaman normal dağılıma yaklaşır. Belirtilen bu şartları sağlamayan örnek genişlikleri için elde edilecek örnekleme dağılımının şekli normal olmayacaktır. Bu durum aşağıda dağılım şekilleri verilerek açıklanmaktadır.

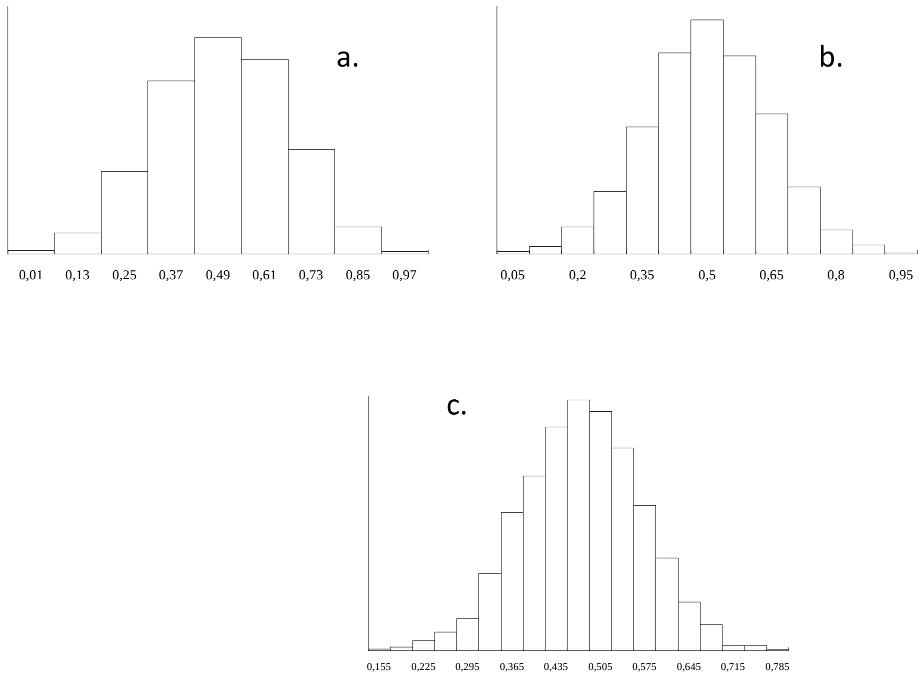
Simulasyonla üzerinde durulan olayın oluş olasılığı π=0.3 olan bir populasyondan örnek genişliği 9, 15, 30 ve 60 olan çok sayıda örnekler çekildiği ve bu örneklerin her birinde üzerinde durulan olayın oluş olasılığı hesaplandığı zaman bu oranlara ait örnekleme dağılımlarının şekli sırası ile Şekil 4.5.a, b, c ve d'de verilmiştir.



ŞEKİL 4.5.(devam). Simulasyon yöntemi ile üretilmiş π=0.3 olan populasyondan çekilen a.)n=9, b.)n=15, c.)n=30 ve d.)n=60 birey içeren örneklerden elde edilen oranlara ait histogramlar

Şekil 4.5'de verildiği gibi oranlara ait örnekleme dağılımının normal dağılıma yaklaşması için gerekli şart ($\pi \leq 1/2$ olduğu için) $n.\pi \geq 10$ dur. Örnek genişliklerinin 9 ve 15 olduğu durumlarda bu şart gerçekleşmemektedir. Ve bu durumda, Şekil 4.5.a ve b'de gösterildiği gibi elde edilen dağılım yatık (çarpık) bir dağılımdır. π=0.3 olduğu zaman, oranlara ait örnekleme dağılımının normal dağılım gösterebilmesi için örnek genişliği en az $n=10/0.3 \cong 33$ olmalıdır. Şekil 4.5.c, örnek genişliği 30 olan örneklerden elde edilen örnekleme dağılımını göstermektedir. Şekilden de görüldüğü gibi örnekleme dağılımının şekli normale yaklaşmıştır. Örnek genişliği 60 olduğu zaman ise oranlara ait örnekleme dağılımı normal dağılıma daha çok yaklaşmıştır (Şekil 4.5.d).

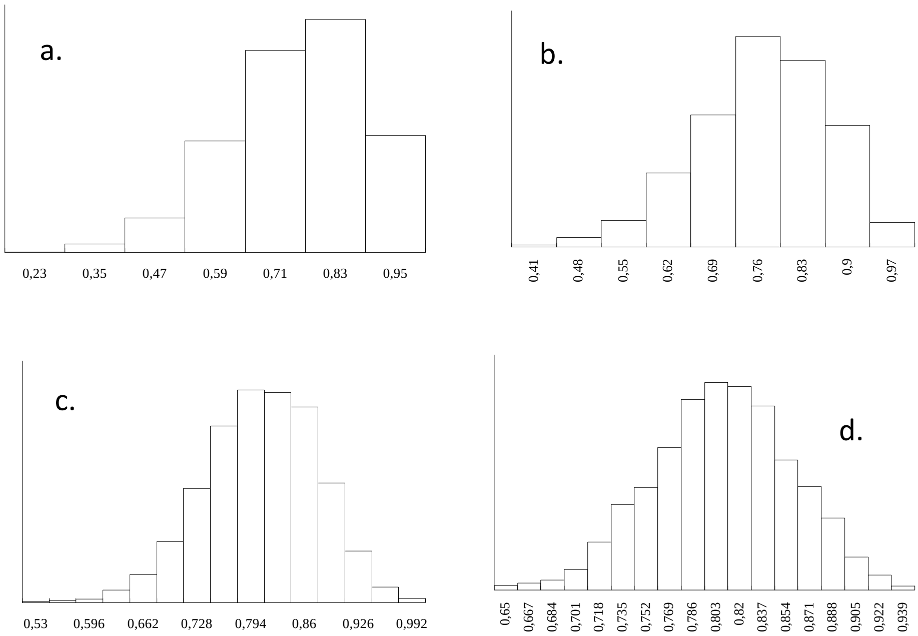
Şekil 4.6. a, b ve c ise π =0.5 olan bir populasyondan çekilen 9, 15 ve 30 bireylik örneklerden elde edilen oranlara ait örnekleme dağılımlarının şeklini göstermektedir.



ŞEKİL 4.6. Simulasyon yöntemi ile üretilmiş π=0.5 olan populasyondan çekilen a.)n=9, b.)n=15 ve c.)n=30 birey içeren örneklerden elde edilen oranlara ait histogramlar

Üzerinde durulan olayın oluş olasılığı, π=0.5 olduğu zaman dağılımın şekli simetriktir. Bu nedenle, söz konusu populasyondan çekilen örneklerin genişliği ne olursa olsun oranlara ait örnekleme dağılımı simetrik bir şekil gösterecektir. Artan örnek genişlikleri ile de dağılım yaklaşık bir normal dağılım olacaktır (Şekil 4.6.a,b,c).

Şekil 4.7 ise simulasyon yöntemi ile üretilmiş ve üzerinde durulan olayın oluş olasılığı 0.5'den büyük olan bir populasyondan 9, 15, 30 ve 60 birey içeren örnekler çekilerek elde edilen oranlara ait örnekleme dağılımlarının şeklini göstermektedir.



ŞEKİL 4.7. Simulasyon yöntemi ile üretilmiş π=0.8 olan populasyondan çekilen a.)n=9, b.)n=15, c.)n=30 ve d.)n=60 birey içeren örneklerden elde edilen oranlara ait histogramlar

Eğer populasyon π'si 1/2'den büyük ise oranlara ait örnekleme dağılımının normal dağılıma yaklaşabilmesi için $n.(1-\pi) \geq 10$ şartının yerine getirilmiş olması gerektiği daha önce belirtilmişti. Söz konusu populasyonda π=0.8 olduğu için örnekleme dağılımının normal dağılıma yaklaşması için örnek genişliği en az n=10/(1-0.8)=50 olmalıdır. Şekil 4.7. a ve b'de gösterildiği gibi küçük örnek genişlikleri için örnekleme dağılımı yatık (çarpık) bir dağılım gösterir. Örnek genişliği 30 olduğu zaman dağılımın şekli normale yaklaşmaktadır (Şekil 4.7.c) ve örnek genişliği 50'den daha büyük olduğu zaman ise oranlara ait örnekleme dağılımı yaklaşık normal bir dağılım göstermektedir (Şekil 4.7.d).

ÖRNEK:

Bir populasyonda üzerinde durulan olayın oluş olasılığının %85 olduğu bilinmekte ise bu populasyondan en az kaçar bireylik örnekler alınmalıdır ki oranlara ait örnekleme dağılımı normal bir dağılıma yaklaşsın?

$\pi \geq 1/2$ ise dağılım şeklinin normale yaklaşması için $n(1-\pi) \geq 10$ olmalıdır. Bu durumda örnek genişliği en az n.(1–0.85)=10'dur. Bu işlem yapıldığı zaman örnek genişliğinin en az 67 olması gerektiği bulunur. Böyle bir populasyondan çekilen örneklerden hesaplanan oranlara ait dağılımın normal dağılıma yaklaşması için en az 67 birey içeren örnekler alınmalıdır.

ÖRNEK:

π=0.7 olan bir populasyondan örnek genişliği 100 olan mümkün olan sayıda rastgele örnekler alınsa, hesaplanan oranların % ne kadarının değeri 0.608 ile 0.792 arasındadır?

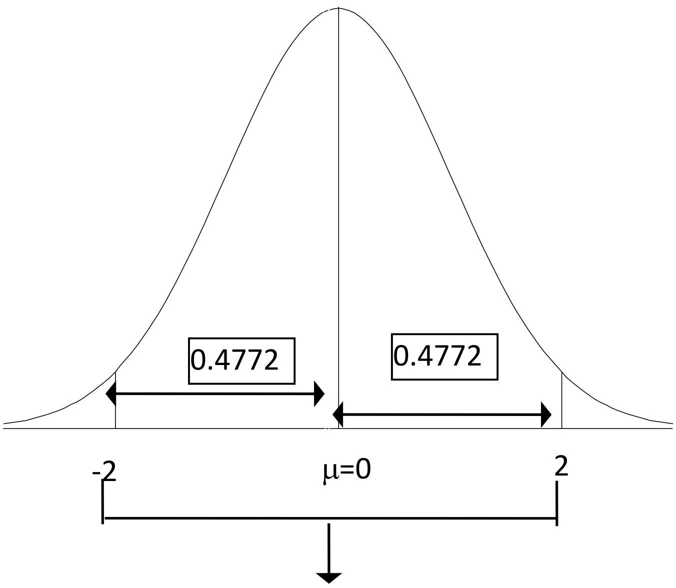
Oranlara ait örnekleme dağılımın parametreleri ortalama, $\mu_p=\pi$ ve varyans $\sigma_p^2 = \frac{\pi(1-\pi)}{n}$ 'dir. İstenen olasılık ise P(0.908<p<0.792)'dir. Bu olasılığı hesaplamak için aşağıda verilen Z eşitliği kullanılarak bu oranlara karşılık gelen Z-değerlerinin bulunması gerekir.

$$Z = \frac{p-\mu_p}{\sigma_p} = \frac{p-\mu_p}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

Yukarıda verilen eşitlik kullanılarak söz konusu oranlara karşılık gelen Z-değerleri aşağıdaki şekilde hesaplanır:

$$P\left(\frac{(0.608-0.7)}{\sqrt{\frac{0.7(1-0.7)}{100}}} < Z < \frac{(0.792-0.7)}{\sqrt{\frac{0.7(1-0.7)}{100}}}\right) = P(-2 < Z < 2)$$

İstenen olasılığın bulunması ise aşağıda şekil üzerinde gösterilmiştir.



Oranların 0.608 ile 0.792 arasında olma olasılığı;

P(-2<Z<0)+P(0<Z<2)= 0.4772+0.4772=0.9544'tür.

Yani hesaplanan oranların %95.44'ünün değeri 0.608 ile 0.792 arasındadır.

4.4. Oranlar arası Farka ait Örnekleme Dağılımı

Üzerinde durulan olayın oluş olasılığı π olan bir populasyondan genişliği n_1 ve n_2 olan örnekler çekilse ve bu örneklerde p_1 ve p_2'ler hesaplansa, daha sonra örneklerden hesaplan olasılıklar tamamen tesadüfen yan yana getirilerek farkları alınsa, hesaplanan farklar bir dağılım gösterecektir ki bu dağılıma "**oranlar arası farka ait örnekleme dağılımı**" denir. Oranlar arası farka ait örnekleme dağılımının ortalaması $\mu_{(p_1-p_2)}=0$ ve varyansı $\sigma^2_{(p_1-p_2)}=\sigma^2_{p_1}+\sigma^2_{p_2}$'dır. Bu da aşağıdaki şekilde gösterilir;

$$\sigma^2_{(p_1-p_2)}=\frac{\pi(1-\pi)}{n_1}+\frac{\pi(1-\pi)}{n_2} \quad ...(4.10)$$

Eğer populasyondaki oluş olasılığı bilinmiyorsa örnekten hesaplanan değerler kullanılır. Bu durumda oranlar arası farkın örnekleme dağılımının tahmini varyansı;

$$S^2_{(p_1-p_2)}=\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2} \quad ...(4.11)$$

Oranlar arası farka ait örnekleme dağılımı, populasyondan çekilen örneklerin genişlikleri yeteri kadar büyükse normal dağılım gösterir.