

ÖRNEKLEME DAĞILIMLARI

Üzerinde çalışılan populasyondan belirli örnek genişliğinde mümkün olan sayıda tesadüf örnekleri alınsa ve üzerinde çalışılan istatistikler hesaplınsa, hesaplanan istatistiklerin her biri parametrenin ayrı ayrı birbirlerinden bağımsız birer tahminidir. Hesaplanan istatistiklerin bazıları parametreye eşit, bazıları parametreden küçük bazıları ise parametreden büyüktür. Örnekler, populasyondan tamamen tesadüfen seçildiği için söz konusu istatistikler örnekten örneğe değişerek bir dağılım gösterir. Bu dağılıma söz konusu istatistiğe ait **örnekleme dağılımı** denir.

Bu şekilde elde edilen örnekleme dağılımları teorik dağılımlardır ve araştırmacıya hipotez kontrolleri için gereklidir. Çünkü hipotez kontrolü, üzerinde çalışılan örneğin hipotezle belirtilen populasyonu temsil etme olasılığının, diğer bir deyişle hipotezle belirtilen populasyondan söz konusu istatistiğe ait örnekleme dağılımının dahil olma olasılığının hesaplanması ve hesaplanan bu olasılık dikkate alınarak karar verilmesi işlemidir.

Üzerinde çalışılan populasyondan elde edilecek örnekleme dağılımının şekli ve parametreleri, populasyonun şekline, parametrelerine ve populasyondan tesadüfen seçilen örneklerin genişliğine göre değişir.

5.1. Ortalamaya ait Örnekleme Dağılımı

Ortalama μ ve standart sapması σ olan normal dağılım gösteren bir populasyondan belirli örnek genişliğinde (n) geriye iadeli olarak mümkün olan sayıda örnekler alınsa ve bu örneklerin her birinde ortalama hesaplınsa, hesaplanan ortalamaların her biri populasyon ortalamasının, yani μ 'nün, birbirlerinden bağımsız birer tahminidir. Hesaplanan ortalamaların bazıları μ 'ye eşit, bazıları μ 'den küçük ve bazıları μ 'den büyüktür. Mümkün olan sayıdaki örnekler populasyondan tamamen tesadüfen seçildiği için hesaplanan ortalamalar örnekten örneğe değişerek bir dağılım gösterir. Bu dağılıma **“ortalamaya ait örnekleme dağılımı”** denir.

Ortalamaya ait örnekleme dağılımının ortalama ($\mu_{\bar{x}}$) ve standart sapma ($\sigma_{\bar{x}}$) olmak üzere iki parametresi vardır. Dağılımın ortalaması populasyon ortalamasına eşittir. Standart sapması ise 5.1 numaralı eşitlikte görüldüğü gibi populasyona ait standart sapmanın örnek genişliğinin kareköküne bölümü olarak hesaplanır.

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \sqrt{\frac{\sigma_x^2}{n}} \quad \dots(5.1)$$

(5.1) numaralı eşitlik kullanılarak hesaplanan standart sapma, ortalamaya ait örnekleme dağılımının populasyondan hesaplanan standart sapmasıdır.

Yapılan çoğu çalışmada üzerinde çalışılan populasyonun standart sapması bilinmez. Bu sebeple populasyonun varyansı örnekten tahmin edilir. Bu durumda ortalamaya ait örnekleme dağılımının standart sapması;

$$S_{\bar{x}} = \frac{S_x}{\sqrt{n}} = \sqrt{\frac{S_x^2}{n}} \quad \dots(5.2)$$

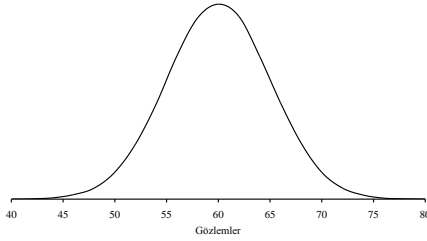
şeklinde tahmin edilir. Bu ifadeye ortalamaya ait örnekleme dağılımının örnekten tahmin edilen standart sapması veya kısaca **ortalamanın standart hatası** denir. Ortalamanın standart hatası hesaplandıktan sonra örnek ortalaması standart hatası ile birlikte $\bar{X} \pm S_{\bar{x}}$ şeklinde verilir.

Ortalamaya ait örnekleme dağılımın şekli örneklerin seçildiği populasyonun şekli ile seçilen örneklerin genişliğine bağlıdır.

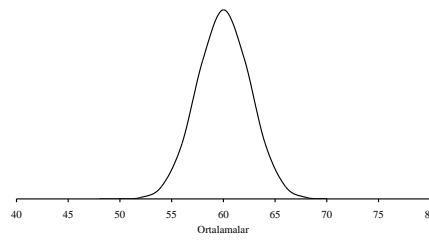
Normal dağılım gösteren bir populasyondan elde edilecek ortalamalara ait örnekleme dağılımı örnek genişliği ne olursa olsun normal dağılım gösterir. Şekil 5.1a'da görüldüğü gibi üzerinde çalışılan populasyon ortalaması, $\mu=60$ ve standart sapması, $\sigma=5$ olan normal bir dağılım gösteriyor olsun. Bu populasyondan tamamen tesadüfen ve mümkün olan sayıda $n=4$, $n=10$ ve $n=30$ olan örnekler alınsa ve ortalamalara ait örnekleme dağılımları oluşturulsa Şekil 5.1b, c ve d'de görüldüğü gibi ortalamalara ait örnekleme dağılımları normal dağılım gösterir.

Normal dağılım gösteren bir populasyondan alınan örneklerin genişliği dağılım şeklini değil dağılımın standart sapmasını etkiler. Şekil 5.1'de görüldüğü gibi populasyondan alınan örneklerin genişliği arttıkça ortalamaya ait örnekleme dağılımının standart sapması azalır. Çünkü örnek genişliği arttıkça tahmin edilen istatistikler parametreye yaklaşır ve örneklerden hesaplanan ortalamalar arasındaki değişim azalır.

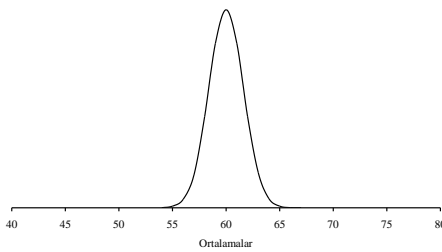
a. Populasyon



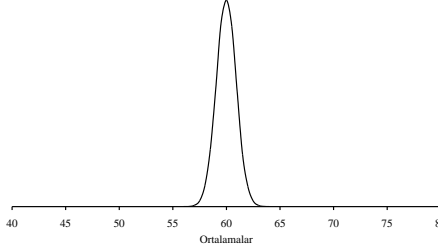
b. n=4 bireylik örnek



c. n=10 bireylik örnek



d. n=30 bireylik örnek



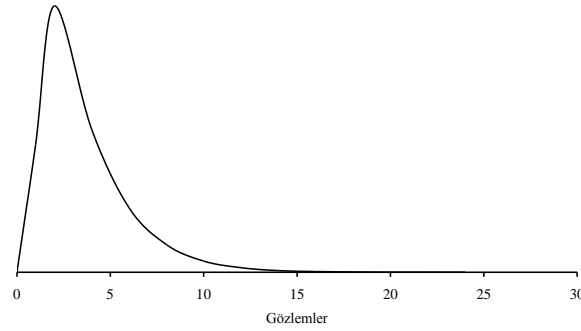
ŞEKİL 5.1. Benzetim (simulasyon) yöntemi ile üretilmiş a. ortalaması, $\mu=60$ ve standart sapması, $\sigma=5$ olan bir normal dağılım, b. $n=4$, c. $n=10$ ve d. $n=30$ bireylik örneklerden elde edilen ortalamaya ait örnekleme dağılımları

Şekil 5.1b'de $n=4$ bireylik örneklerden elde edilen ortalama ait örnekleme dağılımı görülmektedir. Bu dağılımın standart sapması $\sigma_{\bar{x}} = \frac{5}{\sqrt{4}} = 2.5$ olarak hesaplanırken, $n=10$ olduğu zaman $\sigma_{\bar{x}} = \frac{5}{\sqrt{10}} = 1.58$ (Şekil 5.1c) ve $n=30$ olduğu zaman standart sapma $\sigma_{\bar{x}} = \frac{5}{\sqrt{30}} = 0.91$ (Şekil

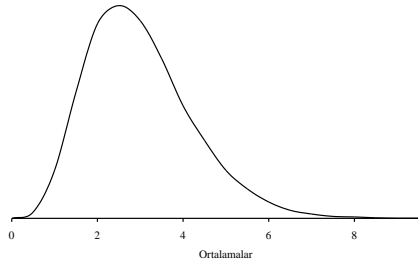
5.1d) olmakta ve görüldüğü gibi örnek genişliği arttıkça ortalamaya ait örnekleme dağılımının standart sapması azalmaktadır.

Üzerinde çalışılan populasyon yatık (sağa veya sola çarpık) bir dağılım gösteriyor ise bu durumda elde edilecek ortalamaya ait örnekleme dağılımının şekli örnek genişliği arttıkça normal dağılıma yaklaşır. Eğer üzerinde çalışılan populasyon Şekil 5.2a'da görüldüğü gibi ise ortalamaya ait örnekleme dağılımının şekli Şekil 5.2 b,c ve d'de görüldüğü gibi örnek genişliği arttıkça normal dağılıma yaklaşacaktır.

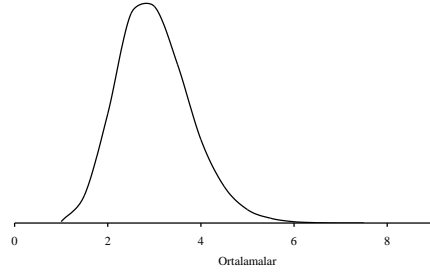
a. Çarpık dağılım



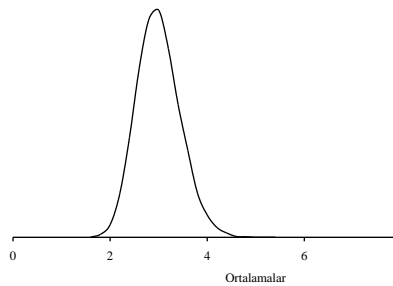
b. n=4 bireylik örnek



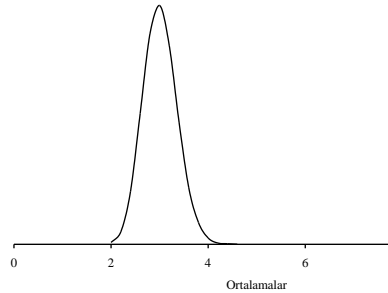
c. n=10 bireylik örnek



d. n=30 bireylik örnek



e. n=50 bireylik örnek



Şekil 5.2. Benzetim yöntemi ile üretilmiş (a) Çarpık dağılım, b. n=4, c. n=10, d. n=30 ve e. n=50 bireylik örneklerden elde edilen ortalamaya ait örnekleme dağılımları

Normal dağılım gösteren ortalamaya ait örnekleme dağılımını oluşturan ortalamalar (5.3) numaralı eşitlik kullanılarak standart normal dağılıma dönüştürülebilir. Böylece üzerinde çalışılan bir örnekten hesaplanacak ortalamanın ilgilenilen ortalamaya ait örnekleme dağılımına dahil olma olasılığı, standart normal dağılıma ilişkin olasılık tablosundan yararlanılarak kolaylıkla hesaplanabilir.

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \quad \dots(5.3)$$

ÖRNEK:

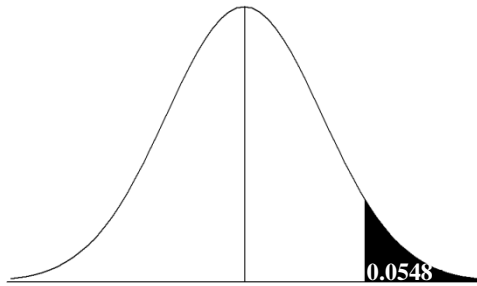
Ortalaması $\mu = 500$ kg ve standart sapması $\sigma = 50$ kg olan normal dağılım gösteren bir popülasyondan her defasında 25 birey bulunan çok fazla sayıda rastgele örnekler seçilse ve seçilen bu örneklerin her birinde ortalama hesaplanırsa;

a. Hesaplanan bu ortalamalar normal bir dağılım gösterir. Çünkü örneklerin seçildiği popülasyonun, ortalaması $\mu=500$ kg ve standart sapması $\sigma=50$ kg olan bir normal dağılım gösterdiği daha önceden bildirilmiştir. Hesaplanan ortalamaların gösterdiği bu dağılıma “**ortalamaya ait örnekleme dağılımı**” denir. Ortalamaya ait örnekleme dağılımının, ortalama, $\mu_{\bar{X}}=500$ kg ve standart sapması, $\sigma_{\bar{X}} = \frac{50}{\sqrt{25}} = 10$ kg olmak üzere iki parametresi vardır.

b. Bu popülasyondan alındığı ileri sürülen 16 bireylik bir örnekte ortalama 520 kg olarak hesaplanmışsa ve bu örneğin söz konusu popülasyonu temsil etme olasılığı hesaplanmak isteniyorsa ilk olarak ortalaması $\mu=500$ kg ve standart sapması $\sigma=50$ kg olan normal bir dağılım gösteren popülasyondan $n=16$ bireylik örneklerden elde edilecek ortalamaya ait örnekleme dağılımının parametreleri, $\mu_{\bar{X}}=500$ kg ve (5.1) numaralı eşitlik kullanılarak standart sapması, $\sigma_{\bar{X}} = \frac{50}{\sqrt{16}} = 12.5$ kg olarak belirlenir. Ortalamaya ait örnekleme dağılımını normal dağılım gösterdiği için örnekten hesaplanan ortalamaya karşılık gelen Z-değeri (5.3) numaralı eşitlikten;

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{520 - 500}{12.5} = 1.6$$

olarak hesaplanır. Z-değeri hesaplandıktan sonra 16 bireylik örneğin söz konusu popülasyondan tesadüfen seçilmiş bir örnek olma olasılığı aşağıdaki şekilde bulunur.



Söz konusu olasılık yandaki grafikte siyah taralı alandır ve $0.5 - P(0 < Z < 1.6)$ olarak hesaplanır ve 0.0548 bulunur, yani $n=16$ bireylik örneğin söz konusu popülasyondan tesadüfen seçilmiş örneklerden biri olma olasılığı %5.48'dir.

5.2. Ortalamalar arası Farka ait Örnekleme Dağılımı

Ortalaması μ ve standart sapması σ olan bir normal dağılım gösteren bir popülasyondan önce n_A ve sonra n_B genişliğinde geriye iadeli olarak mümkün olan sayıda tesadüf örnekleri alınsa ve her birinde ortalama hesaplanırsa, hesaplanan ortalamaların her biri popülasyon ortalamasının, yani μ 'nün, birer tahminidir. n_A ve n_B birey içeren örneklerden hesaplanan ortalamalar tamamen tesadüfen yan yana getirilerek ortalamalar arasındaki farklar bulunsa bu farkların sıfır olması beklenir. Hesaplanan

farkların bazıları 0, bazıları 0'dan küçük ve bazıları ise 0'dan büyüktür. Mümkün olan sayıdaki örnekler popülasyondan tamamen tesadüfen seçildiği ve tesadüfen yan yana getirilerek farklar bulunduğu için örnekten örneğe değişir ve bir dağılım gösterir. Bu dağılıma “**ortalamar arası farka ait örnekleme**” dağılımı denir.

Ortalamar arası farka ait örnekleme dağılımının ortalama, μ_D ve standart sapma, σ_D olmak üzere iki parametresi vardır. Dağılımın ortalaması, $\mu_D = \mu_{\bar{A}} - \mu_{\bar{B}} = 0$ 'dır. Çünkü n_A ve n_B birey içeren örneklerden hesaplanan ortalamalar aynı popülasyon ortalamasının birbirlerinden bağımsız birer tahminidir, yani $\mu_{\bar{A}} = \mu$ ve $\mu_{\bar{B}} = \mu$ 'dür. Ortalamalar arası farka ait örnekleme dağılımının varyansı ise $\sigma_D^2 = \sigma_{\bar{A}}^2 + \sigma_{\bar{B}}^2$ 'dir. n_A ve n_B birey içeren örneklerden hesaplanan varyanslar aynı popülasyonun varyansının tahmini olduğu için $\sigma_{\bar{A}}^2 = \frac{\sigma^2}{n_A}$ ve $\sigma_{\bar{B}}^2 = \frac{\sigma^2}{n_B}$ 'dır. Buradan dağılımın varyansı;

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

ve dağılımın standart sapması;

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

$$\text{eğer } n_A = n_B = n \text{ ise } \sigma_D = \sqrt{2 \frac{\sigma^2}{n}} = \sigma \sqrt{\frac{2}{n}} \quad \dots(5.4)$$

olarak bulunur. Bu ifade ortalamalar arası farka ait örnekleme dağılımının popülasyondan hesaplanan standart sapmasıdır.

ÖRNEK:

Bir tarlada yetişen mısır koçanlarının uzunluklarının, ortalaması $\mu=85$ cm ve standart sapması, $\sigma=8$ cm olan bir normal dağılım gösterdiği bildirilmiştir. Bu koçanlardan $n_A=20$ ve $n_B=10$ koçanlık mümkün olan sayıda tesadüf örnekleri alınsa ve koçan uzunluklarının ortalamaları hesaplanırsa, hesaplanan ortalama koçan uzunlukları aynı popülasyon ortalamasının birbirlerinden bağımsız birer tahminidir.

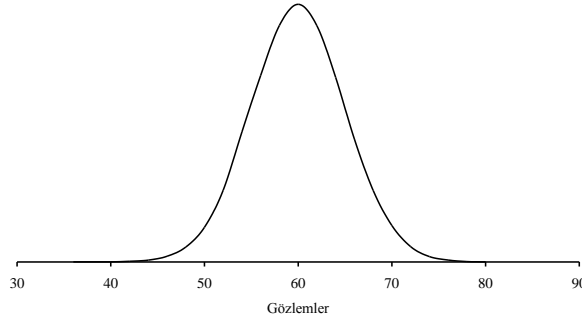
$n_A=20$ ve $n_B=10$ koçanlık örneklerden hesaplanan ortalamalar tamamen tesadüfen yan yana getirilip aralarındaki farklar hesaplanırsa bu farkların bir normal dağılım göstermesi beklenir. Çünkü koçan uzunluklarının normal dağılım gösterdiği daha önceden bilinmektedir. Bu dağılıma “**ortalamar arası farka ait örnekleme dağılımı**” denir. Bu dağılımın, ortalama, μ_D ve standart sapma, σ_D olmak üzere iki parametresi vardır. Dağılımın ortalaması, $\mu_D = \mu_{\bar{A}} - \mu_{\bar{B}} = 0$ 'dır. Standart sapması ise (5.4) numaralı eşitlikten;

$$\sigma_D = \sigma \sqrt{\frac{(n_A + n_B)}{n_A n_B}} = 8 \sqrt{\frac{20+10}{(20)(10)}} = 3.098 \text{ cm}$$

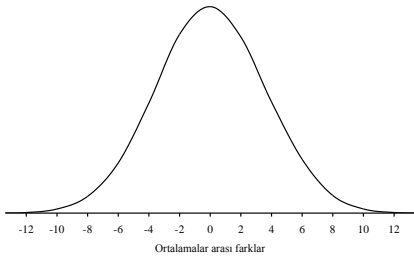
olarak bulunur.

n_A ve n_B birey içeren örnekler, normal dağılım gösteren bir populasyondan alınıyorsa, alınan örneklerin genişliği ne olursa olsun ortalamalar arası farka ait örnekleme dağılımı normal dağılım gösterir. (Şekil 5.3).

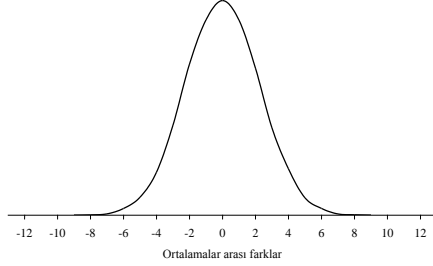
a. Normal dağılım gösteren populasyon



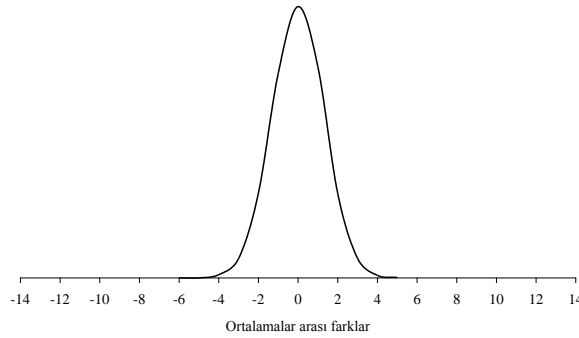
b. $n_A=n_B=4$ bireylik örnek



c. $n_A=n_B=10$ bireylik örnek



d. $n_A=n_B=30$ bireylik örnek



Şekil 5.3. Simulasyon yöntemi ile üretilmiş a. ortalaması 60 ve standart sapması 5 olan normal dağılım ve bu dağılımı gösteren populasyondan b. $n_A= n_B=4$, c. $n_A= n_B=10$ ve d. $n_A= n_B=30$ olan örneklerden elde edilen ortalamalar arası farkların dağılımları

Şekil 5.3a'da görüldüğü gibi üzerinde çalışılan populasyon $\mu=60$ $\sigma=5$ olan bir normal dağılım gösterebilir. Bu populasyondan $n_A= n_B=4$, $n_A= n_B=10$ ve $n_A= n_B=30$ bireylik örneklerden elde edilen ortalamalar arası farka ait örnekleme dağılımlarının şekli, Şekil 5.3.b, c ve d'de görülmektedir. Şekil 5.3.b, c ve d'de görüldüğü gibi eğer örnekler normal dağılım gösteren bir populasyondan alınıyor ise

örnek genişliği ne olursa olsun ortalamalar arası farka ait örnekleme dağılımı normal dağılım gösterir. Şekil 5.3'deki grafiklerden görüldüğü gibi örnek genişliği arttıkça sadece ortalamalar arası farka ait örnekleme dağılımının standart sapması azalmaktadır.

Normal dağılım gösteren ortalamalar arası farka ait örnekleme dağılımını oluşturan ortalamalar arası farklar (5.5) numaralı eşitlik kullanılarak standart normal dağılıma dönüştürülebilir. Böylece üzerinde çalışılan örneklerden hesaplanacak ortalamalar arası farkın ilgilenilen ortalamalar arası farka ait örnekleme dağılımına dahil olma olasılığı hesaplanabilir.

Örneklerden hesaplanacak ortalamalar arası farka karşılık gelen Z-değeri;

$$Z = \frac{(\bar{A} - \bar{B}) - \mu_D}{\sigma_D}$$

şeklinde yazılır. Bu eşitlikte $\mu_D=0$ olduğundan;

$$Z = \frac{(\bar{A} - \bar{B})}{\sigma_D} \quad \dots(5.5)$$

şeklinde ifade edilir.

ÖRNEK:

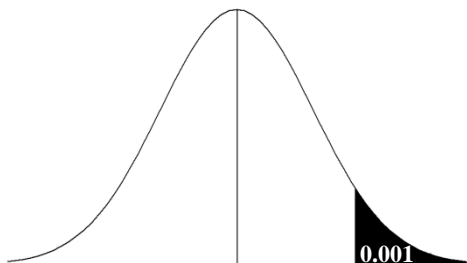
Aynı tarladan hasat edilen fasulye tanelerinin ağırlıklarının ortalaması, $\mu=0.7$ g ve standart sapması, $\sigma=0.08$ g olan bir normal dağılım gösterdiği bildirilmiştir. Bu tarladan hasat edilen $n_A=25$ fasulye tanesinin ortalaması 0.74 g ve $n_B=15$ adet fasulye tanesinin ortalaması 0.66 g olarak bulunmuştur. İki örnek ortalaması arasındaki farkın, bu populasyondan $n_A=25$ ve $n_B=15$ bireylik örneklerden elde edilecek ortalamalar arası farka ait örnekleme dağılımına dahil olma olasılığı, diğer bir deyişle bu örneklerin bu tarladan tesadüfen alınmış örnekler olma olasılığı nedir?

İstenen olasılığın hesaplanması için ilk olarak ortalaması $\mu=0.7$ g ve standart sapması $\sigma=0.08$ g olan normal bir dağılım gösteren populasyondan $n_A=25$ ve $n_B=15$ bireylik örneklerden elde edilecek ortalamalar arası farka ait örnekleme dağılımının parametreleri, $\mu_D=0$ ve (5.4) numaralı eşitlik kullanılarak standart sapması, $\sigma_D = 0.08 \sqrt{\frac{25+15}{(25)(15)}} = 0.026$ g olarak belirlenir. Ortalamalar arası farka

ait örnekleme dağılımı, normal dağılım gösterdiği için örnekten hesaplanan ortalamaya karşılık gelen Z-değeri (5.5) numaralı eşitlikten;

$$Z = \frac{(0.74 - 0.66)}{0.026} \cong 3.08$$

olarak hesaplanır. Z-değeri hesaplandıktan sonra $n_A=25$ ve $n_B=15$ bireylik örneklerden elde edilecek ortalamalar arası farkın, ortalamalar arası farka ait örnekleme dağılımına dahil olma olasılığı, diğer bir deyişle her iki örneğinde aynı tarlada yetiştirilen fasulye tanelerinden seçilmiş örnekler olma olasılığı aşağıdaki şekilde bulunur.



Söz konusu olasılık yandaki grafikte siyah taralı alandır ve $0.5 - P(0 < Z < 3.08)$ olarak hesaplanır ve 0.001 bulunur, yani örneklerin aynı fasulye tarlasında yetişen fasulyelerden alınmış olma olasılığı %0.1'dir.

5.2.1. Toplanmış Varyans

Yapılan çoğu çalışmada üzerinde çalışılan örneklerin alındığı populasyonun varyansı bilinmez. Bu durumda ortalamalar arası farka ait örnekleme dağılımının standart sapması örnekten tahmin edilir.

Varyansı bilinmeyen bir populasyondan birden fazla örnek alınarak populasyon varyansı tahmin edileceği zaman, σ^2 yerine kullanılacak populasyon varyansının en iyi tahmini, birden fazla örneklerden hesaplanan varyansların serbestlik dereceleri ile tartılı ortalaması olup **toplanmış varyans olarak adlandırılır** ve (5.6) numaralı eşitlik kullanılarak hesaplanır.

$$S^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{(n_A - 1) + (n_B - 1)} \quad \dots(5.6)$$

(2.13) numaralı eşitlikten serbestlik derecesi ile varyansın çarpımı kareler toplamını verdiği için, yani $\sum d_A^2 = (n_A - 1)S_A^2$ ve $\sum d_B^2 = (n_B - 1)S_B^2$ olduğundan, toplanmış varyans (5.6) numaralı eşitlik;

$$S^2 = \frac{\sum d_A^2 + \sum d_B^2}{(n_A - 1) + (n_B - 1)}$$

şeklinde de yazılabilir.

Bu durumda ortalamalar arası farkın standart sapmasını hesaplamak için kullanılan (5.4) numaralı eşitlikte σ^2 yerine toplanmış varyans, S^2 , kullanılarak, ortalamalar arası farka ait örnekleme dağılımının standart sapması (5.7) numaralı eşitlik kullanılarak tahmin edilir.

$$\text{Eğer } n_A \neq n_B \text{ ise } S_D = \sqrt{\frac{\sum d_A^2 + \sum d_B^2}{(n_A - 1) + (n_B - 1)} \cdot \frac{(n_A + n_B)}{(n_A n_B)}}$$

$$\text{Eğer } n_A = n_B = n \text{ ise } S_D = \sqrt{S_A^2 + S_B^2} \quad \dots(5.7)$$

şeklinde hesaplanır. Bu ifadeye ortalamalar arası farka ait örnekleme dağılımının örnekten tahmin edilen standart sapması veya kısaca **örnek ortalamaları arası farkın standart hatası denir**. İki örnek ortalaması arasındaki fark hesaplandıktan sonra, ortalamalar arası fark standart hatası ile birlikte $(\bar{A} - \bar{B}) \pm S_D$ şeklinde verilir.

ÖRNEK:

Normal dağılım gösteren bir populasyondan tesadüfen A ve B olmak üzere iki örnek alınmış ve bu örnekler için aşağıdaki bilgiler verilmiştir.

A örneği için:	$n_A=10$	$\bar{A}=25.0$	$S_A=3.21$
B örneği için:	$n_B=20$	$\bar{B}=22.2$	$S_B=3.09$

A ve B örneklerinin ortalamaları arasındaki farkın standart hatası kaçtır?

A ve B örneklerinin ortalamaları arasındaki farkın standart hatası (5.7) numaralı eşitlik kullanılarak hesaplanır. Bu eşitlik kullanılarak ortalamalar farkın standart hatasının bulunabilmesi için ilk olarak A ve B örnekleri için kareler toplamının aşağıdaki şekilde hesaplanması gerekir.

A örneği için standart sapma, $S_A=3.21$ ise A örneğinin varyansı $S_A^2 = (3.21)^2 = 10.3041$ ve $\sum d_A^2 = (n_A - 1)S_A^2$ 'den kareler toplamı $\sum d_A^2 = (10-1)(10.3041) = 92.7369$ olarak hesaplanır.

B örneği için standart sapma, $S_B=3.09$ ise B örneğinin varyansı $S_B^2 = (3.09)^2 = 9.5481$ ve $\sum d_B^2 = (n_B - 1)S_B^2$ 'den kareler toplamı $\sum d_B^2 = (20-1)(9.5481) = 181.4139$ olarak hesaplanır. Kareler toplamları hesaplandıktan sonra ortalamalar arası farkın standart hatası (5.7) numaralı eşitlikten;

$$S_D = \sqrt{\frac{92.7369 + 181.4139}{(10-1) + (20-1)} \cdot \frac{(10+20)}{(10)(20)}} \cong 1.212$$

olarak hesaplanır.

ÖRNEK:

Varyansı bilinmeyen normal dağılım gösteren bir populasyondan tesadüfen alınan 25 bireylik bir örnekte varyans 2.5 ve 40 bireylik bir örnekte varyans 2.1 olarak bulunmuştur. Populasyon varyansının en iyi tahmini nedir? Hesaplayınız.

Varyansı bilinmeyen normal dağılım gösteren bir populasyondan tesadüfen iki örnek alınmış ve varyanslar 2.5 ve 2.1 olarak hesaplanmıştır. Bu durumda populasyon varyansının en iyi tahmini toplanmış varyanstır ve (5.6) numaralı eşitlik kullanılarak;

$$S^2 = \frac{(25-1)2.5 + (40-1)2.1}{(25-1) + (40-1)} \cong 2.25$$

olarak hesaplanır.

5.3. Oranlara ait Örneklem Dağılımı

Yapılan bir araştırma yada denemede üzerinde çalışılan populasyon, istenen olayın oluş olasılığı π olan bir populasyon olabilir. Bu populasyondan belirli örnek genişliğinde (n) mümkün olan sayıda tesadüf örnekleri seçilse ve bu örneklerin her birinde istenen olayın oluş olasılığı hesaplanırsa, hesaplanan bu olasılıkların her biri birbirlerinden bağımsız olarak populasyonda istenen olayın oluş olasılığı olan π 'nin bir tahminidir ve p ile gösterilir. Bu tahminlerin bazıları π 'ye eşit, bazıları π 'den küçük ve bazıları da π 'den büyüktür. Populasyondan mümkün olan sayıdaki örnekler tamamen tesadüfen seçildiği için hesaplanan bu olasılıklar örnekten örneğe değişerek bir dağılım gösterir. Bu dağılıma ait hesaplanan olasılıklar, oran olarak ifade edildiğinden söz konusu dağılıma "**oranlara ait örneklem dağılımı**" denir.

Oranlara ait örneklem dağılımının ortalama, μ_p , ve standart sapma, σ_p , olmak üzere iki parametresi vardır. Oranlara ait örneklem dağılımının ortalaması, $\mu_p = \pi$ 'dir yani istenen olayın oluş olasılığının populasyondan hesaplanan değerine eşittir. Oranlara ait örneklem dağılımının standart sapması ise;

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \quad \dots(5.8)$$

şeklinde hesaplanır ve bu oranlara ait örneklem dağılımının populasyondan hesaplanan standart sapmasıdır.

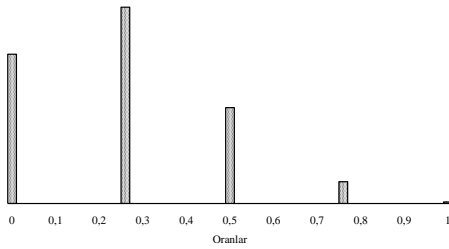
Eğer üzerinde çalışılan populasyona ait π bilinmiyor ise bu oran örnekten tahmin edilir ve (5.8) numaralı eşitlikte π yerine p değeri kullanılarak hesaplama yapılır. Bu durumda oranlara ait örnekleme dağılımının standart sapması (5.9) numaralı eşitlik kullanılarak tahmin edilir ve bu ifade oranlara ait örnekleme dağılımının örnekten tahmin edilen standart sapmasıdır.

$$S_p = \sqrt{\frac{p(1-p)}{n}} \quad \dots(5.9)$$

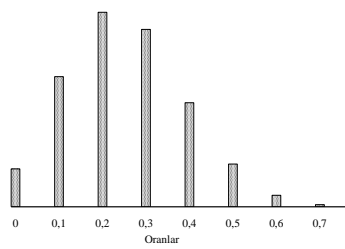
π 'si bilinen bir populasyondan elde edilecek oranlara ait örnekleme dağılımının şekli popülasyonun π 'sine ve popülasyondan çekilen örneklerin genişliğine (n) bağlıdır. Eğer popülasyonda istenen olayın oluş olasılığı, $\pi \leq 1/2$ ise oranlara ait örnekleme dağılımının normal dağılıma yaklaşabilmesi için $n \cdot \pi \geq 10$ olmalıdır. Eğer popülasyonda istenen olayın oluş olasılığı, $\pi > 1/2$ ise oranlara ait örnekleme dağılımının normal dağılıma yaklaşabilmesi için ise $n \cdot (1-\pi) \geq 10$ şartının yerine getirilmiş olması gerekir. Popülasyonun π 'sine bağlı olarak şartlar yerine getirildiği zaman bile oranlara ait örnekleme dağılımı tam olarak normal dağılım olmayıp normal dağılıma yaklaşır. Popülasyondan seçilen örneklerin örnek genişlikleri yukarıda belirtilen şartları sağlamıyorsa, elde edilecek oranlara ait örnekleme dağılımı normal dağılımı göstermez veya dağılım şekli normale yaklaşmaz.

Benzetim tekniği kullanılarak istenen olayın oluş olasılığı $\pi=0.25$ olan bir popülasyondan örnek genişliği 4, 10, 30, 40, 60 ve 100 olmak üzere 25000 örnek alınarak isten olayın oluş olasılığı hesaplanmıştır. Elde edilen oranlara ait örnekleme dağılımlarının şekli sırasıyla Şekil 5.4a, b, c, d, e ve f'de verilmiştir.

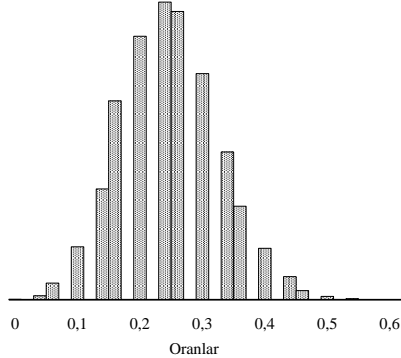
a. n=4 bireylik örnek



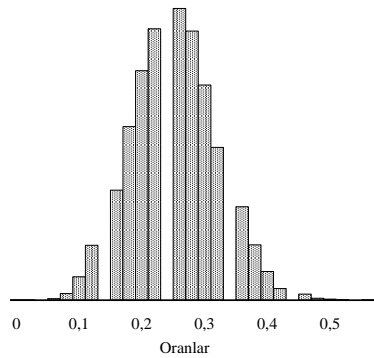
b. n=10 bireylik örnek



c. n=30 bireylik örnek

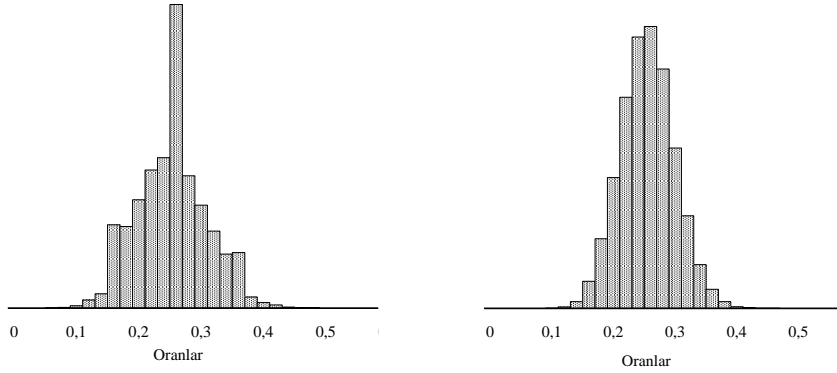


d. n=40 bireylik örnek



e. n=60 bireylik örnek

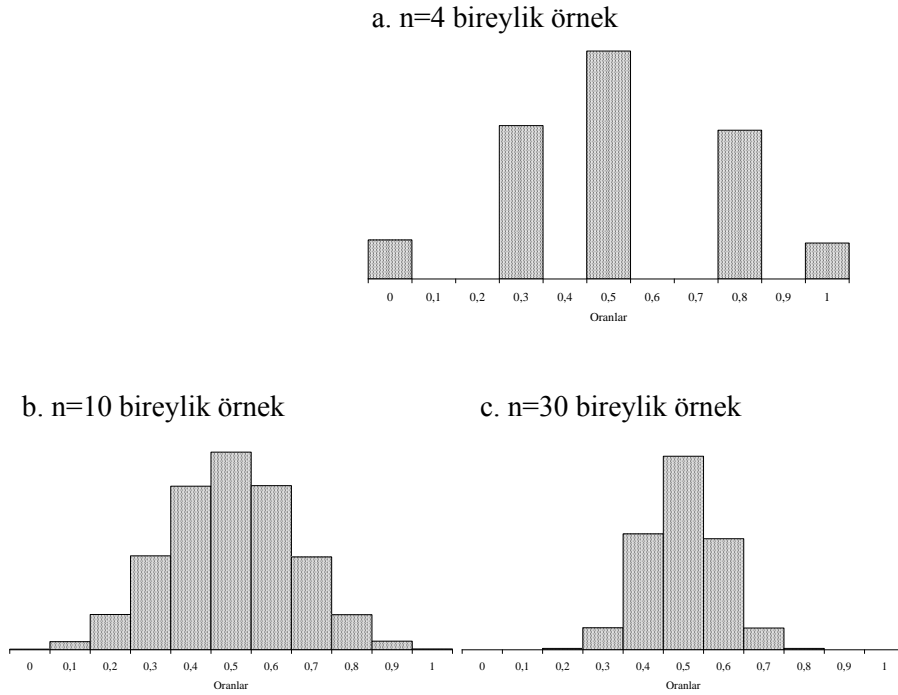
f. n=100 bireylik örnek



Şekil 5.4. Benzetim tekniği ile üretilmiş $\pi=0.25$ olan populasyondan a. $n=4$, b. $n=10$, c. $n=30$, d. $n=40$, e. $n=60$ ve f. $n=100$ bireylik örneklerden elde edilen oranlara ait örnekleme dağılımlarına ait histogramlar

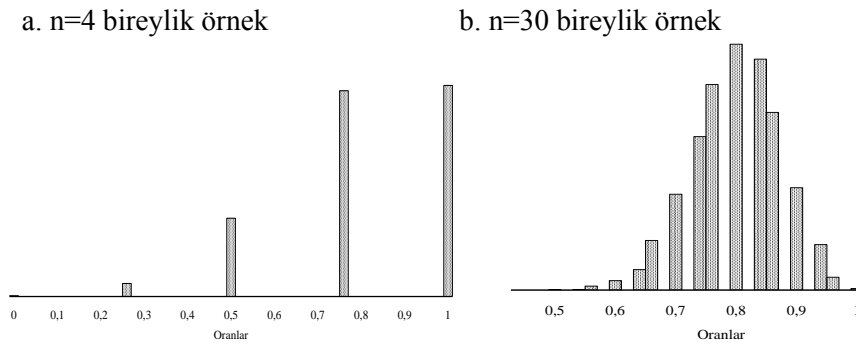
Şekil 5.4'de görüldüğü gibi $\pi=0.25$ olduğu için oranlara ait örnekleme dağılımının normal dağılıma yaklaşması için $n.\pi \geq 10$ olmalıdır. Örnek genişlikleri 4, 10 ve 30 olduğu zaman bu şart yerine gelmemekte ve Şekil 5.4a ve b'de görüldüğü gibi oranlara ait örnekleme dağılımları yatık dağılım göstermektedir. $n=30$ olduğu zaman ise oranlara ait örnekleme dağılımının yatıklığı azalmaktadır. $\pi=0.25$ olduğu için $n.\pi \geq 10$ şartının yerine getirilmiş olması gerekir. Bu sebeple oranlara ait örnekleme dağılımının normal dağılıma yaklaşması için örnek genişliğinin en az $n=10/0.25=40$ olması gerekir. Şekil 5.4d, e ve f sırasıyla örnek genişlikleri 40, 60 ve 100 olan örneklerden elde edilen oranlara ait örnekleme dağılımlarını göstermektedir. Grafiklerde görüldüğü gibi şart yerine getirildiği ve örnek genişliği arttığı zaman oranlara ait örnekleme dağılımının şekli normal dağılıma yaklaşmaktadır.

Binomiyal dağılım gösteren populasyonlarda $\pi=0.5$ olduğu zaman dağılım simetriktir. Bu nedenle Şekil 5.5.a, b ve c'de görüldüğü gibi $\pi=0.5$ olan populasyondan alınan örnek genişliği ne olursa olsun oranlara ait örnekleme dağılımı simetrik dağılım gösterir. $\pi=0.5$ olan populasyondan alınan örneklerin genişliği arttıkça Şekil 5.5.a ve b'de görüldüğü gibi oranlara ait örnekleme dağılımı normal dağılıma daha çok yaklaşacaktır.

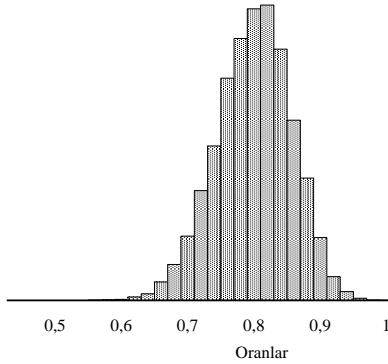


Şekil 5.5. Benzetim tekniği ile üretilmiş $\pi=0.5$ olan populyasyondan a. n=4, b. n=10 ve c. n=30 bireylik örneklerden elde edilen oranlara ait örnekleme dağılımlarına ait histogramlar

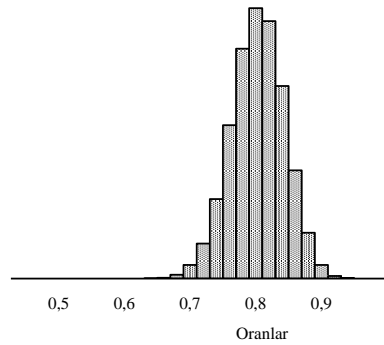
Üzerinde çalışılan populyasyonda, istenen olayın oluş olasılığı $\pi>0.5$ ise elde edilecek oranlara ait örnekleme dağılımının normal dağılıma yaklaşması için $n(1-\pi) \geq 10$ şartı yerine getirilmiş olmalıdır. Şekil 5.6'da $\pi=0.8$ olan populyasyondan 4, 30, 50 ve 100 bireylik örneklerden elde edilen oranlara ait örnekleme dağılımları görülmektedir. Örnek genişliğinin 4 ve 30 (Şekil 5.6a ve b) olması durumunda oranlara ait örnekleme dağılımı sola yatık bir dağılım göstermektedir. $\pi>0.5$ olduğu için $n(1-\pi) \geq 10$ şartı örnek genişliği en az 50 olduğu zaman sağlanmaktadır ki Şekil 5.6.c ve d'de görüldüğü gibi örnek genişliğinin 50 ve 100 olması durumunda oranlara ait örnekleme dağılımı normal dağılıma daha çok yaklaşmaktadır.



c. n=50 bireylik örnek



d. n=100 bireylik örnek



Şekil 5.6 Benzetim tekniği ile üretilmiş $\pi=0.8$ olan populasyondan a. $n=4$, b. $n=10$, c. $n=50$ ve d. $n=100$ bireylik örneklerden elde edilen oranlara ait örnekleme dağılımlarına ait histogramlar

Populasyona ait oluş olasılığının $\pi \leq 1/2$ olması durumunda $n \cdot \pi \geq 10$ ve $\pi > 1/2$ durumunda ise $n \cdot (1-\pi) \geq 10$ şartını sağlayan ve normal dağılıma yaklaşan oranlara ait örnekleme dağılımını oluşturan oranlar (5.10) numaralı eşitlik kullanılarak standart normal dağılıma dönüştürülebilir.

$$Z = \frac{p - \mu_p}{\sigma_p} \quad \dots(5.10)$$

Böylece üzerinde çalışılan örnekten hesaplanan oranın ilgilenilen oranlara ait örnekleme dağılımına dahil olma olasılığı, yani π 'si bilinen söz konusu populasyonu temsil etme olasılığı hesaplanabilir.

ÖRNEK:

Bir fakültekteki öğrencilerin %31'inin sigara içtiği bildirilmiştir. Bu fakülteden en az kaç öğrenci bulunan örnekler seçilmelidir ki oranlara ait örnekleme dağılımı normal dağılıma yaklaşsın?

Üzerinde durulan özellik sigara içenlerle ilgili olduğundan ve söz konusu fakültekteki öğrencilerin %31'i sigara içtiği için bu özelliğe ilişkin oluş olasılığı $\pi=0.31$ 'dir. Bu özelliğe ilişkin oluş olasılığı $\pi < 0.5$ olduğundan elde edilecek oranlara ait örnekleme dağılımının normal dağılıma yaklaşması için $n \cdot \pi \geq 10$ şartının yerine getirilmiş olması gerekir. Bu durumda $n(0.31)=10$ eşitliğinden her bir örnekteki olması gereken en az öğrenci sayısı $n=10/0.31=32.26$ olarak bulunur. Öğrencilerin sigara içme oranı %31 olan bu fakülteden elde edilecek oranlara ait örnekleme dağılımının normal dağılıma yaklaşabilmesi için içlerinde en az 33 öğrenci bulunan örneklerin seçilmesi gerekir.

ÖRNEK:

$\pi=0.55$ olan bir populasyondan örnek genişliği 50 olan mümkün olan sayıda rastgele örnekler alınsa, hesaplanan oranların % ne kadarının değeri 0.405 ile 0.715 arasındadır?

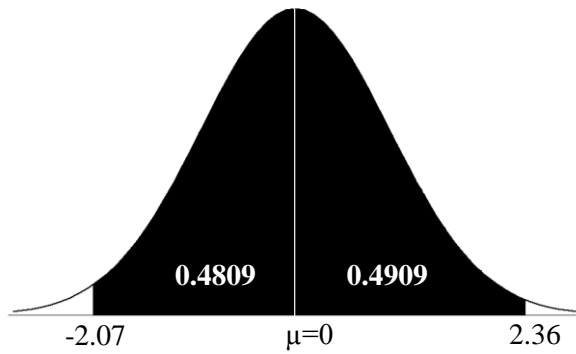
İstenen olasılığın standart normal dağılım kullanılarak hesaplanabilmesi için $\pi > 1/2$ olduğundan $n \cdot (1-\pi) \geq 10$ şartının yerine getirilmiş olması gerekir. Bu örnekte $50 \times (1-0.55)=22.5$ olarak bulunduğu için gerekli şart yerine getirilmiştir. Bu durumda 50 bireylik örneklerden hesaplanan oranlar (5.10) numaralı eşitlik kullanılarak Z-değerine dönüştürülebilir ve standart normal dağılımdan isten olasılık aşağıdaki şekilde hesaplanabilir.

Burada ilgilenilen oranlara ait örnekleme dağılımının parametreleri ortalama, $\mu_p=0.55$ ve standart sapma (5.8) numaralı eşitlikten $\sigma_p = \sqrt{\frac{0.55(1-0.55)}{50}} \cong 0.07$ olarak bulunur.

Parametreleri yukarıda verilen oranlara ait örnekleme dağılımında hesaplanan tüm oranların içinde 0.405 ile 0.715 arasında olanların olasılığı, $P(0.405 < p < 0.715)$ olup (5.10) numaralı eşitlikten;

$$P\left(\frac{(0.405-0.55)}{0.07} < Z < \frac{(0.715-0.55)}{0.07}\right) = P(-2.07 < Z < 2.36)$$

istenen olasılık aşağıdaki şekilde bulunur. Grafikte istenen olasılık taralı alandır. Bu alan $P(-2.7 < Z < 0) + P(0 < Z < 2.36)$ 'ye eşittir. Bu da, $0.4809 + 0.4909 = 0.9799$ olarak hesaplanır.



Hesaplanan bu olasılık, $\pi=0.55$ olan bir popülasyondan örnek genişliği 50 olan mümkün olan sayıda rastgele örnekler alınsa ve her birinde istenen olayın oluş olasılığı hesaplanırsa, hesaplanan olasılıkların %97.99'unun 0.405 ile 0.715 arasında olduğunu ifade eder.

5.4. Oranlar arası Farka ait Örnekleme Dağılımı

Üzerinde çalışılan olayın oluş olasılığı π olan bir popülasyondan önce n_A ve sonra n_B genişliğinde geriye iadeli olarak mümkün olan sayıda tesadüf örnekleri alınsa ve her birinde istenen olayın oluş olasılığı hesaplanırsa, hesaplanan olasılıkların her biri popülasyona ait olasılığın birer tahminidir. n_A ve n_B birey içeren örneklerden hesaplanan olasılıklar tamamen tesadüfen yan yana getirilerek olasılıklar arasındaki farklar bulunsa bu farkların sıfır olması beklenir. Hesaplanan farkların bazıları 0, bazıları 0'dan küçük ve bazıları ise 0'dan büyüktür. Mümkün olan sayıdaki örnekler söz konusu popülasyondan tamamen tesadüfen seçildiği ve tesadüfen yan yana getirilerek aralarındaki farklar bulunduğu için örnekten örneğe değişir ve bir dağılım gösterir. Bu dağılıma "**oranlar arası farka ait örnekleme**" dağılımı adı verilir.

Popülasyona ait oluş olasılığının $\pi \leq 1/2$ olması durumunda $n_A \cdot \pi \geq 10$ ve $n_B \cdot \pi \geq 10$, $\pi > 1/2$ durumunda ise $n_A \cdot (1-\pi) \geq 10$ ve $n_B \cdot (1-\pi) \geq 10$ şartı sağlanıyor ise oranlar arası farka ait örnekleme dağılımı normal dağılıma yaklaşır.

Oranlar arası farka ait örnekleme dağılımının iki parametresi vardır. Dağılımın ortalaması 0, yani, $\mu_{(p_A-p_B)}=0$ dir. Oranlar arası farka ait örnekleme dağılımının standart sapması ise (5.11) numaralı eşitlik kullanılarak hesaplanır.

$$\sigma_{(p_A-p_B)} = \sqrt{\frac{\pi(1-\pi)}{n_A} + \frac{\pi(1-\pi)}{n_B}} \quad \dots(5.11)$$

Örnekler aynı populasyondan tesadüfen alınmış örnekler olduğu zaman $\pi_A = \pi_B = \pi$ olacaktır. Bu durumda π 'nin en iyi tahmini (5.12) numaralı eşitlikte verildiği gibidir;

$$p = \frac{x_A + x_B}{n_A + n_B} \quad \dots(5.12)$$

Bu durumda, n_A ve n_B genişliğinde örneklerden hesaplanan oranlar arası farkın, n_A ve n_B genişliğinde örneklerden elde edilecek oranlar arası farka ait örnekleme dağılımına dahil olma olasılığını hesaplamak için (5.13) numaralı eşitlik kullanılır.

$$Z = \frac{(p_A - p_B) - \mu_{(p_A-pB)}}{\sigma_{(p_A-pB)}} \quad \dots(5.13)$$

(5.13) numaralı eşitlikte $\sigma_{(p_A-pB)}$ ise (5.14) numaralı eşitlikte verildiği şekilde hesaplanır:

$$\sigma_{(p_A-pB)} = \sqrt{p(1-p)\left(\frac{1}{n_A} + \frac{1}{n_B}\right)} \quad \dots(5.14)$$

ÖRNEK:

Aynı tipten ürünün yetiştiriciliğinin yapıldığı İç Anadolu Bölgesinden tesadüfen seçilen 400 çiftçiden 250'sinin yaptığı üretimle ilgili kooperatif üyeliği bulunurken, Doğu Anadolu bölgesinden tesadüfen seçilen 300 çiftçiden 100'ünün kooperatife kayıtlı olduğu saptanmıştır. Bölgeler bakımından bir üretim kooperatifine kayıt olma oranları arasındaki farkın standart sapması kaçta eşittir?

$p = \frac{250+100}{400+300} = 0.5$ olmak üzere, oranlar arasındaki farkın standart sapması (5.11b) sayılı ifadededen,

$$\sigma_{(p_A-pB)} = \sqrt{0.5(1-0.5)\left(\frac{1}{400} + \frac{1}{300}\right)} = 0.0382 \text{ olarak hesaplanır.}$$

5.5. Korelasyon Katsayısına ait Örnekleme Dağılımı

X ve Y özellikleri arasındaki korelasyon katsayısı ρ olan bir populasyondan belirli örnek genişliğinde (n) mümkün olan sayıda tesadüf örnekleri seçilse ve bu örneklerin her birinde söz konusu özellikler arasındaki korelasyon katsayısı hesaplanırsa, hesaplanan korelasyon katsayılarının her biri ρ 'nun bir tahmini olup bazıları ρ 'ya eşit, bazıları ρ 'dan küçük ve bazıları da ρ 'dan büyüktür. Populasyondan mümkün olan sayıdaki örnekler tamamen tesadüfen seçildiği için hesaplanan korelasyon katsayıları örnekten örneğe değişerek bir dağılım gösterir. Bu dağılıma “**korelasyon katsayılarına ait örnekleme dağılımı**” denir.

Korelasyon katsayısına ait örnekleme dağılımının μ_r ve σ_r olmak üzere iki parametresi vardır. Dağılımın ortalaması, μ_r , populasyondaki korelasyon katsayısına eşittir, yani $\mu_r = \rho$ 'dur. Örnekleme dağılımının standart sapması ise (5.15) numaralı eşitlik kullanılarak hesaplanır:

$$\sigma_r = \sqrt{\frac{(1-\rho^2)}{n}} \quad \dots(5.15)$$

(5.15) numaralı eşitlikte $(1-\rho^2)$ değeri, bağımlı değişkendeki (Y) varyansın, bağımsız değişken (X) ile açıklanamayan kısmıdır. Bu sebeple populasyona ait korelasyon katsayısı küçüldükçe, söz

konusu populyasyondan elde edilecek korelasyon katsayısına ait örnekleme dağılımının standart sapması büyür. Örneklerin tesadüfen alındığı populyasyonun korelasyon katsayısı, $\rho=1.0$ ise, populyasyondan alınan bütün örneklerde hesaplanan korelasyon katsayısı 1.0 olarak bulunacak ve korelasyon katsayısına ait örnekleme dağılımının standart sapması 0 olacaktır.

Aynı populyasyondan tesadüfen alınan örneklerin genişliği arttıkça korelasyon katsayısına ait örnekleme dağılımının standart sapması azalır. Çünkü örnek genişliği arttıkça örneklerden hesaplanan korelasyon katsayıları ρ 'ya yaklaşacak ve örneklerden hesaplanan korelasyon katsayıları arasındaki değişim dolayısıyla korelasyon katsayısına ait örnekleme dağılımının standart sapması azalacaktır.

Üzerinde çalışılan populyasyona ait korelasyon katsayısı bilinmiyorsa, bu populyasyondan elde edilecek korelasyon katsayısına ait örnekleme dağılımının standart sapması (5.16) numaralı eşitlik kullanılarak örnekten hesaplanır.

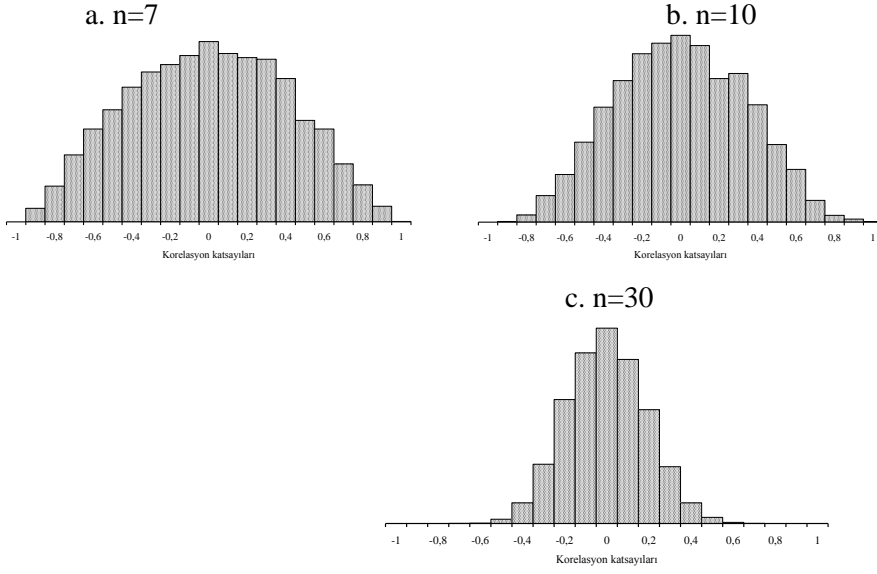
$$S_r = \sqrt{\frac{(1-r^2)}{(n-2)}} \quad \dots(5.16)$$

(5.14) numaralı eşitlik kullanılarak hesaplanan standart sapma, korelasyon katsayısına ait örnekleme dağılımının örnekten tahmin edilen standart sapması veya kısaca korelasyon katsayısının **standart hatasıdır**. Üzerinde çalışılan iki özellik arasındaki korelasyon katsayısı hesaplandıktan sonra, korelasyon katsayısı standart hatası ile birlikte $r \pm S_r$ şeklinde verilir.

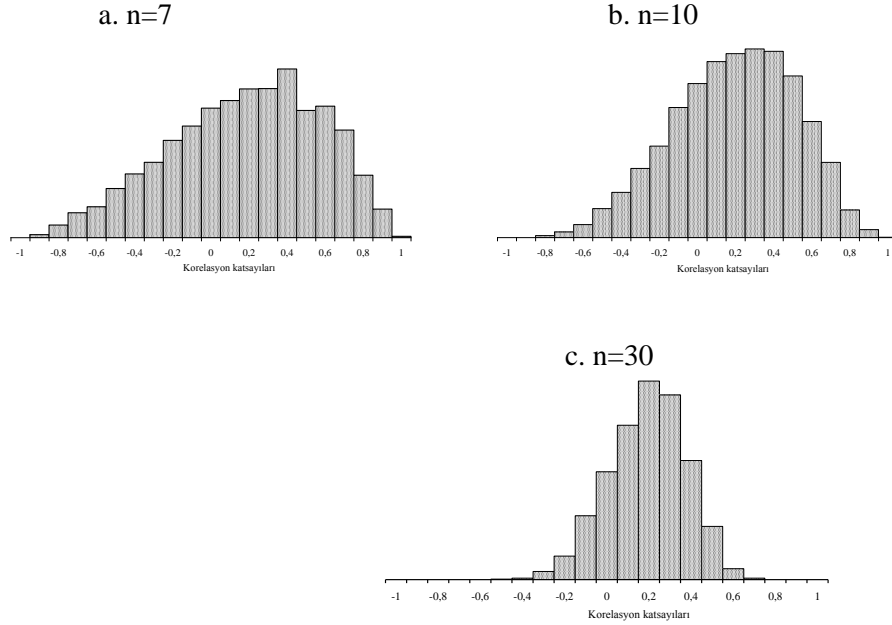
Örneklerin tesadüfen alındığı populyasyonun korelasyon katsayısı, $\rho=0$ ise populyasyondan alınan örneklerin genişliği ne olursa olsun korelasyon katsayısına ait örnekleme dağılımı normal dağılıma yaklaşır. Şekil 5.7a, b ve c'de, $\rho=0$ olan populyasyondan simülasyon yöntemi ile **10000 tane** 7, 10 ve 30 bireylik tesadüf örnekleri alınarak hesaplanan korelasyon katsayılarının dağılımı gösterilmiştir.

Örneklerin alındığı populyasyona ait korelasyon katsayısı sıfırdan farklılaştıkça elde edilecek korelasyon katsayısına ait örnekleme dağılımının şekli normal dağılımdan uzaklaşır. Şekil 5.8'de, $\rho=0.2$ olan populyasyondan simülasyon yöntemi ile **10000 tane** 7, 10 ve 30 bireylik tesadüf örnekleri alınarak hesaplanan korelasyon katsayılarının dağılımı gösterilmiştir.

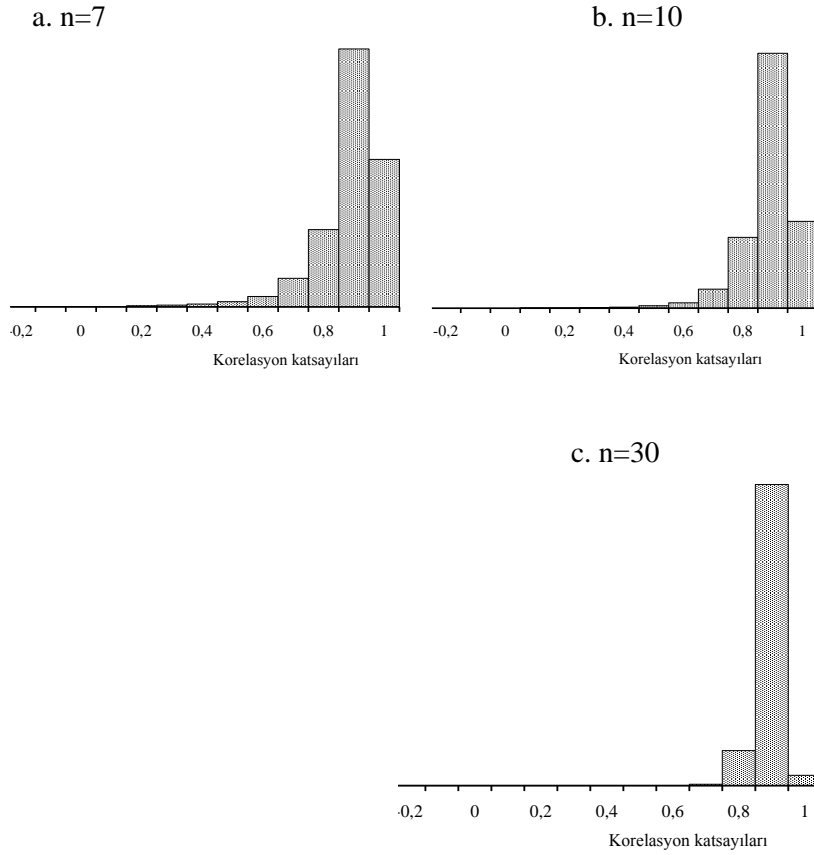
Örneklerin alındığı populyasyona ait korelasyon katsayısı büyüdükçe söz konusu populyasyondan elde edilecek korelasyon katsayısına ait örnekleme dağılımının şekli normal dağılımdan çok uzaklaşır ve çarpık dağılım haline gelir. Şekil 5.9'da, $\rho=0.9$ olan populyasyondan simülasyon yöntemi ile **10000 tane** 7, 10 ve 30 bireylik tesadüf örnekleri alınarak hesaplanan korelasyon katsayılarının dağılımı gösterilmiştir.



Şekil 5.7. Benzetim tekniği ile üretilmiş $\rho=0$ olan populasyondan a. $n=7$, b. $n=10$ ve c. $n=30$ bireylik örneklerden elde edilen korelasyon katsayılarına ait örnekleme dağılımlarının histogramları



Şekil 5.8. Benzetim tekniği ile üretilmiş $\rho=0.2$ olan populasyondan a. $n=7$, b. $n=10$ ve c. $n=30$ bireylik örneklerden elde edilen korelasyon katsayılarına ait örnekleme dağılımlarının histogramları



Şekil 5.9 Benzetim tekniği ile üretilmiş $\rho=0.9$ olan populasyondan a. $n=7$, b. $n=10$ ve c. $n=30$ bireylik örneklerden elde edilen korelasyon katsayılarına ait örnekleme dağılımlarının histogramları

Örneklerin alındığı populasyona ait korelasyon katsayısı, $\rho \neq 0$ olduğu zaman bu populasyondan elde edilecek korelasyon katsayısına ait örnekleme dağılımının normal dağılım göstermesi için her bir örnekten hesaplanan korelasyon katsayısının (5.17) numaralı eşitlik kullanılarak Z_r -değerlerine dönüştürülmesi gerekir.

$$Z_r = \frac{1}{2} \log_e \frac{(1+r)}{(1-r)} = 1.1513 \log \frac{(1+r)}{(1-r)} \quad \dots(5.17)$$

Hesaplanan Z_r -değerlerinin, ortalaması;

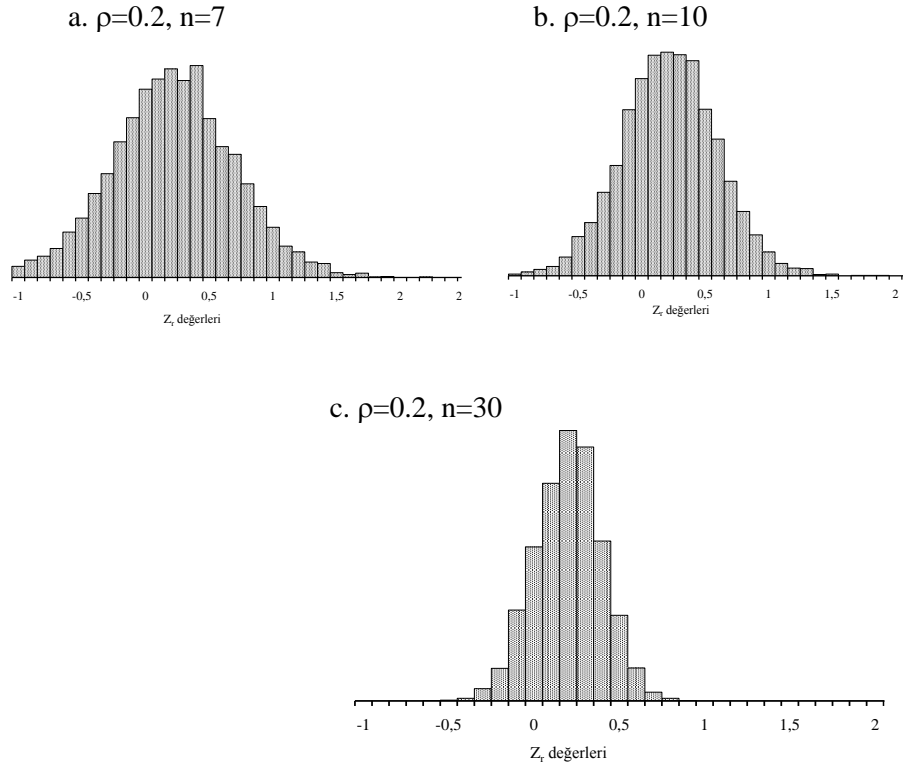
$$\mu_{Z_r} = \frac{1}{2} \log_e \frac{(1+\rho)}{(1-\rho)} \quad \dots(5.18)$$

ve standart sapması;

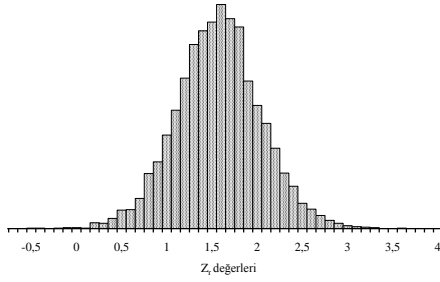
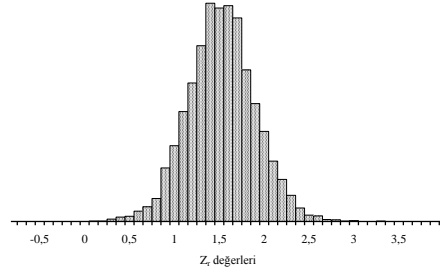
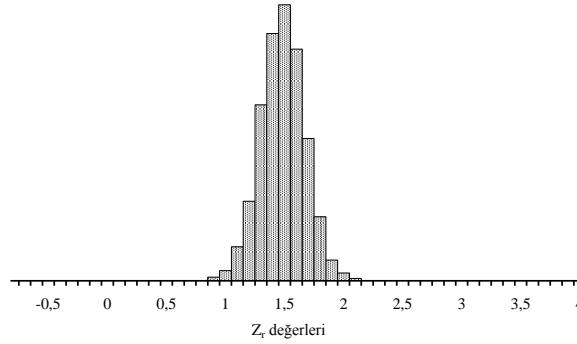
$$\sigma_{Z_r} = \sqrt{\frac{1}{(n-3)}} \quad \dots(5.19)$$

şeklinde hesaplanır. Her bir örnekten bu şekilde hesaplanan korelasyon katsayısına karşılık gelen Z_r -değerleri, populasyona ait korelasyon katsayısı ne olursa olsun normal dağılıma yaklaşır. $\rho=0.2$ ve

$\rho=0.9$ olan populasyonlardan, 7, 10 ve 30 bireylik tesadüf örnekleri alındığında korelasyon katsayılarına karşılık gelen Z_r -değerlerinin dağılımları Şekil 5.10 ve Şekil 5.11’de görülmektedir.



Şekil 5.10 Benzetim tekniği ile üretilmiş $\rho=0.2$ olan populasyondan a. $n=7$, b. $n=10$ ve c. $n=30$ bireylik örneklerden elde edilen Z_r -değerlerinin dağılımlarının histogramları

a. $\rho=0.9$, $n=7$ b. $\rho=0.9$, $n=10$ c. $\rho=0.9$, $n=30$ 

Şekil 5.11 Benzetim tekniği ile üretilmiş $\rho=0.9$ olan populasyondan a. $n=7$, b. $n=10$ ve c. $n=30$ bireylik örneklerden elde edilen Z_r -değerlerinin dağılımlarının histogramları

Korelasyon katsayısı, $\rho \neq 0$ olan bir populasyondan alınan örneklerden hesaplanan korelasyon katsayıları Z_r -değerlerine dönüştürülerek korelasyon katsayılarına ait örnekleme dağılımının şeklinin normal dağılıma yaklaşması sağlandıktan sonra korelasyon katsayısına ait örnekleme dağılımını oluşturan korelasyon katsayıları (5.20) numaralı eşitlik kullanılarak standart normal dağılıma dönüştürülebilir.

$$Z = \frac{Z_r - \mu_{Z_r}}{\sigma_{Z_r}} \quad \dots(5.20)$$

Böylece üzerinde çalışılan örnekten hesaplanan korelasyon katsayısının ilgilenilen korelasyon katsayısına ait örnekleme dağılımına dahil olma olasılığı, yani ρ 'su bilinen söz konusu populasyonu temsil etme olasılığı hesaplanabilir.

ÖRNEK:

Bir bölgede yaşayan sağlıklı yetişkinlerde yaş ile kan basıncı arasındaki korelasyon katsayısının 0.75 olduğu bildirilmiştir. Bu bölgeden tesadüfen seçildiği ileri sürülen 150 bireyde yaş ile kan basıncı arasındaki korelasyon katsayısı 0.82 olarak hesaplanmıştır. Söz konusu 150 bireyin yaş ile kan basıncı arasındaki korelasyon katsayısının 0.75 olduğu bildirilen bölgeden tesadüfen seçilmiş bireyler olma olasılığı nedir?

Bir bölgede yaşayan sağlıklı bireylerde yaş ile kan basıncı arasındaki korelasyon katsayısının 0.75 olduğu bildirilmiştir. 150 bireyin bu populasyondan tesadüfen alınmış bireyler olduğu öne sürüldüğüne göre soruda, 150 bireylik örneğin % kaç olasılıkla $\rho \neq 0$ olan söz konusu populasyondan alınmış olduğu araştırılmaktadır. Daha önce açıklandığı gibi $\rho \neq 0$ olduğu zaman elde edilecek korelasyon katsayısına ait örnekleme dağılımının normal dağılım göstermesi için her bir örnekten hesaplanan korelasyon katsayısının (5.17) numaralı eşitlik kullanılarak Z_r -değerine dönüştürülmesi gerekmektedir.

$$Z_r = \frac{1}{2} \log_e \frac{(1+r)}{(1-r)} = \frac{1}{2} \log_e \frac{(1+0.82)}{(1-0.82)} = 1.15682$$

Hesaplanan Z_r -değerlerinin, ortalaması ve standart sapması (5.18) ve (5.19) numaralı eşitliklerden;

$$\mu_{Z_r} = \frac{1}{2} \log_e \frac{(1+\rho)}{(1-\rho)} = \frac{1}{2} \log_e \frac{(1+0.75)}{(1-0.75)} = 0.97296$$

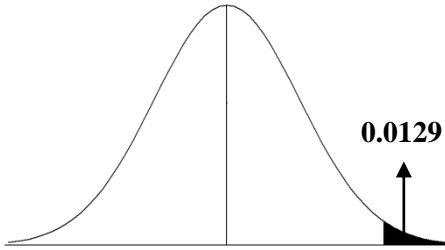
ve

$$\sigma_{Z_r} = \sqrt{\frac{1}{(n-3)}} = \sqrt{\frac{1}{(150-3)}} = 0.0825$$

olarak hesaplanır. Daha sonra (5.20) numaralı eşitlik kullanılarak Z -değeri aşağıdaki şekilde hesaplanır.

$$Z = \frac{1.15682 - 0.97296}{0.0825} = 2.2286 \approx 2.23$$

150 bireylik örneğin söz konusu populasyondan tesadüfen alınmış olma olasılığı $P(Z > 2.23)$ yani aşağıdaki grafikte istenen olasılık taralı alandır.



Bu alan, $0.5 - P(0 < Z < 2.23)$ 'ye eşittir. Z -değerlerinin 0.4871'i 0 ile 2.23 arasındadır (Tablo A) ve taralı alan $0.5 - 0.4871 = 0.0129$ olarak hesaplanır. Bu olasılık, yaş ile kan basıncı arasındaki korelasyon katsayısı 0.82 olarak hesaplanan 150 bireylik örneğin $\rho = 0.75$ olan populasyondan tesadüfen seçilmiş bir örnek olma olasılığının %1.29 olduğunu gösterir.

ÖRNEK:

Bir çeşit ayçiçeği bitkisinde çiçek çapı ile tane sayısı arasındaki korelasyon katsayısının 0.70 olduğu ileri sürülmüştür. Bu çeşitten olduğu ileri sürülen 15 ayçiçeği bitkisinde çiçek çapı ile tane sayısı arasındaki korelasyon katsayısı 0.61 olarak bulunmuştur. 15 ayçiçeği bitkisinde hesaplanan bu korelasyon katsayısının, çiçek çapı ile tane sayısı arasındaki korelasyon katsayısının 0.70 olduğu ileri sürülen çeşitten olma olasılığı nedir?

Bir çeşit ayçiçeği bitkisinde çiçek çapı ile tane sayısı arasındaki korelasyon katsayısının 0.70 olduğu ileri sürülmüştür. 15 ayçiçeği bitkisinin söz konusu çeşitten olduğu öne sürüldüğüne göre 15

bitkilik örneğin % kaç olasılıkla $\rho \neq 0$ olan populasyondan alınmış olduğu araştırılmaktadır. Daha önce açıklandığı gibi $\rho \neq 0$ olduğu zaman elde edilecek korelasyon katsayısına ait örnekleme dağılımının normal dağılım göstermesi için her bir örnekten hesaplanan korelasyon katsayısının (5.17) numaralı eşitlik kullanılarak Z_r -değerine dönüştürülmesi gerekmektedir.

$$Z_r = \frac{1}{2} \log_e \frac{(1+r)}{(1-r)} = \frac{1}{2} \log_e \frac{(1+0.61)}{(1-0.61)} = 0.7089$$

Hesaplanan Z_r -değerlerinin, ortalaması ve standart sapması (5.18) ve (5.19) numaralı eşitliklerden;

$$\mu_{Z_r} = \frac{1}{2} \log_e \frac{(1+\rho)}{(1-\rho)} = \frac{1}{2} \log_e \frac{(1+0.7)}{(1-0.7)} = 0.8673$$

ve

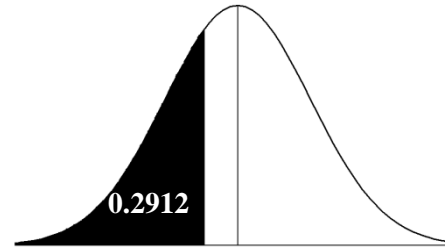
$$\sigma_{Z_r} = \sqrt{\frac{1}{(n-3)}} = \sqrt{\frac{1}{(15-3)}} = 0.2887$$

olarak hesaplanır. Daha sonra (5.20) numaralı eşitlik kullanılarak Z-değeri aşağıdaki şekilde hesaplanır.

$$Z = \frac{0.7089 - 0.8673}{0.2887} = -0.5487 \cong -0.55$$

15 bitkilik örneğin söz konusu çeşitten tesadüfen alınmış olma olasılığı $P(Z < -0.55)$ yani aşağıdaki grafikte istenen olasılık taralı alandır.

Bu alan, $0.5 - P(-0.55 < Z < 0)$ olarak hesaplanır. Z-değerlerinin 0.2088'i -0.55 ile 0 arasındadır (Tablo A). Dahil olma olasılığı, yani taralı alan $0.5 - 0.2088 = 0.2912$ olarak bulunur. Bu olasılık, 15 ayçiçeği bitkisinin, çiçek çapı ile tane sayısı arasındaki korelasyon katsayısının 0.70 olduğu ileri sürülen çeşitten olma olasılığının %29.12 olduğunu gösterir.



5.6. Korelasyon Katsayıları Arasındaki Farka Ait Örnekleme Dağılımı

Üzerinde çalışılan iki özellik arasındaki korelasyon katsayısının ρ olduğu bir populasyondan; önce n_A , sonra da n_B genişliğinde geriye iadeli olmak üzere mümkün olan sayıda tesadüf örnekleri alınsa ve her bir örnekte bu iki özellik arasındaki korelasyon katsayısı hesaplınsa, hesaplanan korelasyon katsayılarının her biri populasyona ait korelasyon katsayısının, yani ρ 'nun, birer tahmini olup, ρ kadar çıkmaları beklenir. Dolayısıyla, n_A ve n_B birey içeren örneklerden hesaplanan bu korelasyon katsayıları tamamen tesadüfen yan yana getirilerek aralarındaki farklar bulınsa, bu farkların sıfır olması beklenir. Hesaplanan farkların bazıları 0, bazıları 0'dan küçük ve bazıları ise 0'dan büyüktür. Mümkün olan sayıdaki örnekler, söz konusu populasyondan tamamen tesadüfen seçildiği ve tesadüfen yan yana getirilerek farklar bulunduğu için örnekten örneğe değişerek bir dağılım gösterir. Bu dağılıma "korelasyon katsayıları arasındaki farka ait örnekleme" dağılımı denir.

Bu dağılımın parametrelerinin belirlenebilmesi için her bir örnekten hesaplanmış korelasyon katsayısına karşılık gelen Z_r -değerinin hesaplanması gerekir. Böylece korelasyon katsayıları arasındaki farka ait örnekleme dağılımının normal dağılıma yaklaşması sağlanır. Korelasyon katsayıları arasındaki farka ait örnekleme dağılımının iki parametresi vardır. **Bu dağılımın ortalaması 0, yani, $\mu_{(Z_{r1}-Z_{r2})} = 0$** ve standart sapması (5.21) numaralı ifadeye göre hesaplanır.

$$\sigma_{(Z_{rA}-Z_{rB})} = \sqrt{\frac{1}{(n_A - 3)} + \frac{1}{(n_B - 3)}} \quad \dots(5.21)$$

Korelasyon katsayısı ρ olan bir populasyondan alınan n_A ve n_B genişliğindeki örneklerden hesaplanan korelasyon katsayıları Z_r -değerlerine dönüştürülerek korelasyon katsayıları arasındaki farka ait örnekleme dağılımının şeklinin normal dağılıma yaklaşması sağlandıktan sonra, korelasyon katsayıları arasındaki farka ait örnekleme dağılımını oluşturan farklar (5.22) numaralı ifade kullanılarak standart normal dağılıma (Z-Dağılımına) dönüştürülür. Hesaplanan bu Z-İstatistiğine göre de iki korelasyon katsayısı arasındaki farkın tesadüften ileri gelip gelmediğine karar verilir.

$$Z = \frac{Z_{rA} - Z_{rB}}{\sqrt{\frac{1}{(n_A - 3)} + \frac{1}{(n_B - 3)}}} \quad \dots(5.22)$$

ÖRNEK :

Bir büyükbaş sürüsünden tesadüfen seçilen 2 yaşlı $n_A = 20$ inekte gövde derinliği ile canlı ağırlık arasındaki korelasyon katsayısı, $r_A = 0.60$ olarak, aynı süreden tesadüfen seçilen 3 yaşlı $n_B = 30$ inekte ise söz konusu iki özellik arasındaki korelasyon katsayısı $r_B = 0.69$ olarak hesaplanmıştır. Korelasyon katsayıları arasındaki farkın, aynı sürüden söz konusu örnek genişlikleri ile elde edilecek korelasyon katsayıları arasındaki farka ait örnekleme dağılımına dahil olma olasılığı nedir?

Korelasyon katsayıları arasındaki farka ait örnekleme dağılımının normal dağılıma yaklaşması için n_A ve n_B genişliğindeki örneklerden hesaplanan korelasyon katsayılarının (5.17) numaralı eşitlik kullanılarak aşağıda görüldüğü gibi Z_r -değerlerine dönüştürülmesi gerekir.

$$Z_{rA} = \frac{1}{2} \log_e \frac{(1+0.60)}{(1-0.60)} = 1.1513 \log \frac{(1+0.60)}{(1-0.60)} = 0.6932$$

$$Z_{rB} = \frac{1}{2} \log_e \frac{(1+0.69)}{(1-0.69)} = 1.1513 \log \frac{(1+0.69)}{(1-0.69)} = 0.8480$$

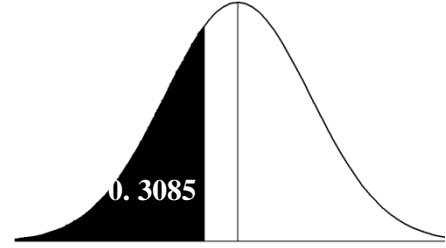
Korelasyon katsayılarına karşılık gelen Z_r -değerleri hesaplandıktan sonra (5.22) numaralı eşitlik kullanılarak, korelasyon katsayıları arasındaki farka karşılık gelen Z-değeri;

$$Z = \frac{Z_{rA} - Z_{rB}}{\sqrt{\frac{1}{(n_A - 3)} + \frac{1}{(n_B - 3)}}} = \frac{0.6932 - 0.8480}{\sqrt{\frac{1}{(20-3)} + \frac{1}{(30-3)}}} = 0.4999 \cong -0.5$$

olarak hesaplanır.

İki korelasyon katsayısı arasındaki farkın aynı sürüden elde edilecek korelasyon katsayıları arasındaki farka ait örnekleme dağılımına dahil olma olasılığı $P(Z < -0.5)$ aşağıdaki grafikte taralı olarak gösterilen alandır.

Bu alan, $0.5 - P(-0.5 < Z < 0)$ olarak hesaplanır. Z-değerlerinin 0.1915'i -0.5 ile 0 arasındadır (Tablo A). Korelasyon katsayıları arasındaki farkın, söz konusu örnekleme dağılımına dahil olma olasılığı, yani grafikte taralı alan $0.5 - 0.1915 = 0.3085$ olarak hesaplanır. Dolayısıyla 2 ve 3 yaşlı ineklerde söz konusu iki özellik arasındaki korelasyon katsayıları arasındaki farkın, aynı sürüden aynı özellikler için elde edilebilecek sonsuz sayıda korelasyon katsayıları arasındaki farka ait örnekleme dağılımına dahil olma olasılığının %30.85 olduğuna karar verilebilir.



5.7. Regresyon Katsayısına ait Örnekleme Dağılımı

Y özelliğinin X özelliğine göre regresyon katsayısı β_{yx} olan bir populasyondan belirli örnek genişliğinde (n) mümkün olan sayıda tesadüf örnekleri seçilse ve bu örneklerin her birinde b_{yx} hesaplanırsa, hesaplanan regresyon katsayılarının her biri β_{yx} 'in bir tahminidir. Bunlardan bazıları β_{yx} 'e eşit, bazıları β_{yx} 'den küçük ve bazıları da β_{yx} 'den büyüktür. Söz konusu populasyondan mümkün olan sayıdaki örnekler tamamen tesadüfen seçildiği için hesaplanan regresyon katsayıları örnekten örneğe değişerek bir dağılım gösterir. Bu dağılıma “**regresyon katsayılarına ait örnekleme dağılımı**” denir.

Regresyon katsayısına ait örnekleme dağılımının ortalama ve standart sapma olmak üzere iki parametresi vardır. Bu dağılımın ortalaması β_{yx} 'ye eşittir. Varyansı ise (5.23) numaralı ifade kullanılarak hesaplanır.

$$\sigma_b^2 = \frac{\sigma_e^2}{\sum d_x^2} \quad \dots(5.23)$$

(5.23) numaralı ifadede σ_e^2 , Y bağımlı değişkenine ait gözlem değerlerinin regresyon doğrusundan sapma kareler ortalamasıdır. Eğer populasyona ait, $Y = \alpha + \beta X + e$ modelinin α ve β parametreleri biliniyorsa σ_e^2 , (5.24) numaralı eşitlik kullanılarak hesaplanır.

$$\sigma_e^2 = \frac{\sum e^2}{N} = \frac{\sum (Y - \hat{Y})^2}{N} \quad \dots(5.24)$$

(5.24) numaralı eşitlikte $\sum e^2$, gözlemlerin regresyon doğrusundan sapma kareler toplamı olup (5.25) numaralı eşitlikte verildiği gibidir.

$$\sum (Y - \hat{Y})^2 = \sum e^2 = \sum d_y^2 - \beta_{yx} \sum d_x d_y \quad \dots(5.25)$$

(5.25) numaralı eşitlikte, β_{yx} yerine, eşiti olan $\frac{\sum d_x d_y}{\sum d_x^2}$ yazılarak (5.25) numaralı eşitlik (5.26)

numaralı eşitlikte verildiği gibi düzenlenebilir.

$$\sum e^2 = \sum d_y^2 - \frac{(\sum d_x d_y)^2}{\sum d_x^2} \quad \dots(5.26)$$

Üzerinde çalışılan populasyona ait α ve β parametreleri bilinmiyorsa, regresyon katsayısına ait varyans (5.27) numaralı eşitlikte verildiği gibi hesaplanır.

$$S_b^2 = \frac{S_e^2}{\sum d_x^2} \quad \dots(5.27)$$

(5.27) numaralı eşitlikte S_e^2 , regresyon doğrusundan sapma kareler ortalamasıdır ve (5.28) numaralı eşitlik kullanılarak hesaplanır.

$$S_e^2 = \frac{\sum (Y - \hat{Y})^2}{n - 2} = \frac{\sum d_y^2 - b_{yx} \sum d_x d_y}{n - 2} = \frac{\sum d_y^2 - \frac{(\sum d_x d_y)^2}{\sum d_x^2}}{n - 2} \quad \dots(5.28)$$

Regresyon katsayısının standart hatası ise (5.29) numaralı eşitlikte verildiği gibidir.

$$S_b = \sqrt{\frac{S_e^2}{\sum d_x^2}} \quad \dots(5.29)$$

5.8. Regresyon Katsayıları Arasındaki Farka ait Örneklemeye Dağılımı

Y'nin X'e göre regresyon katsayısı β_{yx} olan bir populasyondan n_A ve n_B örnek genişliğinde geriye iadeli olarak mümkün olan sayıda tesadüf örnekleri alınsa ve her birinde Y'nin X'e göre regresyon katsayıları hesaplanırsa, hesaplanan regresyon katsayılarının her biri populasyona ait regresyon katsayısının, yani β_{yx} 'nin, birer tahminidir. n_A ve n_B birey içeren örneklerden hesaplanan regresyon katsayıları tamamen tesadüfen yan yana getirilerek katsayılar arasındaki farklar bulunsu bu farkların sıfır olması beklenir. Hesaplanan farkların bazıları 0, bazıları 0'dan küçük ve bazıları ise 0'dan büyüktür. Mümkün olan sayıdaki örnekler, söz konusu populasyondan tamamen tesadüfen seçildiği ve tesadüfen yan yana getirilerek farklar bulunduğu için örnekten örneğe değişerek bir dağılım gösterir. Bu dağılıma "**regresyon katsayıları arasındaki farka ait örnekleme**" dağılımı denir.

Regresyon katsayıları arasındaki farka ait örnekleme dağılımının ortalaması, $\mu_{(b_A-b_B)}=0$ ve varyansı (5.30) numaralı eşitlikte verildiği gibidir.

$$\sigma_{(b_1-b_2)}^2 = \sigma_{b_A}^2 + \sigma_{b_B}^2 \quad \dots(5.30)$$

(5.30) numaralı eşitlikte $\sigma_{b_A}^2$ ve $\sigma_{b_B}^2$ (5.23) numaralı eşitlik kullanılarak hesaplanır. $\sigma_{b_A}^2$ ve $\sigma_{b_B}^2$ 'nin (5.23) numaralı eşitlikteki değerleri kullanılarak (5.30) numaralı eşitlik (5.31) numaralı eşitlikte verildiği gibi düzenlenebilir.

$$\sigma_{(b_1-b_2)}^2 = \left(\frac{\sigma_e^2}{\sum d_{xA}^2} + \frac{\sigma_e^2}{\sum d_{xB}^2} \right) \quad \dots(5.31)$$

n_A ve n_B örnek genişliğinde örnekler aynı populasyondan tesadüfen alındığı için σ_e^2 'ler aynıdır. Bu sebeple (5.31) numaralı eşitlikte σ_e^2 ortak parantezine alınabilir. (5.31) numaralı eşitlikte verilen regresyon katsayıları arasındaki farka ait örnekleme dağılımının varyansının karekökü, dağılımın standart sapmasıdır ve (5.32) numaralı eşitlikte verilmiştir.

$$\sigma_{(b_A-b_B)} = \sqrt{\sigma_e^2 \left(\frac{1}{\sum d_{xA}^2} + \frac{1}{\sum d_{xB}^2} \right)} \quad \dots(5.32)$$

n_A ve n_B örnek genişliğinde örneklerin alındığı populasyona ait β_{yx} bilinmiyor ise regresyon katsayıları arasındaki farka ait örnekleme dağılımının varyansı (5.33) numaralı eşitlikte verildiği gibidir.

$$S_{(b_A-b_B)}^2 = \frac{S_e^2}{\sum d_{xA}^2} + \frac{S_e^2}{\sum d_{xB}^2} \quad \dots(5.33)$$

(5.33) numaralı eşitlik düzenlenerek ve karekökü alınarak dağılımın standart sapması, yani regresyon katsayıları arasındaki farkın standart hatası (5.34) numaralı eşitlikteki gibi bulunur.

$$S_{(b_A-b_B)} = \sqrt{S_e^2 \left(\frac{1}{\sum d_{xA}^2} + \frac{1}{\sum d_{xB}^2} \right)} \quad (5.34)$$

(5.34) numaralı eşitlikte S_e^2 , n_A ve n_B örnek genişliğinde örneklerden (5.28) numaralı eşitlik kullanılarak hesaplanan regresyondan sapma kareler ortalamalarının serbestlik dereceleri ile tartılı ortalamasıdır ve (5.35) numaralı eşitlik kullanılarak hesaplanır. S_e^2 , (5.32) numaralı eşitlikteki populasyona ait regresyondan sapma kareler ortalamasının, σ_e^2 , en iyi tahminidir.

$$S_e^2 = \frac{\sum d_{eA}^2 + \sum d_{eB}^2}{(n_A - 2) + (n_B - 2)} \quad \dots(5.35)$$

5.8. SORULAR

1. Örnekleme dağılımı nedir? Açıklayınız.
2. Örnekleme dağılımları ne amaçla kullanılır? Açıklayınız.
3. Oranlara ve korelasyon katsayılarına ait örnekleme dağılımının parametreleri ve dağılım şekli hakkında ne söylenebilir? Açıklayınız.
4. $\pi=0.6$ olan bir populasyondan örnek genişliği 120 olan mümkün olan sayıda rasgele örnekler alınsa;
 - a. Hesaplanan oranların % ne kadarının değeri 0.5 ile 0.75 arasındadır?
 - b. Hesaplanan oranların % ne kadarının değeri 0.7'den daha düşüktür?
5. Ortalaması $\mu=60$, standart sapması da $\sigma=6.3$ olan bir populasyondan alınıp alınmadığı bilinmeyen 16 kişilik bir örnekte ortalama 56.4 olarak hesaplanmış ise bu örneğin ortalamalara ait örnekleme dağılımına dahil olma olasılığı nedir?
6. Ortalaması 40 ve standart sapması 3 olan normal dağılım gösteren bir populasyondan tesadüfen alındığı ileri sürülen 36 bireylik bir örneğin ortalaması 41,2 olarak hesaplanmıştır. Bu örneğin ortalamalara ait örnekleme dağılımına dahil olma olasılığı nedir?

7. Ortalaması $\mu = 500$ kg, varyansı $\sigma^2 = 2500$ kg olan normal dağılım gösteren bir populasyondan her defasında 25 birey bulunan çok fazla sayıda rastgele örnekler çekilse ve çekilen bu örneklerin her birinde ortalama hesaplanırsa, bu hesaplanan ortalamaların ortalamasının ve standart sapmasının kaç olması beklenir? Nasıl bir dağılım gösterirler?
8. $\pi = 0.4$ olan binomiyal bir populasyondan her defasında 25 birey bulunan rastgele örnekler çekilse ve bu örneklerin her birinde p'ler hesaplanırsa, bu p'lerin gösterdiği dağılımın ortalama ve standart sapması nedir? Hesaplayınız.
9. $\pi = 0.75$ olan binomiyal bir populasyondan her defasında en az kaç birey bulunan örnekler seçilmelidir ki oranlara ait örnekleme dağılımı normal dağılıma yaklaşsın? Açıklayınız.
10. $\pi = 0.4$ olan binomiyal bir populasyondan her defasında 25 birey bulunan rastgele örnekler çekilse ve bu örneklerin her birinde p'ler hesaplanırsa, bu p'lerin gösterdiği dağılıma ne isim verilir ve parametreleri nelerdir?
11. Varyansı bilinmeyen normal dağılım gösteren bir populasyondan tesadüfen alınan 25 bireylik bir örnekte varyans 2.5 ve 40 bireylik bir örnekte varyans 2.1 olarak bulunmuştur. Populasyon varyansının en iyi tahmini nedir? Hesaplayınız.
- 12) Ortalaması $\mu = 500$ kg ve standart sapması $\sigma = 50$ kg olan normal dağılım gösteren bir populasyondan her defasında 25 birey bulunan çok fazla sayıda rastgele örnekler çekilse ve çekilen bu örneklerin her birinde ortalama hesaplanırsa;
 - a. Bu hesaplanan ortalamalar nasıl bir dağılım gösterir ve parametreleri nelerdir?
 - b. Bu populasyondan alındığı ileri sürülen 16 bireylik bir örnekte ortalama 520 kg olarak hesaplanmışsa bu örneğin söz konusu populasyonu temsil etme olasılığı nedir?
- 13) A ve B gibi iki muamelenin denendiği bir deneme sonunda, $n_A = n_B = 10$, $\bar{A} = 25$, $S_A = 3.21$ ve $\bar{B} = 22.2$, $S_B = 3.09$ olarak bulunmuştur. Bu iki ortalama arasındaki farkın standart hatası kaçtır?

HİPOTEZ KONTROLLERİ

6.1. GİRİŞ

Yapılan bütün araştırmaların en önemli amacı, üzerinde çalışılan populasyonlar hakkında güvenilir ve yorumlanabilir sonuçlara ulaşmaktır. İstatistik bilimi, araştırmalardan elde edilen verilere dayanarak, yani örneklerden hesaplanan istatistikleri kullanarak populasyon parametreleri hakkındaki sonuçlara (**making inference**), populasyon parametrelerinin tahmini veya populasyon parametreleri ile ilgili öne sürülen hipotezlerin kontrol edilmesi ile ulaşır.

Populasyonu temsil eden örneklerden hesaplanan istatistikler populasyon parametrelerinin tahminidir. Bu tahminlere nokta tahminleri de denir. Eğer araştırmacı arzu ederse fonksiyonları bilinen örnekleme dağılımlarından yararlanarak ilgilendiği populasyon parametresine ilişkin aralık tahminleri de yapabilir.

Populasyon parametreleri hakkında bilgi edinmenin diğer bir yolu ise **hipotez kontrolü**, yani populasyon parametreleri ile ilgili öne sürülen hipotezlerin kontrol edilmesidir. Araştırmacı ilgilendiği konu ile ilgili bir hipotez öne sürer, yürüttüğü araştırma sonunda veri elde eder ve öne sürdüğü hipotezi elde ettiği bu veriler yardımıyla test eder. Yaptığı hipotez kontrolü sonucunda, öne sürdüğü hipotez ile topladığı veriler arasında bir uyum olduğuna karar verirse öne sürdüğü hipotezin geçerli olduğuna karar verir ve hipotezi kabul eder. Bu yaklaşıma genel olarak TÜMEVARIM (parçadan bütüne ulaşma) adı verilmektedir.

Hipotez kontrolü: n bireylik bir örnekten hesaplanan bir istatistiğin, bu örneğin ait olduğu populasyondan yine her birinde n birey bulunan mümkün olan bütün örneklerde hesaplanabilecek aynı istatistiğe ait örnekleme dağılımına dahil olma veya ait olma olasılığına dayanarak karar verme ve genelleme yapma tekniğidir.

Hipotez kontrolü bir işlemler dizisidir. Öne sürülen bir hipotezin kabul edilip edilmeyeceğine karar vermek için bu işlemlerin yerine getirilmesi gerekir. Bu sebeple birçok kitapta “**hipotez kontrolü**” yerine “**istatistik kontrol**” deyimini de kullanılır.

6.2. Hipotez Kontrolünün Adımları

Belirli bir konuda yapılan araştırma sonucu elde edilen veriler yardımıyla karar vermede öncelikle kontrol edilebilir (istatistik hesaplamalarla test edilebilir) bir hipotezin oluşturulması gerekir. İstatistik hesaplamalar sonucu bu kontrol edilebilir hipoteze “**Kontrol Hipotezi**” veya “**Test Hipotezi**” adı verilir. Kontrol Hipotezinin ret edilmesi halinde neyin geçerli olacağını veya genellenmenin ne olacağını belirlemek üzere Kontrol Hipotezindeki iddiaların aksinin belirtildiği hipoteze ise “**Karşıt Hipotez**” veya “**Alternatif Hipotez**” adı verilir. Kontrol ve Karşıt hipotezlerin oluşturdukları bu ikili yapıya ise “**Hipotez Takımı**” denir. Hipotez Kontrolü ile Kontrol Hipotezi “Kabul” ve/veya “Ret” edilir. Karşıt Hipotezin ise kontrol edilmesi, dolayısıyla kabul ve/veya ret edilmesi mümkün değildir. Karşıt hipotez sadece Kontrol Hipotezinde ileri sürülen iddia veya görüşlerin ret edilmesi halinde neyin geçerli olabileceğini belirtmektedir. Kontrol Hipotezinin kabul veya reddedilmesine karar vermek için yerine getirilmesi gereken adımlar aşağıdaki gibi sıralanabilir:

1. Hipotez Takımının oluşturulması (Kontrol ve Karşıt),
2. Test istatistiğinin hesaplanması,
3. I. tip hata olasılığının belirlenmesi,

4. Kontrol hipotezinin kabul ve ret bölgelerinin belirlenmesi,
5. Eğer gerekiyorsa testin gücünün hesaplanması.

6.2.1. Hipotez Takımının Oluşturulması

Hipotez kontrolünün ilk adımı hipotez takımının oluşturulmasıdır. Kontrol ve Karşıt hipotezlerin oluşturdukları yapıya Hipotez Takımı adı verilmektedir. Kontrol hipotezi sembolik olarak H_0 ile karşıt hipotez de H_1 ile gösterilir. Karşıt hipotez bazı kitaplarda, özellikle İngilizce kitaplarda, alternatif hipotez olarak da isimlendirilir ve H_A ile gösterilir.

Kontrol hipotezi araştırmacının, araştırmasına başlarken öne sürdüğü ve kontrol etmek istediği hipotezdir. Kontrol hipotezinin kontrol edilebilir olması, bir anlam taşıması ve mutlaka bir karşıtının olması gerekir. Kontrol hipotezinin bir karşıtının olması zorunludur, çünkü yapılan hipotez kontrolü sonunda bu hipotezin reddedilmesine karar verilmesi durumunda karşıt hipotez geçerli olur.

Kontrol hipotezi ile araştırmacı, üzerinde çalıştığı populasyon parametresi ile örnekten hesapladığı istatistik arasındaki farkın tamamen tesadüften ileri geldiğini ileri sürer. Populasyon parametresi ile örnekten hesaplanan istatistik arasındaki farkın tamamen tesadüften ileri geliyor olması; bu farkın kontrol altına alınamayan faktörlerden kaynaklandığını, dolayısıyla örneğinin bu populasyona ait olduğunu ve parametre ile istatistik arasındaki bu farkın da sıfır olarak kabul edilebileceğini gösterir.

ÖRNEK 1:

Bir patates çeşidinin dekara veriminin ortalaması, $\mu=3800$ kg ve standart sapmasının da, $\sigma=300$ kg olan normal dağılım gösterdiğini bilen bir araştırmacı, patateste belirli bir yaprak zararlısı ile mücadele için kullanılacak ilacın, patatesin verimini artıracakını öne sürmüş olsun. Araştırmacı, öne sürdüğü bu hipotezin doğruluğunu kontrol etmek için 50 parselde yaprak zararlısına karşı ilaçlama yaparak yürüttüğü deneme sonunda verim ortalamasını 3900 kg olarak bulmuş olsun. Bu durumda araştırmacı, “patateste söz konusu yaprak zararlısı ile mücadele ilacının, patatesin verimini artırıp artırmadığını” hipotez kontrolü yaklaşımı ile kontrol ederek karar vermelidir.

Hipotez kontrolünün ilk adımı hipotez takımını oluşturmaktır. Yani kontrol ve karşıt hipotezleri kurmalıdır. Kontrol hipotezinde, populasyona ait parametre ile örnekten hesaplanan istatistik arasındaki farkın tesadüften ileri geldiği ve sıfır kabul edilebileceği öne sürülür. Bu sebeple, verilen örnekte hipotez kontrolüne aşağıdaki kontrol hipotezi kurularak başlanır. Böylece örneğin alınmış olduğu populasyonun ortalamasının $\mu_x = 3800$ olduğu kontrol hipotezi gereği kabul edilmiş ve belirlenmiş olur. Yani kontrol hipotezi:

H_0 : Örnek ortalaması ile populasyon ortalaması arasındaki fark tesadüften ileri gelmektedir, söz konusu fark sıfır kabul edilebilir. Yani, ilaçlama dekara patates verimini artırmamıştır. (Kısaca $\mu_x = 3800$ veya $\mu_0 = 3800$ kg'dır.)

şeklinde ifade edilir.

ÖRNEK 2:

Klasik yem karması ile beslenen etlik piliçlerde 42. gün canlı ağırlık ortalamasının 2200 g olduğu ve normal dağılım gösterdiği bildirilmiştir. Yeni hazırlanan bir yem karmasının canlı ağırlık üzerine etkisinin olduğu öne sürülmüştür. Bunun doğruluğunu kontrol etmek için tesadüfen alınan 100 adet etlik piliç yeni yem karması ile beslenmiş ve canlı ağırlık ortalaması 2240 g ve standart sapması da 240 g olarak bulunmuştur. Yeni yem karmasının canlı ağırlık üzerine etki bakımından klasik yem karmasından farklı olduğu söylenebilir mi?

Burada araştırmacı, canlı ağırlık üzerine etki bakımından yeni yem karmasının klasik yem karmasından farklı olup olmadığını araştırmak istemektedir. Yapılacak hipotez kontrolünde, kontrol hipotezi aşağıdaki şekilde kurulmalıdır:

H₀: Örnek ortalaması ile populasyon ortalaması arasındaki fark tesadüften ileri gelmektedir, söz konusu fark sıfır kabul edilebilir. Yani yeni yem karması canlı ağırlığa etki bakımından klasik yem karmasından farklı değildir. (Kısaca $\mu_x = 2200$ veya $\mu_0 = 2200$ kg'dır.)

Kontrol hipotezi kurulduktan sonra karşıt hipotezin kurulması gerekir. Karşıt hipotez, yapılacak hipotez kontrolünün tek veya çift taraflı hipotez kontrolü olduğunu belirleyen ve kontrol hipotezinin reddedilmesi durumunda kabul edilecek olan hipotezdir.

Eğer araştırmacı, örnekten hesaplanan istatistiğin, populasyona ait parametreden farklı olup olmadığı ile ilgileniyorsa, bu durumda yapılması gereken çift taraflı hipotez kontrolüdür. Çift taraflı hipotez kontrolünde, örnekten hesaplanan istatistiğin, söz konusu istatistiğe ait örnekleme dağılımına dahil olup olmadığı ile ilgilenilir. Bu sebeple de, örnekleme dağılımının ortalamasından hem büyük hem de küçük tarafa (her iki tarafa da) sapanların ihtimali bulunur. Yani uygulanan muamelenin (faktörün) üzerinde durulan özelliği sadece etkileyip etkilemediği (etki azaltma yönünde de olabilir artırma yönünde de olabilir) araştırılıyorsa çift taraflı hipotez kontrolüne başvurulur.

Çift taraflı hipotez kontrolünde karşıt hipotez, elimizdeki örnekten hesaplanan istatistiğin (ortalamanın) kontrol hipotezi ile tanımlanan populasyona ait parametreden (populasyon ortalamasından $\mu = a$) farklıdır veya değişmiştir, yani **H₁:** $\mu \neq a$ şeklinde kurulur.

Bazı durumlarda ise araştırmacı örnekten hesaplanan istatistiğin, populasyona ait parametreden büyük veya küçük oluşu ile ilgileniyor olabilir. Bu durumda tek taraflı hipotez kontrolünün yapılması gerekir. Tek taraflı hipotez kontrolünde, örnekten hesaplanan istatistiğin, örnekleme dağılımının ortalamasından büyük veya küçük oluşu, yani örnekten hesaplanan istatistik kadar ve daha fazla sapanların ihtimali hesaplanır.

Tek taraflı hipotez kontrolünde karşıt hipotez, üzerinde çalışılan populasyona ait parametre, bildirilen değerden ($\mu = a$) büyüktür veya küçüktür şeklinde aşağıda gösterildiği gibi kurulur.

$$\begin{array}{ll} H_0: \mu = a & \text{veya} & H_0: \mu = a \\ H_1: \mu < a & & H_1: \mu > a \end{array}$$

ÖRNEK 1:

50 parselde yaprak zararlısına karşı ilaçlama yapan araştırmacının amacı, ilaçlamanın dekara patates verimini artırıp artırmadığını araştırmaktır. Bu durumda araştırmacı, örnekten hesaplanan istatistiğin, ortalamaya ait örnekleme dağılımının ortalamasından büyük tarafa sapması ile ilgilenmektedir. Bu nedenle de tek taraflı hipotez kontrolünün yapılması gerekir ve karşıt hipotez aşağıda şekilde kurulur:

H₁: Populasyon ortalaması ile örnek ortalaması arasındaki fark tesadüften ileri gelmemektedir, söz konusu fark sıfır kabul edilemez. Yani üzerinde çalışılan örnek ortalaması söz konusu populasyonun ortalamasından daha büyük olan bir populasyondan tesadüfen alınmış bir örnektir. Başka bir deyişle, yapılan ilaçlama dekara patates verimini artırmıştır. Kısaca, $\mu_x > 3800\text{kg}$ 'dır.

Böylece, hipotez kontrolünde kontrol edilecek hipotez ve karşıt hipotez :

H₀: Örnek ortalaması ile populasyon ortalaması arasındaki fark tesadüften ileri gelmemektedir, söz konusu fark sıfır kabul edilebilir. Yani, ilaçlama dekara patates verimini artırmamıştır. (Kısaca $\mu_x = 3800$ veya $\mu_0 = 3800\text{kg}$ 'dır.)

H₁: Populasyon ortalaması ile örnek ortalaması arasındaki fark tesadüften ileri gelmemektedir, söz konusu fark sıfır kabul edilemez. Yani üzerinde çalışılan örnek ortalaması söz konusu populasyonun ortalamasından daha büyük olan bir populasyondan tesadüfen alınmış bir örnektir. Başka bir deyişle, yapılan ilaçlama dekara patates verimini artırmıştır. (Kısaca, $\mu_x > 3800\text{kg}$ 'dır.)

ÖRNEK 2:

Araştırmacı, canlı ağırlık üzerine etki bakımından yeni yem karmasının klasik yem karmasından farklı olup olmadığını araştırmak istemektedir. Bu sebeple çift taraflı hipotez kontrolü yapılması gerekir. Çünkü örnekten hesaplanan ortalamanın söz konusu populasyondan çekilen örneklerden elde edilecek ortalamaya ait örnekleme dağılımına dahil olup olmadığı araştırılmaktadır. Yapılacak hipotez kontrolünde karşıt hipotezin aşağıdaki şekilde kurulması gerekir.

H₁: Örnek ortalaması ile populasyon ortalaması arasındaki fark tesadüften ileri gelmemektedir, söz konusu fark sıfır kabul edilemez. Yani yeni yem karması canlı ağırlığa etki bakımından klasik yem karmasından farklıdır. Kısaca, $\mu_x \neq 2200\text{g}$ 'dır.

Bu durumda kontrol edilecek hipotez ve karşıt hipotez;

H₀: Örnek ortalaması ile populasyon ortalaması arasındaki fark tesadüften ileri gelmemektedir, söz konusu fark sıfır kabul edilebilir. Yani yeni yem karması canlı ağırlığa etki bakımından klasik yem karmasından farklı değildir. (Kısaca, $\mu_x = 2200$ veya $\mu_0 = 2200\text{kg}$ 'dır.)

H₁: Örnek ortalaması ile populasyon ortalaması arasındaki fark tesadüften ileri gelmemektedir, söz konusu fark sıfır kabul edilemez. Yani yeni yem karması canlı ağırlığa etki bakımından klasik yem karmasından farklıdır. (Kısaca, $\mu_x \neq 2200\text{g}$ 'dır.)

6.2.2. Test İstatistiği

Hipotez kontrolünde bir sonraki adım test istatistiğinin belirlenmesi ve hesaplanmasıdır. Çünkü hipotez kontrolü yapmak, tesadüf örneğinden hesaplanan istatistiğin, üzerinde çalışılan populasyona ait aynı örnek genişliğindeki tesadüf örneklerinden elde edilecek söz konusu istatistiğe ait örnekleme dağılımına dahil olma olasılığını bulmak ve hesaplanan olasılığa göre karar vermektir. Yapılan hipotez kontrolü sonunda verilen kararın güvenilir olması için test istatistiğinin doğru olarak

belirlenmesi çok önemlidir. Hipotez kontrolünün ne amaçla yapıldığına ve populasyon parametresinin bilinip bilinmediğine bağlı olarak kullanılması gereken test istatistikleri Tablo 6.1’de verilmiştir.

Tablo 6.1. Kontrol edilecek hipotez ve kullanılması gereken test istatistiği

Hipotez kontrolü	Test istatistiği
1. Örnek ortalamasının bilinen bir ortalama ile karşılaştırılması	σ^2 biliniyor ise; $\pm Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$
	σ^2 bilinmiyor ise; $\pm t = \frac{\bar{X} - \mu_{\bar{X}}}{S_{\bar{X}}}$
2. Bağımsız iki grubun karşılaştırılması	σ^2 biliniyor ise; $Z = \frac{(\bar{A} - \bar{B}) - (\mu_{\bar{A}} - \mu_{\bar{B}})}{\sigma_{(\bar{A}-\bar{B})}}$
	σ^2 bilinmiyor ise; $t = \frac{(\bar{A} - \bar{B}) - \mu_D}{S_{(\bar{A}-\bar{B})}}$
3. Bağımlı iki grubun karşılaştırılması	Eş-yapma t-testi $t = \frac{\bar{D} - \mu_{\bar{D}}}{S_{\bar{D}}}$
4. Örnekten hesaplanan oranın bilinen bir oran ile karşılaştırılması	π biliniyorsa; $Z = \frac{(p - \mu_p)}{\sigma_p}$
5. İki oranın karşılaştırılması	π 'nin en iyi tahmini; $p = \frac{x_1 + x_2}{n_1 + n_2}$ 'dir. Test istatistiği; $Z = \frac{p_1 - p_2}{\sigma_{(p_1-p_2)}}$
6. Örnekten hesaplanan korelasyon katsayısının bilinen bir korelasyon katsayısı ile karşılaştırılması	$Z = \frac{Z_r - \mu_{Z_r}}{\sigma_{Z_r}}$
7. Örnekten hesaplanan korelasyon katsayısının sıfırdan farklı olup olmadığının kontrol edilmesi	$t = \frac{r - \rho}{S_r}$
Hipotez kontrolü	Test istatistiği

8. İki korelasyon katsayısının karşılaştırılması	$Z = \frac{Z_{r1} - Z_{r2}}{\sqrt{\frac{1}{(n_1 - 3)} + \frac{1}{(n_2 - 3)}}}$
9. Örnekten hesaplanan regresyon katsayısının sıfırdan farklı olup olmadığının kontrol edilmesi	$t = \frac{b_{yx} - \beta_{yx}}{S_b}$
10. İki regresyon katsayısının karşılaştırılması	$t = \frac{(b_A - b_B)}{S_{(b_A - b_B)}}$

Tablo 6.1.(dvm). Kontrol edilecek hipotez ve kullanılması gereken test istatistiği

Test istatistiği hesaplandıktan sonra kullanılacak olan test dağılımını hesaplanan test istatistiği belirler. Örneğin, hipotez kontrolü için hesaplanan test istatistiği Z-değeri ise kullanılması gereken test dağılımı standart normal (Z dağılımı) dağılımdır. Eğer test istatistiği olarak t-değeri hesaplanmış ise kullanılması gereken test dağılımı t-dağılımdır (Student's t-distribution).

6.2.3. I. ve II. Tip Hata Olasılıkları

Hipotez kontrolü yapılarak hangi hipotezin doğru olduğuna karar vermede iki tip hata yapma olasılığı vardır.

1. Gerçekte H_0 'ın doğru olmasına rağmen hipotez kontrolü sonucu H_0 hipotezinin ret edilerek yanlış olduğuna karar verilmesiyle yapılacak olan hatadır. Böyle durumlardaki yanlış olma olasılığına I. Tip Hata Olasılığı adı verilir ve α ile gösterilir.

2. Gerçekte H_0 'ın yanlış olmasına rağmen hipotez kontrolü sonucu H_0 hipotezinin kabul edilerek doğru olduğuna karar verilmesiyle yapılacak olan hatadır. Böyle durumlardaki yanlış olma olasılığına da II. Tip Hata Olasılığı adı verilir ve β ile gösterilir.

Yapılan kontroller sonucunda verilen kararların güvenilir olması veya olmaması α ve β olasılıklarına bağlıdır. Verilen kararların isabetli veya hatalı olması ve bu durumlardaki hatalar Tablo 6.2'de gösterilmektedir.

TABLO 6.2. Hipotez kontrollerinde isabetli ve hatalı kararlar ve ihtimalleri

Gerçek durum	Kontrol sonunda verilen karar	
	H_0 hipotezi kabul	H_0 hipotezi ret
H_0 doğru	Doğru karar ($1-\alpha$)	Yanlış karar (α)
H_0 yanlış	Yanlış karar (β)	Doğru karar ($1-\beta$)

Gerçekte doğru olan H_0 hipotezinin kontrol sonunda da kabul edilmesi verilen kararın doğru olduğunu gösterir ve bunun olasılığı $(1-\alpha)$ 'dır (Tablo 6.2). Benzer şekilde gerçekte yanlış olan H_0

hipotezinin ret edilmesi ile yine doğru bir karar verilmiş olur ve bunun olasılığı $(1-\beta)$ 'dir. Bu olasılığa “**Testin Gücü**” denir ve mümkün olduğunca yüksek (1 'e yakın) olması istenir.

Hipotez kontrolüne başlanırken araştırmacının I. tip hata olasılığını, yani α -değerini, belirlemesi gerekir. α -değeri, biyolojik, fen ve uygulamalı bilimlerde genel olarak %1 ve/veya %5 olarak belirlenir. α -değerinin belirlenmesi konusunda herhangi bir kural yoktur. Araştırmacı α -değerini belirlerken, α -değeri küçüldükçe verilen kararların güvenilirliğinin artacağını unutmamalıdır. Eğer yapılan hipotez kontrolü sonucu varılacak karar sağlık, çevre kirliliği vb. gibi konular açısından büyük riskler taşıyor ise I. tip hata olasılığının $\alpha=0.01$ 'den daha küçük olarak belirlenmesi gerekir.

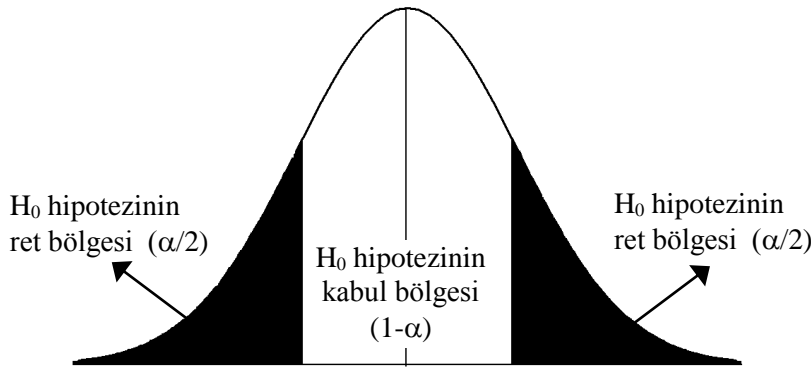
I. ve II. tip hata olasılıkları arasında ters ilişki vardır. Hipotez kontrolü yapılırken belirlenecek olan α -değerinin küçülmesi, II. tip hata olasılığının artmasına yol açar.

6.2.4. Kabul ve Ret Bölgelerinin Belirlenmesi

Hipotez kontrolüne başlanırken belirlenen I. tip hata olasılığı, yani α -değeri, test dağılımını H_0 hipotezini **kabul bölgesi** ve H_0 hipotezini **ret bölgesi (veya kritik bölge)** olmak üzere iki bölgeye ayırır. Hesaplanan test istatistiği “ **kabul** ” bölgesinde yer alıyorsa H_0 hipotezi kabul edilir. Hesaplanan test istatistiğinin değeri “ **kabul** ” bölgesinde dışında kalıyorsa H_0 hipotezi ret edilir ve karşıt hipotez (H_1) geçerli olur.

Belirlenen I. tip hata olasılığı doğrultusunda kontrol hipotezinin kabul ve ret (kritik) bölgelerinin nasıl belirleneceği, hipotez kontrolünün tek veya çift taraflı olmasına göre değişir.

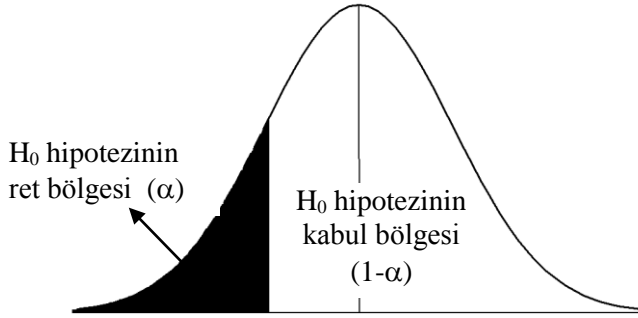
Çift taraflı hipotez kontrolünde araştırmacı, örnekten hesaplanan istatistiğin belirli bir örnekleme dağılımına dahil olup olmadığı ile ilgilenir. Bu durumda örnekten hesaplanan istatistik, örnekleme dağılımının ortalamasından negatif veya pozitif tarafa sapabilir. Dolayısıyla çift taraflı hipotez kontrolünde örnekleme dağılımının ortalamasından büyük veya küçük tarafa, örnekten hesaplanan değer veya daha fazla sapanların ihtimali bulunur. Şekil 6.1, önceden belirlenen I. tip hata olasılığına göre çift taraflı hipotez kontrolünde ret ve kabul bölgelerini göstermektedir.



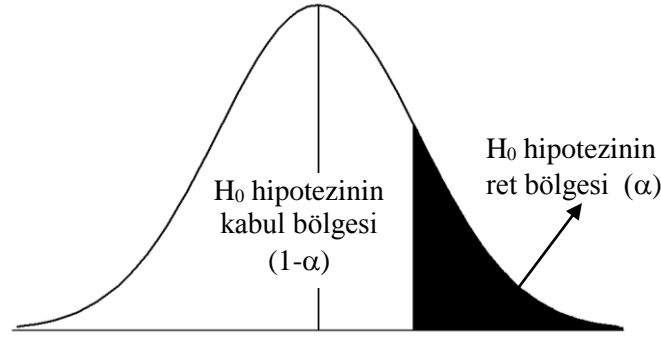
ŞEKİL 6.1. Çift taraflı hipotez kontrolünde kontrol hipotezinin ret ve kabul bölgeleri

Tek taraflı hipotez kontrolünde, örnekten hesaplanan istatistiğin, örnekleme dağılımının ortalamadan ya büyük ya da küçük tarafa sapanması ile ilgilenilir. Bu sebeple, tek taraflı hipotez kontrolünde, örnekleme dağılımının ortalamasından, örnekten hesaplanan istatistik kadar ve daha küçük ya da daha büyük tarafa sapanlarının olasılığı bulunur. Şekil 6.2, önceden belirlenen I. tip hata olasılığına göre tek taraflı hipotez kontrolünde, ret ve kabul bölgelerini göstermektedir.

a. $H_0: \mu=a$ kontrol hipotezi $H_1: \mu < a$ hipotezine karşı test ediliyorsa



b. $H_0: \mu=a$ kontrol hipotezi $H_1: \mu > a$ hipotezine karşı test ediliyorsa



ŞEKİL 6.2. Tek taraflı hipotez kontrolünde kontrol hipotezinin ret ve kabul bölgeleri

Hipotez kontrolünün adımları tamamlandıktan sonra araştırmacının hangi hipotezi kabul edeceğine karar vermesi gerekir. Örnekten hesaplanan istatistiğin söz konusu örnekleme dağılımına dahil olma olasılığı I. tip hata olasılığından büyük ise, yani hesaplanan test istatistiği, kontrol hipotezinin (H_0) kabul bölgesinde yer alıyorsa kontrol hipotezi (H_0) kabul edilir. Eğer örnekten hesaplanan istatistiğin söz konusu örnekleme dağılımına dahil olma olasılığı I. tip hata olasılığından küçük ise, yani hesaplanan test istatistiği kontrol hipotezinin (H_0) kabul bölgesinde yer almıyorsa kontrol hipotezi (H_0) reddedilir.

6.2.5. Testin Gücü

Hipotez kontrolü tamamlandıktan sonra testin gücü de hesaplanabilir. Testin gücü, daha önce de belirtildiği gibi $(1-\beta)$ 'dir ve β değeri küçüldükçe testin gücü artar.

ÖRNEK:

Bir meyve suyu fabrikasında üretilen portakal sularındaki C-vitamini miktarının ortalaması, $\mu=17$ mg/L ve standart sapmasının, $\sigma=1.5$ mg/L olan normal dağılım gösterdiği bildirilmiştir.

Bu fabrikada üretilen portakal sularından rastgele 16 şişelik örnekler alınsa ve bunların her birinde C-vitamini tayin edilse, üretimde herhangi bir değişiklik olmadığı sürece $n=16$ şişelik

örneklerde C-vitamini miktarı ortalaması, $\mu_{\bar{x}} = 17$ mg/L ve standart sapması, $\sigma_{\bar{x}} = \frac{1.5}{\sqrt{16}} = 0.375$

mg/L olan normal dağılım gösterir.

Araştırmacının yaptığı hipotez kontrolünde, kontrol ve karşıt hipotezler aşağıdaki gibi olsun:

$$H_0 : \mu_X = 17 \text{ mg/L}$$

$$H_1 : \mu_X < 17 \text{ mg/L}$$

Fabrikada üretilen portakal sularındaki C-vitamini miktarına ait standart sapmanın, $\sigma=1.5$ mg/L olduğu bilindiği için test dağılımı standart normal dağılımdır.

Birinci tip hata olasılığının önceden $\alpha=0.05$ olarak belirlendiği hipotez kontrolünde, herhangi bir örnek ortalaması için Z-değeri -1.645 (Tablo A) ve daha küçük ise kontrol hipotezi reddedilerek karşıt hipotez kabul edilecektir. Söz konusu meyve suyu fabrikasından tesadüfen alınan belirli örnek genişliğindeki örneklerden hesaplanacak ortalamanın $\bar{X}=16.383$ ve 16.383'ten küçük olduğu her

durumda kontrol hipotezi reddedilecektir. Bu örnek ortalaması; $Z_i = \frac{X_i - \mu}{\sigma}$ eşitliğinden

hesaplanmıştır. Tablo A'dan standart normal dağılımda %5'lik alanın $Z=1.645$ değerinden başladığı

görülmür. Bu durumda $-1.645 = \frac{\bar{X} - 17}{0.375}$ şeklinde yazılarak ortalama $\bar{X}=16.383$ olarak hesaplanır. Bu,

meyve suyu fabrikasının ürettiği portakal sularından tesadüfen alınan 16 şişede C-vitamini ortalamasının "<16.383 mg/L" olarak bulunması, meyve sularındaki C-vitamini ortalamasının 17 mg/L olduğu kontrol hipotezinin reddedileceğini gösterir.

Söz konusu meyve suyu fabrikasında üretilen portakal sularındaki C-vitamini ortalamasının 17 mg/L olmasına rağmen kontrol hipotezinin yapılan hipotez kontrolü sonunda yanlış olarak reddedilme olasılığı $\alpha=0.05$ 'tir, yani I. tip hata olasılığıdır.

Meyve suyu fabrikasında üretilen portakal sularındaki ortalama C-vitamininin 17 mg/L'den daha düşük olduğu durumlarda kontrol hipotezinin, yani $\mu=17$ mg/L hipotezinin kabul edilme olasılığı ise II. tip hata olasılığı (β)'dir.

Hipotez kontrolünde II. tip hata olasılığının hesaplanabilmesi için karşıt hipotezde dikkate alınan değerlerin bilinmesi gerekir. Yapılan hipotez kontrolünde karşıt hipotez, $H_1 : \mu_X < 17 \text{ mg/L}$ idi ve bu ortalamanın 17 mg'dan küçük olduğu bütün değerleri kapsar. $H_1 : \mu_{\bar{x}} < 17 \text{ mg/L}$ hipotezi ile kapsanan değerlerin her biri için II. tip hata olasılığı hesaplanabilir.

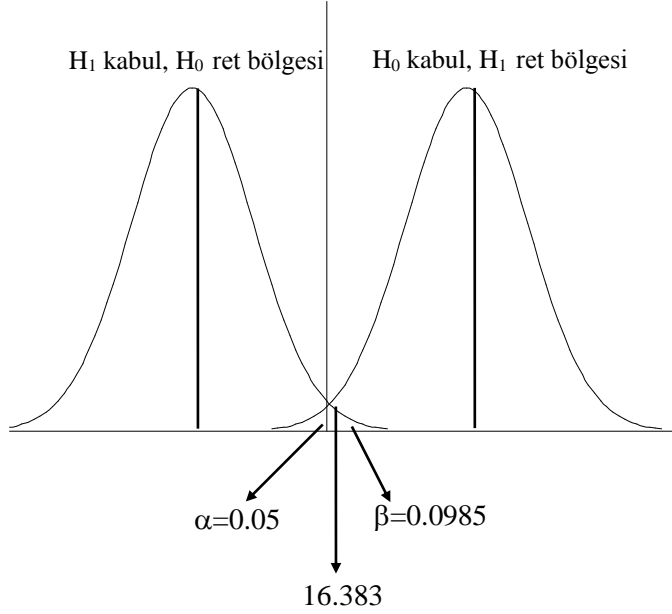
Eğer araştırmacı karşıt hipotezini örneğin $H_1 : \mu_x=15.9$ olarak kurmuş ise bunun için II. tip hata olasılığı (β) hesaplanabilir. Burada bulunması gereken olasılık ortalaması $\mu=15.9$ mg ve standart sapması 0.375 mg olan normal dağılımda 16.383 mg'dan daha büyüklerin oranıdır. Bu oran da 16.383'e karşılık gelen Z-değeri hesaplanarak bulunur.

$$Z = \frac{16.383 - 15.9}{0.375} = 1.288 \cong 1.29$$

Standart normal dağılım (Z dağılımı) tablosundan (Tablo A) 0 ile 1.29 arasındaki Z'lerin oranı 0.4015 olarak bulunur. Z-değerlerinin 1.29 ve daha büyük değere sahip olma olasılıkları ise $0.5 - 0.4015 = 0.0985$ 'dir ve bu da $\alpha=0.05$ olduğunda, karşıt hipotezin ortalaması 15.9 mg ise II. tip hata olasılığıdır, yani $\beta=0.0985$ 'dir. Bu olasılık, gerçekte ortalama C-vitamini 15.9 mg iken 17 mg olarak

kabul edilme olasılığıdır. Bu durumda yapılan hipotez kontrolünün gücü, $1-\beta = 1-0.0985=0.9015$ 'dir. Testin gücü, $\alpha=0.05$ ve karşıt hipotez ile belirtilen ortalama, $\mu=15.9$ mg olduğu durumda karşıt hipotezin kabul edilme olasılığıdır. Başka bir deyişle, doğru karar verme olasılığı %90.15'dir. Şekil 6.3, bu hipotez kontrolü için I. ve II. tip hata olasılıklarını göstermektedir.

Hipotez kontrollerinde karşıt hipotez ile belirtilen parametre, kontrol edilen hipotezin parametresine yaklaştıkça testin gücü azalır. Buna karşın örnek genişliği arttıkça aynı H_1 karşıt hipotezlerine karşı yapılacak testin gücü artar.



ŞEKİL 6.3. $H_0: \mu_x = 17$ mg kontrol hipotezinin $H_1: \mu_x = 15.9$ mg hipotezine karşı kontrolünde I. ve II. tip hata olasılıkları.

Örneğin, araştırmacı karşıt hipotezini $H_1: \mu_x=16.2$ olarak kurmuş ise bunun reddedilme olasılığı $Z = \frac{16.383-16.2}{0.375} = 0.488 \approx 0.49$ ve bu değerden daha çok sapanların olasılığıdır. Bu olasılık ise $0.5-0.1879=0.3121$ olarak hesaplanır. Bu da $\alpha=0.05$ olduğunda, karşıt hipotezin ortalaması 16.2 mg ise II.tip hata olasılığıdır, yani $\beta=0.3121$ 'dir. Bu olasılık, gerçekte ortalama C-vitamini 16.2 mg iken bunun 17 mg olarak kabul edilme olasılığıdır. Bu durumda testin gücü $1-0.3121=0.6879$ 'dur. Testin gücü, $\alpha=0.05$ ve karşıt hipotez ile belirtilen ortalama, $\mu=16.2$ mg olduğu durumda karşıt hipotezin kabul edilme olasılığıdır. Başka bir deyişle, doğru karar verme olasılığı %68.79'dur.

6.3. SORULAR

1. Hipotez kontrolünün amacı nedir? Açıklayınız.
2. Kontrol ve karşıt hipotez nedir? Örnek vererek açıklayınız.
3. I. ve II. Tip hata nedir? Açıklayınız.
4. Testin gücü nedir? Örnek vererek açıklayınız.

5. Türkiye’de ki hastanelerinin mikrobiyoloji laboratuvarlarında yapılan testlerdeki hata oranının %10 olduğu bilinmektedir. Yeni açılan bir hastanenin mikrobiyoloji laboratuvarında yapılan ölçümlerden rasgele olarak seçilen 50 sonuçtan 13 tanesinin hatalı olduğu gözlemiştir. Yeni açılan hastanenin laboratuvar sonuçlarındaki hata oranının normalin üzerinde olduğu söylenebilir mi?
6. Tek ve çift taraflı hipotez kontrolünü açıklayınız.