ÖRNEKLEME DAĞILIMLARI

Üzerinde çalışılan populasyondan belirli örnek genişliğinde mümkün olan sayıda tesadüf örnekleri alınsa ve üzerinde çalışılan istatistikler hesaplansa, 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 araştırıcıya 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

Ortalaması 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 hesaplansa, 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.

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

(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;

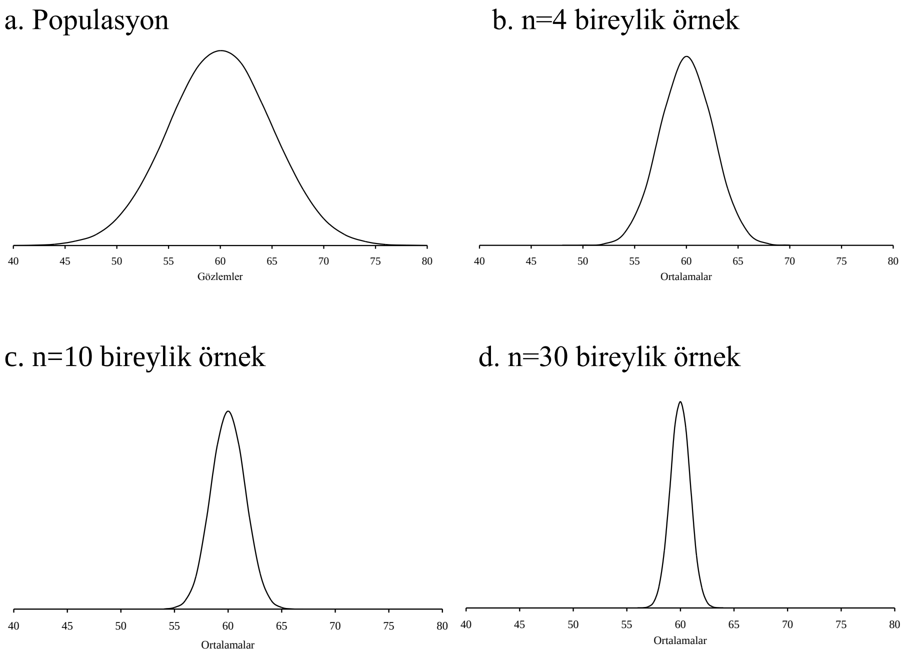
$$S_{\bar{x}} = \frac{S_x}{\sqrt{n}} = \sqrt{\frac{S_x^2}{n}} \qquad \ldots(5.2)$$

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 $\overline{X} \pm S_{\overline{X}}$ 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, μ=60 ve standart sapması, σ=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.

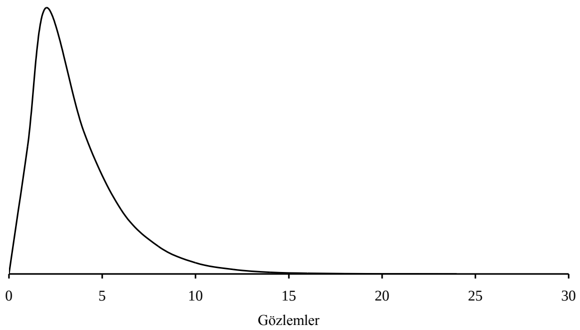


ŞEKİL 5.1. Benzetim (simulasyon) yöntemi ile üretilmiş a. ortalaması, μ=60 ve standart sapması, σ=5 olan bir normal dağılım, b. n=4, c. n=10 ve d. n=30 bireylik örneklerden elde edilen ortalamaya ait örnekleme dağılımları

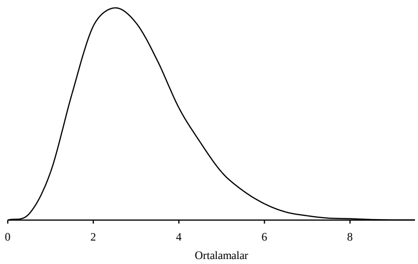
Şekil 5.1b'de n=4 bireylik örneklerden elde edilen ortalama ait örnekleme dağılımı görülmektedir. Bu dağılımın standart sapması $\sigma_{\bar{x}} = \frac{5}{\sqrt{4}} = 2.5$ olarak hesaplanırken, n=10 olduğu zaman $\sigma_{\bar{x}} = \frac{5}{\sqrt{10}} = 1.58$ (Şekil 5.1c) ve n=30 olduğu zaman standart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

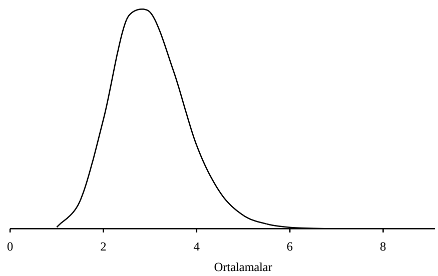
a. Çarpık dağılım



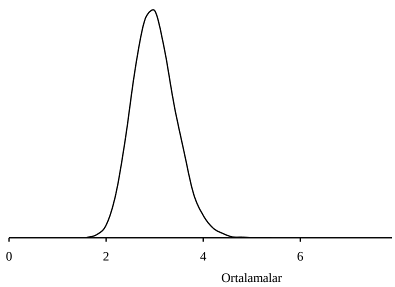
b. n=4 bireylik örnek



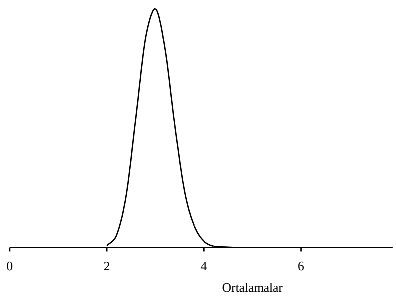
c. n=10 bireylik örnek



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.

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} \qquad \ldots(5.3)$$

ÖRNEK:

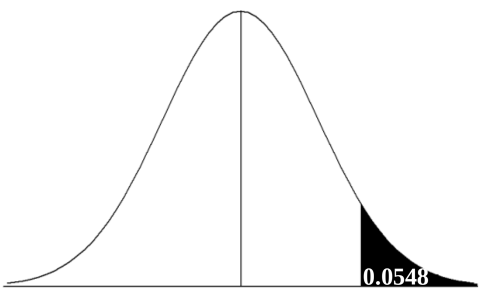
Ortalaması $\mu = 500$ kg ve standart sapması $\sigma = 50$ kg olan normal dağılım gösteren bir populasyondan her defasında 25 birey bulunan çok fazla sayıda rastgele örnekler seçilse ve seçilen bu örneklerin her birinde ortalama hesaplansa;

a. Hesaplanan bu ortalamalar normal bir dağılım gösterir. Çünkü örneklerin seçildiği populasyonun, ortalaması μ=500 kg ve standart sapması σ=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, $\mu_{\overline{x}}$=500 kg ve standart sapması, $\sigma_{\overline{x}} = \frac{50}{\sqrt{25}} = 10$ kg olmak üzere iki parametresi vardır.

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

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

olarak hesaplanır. Z-değeri hesaplandıktan sonra 16 bireylik örneğin söz konusu populasyondan 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 populasyondan 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 populasyondan önce n_A ve sonra n_B genişliğinde geriye iadeli olarak mümkün olan sayıda tesadüf örnekleri alınsa ve her birinde ortalama hesaplansa, hesaplanan ortalamaların her biri populasyon 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 populasyondan tamamen tesadüfen seçildiği ve tesadüfen yan yana getirilerek farklar bulunduğu için örnekten örneğe değişir ve bir dağılım gösterir. Bu dağılıma "**ortalamalar arası farka ait örnekleme**" dağılımı denir.

Ortalamalar 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 populasyon 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 populasyonun varyansının tahmini olduğu için $\sigma_{\bar{A}}^2=\dfrac{\sigma^2}{n_A}$ ve $\sigma_{\bar{B}}^2=\dfrac{\sigma^2}{n_B}$ '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;

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

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

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

ÖRNEK:

Bir tarlada yetişen mısır koçanlarının uzunluklarının, ortalaması μ=85 cm ve standart sapması, σ=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 hesaplansa, hesaplanan ortalama koçan uzunlukları aynı populasyon ortalamasının birbirlerinden bağımsız birer tahminidir.

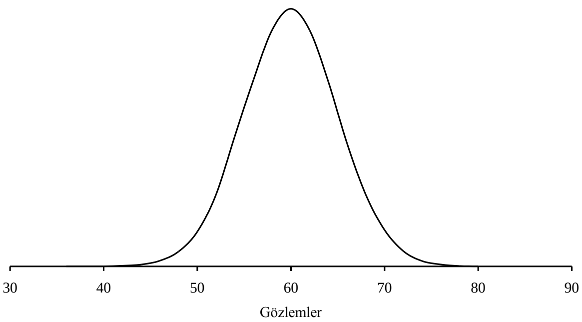
n_A=20 ve n_B=10 koçanlık örneklerden hesaplanan ortalamalar tamamen tesadüfen yan yana getirilip aralarındaki farklar hesaplansa bu farkların bir normal dağılım göstermesi beklenir. Çünkü koçan uzunluklarının normal dağılım gösterdiği daha önceden bilinmektedir. Bu dağılıma "**ortalamalar arası farka 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;

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

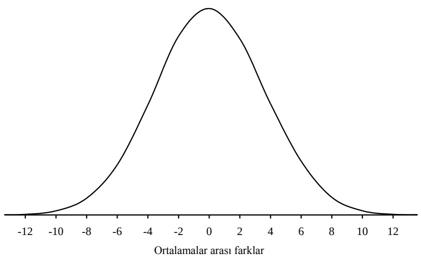
olarak bulunur.

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).

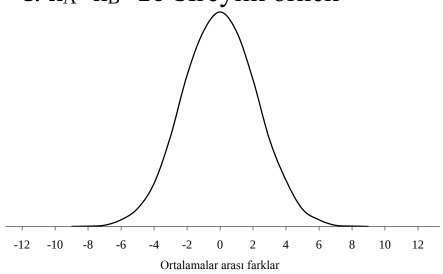
a. Normal dağılım gösteren populasyon



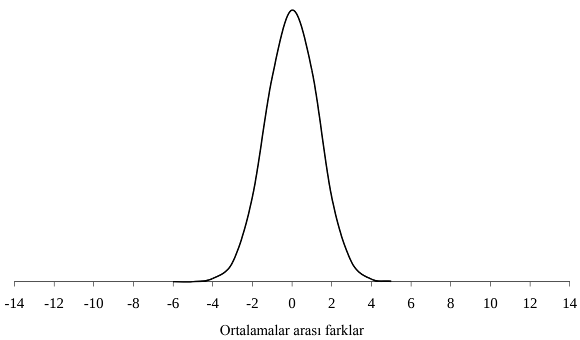
b. n_A=n_B=4 bireylik örnek



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 μ=60 σ=5 olan bir normal dağılım göstersin. 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 μ_D=0 olduğundan;

$$Z = \frac{(\bar{A} - \bar{B})}{\sigma_D} \qquad \ldots(5.5)$$

şeklinde ifade edilir.

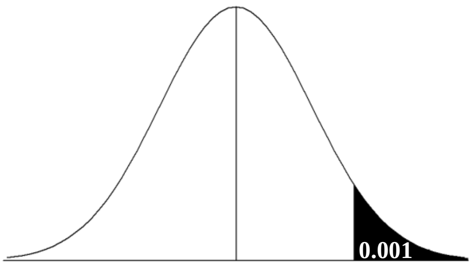
ÖRNEK:

Aynı tarladan hasat edilen fasulye tanelerinin ağırlıklarının ortalaması, μ=0.7 g ve standart sapması, σ=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 μ=0.7 g ve standart sapması σ=0.08g olan normal bir dağılım gösteren populasyondan n_A=25 ve n_B=15 bireylik örneklerden elde edilecek ortalamalar arası farka ait örnekleme dağılımının parametreleri, μ_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.

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

(2.13) numaralı eşitlikten serbestlik derecesi ile varyansın çarpımı kareler toplamını verdiğinden, 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 arasın farka ait örnekleme dağılımının standart sapması (5.7) numaralı eşitlik kullanılarak tahmin edilir.

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

$$\text{Eğer } n_A = n_B = n \text{ ise } S_D = \sqrt{S_{\bar{A}}^2 + S_{\bar{B}}^2} \qquad \ldots(5.7)$$

şeklinde hesaplanır. Bu ifadeye ortalamalar arası farka ait örnekleme dağılımın 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 $\overline{A}$=25.0 S_A=3.21

B örneği için: n_B=20 $\overline{B}$=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 $\sum d_A^2=(n_A-1)S_A^2$'dan 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 Örnekleme 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 hesaplansa, 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 örnekleme dağılımı**" denir.

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

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \qquad \ldots(5.8)$$

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

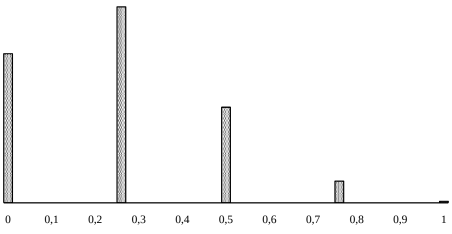
Eğer üzerinde çalışılan populasyona ait π bilinmiyor ise bu oran örnekten tahmin edilir ve (5.8) numaralı eşitlikte π yerine p değeri kullanılarak hesaplama yapılır. Bu durumda oranlara ait örnekleme dağılımının standart sapması (5.9) numaralı eşitlik kullanılarak tahmin edilir ve bu ifade oranlara ait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

$$S_p = \sqrt{\frac{p(1-p)}{n}} \qquad \ldots(5.9)$$

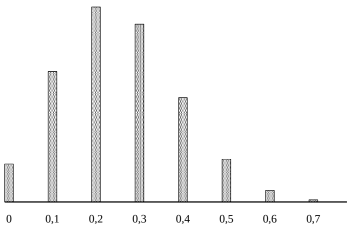
π'si bilinen bir populasyondan elde edilecek oranlara ait örnekleme dağılımının şekli populasyonun π'sine ve populasyondan çekilen örneklerin genişliğine (n) bağlıdır. Eğer populasyonda istenen olayın oluş olasılığı, $\pi \leq 1/2$ ise oranlara ait örnekleme dağılımının normal dağılıma yaklaşabilmesi için $n.\pi \geq 10$ olmalıdır. Eğer populasyonda istenen olayın oluş olasılığı, $\pi > 1/2$ ise oranlara ait örnekleme dağılımının normal dağılıma yaklaşabilmesi için ise $n.(1-\pi) \geq 10$ şartının yerine getirilmiş olması gerekir. Populasyonun π'sine bağlı olarak şartlar yerine getirildiği zaman bile oranlara ait örnekleme dağılımı tam olarak normal dağılım olmayıp normal dağılıma yaklaşır. Populasyondan 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 π=0.25 olan bir populasyondan 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

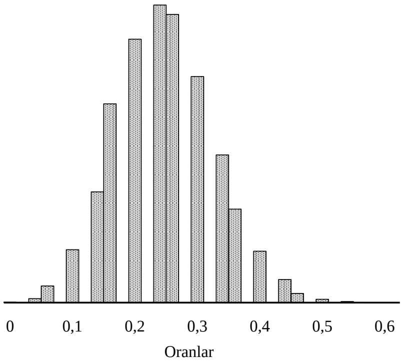


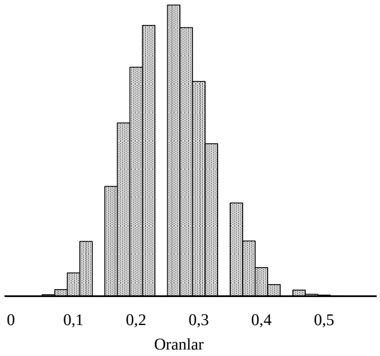
b. n=10 bireylik örnek



c. n=30 bireylik örnek

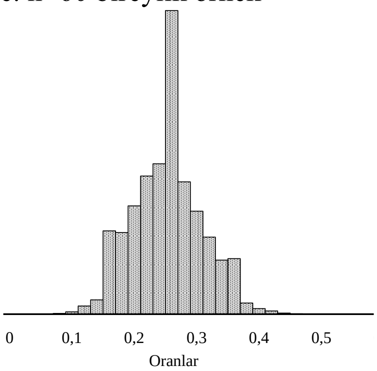
d. n=40 bireylik örnek

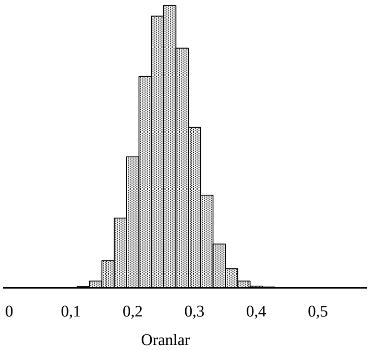




e. n=60 bireylik örnek

f. n=100 bireylik örnek

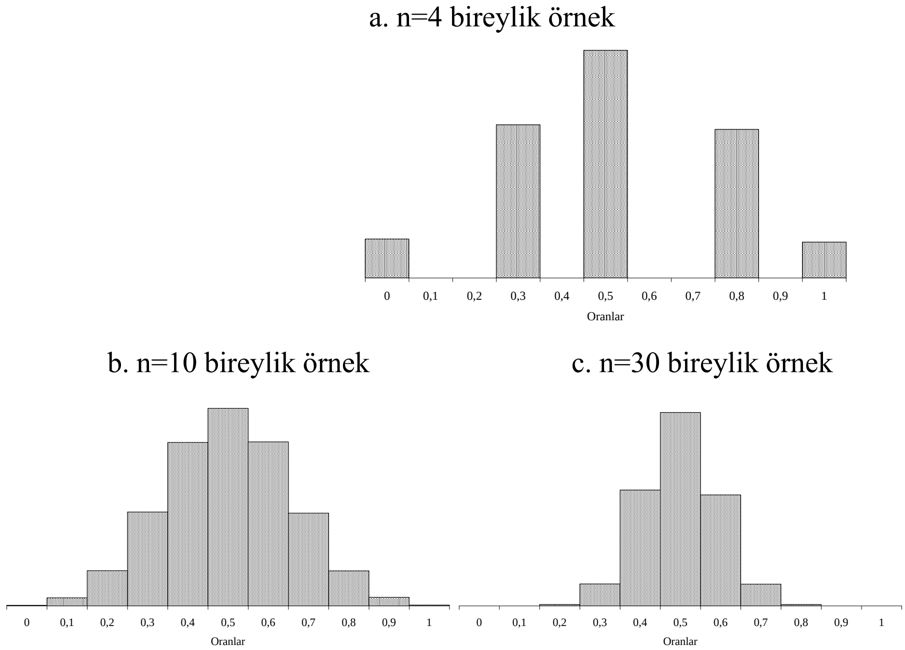




Şekil 5.4. Benzetim tekniği ile üretilmiş π=0.25 olan populasyondan a. n=4, b. n=10, c. n=30, d. n=40, e. n=60 ve f. n=100 bireylik örneklerden elde edilen oranlara ait örnekleme dağılımlarına ait histogramlar

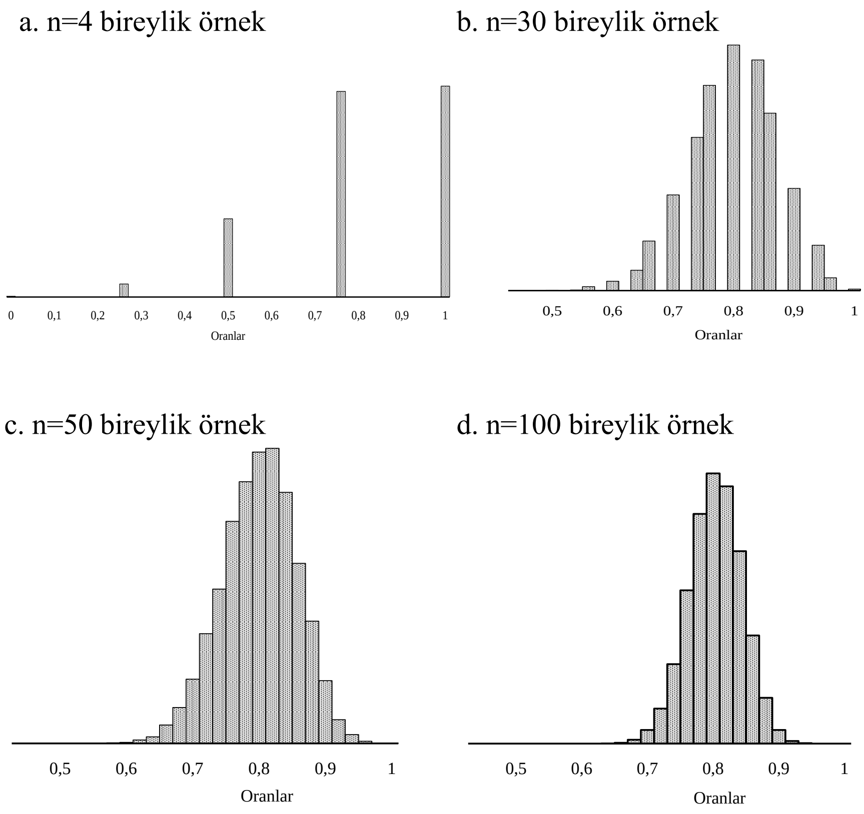
Şekil 5.4'de görüldüğü gibi π=0.25 olduğu için oranlara ait örnekleme dağılımının normal dağılıma yaklaşması için $n.\pi \geq 10$ olmalıdır. Örnek genişlikleri 4, 10 ve 30 olduğu zaman bu şart yerine gelmemekte ve Şekil 5.4a ve b'de görüldüğü gibi oranlara ait örnekleme dağılımları yatık dağılım göstermektedir. n=30 olduğu zaman ise oranlara ait örnekleme dağılımının yatıklığı azalmaktadır. π=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 π=0.5 olduğu zaman dağılım simetriktir. Bu nedenle Şekil 5.5.a, b ve c'de görüldüğü gibi π=0.5 olan populasyondan alınan örnek genişliği ne olursa olsun oranlara ait örnekleme dağılımı simetrik dağılım gösterir. π=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 π=0.5 olan populasyondan a. n=4, b. n=10 ve c. n=30 bireylik örneklerden elde edilen oranlara ait örnekleme dağılımlarına ait histogramlar

Üzerinde çalışılan populasyonda, istenen olayın oluş olasılığı π>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 π=0.8 olan populasyondan 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. π>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.



Şekil 5.6 Benzetim tekniği ile üretilmiş π=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 $\pi \le 1/2$ olması durumunda $n.\pi \ge 10$ ve $\pi > 1/2$ durumunda ise $n.(1-\pi) \ge 10$ şartını sağlayan ve normal dağılıma yaklaşan oranlara ait örnekleme dağılımını oluşturan oranlar (5.10) numaralı eşitlik kullanılarak standart normal dağılıma dönüştürülebilir.

$$Z = \frac{p - \mu_p}{\sigma_p} \qquad \ldots(5.10)$$

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 fakültedeki 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 fakültedeki öğrencilerin %31'i sigara içtiği için bu özelliğe ilişkin oluş olasılığı π=0.31'dir. Bu özelliğe ilişkin oluş olasılığı π< 0.5 olduğundan elde edilecek oranlara ait örnekleme dağılımının normal dağılıma yaklaşması için $n.\pi \ge 10$ 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:

π=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?

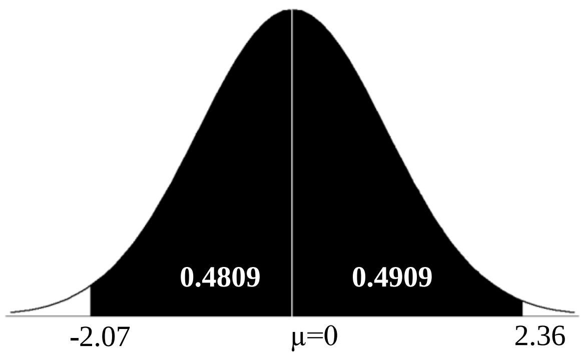
İstenen olasılığın standart normal dağılım kullanılarak hesaplanabilmesi için $\pi>1/2$ olduğundan $n.(1-\pi) \geq 10$ şartının yerine getirilmiş olması gerekir. Bu örnekte 50x(1-0.55)=22.5 olarak bulunduğu için gerekli şart yerine getirilmiştir. Bu durumda 50 bireylik örneklerden hesaplanan oranlar (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, μ_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, π=0.55 olan bir populasyondan örnek genişliği 50 olan mümkün olan sayıda rastgele örnekler alınsa ve her birinde istenen olayın oluş olasılığı hesaplansa, 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 populasyondan önce n_A ve sonra n_B genişliğinde geriye iadeli olarak mümkün olan sayıda tesadüf örnekleri alınsa ve her birinde istenen olayın oluş olasılığı hesaplansa, hesaplanan olasılıkların her biri populasyona 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 populasyondan 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.

Populasyona ait oluş olasılığının $\pi \leq 1/2$ olması durumunda $n_A.\pi \geq 10$ ve $n_B.\pi \geq 10$, $\pi > 1/2$ durumunda ise $n_A.(1-\pi) \geq 10$ ve $n_B.(1-\pi) \geq 10$ 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, $\mu_{(pA-pB)}=0$ dır. Oranlar arası farka ait örnekleme dağılımının standart sapması ise (5.11) numaralı eşitlik kullanılarak hesaplanır.

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

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;

$$p = \frac{x_A + x_B}{n_A + n_B} \qquad \ldots(5.12)$$

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.

$$Z = \frac{(p_A - p_B) - \mu_{(pA-pB)}}{\sigma_{(p_A - p_B)}} \qquad \ldots(5.13)$$

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

$$\sigma_{(p_A - p_B)} = \sqrt{p(1-p)(\frac{1}{n_A} + \frac{1}{n_B})} \qquad \ldots(5.14)$$

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 kaça eşittir?

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

$\sigma_{(p_A - p_B)} = \sqrt{0.5(1-0.5)(\frac{1}{400} + \frac{1}{300})} = 0.0382$ 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 hesaplansa, 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:

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

(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 populasyondan elde edilecek korelasyon katsayısına ait örnekleme dağılımının standart sapması büyür. Örneklerin tesadüfen alındığı populasyonun korelasyon katsayısı, ρ=1.0 ise, populasyondan alınan bütün örneklerde hesaplanan korelasyon katsayısı 1.0 olarak bulunacak ve korelasyon katsayına ait örnekleme dağılımının standart sapması 0 olacaktır.

Aynı populasyondan 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 dolaysıyla korelasyon katsayısına ait örnekleme dağılımının standart sapması azalacaktır.

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

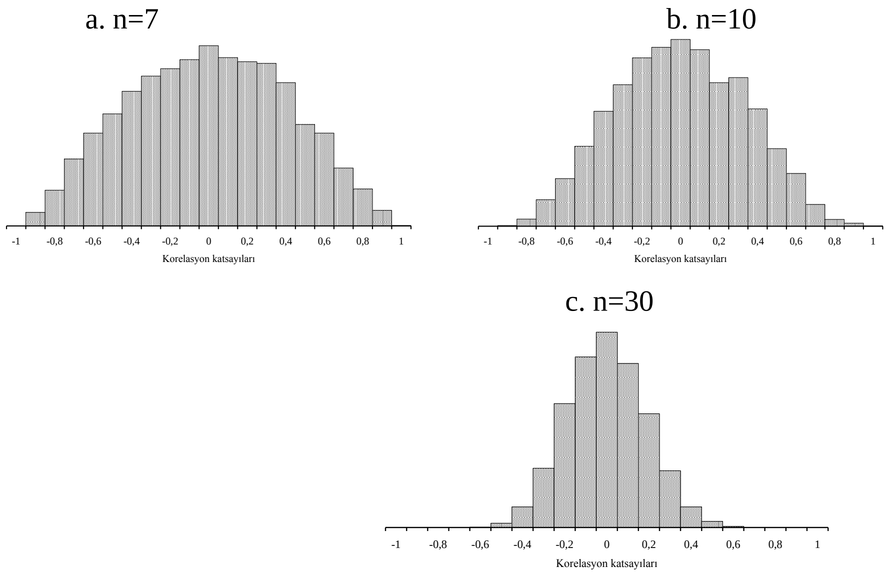
$$S_r = \sqrt{\frac{(1-r^2)}{(n-2)}} \qquad \ldots(5.16)$$

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

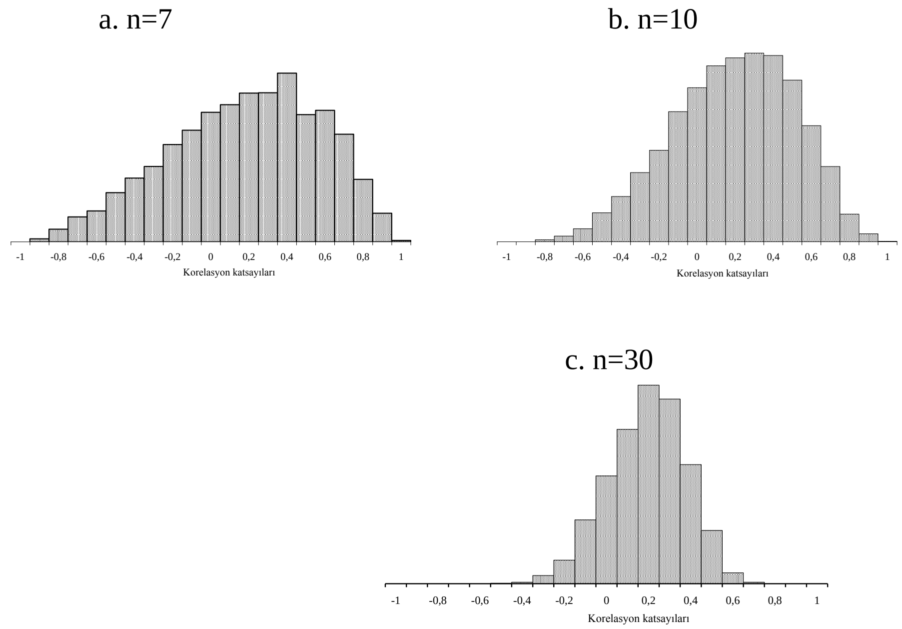
Örneklerin tesadüfen alındığı populasyonun korelasyon katsayısı, ρ=0 ise populasyondan alınan örneklerin genişliği ne olursa olsun korelasyon katsayına ait örnekleme dağılımı normal dağılıma yaklaşır. Şekil 5.7a, b ve c’de, ρ=0 olan populasyondan simulasyon yöntemi ile **10000 tane** 7, 10 ve 30 bireylik tesadüf örnekleri alınarak hesaplanan korelasyon katsayılarının dağılımı gösterilmiştir.

Örneklerin alındığı populasyona ait korelasyon katsayısı sıfırdan farklılaştıkça elde edilecek korelasyon katsayısına ait örnekleme dağılımının şekli normal dağılımdan uzaklaşır. Şekil 5.8’de, ρ=0.2 olan populasyondan simulasyon yöntemi ile **10000 tane** 7, 10 ve 30 bireylik tesadüf örnekleri alınarak hesaplanan korelasyon katsayılarının dağılımı gösterilmiştir.

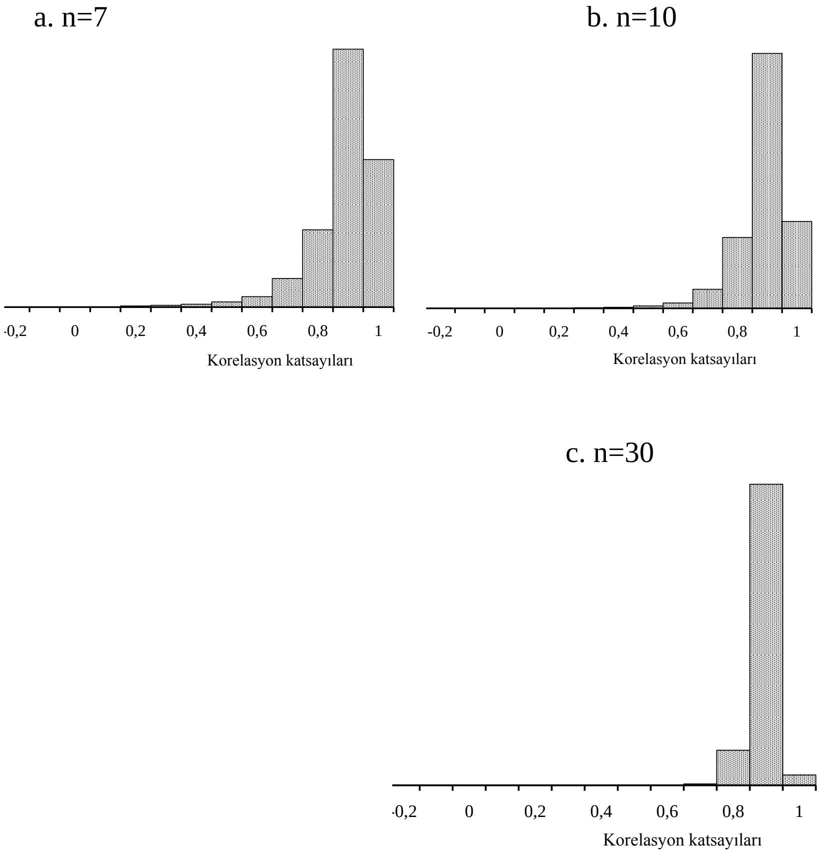
Örneklerin alındığı populasyona ait korelasyon katsayısı büyüdükçe söz konusu populasyondan elde edilecek korelasyon katsayısına ait örnekleme dağılımının şekli normal dağılımdan çok uzaklaşır ve çarpık dağılım haline gelir. . Şekil 5.9'da, ρ=0.9 olan populasyondan simulasyon yöntemi ile **10000 tane**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



Şekil 5.7. Benzetim tekniği ile üretilmiş ρ=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 ρ=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 ρ=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.

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

Hesaplanan Z_r-değerlerinin, ortalaması;

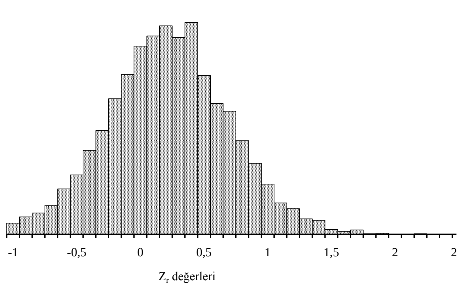
$$\mu_{Z_r} = \frac{1}{2}\log_e\frac{(1+\rho)}{(1-\rho)} \qquad \ldots(5.18)$$

ve standart sapması;

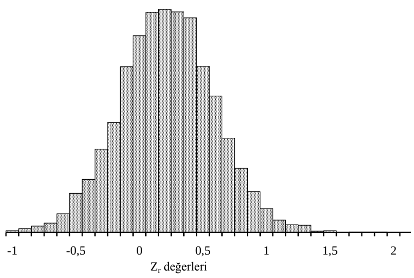
$$\sigma_{Z_r} = \sqrt{\frac{1}{(n-3)}} \qquad \ldots(5.19)$$

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. ρ=0.2 ve ρ=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.

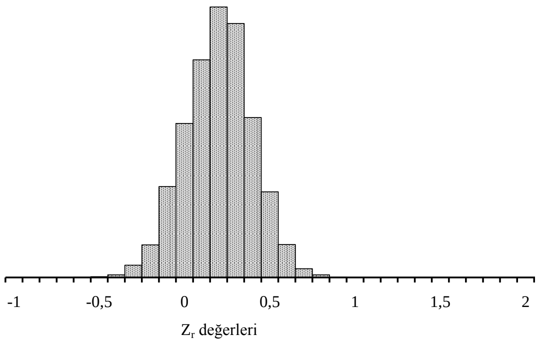
a. ρ=0.2, n=7



b. ρ=0.2, n=10

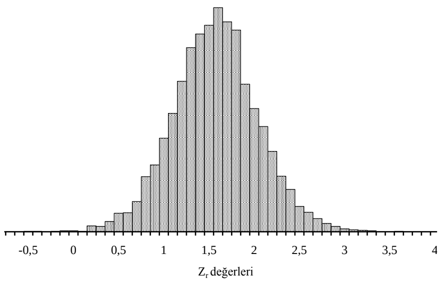


c. ρ=0.2, n=30

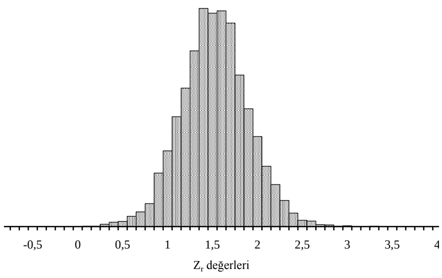


Şekil 5.10 Benzetim tekniği ile üretilmiş ρ=0.2 olan populasyondan a. n=7, b. n=10 ve c. n=30 bireylik örneklerden elde edilen Z_r-değerlerinin dağılımlarının histogramları

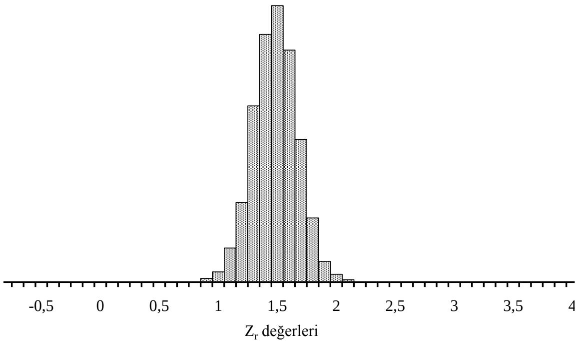
a. ρ=0.9, n=7



b. ρ=0.9, n=10



c. ρ=0.9, n=30



Şekil 5.11 Benzetim tekniği ile üretilmiş ρ=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, ρ≠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 katsayına ait örnekleme dağılımını oluşturan korelasyon katsayıları (5.20) numaralı eşitlik kullanılarak standart normal dağılıma dönüştürülebilir.

$$Z = \frac{Z_r - \mu_{Z_r}}{\sigma_{Z_r}} \qquad \ldots(5.20)$$

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 ρ≠0 olan söz konusu populasyondan alınmış olduğu araştırılmaktadır. Daha önce açıklandığı gibi ρ≠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;

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

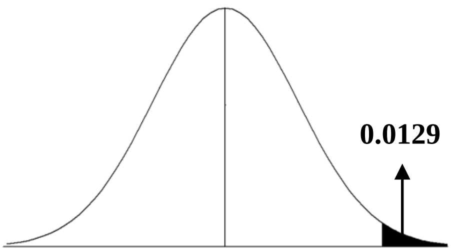
ve

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

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

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

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= olarak hesaplanır. Bu olasılık, yaş ile kan basıncı arasındaki korelasyon katsayısı 0.82 olarak hesaplanan 150 bireylik örneğin ρ=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;

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

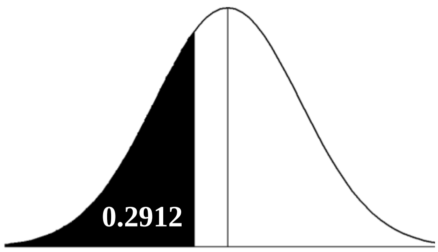
ve

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

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=0.2912olarak 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 hesaplansa, 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 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 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.

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

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.

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

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 Zr-değerlerine dönüştürülmesi gerekir.

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

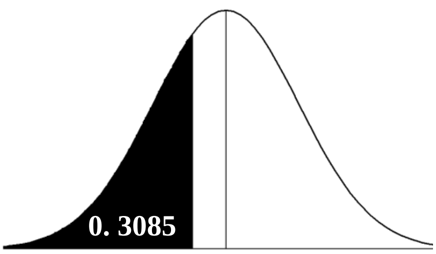
$$Z_{rB}=\frac{1}{2}\log_e\frac{(1+0.69)}{(1-0.69)}=1.1513\log\frac{(1+0.69)}{(1-0.69)}=0.8480$$

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 sayıdaki 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} hesaplansa, 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 katsayına 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.

$$\sigma_b^2 = \frac{\sigma_e^2}{\sum d_x^2} \qquad \ldots(5.23)$$

(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.

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

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

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

(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.

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

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

$$S_b^2 = \frac{S_e^2}{\sum d_x^2} \qquad \ldots(5.27)$$

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

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

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

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

5.8. Regresyon Katsayıları Arasındaki Farka ait Örnekleme 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 hesaplansa, 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 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 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, $\mu_{(bA-bB)}=0$ ve varyansı (5.30) numaralı eşitlikte verildiği gibidir.

$$\sigma^2_{(b1-b2)} = \sigma^2_{bA} + \sigma^2_{bB} \quad \ldots(5.30)$$

(5.30) numaralı eşitlikte σ^2_{bA} ve σ^2_{bB} (5.23) numaralı eşitlik kullanılarak hesaplanır. σ^2_{bA} ve σ^2_{bB}'nin (5.23) numaralı eşitlikteki değerleri kullanılarak (5.30) numaralı eşitlik (5.31) numaralı eşitlikte verildiği gibi düzenlenebilir.

$$\sigma^2_{(b1-b2)} = \left(\frac{\sigma_e^2}{\sum d_{xA}^2} + \frac{\sigma_e^2}{\sum d_{xB}^2}\right) \quad \ldots(5.31)$$

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.

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

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.

$$S^2_{(bA-bB)} = \frac{S^2_e}{\sum d^2_{xA}} + \frac{S^2_e}{\sum d^2_{xB}} \qquad \text{...(5.33)}$$

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

$$S_{(bA-bB)} = \sqrt{S^2_e\left(\frac{1}{\sum d^2_{xA}} + \frac{1}{\sum d^2_{xB}}\right)} \qquad (5.34)$$

(5.34) numaralı eşitlikte S^2_e, 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^2_e, (5.32) numaralı eşitlikteki populasyona ait regresyondan sapma kareler ortalamasının, σ^2_e, en iyi tahminidir.

$$S^2_e = \frac{\sum d^2_{eA} + \sum d^2_{eB}}{(n_A - 2) + (n_B - 2)} \qquad \text{...(5.35)}$$

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. π=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 μ=60, standart sapması da σ=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 μ = 500 kg, varyansı σ^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 hesaplansa, bu hesaplanan ortalamaların ortalamasının ve standart sapmasının kaç olması beklenir? Nasıl bir dağılım gösterirler?
8. π = 0.4 olan binomiyal bir populasyondan her defasında 25 birey bulunan rastgele örnekler çekilse ve bu örneklerin her birinde p'ler hesaplansa, bu p'lerin gösterdiği dağılımın ortalama ve standart sapması nedir? Hesaplayınız.

9. π = 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. π = 0.4 olan binomiyal bir populasyondan her defasında 25 birey bulunan rastgele örnekler çekilse ve bu örneklerin her birinde p'ler hesaplansa, 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 μ = 500 kg ve standart sapması σ = 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 hesaplansa;
   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, $n_A = n_B = 10$, $\overline{A}$ =25, S_A=3.21 ve $\overline{B}$ =22.2, S_B=3.09 olarak bulunmuştur. Bu iki ortalama arasındaki farkın standart hatası kaçtır?

HİPOTEZ KONTROLLERİ 1: Z-KONTROLLERİ

7.1. Giriş

Bir araştırıcı yürüttüğü bir denemeden elde ettiği verileri kullanarak populasyon parametrelerini tahmin ettikten sonra populasyon parametreleri ile ilgili öne sürdüğü hipotezini kontrol edebilir.

Üzerinde çalışılan özelliklere ait toplanan veriler, çalışılan özelliğe bağlı olarak farklı dağılımlar gösterebilir. Ancak, ölçüm, tartım veya analiz edilerek elde edilen özellikler için en sık rastlanan dağılım normal dağılımdır.

Elde edilen verilerin istatistik olarak analiz edilmesi sonucunda güvenilir sonuçların elde edilmesi, verilerin dağılımı hakkında bilgi sahibi olmayı ve doğru istatistik metodunun seçimini gerektirir.

İstatistik metotlar parametrik ve parametrik olmayan metotlar olarak iki gruba ayrılır. Parametrik yöntemlerin ön şartlarından biri, analiz edilecek verilerin normal dağılım göstermesidir. Analiz edilecek verilerde en sık rastlanan dağılım, normal dağılımdır. Bu sebeple bu kitabın kapsamında “verilerin normal dağılım göstermesi” ön şartının sağlandığı varsayımı altında parametrik hipotez kontrolleri örnekler ile açıklanacaktır.

Normal dağılım gösterdiği varsayılan bir özelliğin ait olduğu populasyonun ortalaması ve standart sapması biliniyorsa bu normal dağılım da diğer bütün normal dağılımlar gibi standart normal dağılıma dönüştürülebilir. Böylece istenen olasılıklar hesaplanabilir ve hipotez kontrolü yapılabilir. Hipotez kontrolü, normal dağılım gösteren populasyondan alındığı ileri sürülen “n” hacimli bir örnekten hesaplanan herhangi bir istatistiğin, söz konusu populasyondan “n” hacimli örneklerden elde edilecek söz konusu istatistiğe ait örnekleme dağılımına dahil olma olasılığının hesaplanmasıdır. Bunun için, örnekten hesaplanan istatistik ortalama ise hesaplanan ortalamaya karşılık gelen Z-değeri (5.3) numaralı eşitlik kullanılarak $Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}}$ şeklinde hesaplanarak standart normal dağılıma dönüştürülür.

Bölüm 4.4’te açıklandığı gibi standart normal dağılımın olasılık yoğunluk fonksiyonu (4.7 numaralı eşitlik);

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

şeklindedir. Ortalama ile belirli Z-değerleri arasında kalan alanlar olasılık yoğunluk fonksiyonunun integrali alınarak hesaplanmış ve tablo olarak düzenlenmiştir (Tablo A).

Örnekten hesaplanan ortalamadan daha büyük veya daha küçük değerlerin ortalamaya ait örnekleme dağılımına dahil olma olasılığı Tablo A’da verilen standart normal dağılım tablosundan daha önce açıklandığı şekilde (Bölüm 5.1) hesaplanır. Araştırıcı, hesaplanan olasılık doğrultusunda öne sürdüğü hipotez hakkında karar verdiği zaman hipotez kontrolü tamamlanmış olur.

7.2. Z-Kontrolleri

Hipotez kontrolleri yapılırken izlenmesi gereken adımlar Bölüm VI'da açıklanmıştı. Üzerinde çalışılan örneğin temsil ettiği populasyona ait dağılımın bilindiği durumlarda test dağılımı olarak Z-dağılımı ve test istatistiği olarak Tablo 6.1'de de verildiği gibi Z-değeri hesaplanır. Örnekten hesaplanan çeşitli istatistikler için Z-dağılımı kullanılarak hipotez kontrolleri bu bölümde örnekler ile açıklanacaktır.

7.2.1. Ortalamaya ait Hipotez Kontrolü

Parametreleri bilinen (ortalama ve standart sapma) normal dağılımlı bir populasyondan tamamen tesadüfen alındığı ileri sürülen bir örnekten hesaplanan ortalama ile populasyonun bilinen ortalaması arasındaki farkın tesadüften ileri gelip gelmediğine ortalamaya ait hipotez kontrolü yapılarak karar verilir.

ÖRNEK 1:

Kafkas ırkı arılarda dil uzunluğunun, ortalaması 7.2 mm (μ=7.2) ve standart sapması da 0.3 mm (σ = 0.3) olan normal dağılım gösterdiği bilinmektedir. Tesadüfen alınan 25 arıda dil uzunluğu ortalaması 7.05 mm olarak bulunmuştur. Ölçümü yapılan bu arı örneği, Kafkas ırkı arı populasyonuna mı aittir?

Bu örnekte araştırıcının amacı, tesadüfen alınan 25 arının dil uzunluğu bakımından Kafkas ırkı arı populasyonuna ait olup olmadığını kontrol etmektir. Bunun için hipotez kontrolü yapılması gerekir ve ilk olarak aşağıdaki şekilde hipotez takımı oluşturulur.

H_0: Populasyon ortalaması ile örnek ortalaması arasındaki fark tesadüften ileri gelmektedir. Gözlenen 0.15 mm'lik fark istatistik olarak önemli değildir ve sıfır kabul edilebilir. Yani tesadüfen alınan 25 arının Kafkas ırkı arı populasyonuna ait olduğu söylenebilir. Kısaca, μ_x=7.2 mm'dir.

H_1: Populasyon ortalaması ile örnek ortalaması arasındaki fark tesadüften ileri gelmemektedir. Gözlenen 0.15 mm'lik fark istatistik olarak önemlidir ve sıfır kabul edilemez. Yani tesadüfen alınan 25 arının Kafkas ırkı arı populasyonuna ait olduğu söylenemez. Kısaca, $\mu_x \neq 7.2$ mm'dir.

Araştırıcı, tesadüfen alınan 25 arının Kafkas ırkı arı populasyonuna ait olup olmadığını araştırdığı ve dil uzunluğu ortalamasının populasyon ortalamasından farklı olup olmadığı ile ilgilendiği başka bir deyişle örnek ortalamasının, populasyon ortalamasından küçük yada büyük olmasıyla ilgilenmediği için hipotez kontrolü çift taraflıdır. Yapılan hipotez kontrolünün çift taraflı kontrol olduğunu karşıt hipotez belirler ve bunun için "$\mu_x \neq 7.2$ mm" şeklinde kurulmuştur.

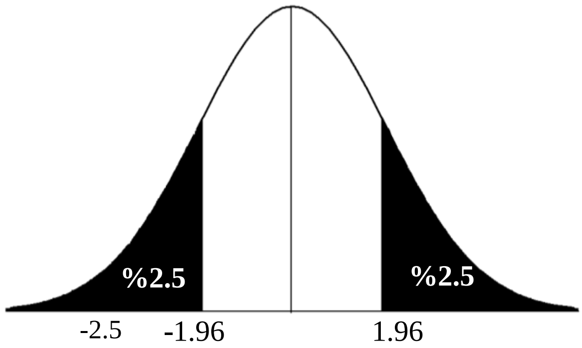
Kafkas ırkı arılarda dil uzunluğuna ait populasyonun μ=7.2 ve σ = 0.2 mm olan normal dağılım gösterdiği bildirilmişti. Bu populasyondan n= 25 arılık çok sayıda örnekler alınsa ve ortalamalar hesaplansa, hesaplanan bu ortalamalar örnekten örneğe değişerek bir dağılım gösterir. Bu dağılıma, ortalamalara ait örnekleme dağılımı adı verilir.

Bu dağılımın parametreleri, $\mu_{\bar{x}} = \mu_x$ =7.2 mm ve standart sapması da $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{0.3}{\sqrt{25}}$ =0.06 mm olup, normal dağılım gösterir.

Kafkas ırkı arılarda dil uzunluğu populasyonunun standart sapması bilindiği için yapılacak hipotez kontrolünde test dağılımı olarak Z-dağılımı ve test istatistiği olarak ta Z-değeri kullanılır. Örnekten hesaplanan istatistik ortalama ise hesaplanan ortalamaya karşılık gelen Z-değeri (5.3) numaralı eşitlik kullanılarak;

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{7.05 - 7.2}{0.06} = -2.5 \text{ olarak hesaplanır.}$$

Eğer I. tip hata olasılığı (α) %5 olarak belirlenmiş ise çift taraflı hipotez kontrolü yapıldığı için I. tip hata olasılığının yarısı, ortalamadan büyük değerlerin bulunduğu diğer yarısı ise ortalamadan küçük değerlerin bulunduğu tarafta alınır ve böylece test dağılımında kontrol hipotezini kabul ve ret bölgeleri Şekil 7.1'de görüldüğü gibi belirlenmiş olur.



Şekil 7.1. Test dağılımında ret ve kabul bölgeleri

Şekil 7.1'de siyah taralı alanlar, çift taraflı hipotez kontrolünde kontrol hipotezinin ret bölgeleri olup bu bölgelerin başlangıç noktaları da ± 1.96 değerleridir. Yukarıda verilen örneğimizden hesaplanan ortalamaya karşılık gelen Z-değerinin, yani test istatistiğinin değeri -2.5 olup, kontrol hipotezini ret bölgesine düştüğü için kontrol hipotezi (H_0) ret edilir ve karşıt hipotez geçerli olur. Bu durumda, tesadüfen alınan 25 arının sadece dil uzunluğu dikkate alındığında Kafkas ırkı arı populasyonuna ait olduğu söylenemez. Çünkü populasyon ortalaması ile örnek ortalaması arasındaki fark tesadüfi olmayıp aksine bu fark örneği oluşturan arıların başka bir ırka ait arılar olmasından kaynaklanmaktadır.

ÖRNEK 2:

Karayaka ırkı koyunlarda laktasyon süt veriminin, ortalaması 40 kg (μ=40 kg) ve standart sapması da 8 kg (σ = 8 kg) olan normal dağılım gösterdiği bilinmektedir. Tesadüfen alınan 64 adet koyunda laktasyon süt verimi ortalaması 38 kg olarak bulunmuştur. Acaba tesadüfen alınan 64 koyun laktasyon süt verimi 40 kg'dan daha az olan bir ırktan mıdır?

Burada, tesadüfen alınan 64 koyunun laktasyon süt verimi 40 kg'dan daha az olup olmadığı araştırıldığı için yapılması gereken tek taraflı hipotez kontrolüdür ve hipotezlerin aşağıdaki şekilde kurulması gerekir.

H_0: Populasyon ortalaması ile örnek ortalaması arasındaki fark tesadüften ileri gelmektedir. Gözlenen 2 kg'lık fark istatistik olarak önemli değildir, sıfır kabul edilebilir. Tesadüfen alınan 64 koyunun Karayaka ırkı koyunlar olduğu söylenebilir. Kısaca, μ_x=40 kg'dır.

H_1: Populasyon ortalaması ile örnek ortalaması arasındaki fark tesadüften ileri gelmemektedir. Gözlenen 2 kg'lık fark istatistik olarak önemlidir ve sıfır kabul edilemez. Tesadüfen alınan 64 koyun, süt verimi, Karayaka ırkı koyunlarının süt veriminden daha düşük bir ırktandır. Kısaca, μ_x<40 kg'dır.

Hipotezler kurulduktan sonra hipotez kontrolünün tamamlanması için gerekli işlemler ve verilen karar Tablo 7.1'de açıklanmıştır.

Tablo 7.1. Tesadüfen alınan 64 koyunun laktasyon süt veriminin 40 kg'dan daha az olup olmadığına ilişkin hipotez kontrolü

Örnekten hesaplanan istatistik, ortalama olduğu için kullanılması gereken örnekleme dağılımı, ortalamalara ait örnekleme dağılımıdır. Karayaka ırkı koyunlarda laktasyon süt verimi populasyonun normal dağılım gösterdiği bildirildiği için bu populasyondan elde edilecek ortalamaya ait örnekleme dağılımı, parametreleri $\mu_{\bar{x}} = \mu_x$ =40 kg ve $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{8}{\sqrt{64}}$ =1 kg olan normal dağılım gösterir.

Populasyona ait standart sapma bilindiği için hipotez kontrolünde kullanılması gereken test dağılımı **Z-dağılımı** ve test istatistiği **Z-değeridir**.

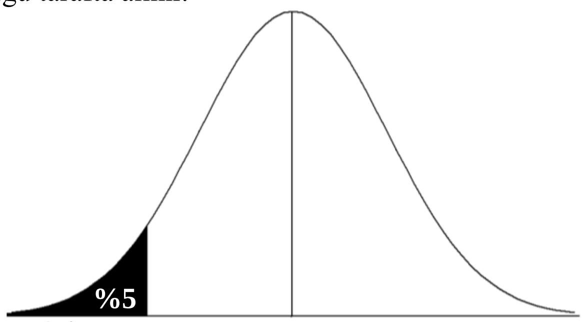
Üzerinde çalışılan örneğin, Karayaka ırkı koyunlarda laktasyon süt verimi populasyonunu temsil etme olasılığını bulmak için hesaplanan ortalamaya karşılık gelen Z-değeri $Z = \frac{38-40}{1} = -2.0$ olarak bulunur.

Tablo 7.1. devam

Z-dağılımında bütün Z-değerlerinin -2.0 ve daha küçük olma olasılığı, Tablo A'dan yararlanılarak, Bölüm 5.1 de açıklandığı şekilde P(Z≤-2.0)=0.0228 olarak hesaplanır.

Eğer araştırıcı I. tip hata olasılığını α= %5 olarak belirlemiş ise, 64 koyunluk laktasyon süt verimlerini içeren söz konusu örneğin Karayaka ırkı koyunlarda laktasyon süt verimine ait populasyonu temsil etme olasılığı %5'ten düşüktür. Bu durumda kontrol hipotezi ret edilir. Karşıt hipotez geçerli olur. Yani söz konusu örneğin laktasyon süt verimi ortalaması 40 kg'dan daha az olan bir ırkı temsil ettiği söylenebilir.

Karar verme aşamasında üzerinde çalışılan örneğin populasyonu temsil etme olasılığı kullanılarak karar verilebileceği gibi şu şekilde de karar verilebilir: Z-dağılımında kontrol hipotezinin ret bölgesi, yani %5'lik alan, tek taraflı kontroller için ya 1.645 yada -1,645 değerinden başlamaktadır. Bu örnekte tesadüfen alınan 64 koyunun laktasyon süt veriminin 40 kg'dan daha az olup olmadığı kontrol edildiği için Şekil 7.2'de görüldüğü gibi %5'lik alan ortalamadan küçük değerlerin bulunduğu tarafta alınır.



Şekil 7.2. Test dağılımında ret ve kabul bölgeleri

Şekil 7.2'de görüldüğü gibi hesaplanan test istatistiğinin değeri -2.0, -1.645'ten küçük olduğu için, kontrol hipotezinin ret bölgesine düştüğünden kontrol hipotezi ret edilir.

ÖRNEK 3:

Belirli bir dersten öğrencilerin aldıkları notların, ortalaması 60 (μ=60) ve standart sapması da 12 (σ = 12) olan normal dağılım gösterdiği bilinmektedir. Tesadüfen alınan 36 adet öğrenciye yeni bir öğretim metodu uygulanmış ve başarı ortalaması 63 olarak bulunmuştur. Yeni öğretim metodunun öğrencilerde başarıyı artırdığı söylenebilir mi?

Burada araştırıcı bir dersten öğrencilere uygulanan yeni öğretim metodunun başarıyı artırıp artırmadığı ile ilgilenmektedir. Yeni öğretim metodunun başarıyı artırdığının söylenebilmesi için yeni öğretim metodu uygulanan 36 öğrenciden hesaplanan ortalamanın, bildirilen populasyon ortalamasından daha büyük olması gerekmektedir. Dolayısıyla hipotez kontrolü tek taraflıdır ve hipotezler aşağıdaki şekliyle kurulur.

H_0: Populasyon ortalaması ile yeni öğretim metodu uygulanan öğrencilerin not ortalaması arasındaki 3 puanlık fark tesadüften ileri gelmiştir. Bu fark istatistik olarak önemli değildir ve sıfır kabul edilebilir. 36 öğrenciye uygulanan yeni öğretim metodunun başarıyı artırdığı söylenemez. Kısaca, μ_x=60'dır.

H_1: Populasyon ortalaması ile yeni öğretim metodu uygulanan öğrencilerin not ortalaması arasındaki 3 puanlık fark tesadüften ileri gelmemiştir. Bu fark istatistik olarak önemlidir ve sıfır kabul edilemez. 36 öğrenciye uygulanan yeni öğretim metodunun başarıyı artırdığı söylenebilir. Kısaca, μ_x>60'dır.

Hipotezler kurulduktan sonra hipotez kontrolünün tamamlanması için gerekli işlemler ve verilen karar Tablo 7.2'de açıklanmıştır.

Tablo 7.2. 36 öğrenciye uygulanan yeni öğretim metodunun başarıyı artırıp artırmadığına ilişkin hipotez kontrolü

Örnekten hesaplanan istatistik, ortalama olduğu için kullanılması gereken örnekleme dağılımı, ortalamalara ait örnekleme dağılımıdır. Belirli bir dersten öğrencilerin aldıkları notların normal dağılım gösterdiği bildirildiği için bu populasyondan elde edilecek ortalamalara ait örnekleme dağılımı, parametreleri $\mu_{\bar{x}} = \mu_x$ =60 kg ve $\sigma_{\bar{x}} = \frac{12}{\sqrt{36}}$ =2 olan normal dağılım gösterir.

Populasyona ait standart sapma bilindiği için hipotez kontrolünde kullanılması gereken test dağılımı **Z-dağılımı** ve test istatistiği **Z-değeridir**.

Tablo 7.2.devam

Yeni öğretim metodu uygulanan 36 öğrencilik örneğin, ortalaması 60 ve standart sapması da 12 olan ve normal dağılım gösteren final notu populasyonunu temsil etme olasılığını bulmak için hesaplanan ortalamaya karşılık gelen Z-değeri $Z = \frac{63-60}{2} = 1.5$ olarak bulunur.

Z-dağılımında, bütün Z-değerlerinin 1.5 ve daha büyük olma olasılığı, P(Z≥1.5)=0.0668'dir. (% 6.68)

Eğer araştırıcı I. tip hata olasılığını α= %1 olarak belirlemiş ise, 36 öğrencilik örneğin, ortalaması 60 ve standart sapması da 12 olan normal dağılım gösteren final notu populasyonunu temsil etme olasılığı %1'den büyüktür. Bu durumda kontrol hipotezi kabul edilir, yani tesadüfen seçilen 36 öğrenciye uygulanan yeni öğretim metodunun başarıyı arttırdığı söylenemez.

Karar verme aşamasında, üzerinde çalışılan örneğin populasyonu temsil etme olasılığı kullanılarak karar verilebileceği gibi şu şekilde de karar verilebilir: Z-dağılımında kontrol hipotezinin ret bölgesi, yani %1'lik alan, tek taraflı hipotez kontrolleri için ya -2.33 ya da 2.33 değerinden başlamaktadır. Hesaplanan test istatistiğinin değeri -2.0,-2.33'ten büyük olduğu için, kontrol hipotezinin ret bölgesine düştüğünden kontrol hipotezi ret edilir. Dolayısıyla karşıt hipotez geçerli olur.

ÖRNEK 4:

Sağlıklı yetişkin bireylerde sistolik kan basıncı ortalaması 125 mmHg, standart sapması 18 mmHg olan normal dağılım göstermektedir. Tesadüfen seçilen 64 bireyde sistolik kan basıncı ortalaması 132 mmHg olarak bulunmuştur. Bu 64 bireylik örneğin, söz konusu populasyondan tesadüfen alınan örneklerden biri olduğu söylenebilir mi?

Tesadüfen seçilen 64 bireyin sağlıklı bireyler populasyonunu temsil ettiğini söyleyebilmek için, örnekten hesaplanan sistolik kan basıncı ortalaması ile sağlıklı bireylere ait sistolojik kan basıncı ortalaması arasındaki farkın tesadüften ileri geliyor olması gerekir. Çünkü sistolik kan basıncı ortalamasının bildirilen ortalamadan büyük veya küçük olması herhangi bir sağlık sorunun göstergesi olabilir. Bunun için de yapılması gereken çift taraflı hipotez kontrolüdür ve bu kontrol için gerekli olan hipotezler aşağıdaki şekilde kurulur.

H_0: Populasyon ortalaması ile örnek ortalaması arasındaki fark tesadüften ileri gelmektedir. Gözlenen 7 mmHg'lık fark istatistik olarak önemli değildir, sıfır kabul edilebilir. Tesadüfen seçilen 64 bireyin sağlıklı bireyler olduğu söylenebilir. Kısaca, μ_x=125 mmHg'dır.

H_1: Populasyon ortalaması ile örnek ortalaması arasındaki fark tesadüften ileri gelmemektedir. Gözlenen 7 mmHg'lık fark istatistik olarak önemlidir ve sıfır kabul edilemez. Tesadüfen seçilen 64 bireyin, söz konusu populasyondan seçilen sağlıklı bireyler olduğu söylenemez. Kısaca, $\mu_x \neq$125 mmHg'dır.

Hipotezler kurulduktan sonra hipotez kontrolünün tamamlanması için gerekli işlemler ve verilen karar Tablo 7.3'te açıklanmıştır.

Tablo 7.3. 81 bireylik örneğin sistolik kan basıncı bakımından sağlıklı bireyler olup olmadığına ilişkin hipotez kontrolü

Populasyona ait standart sapma bilindiği için hipotez kontrolünde kullanılması gereken test dağılımı **Z-dağılımı** ve test istatistiği **Z-değeridir**. Sağlıklı bireylerden mümkün olan sayıda 81 bireylik örnekler seçilse ve her birinde sistolik kan basıncı ortalaması hesaplansa, hesaplanan bu ortalamalar, ortalaması $\mu_{\bar{x}} = \mu_x$ =125 mmHg ve standart sapması $\sigma_{\bar{x}} = \frac{18}{\sqrt{64}}$ =2.25 mmHg olan normal dağılım gösterir. Üzerinde çalışılan örnekten hesaplanan ortalamaya karşılık gelen Z-değeri $Z = \frac{132-125}{2.25} = 3.11$ olarak bulunur. Eğer I. tip hata olasılığı %1 olarak belirlenmiş ise Z-dağılımında kontrol hipotezinin çift taraflı hipotez kontrolü için ret bölgesi, yani %0.1'lik alan ±2.575 değerinden başlamaktadır. Hesaplanan test istatistiğinin değeri 3.11 olup 2.575 değerinden daha büyüktür ve kontrol

hipotezinin ret bölgesine düşmektedir. Bu sebeple de kontrol hipotezi ret edilir ve seçilen 64 bireyin sağlıklı bireyler olduğu söylenemez.

7.2.2. Ortalamalar arası Farka ait Hipotez Kontrolü

Belirli bir özellik bakımından iki ayrı muameleye (faktöre, gruba) tabi tutulan deney ünitelerinin ortalamaları arasında gözlenen farkın istatistik olarak önemli olup olmadığının test edilmesi amaçlanmış ise ortalamalar arası farka ait hipotez kontrolünün yapılması gerekir. Ortalamalar arası farka ait hipotez kontrolü yapılırken de bir önceki bölümde açıklanan işlem basamakları aynen yerine getirilir. Ortalamalar arası farkların hipotez kontrolü, aşağıda verilen örneklerde detaylı olarak açıklanmıştır.

ÖRNEK 1:

Bir fakültede okutulan istatistik dersi final notlarının standart sapması 8 olan normal dağılım gösterdiği bilinmektedir. Listeden tesadüfen seçilen 16 erkek öğrencinin notlarının ortalaması 75, 25 kız öğrencinin notlarının ortalaması ise 78 olarak bulunmuştur. Kız ve erkek öğrencilerin aldıkları notlar arasındaki fark istatistik olarak önemli midir?

Araştırıcı istatistik dersi final sınavından alınan notlar bakımından kız ve erkek öğrencilerin (iki grup) not ortalamaları arasında gözlenen farkın istatistik olarak önemli olup olmadığını kontrol etmek istemektedir. Bu durumda çift taraflı hipotez kontrolünün yapılması ve hipotezlerin aşağıdaki şekilde kurulması gerekir.

H_0: İstatistik dersi final sınavından alınan notlar bakımından kız ve erkek öğrencilerin not ortalamaları arasındaki fark tesadüften ileri gelmektedir. Kız ve erkek öğrencilerin not ortalamaları arasında gözlenen 3 puanlık fark istatistik olarak önemli değildir ve sıfır kabul edilebilir. Kısaca, $\mu_{\bar{E}} - \mu_{\bar{K}} = 0$ veya $\mu_{\bar{E}} = \mu_{\bar{K}}$ 'dır.

H_1: İstatistik dersi final sınavından alınan notlar bakımından kız ve erkek öğrencilerin not ortalamaları arasındaki fark tesadüften ileri gelmemektedir. Kız ve erkek öğrencilerin not ortalamaları arasında gözlenen 3 puanlık fark istatistik olarak önemlidir ve sıfır kabul edilemez. Kısaca, $\mu_{\bar{E}} - \mu_{\bar{K}} \neq 0$ veya $\mu_{\bar{E}} \neq \mu_{\bar{K}}$ 'dır.

Eğer kontrol hipotezi doğru ise finalde alınan notlar bakımından kız ve erkek öğrencilerin not ortalamaları arasında gözlenen farklar, ortalaması sıfır ve standart sapması (5.4) numaralı eşitlikte verildiği gibi olan bir normal dağılım gösterecektir.

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.

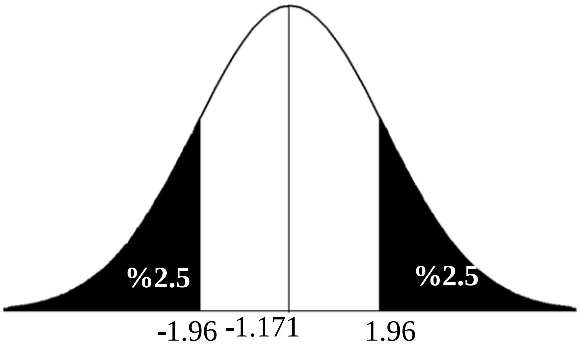
Bir fakültede okutulan istatistik dersi final notlarının standart sapması 8 olan normal dağılım gösterdiği ve listeden tesadüfen seçilen 16 erkek öğrencinin notlarının ortalaması 75, 25 kız öğrencinin notlarının ortalaması ise 78 olarak bulunduğuna göre, ortalamalar arası farka ait örnekleme dağılımının standart sapması (5.4) numaralı eşitlik kullanılarak;

$$\sigma_D = \sigma\sqrt{\frac{(n_A + n_B)}{n_A n_B}} = 8\sqrt{\frac{(25+16)}{(25)(16)}}$$ =2.561 olarak bulunur.

Test istatistiği (Z-değeri) ise (5.5) numaralı eşitlikten;

$$Z = \frac{(\bar{A} - \bar{B})}{\sigma_D} = \frac{(75-78)}{2.561}$$ =-1.171 olarak hesaplanır.

Yapılan hipotez kontrolünde I. tip hata olasılığı önceden (α=0.05) %5 olarak belirlenmiş ise yapılan hipotez kontrolü çift taraflı olduğu için bu ihtimalin yarısı (%2.5'u) ortalamadan küçük değerlerin bulunduğu tarafta, diğer yarısı ise ortalamadan büyük değerlerin bulunduğu tarafta alınır. Şekil 7.3'de görüldüğü gibi kontrol hipotezinin ret bölgesinin ±1.96 değerinden başladığı belirlenmiş olur.



Şekil 7.3. Test dağılımında ret ve kabul bölgeleri

Örnekten hesaplanan ortalamalar arası farka karşılık gelen Z-değeri, yani test istatistiği -1.171olduğu için bu değer -1.96'dan büyüktür ve kontrol hipotezinin kabul bölgesine düşer. Yani, $|Z|<Z_\alpha$ veya |-1.171|<1.96 olduğundan kontrol hipotezi kabul edilir. Finalde alınan notlar bakımından kız ve erkek öğrencilerin not ortalamaları arasında gözlenen farkın, ortalamalar arası farka ait örnekleme dağılımına dahil olma olasılığı %5'ten büyüktür. Bu durumda, finalde alınan notlar bakımından kız ve erkek öğrencilerin not ortalamaları arasında gözlenen farkın tesadüften ileri geldiği ve cinsiyetler arasında başarı bakımından fark olmadığı kararına varılır.

ÖRNEK 2:

Yeni doğan erkek bebeklerin ilk dört ayın sonundaki ağırlıklarının gösterdiği dağılımın normal dağılım, bu dağılıma ilişkin varyansında 2.9 olduğu bilinmektedir. Yeni doğmuş ve aynı ağırlıktaki 40 bebek tesadüfen iki gruba ayrılmış ve 15 bebeğe A maması, 25 bebeğe ise B maması verilmiştir. Dördüncü ay sonunda A maması ile beslenen bebeklerin ağırlıklarının ortalaması 6.5 kg, B maması ile beslenen bebeklerin ağırlıklarının ortalaması ise 7.3 kg olarak bulunmuştur. Dördüncü ay sonundaki ağırlıklarına etki bakımından B mamasının A mamasından daha iyi olduğu söylenebilir mi?

Erkek bebeklerin dördüncü ay ağırlığına etki bakımından B mamasının A mamasından daha iyi olup olmadığı araştırıldığı için tek taraflı hipotez kontrolü yapmak gerekir. Buna ilişkin hipotez takımı aşağıdaki gibi kurulur.

H₀: A ve B maması ile beslenen erkek bebeklerin 4. ay sonundaki ağırlık ortalamaları arasında gözlenen fark tesadüften ileri gelmiştir. 4. ay ağırlıkları bakımından mamalar arasında bir fark yoktur. Kısaca $\mu_{\bar{A}} - \mu_{\bar{B}} = 0$ veya $\mu_{\bar{A}} = \mu_{\bar{B}}$'dir.

H₁: A ve B maması ile beslenen erkek bebeklerin 4. ay sonundaki ağırlık ortalamaları arasında gözlenen fark tesadüften ileri gelmemiştir ve istatistik olarak önemlidir. Yani, 4. ay ağırlığı bakımından B mamasının, A mamasından daha iyi olduğu söylenebilir. Kısaca $\mu_{\bar{B}} - \mu_{\bar{A}} > 0$ veya $\mu_{\bar{B}} > \mu_{\bar{A}}$'dir.

Hipotezler takımı oluşturulduktan sonra hipotez kontrolünün tamamlanması için gerekli işlemler ve verilen karar Tablo 7.5'te açıklanmıştır.

Tablo 7.5. Dördüncü ay sonunda erkek bebeklerin ağırlığına etki bakımından B mamasının A mamasından daha iyi olup olmadığına ilişkin hipotez kontrolü

$n_A = 15$ $\bar{A} = 6.5$ kg
$n_B = 25$ $\bar{B} = 7.3$ kg

olarak bulunduğuna göre ortalamalar arası farka ait örnekleme dağılımının standart sapması (5.4) numaralı eşitlik kullanılarak;

$\sigma_D = \sqrt{(2.9)\frac{(15+25)}{(15)(25)}} = 0.556$ kg ve test istatistiği de

(5.5) numaralı eşitlikten; $Z = \frac{(7.3-6.5)}{0.556} = 1.439$ olarak hesaplanır.

Eğer I. tip hata olasılığı %5 olarak belirlenmiş ise Z-dağılımında kontrol hipotezinin tek taraflı hipotez kontrolü için ret bölgesi 1.645 değerinden başlamaktadır. Hesaplanan test istatistiğinin değeri 1.439 olup 1.645 değerinden küçüktür ve kontrol hipotezini kabul bölgesine düşer. Dolayısıyla kontrol hipotezi kabul edilir ve dördüncü ay sonunda erkek bebeklerin ağırlığına etki bakımından B mamasının A mamasından daha iyi olduğu söylenemez.

7.2.3. Oranlara ait Hipotez Kontrolü

Üzerinde çalışılan kesikli dağılım gösteren özelliklerin bazılarında iki durum söz konusudur ve binomiyal dağılıma uygun bir dağılım gösterirler. İstenen olayın (üzerinde durulan olay) oluş olasılığı Bölüm 4.1'de açıklandığı şekilde hesaplanır. Binomiyal dağılım gösteren bir populasyondan alınmış olduğu ileri sürülen bir örnekten hesaplanan istenen olayın oluş olasılığı (p) ile söz konusu olaya ilişkin bildirilen olasılıklar arasındaki farkın önemli olup olmadığı kontrol edilmek istendiğinde oranlara ait hipotez kontrolünden yararlanılır.

Oranlara ait hipotez kontrolü yapılırken örneklerin binomiyal dağılım gösteren bir populasyon alınmış olması gerekir. Yapılan hipotez kontrolü sonucunda verilen kararın güvenilir olması için üzerinde çalışılan örneğin genişliği, örnekten hesaplanan oranların yaklaşık normal dağılım göstermeleri için yeterli büyüklükte olmalıdır. Bunun için de Bölüm 5.3'te açıklanan $\pi \leq 1/2$ ise oranlara ait örnekleme dağılımının normal dağılıma yaklaşabilmesi için $n.\pi \geq 10$ veya $\pi > 1/2$ ise oranlara ait örnekleme dağılımının normal dağılıma yaklaşabilmesi için $n.(1-\pi) \geq 10$ şartı sağlanmalıdır.

Oranlara ait hipotez kontrolüne geçmeden önce örnek genişliğinin, örnekten hesaplanan oranların yaklaşık normal dağılım göstermesi için yeterli büyüklükte olup olmadığının mutlaka kontrol edilmesi gereklidir. Oranlara ait hipotez kontrolü aşağıda verilen örnekler ile açıklanmıştır.

ÖRNEK 1:

Sarılık tedavisinde yaygın olarak kullanılan bir ilacın hastalığı iyileştirme oranının π=0.80 olduğu bildirilmiştir. Bir ilaç firması bu hastalığın tedavisinde kullanılmak üzere yeni bir ilaç geliştirdiğini belirtmiştir. Bir doktor, yeni geliştirilen ilacın yaygın olarak kullanılan ilaç yerine kullanılıp kullanılamayacağını araştırmak üzere 200 sarılık hastasını yeni ilaç ile tedavi etmiş ve bu hastalardan 170 tanesinin iyileştiğini saptamıştır. Yeni geliştirilen ilaç, yaygın olarak kullanılan ilaç yerine kullanılabilir mi, yani sarılık hastalığını iyileştirme oranı bakımından yeni ilacın daha etkili olduğu söylenebilir mi?

Bu araştırmayı yürüten araştırıcının amacı yeni ilacın kullanılan ilaçtan daha etkili olup olmadığını araştırmaktır. Yeni ilacın daha etkili olduğunun söylenebilmesi için örnekten hesaplanan oranın bildirilen hastalığı iyileştirme oranından daha yüksek olması ve bu yüksekliğinde istatistik olarak önemli olması gerekir. Bu sebeple de tek taraflı hipotez kontrolüne başvurulmalıdır. Buna ilişkin hipotez takımı aşağıdaki gibi oluşturulur.

H_0: Populasyona ait sarılık hastalığını iyileştirme oranı ile örnekten hesaplanan iyileştirme oranı arasındaki fark tesadüften ileri gelmektedir ve sıfır kabul edilebilir. Yeni ilacın, sarılık hastalığını tedavide kullanılan ilaçtan daha etkili olduğu söylenemez. Kısaca, μ_p=0.80 veya π=0.80'dır.

H_1: Populasyona ait sarılık hastalığını iyileştirme oranı ile örnekten hesaplanan iyileştirme oranı arasındaki fark tesadüften ileri gelmemektedir ve sıfır kabul edilemez. Yeni ilacın, sarılık hastalığını tedavide kullanılan ilaçtan daha etkili olduğu söylenebilir. Kısaca, μ_p>0.80 veya π>0.80'dır.

Oranlara ait hipotez kontrolüne geçilmeden önce n=200 hastadan oluşan örneklerden hesaplanacak iyileşme oranlarının normal dağılıma yaklaşıp yaklaşmadığı kontrol edilmelidir. Eğer yeni ilaç ile kullanılan ilacın hastalığı iyileştirme bakımından aralarındaki fark tesadüften ileri geliyorsa, yeni ilaç ile tedavi edilen hastalar da π=0.80 olan binomiyal dağılım gösterecektir. π=0.80 olduğu için $n(1-\pi) \geq 10$ şartının yerine getirilmesi gerekir. $n(1-\pi) = 200(1-0.80) = 20$ ve $20 \geq 10$ olduğu için örnek genişliği yeterlidir ve hipotez kontrolüne geçilebilir.

Eğer kontrol hipotezi doğru ise 200 hastanın iyileşme oranının, sarılık hastalığını iyileştirme oranı π=0.80 olan populasyondan her birinde n=200 birey bulunan örneklerden elde edilecek oranlara ait örnekleme dağılımını oluşturan örneklerden birisi olması gerekir. Araştırıcıyı ilgilendiren oranlara ait örnekleme dağılımı, ortalaması, μ_p=0.80 ve standart sapması (5.8) numaralı eşitlikten $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.8(1-0.8)}{200}}$ =0.0283 olarak hesaplanan yaklaşık normal dağılım gösterir.

Bu durumda, normal dağılıma yaklaşan oranlara ait örnekleme dağılımını oluşturan oranlar (5.10) numaralı eşitlik kullanılarak standart normal dağılıma dönüştürülebilir, yani n=200 sarılık hastasından $p = \frac{170}{200}$ =0.85 olarak hesaplanan iyileşme oranına karşılık gelen Z-değeri;

$$Z = \frac{p-\mu_p}{\sigma_p} = \frac{0.85-0.80}{0.0283} = 3.00$$

olarak hesaplanır.

Eğer I. tip hata olasılığı %5 olarak belirlenmiş ise Z-dağılımında kontrol hipotezinin ret bölgesi, yani %5'lik alan ±1.645 değerinden başlamaktadır. Hesaplanan test istatistiğinin değeri 3.00 olup 1.645 değerinden daha büyüktür ve kontrol hipotezinin ret bölgesine düşmektedir. Dolayısıyla kontrol hipotezi reddedilerek, sarılık hastalığının tedavisinde yeni geliştirilen ilacın, yaygın olarak kullanılan ilaçtan daha etkili olduğu ve bu ilacın yerine kullanılmasının daha akılcı olduğu kararına varılır.

ÖRNEK 2:

Bir laboratuvardaki farelerin %40'ının koyu renkli geri kalanın da açık renkli olduğu bilinmektedir. Bir genetik çalışmada koyu renkli farelerin oranının arttırılması amaçlanmış ve bir yıl sonra tesadüfen alınan 400 fareden 220 tanesinin koyu renkli olduğu gözlenmiştir. Bu genetik çalışmanın koyu renkli farelerin oranını arttırdığı söylenebilir mi?

Bir laboratuvarda yapılan genetik çalışmasında koyu renkli farelerin oranının arttırılması amaçlanmış ve 400 fareden 220 tanesinin koyu renkli olduğu gözlenmiştir. Yürütülen genetik çalışmanın koyu renkli fare oranını artırıp artırmadığına karar vermek için tek taraflı hipotez kontrolü yapılmalıdır. Bu kontrol için gerekli hipotez takımı aşağıdaki gibi oluşturulur.

H_0: Koyu renkli fare oranları arasında gözlenen fark tesadüften ileri gelmiştir ve sıfır kabul edilebilir. Yani, uygulanan genetik çalışmanın koyu renkli fare oranını artırdığı söylenemez. Kısaca, μ_p=0.40 veya π=0.40'dır.

H_1: Koyu renkli fare oranları arasında gözlenen fark tesadüften ileri gelmeyip istatistik olarak önemlidir. Yani, uygulanan genetik çalışmanın koyu renkli fare oranını artırdığı söylenebilir. Kısaca, μ_p>0.40 veya π>0.40'dır.

Oranlara ait hipotez kontrolüne geçilmeden önce, $\pi<0.5$ olduğu için $n(\pi)\geq 10$ şartı kontrol edilmelidir. $n(\pi) = 400(0.40) = 160$ ve $160>10$ olduğu için örnek genişliği normal dağılım için yeterlidir ve hipotez kontrolüne geçilebilir.

Hipotezler kurulduktan sonra hipotez kontrolünün tamamlanması için gerekli işlemler ve verilen karar Tablo 7.6'da açıklanmıştır.

Tablo 7.6. Bir genetik çalışmanın koyu renkli fare oranını artırıp artırmadığına ilişkin hipotez kontrolü

Araştırıcıyı ilgilendiren oranlara ait örnekleme dağılımı, ortalaması μ_p=0.40 ve standart sapması $\sigma_p = \sqrt{\frac{0.4(1-0.6)}{400}}$ =0.0245 olan yaklaşık normal dağılım gösterir. n=400 fareden hesaplanan koyu fare oranı $p = \frac{220}{400}$ =0.55'tir ve buna karşılık gelen Z-değeri de; $$Z = \frac{0.55-0.40}{0.0245} = 6.122$$ dır. I. tip hata olasılığı %1 olarak belirlenmiş ise Z-dağılımında kontrol hipotezinin tek taraflı ret bölgesi, 2.33 değerinden başlamaktadır. Hesaplanan test istatistiğinin değeri 6.122 olup 2.33 değerinden büyüktür ve kontrol hipotezinin ret bölgesine düşer. Dolayısıyla, kontrol hipotezi ret edilir ve uygulanan genetik çalışmanın koyu renkli fare oranını artırdığı söylenebilir.

7.2.4. Oranlara ait Hipotez Kontrolünde Örnek Genişliği

Çalışılan populasyondan alınacak örnekte, istenen olayın (p) ve istenmeyen olayın (q) oluş olasılığının 0 ve 1'e yakın olmaması durumunda populasyona ait oran ile örnekten hesaplanan oran arasındaki farkın belirlenen fark (δ, duyarlılık) kadar küçük olacak şekilde tahmin edilebilmesi için örnek genişliği (7.1) numaralı eşitlik kullanılarak hesaplanabilir.

$$n = \frac{Z_\alpha^2 (pq)}{\delta^2} \qquad \ldots(7.1)$$

Eşitlikte, Z_α^2, belirlenen α seviyesindeki tek veya çift taraflı Z-değerini, δ ise populasyon için bildirilen oran ile örnekten hesaplanan oran arasında olması istenen en küçük farkı ifade etmektedir. Aynı zamanda (7.1) numaralı eşitlik kullanılarak tahmin edilen örnek genişliği, δ kadar olabilecek küçük farklılıkları ortaya çıkarmak için de yeterli örnek genişliğidir (Zar, 1999).

ÖRNEK:

Bir genetik laboratuvarındaki farelerden tesadüfen seçilen 200 fareden 70 tanesinin koyu renkli olduğu gözlenmiştir. H_0: π-p=0 kontrol hipotezinin, H_1: π-p$\neq$0 karşıt hipotezine karşı α=0.01 seviyesinde test edildiğinde H_0 hipotezinin δ=0.1 duyarlılıkla reddedilebilmesi için örnek genişliği en az kaç olmalıdır?

En az örnek genişliği (7.1) numaralı eşitlik kullanılarak tahmin edilebilir. Bu durumda yapılacak hipotez kontrolü çift taraflı hipotez kontrolüdür. α=0.01 olarak belirlendiği için çift taraflı Z_α değeri 2.575'dir.

200 fareden 70 tanesi koyu renkli olduğuna göre farelerin koyu renkli olma olasılığı, p=70/200=0.35'tir. Olması gereken en az örnek genişliği (7.1) numaralı eşitlikten;

$$n = \frac{Z_\alpha^2 (pq)}{\delta^2} = \frac{(2.575)^2(0.35x0.65)}{(0.1)^2} = 150.8 \cong 151$$

olarak hesaplanır. H_0: π-p=0 kontrol hipotezi, H_1: π-p$\neq$0 karşıt hipotezine karşı α=0.01 seviyesinde test edildiğinde H_0 hipotezinin δ=0.1 duyarlılıkla ret edilebilmesi için en az 151 bireylik örnek seçilmelidir.

7.2.5. Oranlar arası Farka ait Hipotez Kontrolü

Binomiyal dağılıma uygun dağılım gösteren bir özelliğin dikkate alındığı çalışmalarda araştırıcı, söz konusu özelliğin istenen halinin gözlenme olasılığı bakımından iki farklı muameleyi, bölgeyi, cinsiyeti, yöntemi vb. karşılaştırmak isteyebilir. Bu durumda oranlar arası farka ait hipotez kontrolü yapılması gerekir.

Oranlar arası farka ait hipotez kontrolünün yapılabilmesi için üzerinde çalışılan örneklerin binomiyal dağılım gösteren populasyondan tamamen tesadüfen alınmış birbirinden bağımsız örnekler olması gerekir.

Yapılan hipotez kontrolü sonucunda verilecek olan kararın güvenilir olması için üzerinde çalışılan örneklerin örnek genişliği, örnekten hesaplanan oranların yaklaşık normal dağılım göstermeleri için yeterli büyüklükte olmalıdır. Bunun için de örnek genişlikleri, $\pi \leq 1/2$ ise $n.\pi \geq 10$ veya $\pi > 1/2$ ise oranlara ait örnekleme dağılımının normal dağılıma yaklaşabilmesi için ise $n.(1-\pi) \geq 10$ şartı sağlanmalıdır.

Gerekli şartlar yerine getirilmiş ise hipotez kontrolü yapılabilir. Hipotez kontrolünde, hipotez takımı aşağıdaki gibi oluşturulmalıdır.

H_0: Örneklerden hesaplanan istenen olayın oluş olasılıkları arasındaki fark tesadüften ileri gelmektedir. Bu fark sıfır kabul edilebilir. Kısaca $\mu_{pA}=\mu_{pB}$ veya $\mu_{pA}-\mu_{pB}=0$ ($\pi_A = \pi_B$ veya $\pi_A - \pi_B = 0$)'dır.

Yapılan çalışmanın amacı doğrultusunda karşıt hipotez 3 farklı şekilde kurulabilir.

1. **H_1:** Örneklerden hesaplanan istenen olayın oluş olasılıkları arasındaki fark tesadüften ileri gelmemektedir. Bu fark sıfır kabul edilemez. Kısaca, $\mu_{pA} \neq \mu_{pB}$ veya $\mu_{pA}-\mu_{pB} \neq 0$ ($\pi_A \neq \pi_B$ veya $\pi_A - \pi_B \neq 0$)'dır.

2. **H_1:** A örneğinden hesaplanan istenen olayın oluş olasılığı, B örneğinden hesaplanan istenen olayın oluş olasılığından daha büyüktür. Kısaca, $\mu_{pA} > \mu_{pB}$ veya $\mu_{pA}-\mu_{pB} > 0$ ($\pi_A > \pi_B$ veya $\pi_A - \pi_B > 0$)'dır.

3. **H_1:** A örneğinden hesaplanan istenen olayın oluş olasılığı, B örneğinden hesaplanan istenen olayın oluş olasılığından daha küçüktür. Kısaca, $\mu_{pA} < \mu_{pB}$ veya $\mu_{pA}-\mu_{pB} < 0$ ($\pi_A < \pi_B$ veya $\pi_A - \pi_B < 0$)'dır.

Örneklerden hesaplanan istenen olayın oluş olasılıkları arasındaki fark tesadüften ileri geliyorsa, yani örnekler π'si bilinen aynı populasyondan alınmış ise $\pi_A = \pi_B = \pi$ olacaktır. Bu durumda populasyona ait π'nin en iyi tahmini (5.12) numaralı eşitlikte verildiği gibi;

$$p = \frac{x_A + x_B}{n_A + n_B} \text{ olacaktır.}$$

Bu durumda, n_A ve n_B genişliğinde örneklerden hesaplanan oranlar arası farkın, n_A ve n_B genişliğindeki örneklerden elde edilecek oranlar arası farka ait örnekleme dağılımına dahil olma olasılığını hesaplamada, (5.13) numaralı eşitlik ve bu oranlar arası farka ait standart sapma da ($\sigma_{(p_A - p_B)}$) (5.14) numaralı eşitlik kullanılarak hesaplanır.

ÖRNEK 1:

Aynı tipten ürün yetiştiriciliği yapan çiftçiler arasından, İç Anadolu bölgesinden (A) tesadüfen seçilen 400 çiftçiden 250'sinin, yaptığı üretimle ilgili kooperatif üyeliği bulunurken, Doğu Anadolu bölgesinden (B) tesadüfen seçilen 300 çiftçiden yaptığı üretimle ilgili 100'ünün kooperatife kayıtlı olduğu saptanmıştır. Bir üretim kooperatifine kayıt olma oranları bakımından bölgeler arasındaki farkın istatistik olarak önemli olduğu söylenebilir mi?

Bu örnekte yapılan çalışmada İç Anadolu ve Doğu Anadolu bölgeleri arasında bir üretim kooperatifine kayıt olma oranları arasındaki farkın önemli olup olmadığı araştırılmak istenmektedir. Hipotez kontrolü için önce İç Anadolu ve Doğu Anadolu bölgelerindeki üretim kooperatifine kayıt olma oranları hesaplanır:

$$p_A = \frac{250}{400} = 0.625 \text{ ve } p_B = \frac{100}{300} \cong 0.333$$

Eğer bölgeler arasında üretim kooperatifine kayıt olma oranları bakımından fark tesadüften ileri geliyorsa iki bölgede, istenen özelliğin olma olasılığı aynı olan (π) populasyonu temsil eder ve populasyona ait üretim kooperatifine kayıt olma oranının (π'nin) en iyi tahmini (5.12) numaralı eşitlikten;

$$p = \frac{250+100}{400+300} \cong 0.5 \text{olarak hesaplanır.}$$

Yapılan hipotez kontrolü sonucunda verilecek olan kararın güvenilir olması için üzerinde çalışılan örneklerin örnek genişliği, örnekten hesaplanan oranların yaklaşık normal dağılım göstermeleri için yeterli büyüklükte olmalıdır. A ve B grupları için $n_A\pi=400(0.5)>10$ ve $n_B\pi=300(0.5)>10$ olduğundan hipotez kontrolüne geçilebilir.

İç Anadolu ve Doğu Anadolu bölgeleri arasında bir üretim kooperatifine kayıt olma oranları arasındaki farkın önemli olup olmadığını araştırmak için uygulanan hipotez kontrolü Tablo 7.7'de verilmiştir.

Tablo 7.7. İç Anadolu ve Doğu Anadolu bölgeleri arasında bir üretim kooperatifine kayıt olma oranları arasındaki farkın önemli olup olmadığını araştırmak için uygulanan hipotez kontrolü

H_0: Bir üretim kooperatifine üye olma oranı bakımından İç Anadolu ve Doğu Anadolu bölgeleri arasında gözlenen fark tesadüften ileri gelmiştir ve sıfır kabul edilebilir. Kısaca, $\pi_A-\pi_B=0$'dır.

H_1: Bir üretim kooperatifine üye olma oranı bakımından İç Anadolu ve Doğu Anadolu bölgeleri arasında gözlenen fark tesadüften ileri gelmemiştir ve sıfır kabul edilemez. Kısaca, $\pi_A-\pi_B\neq0$'dır.

İç Anadolu ve Doğu Anadolu bölgelerindeki kayıt olma oranları:

$$p_A = \frac{250}{400} = 0.625 \text{ ve } p_B = \frac{100}{300} \cong 0.333$$

Kayıt olma oranının, π'nin, en iyi tahmini (5.12) numaralı eşitlikten;

$$p = \frac{250+100}{400+300} \cong 0.5\text{'dir.}$$

Oranlar arasındaki farkın standart sapması (5.14) sayılı formülden;

$$\sigma_{(p_A-p_B)} = \sqrt{0.5(1-0.5)(\frac{1}{400}+\frac{1}{300})} = 0.0382\text{'dir.}$$

(5.13) numaralı eşitlik kullanılarak oranlar arası farka karşılık gelen Z-değeri (test istatistiği);

$$Z = \frac{(0.625-0.333)}{0.0382} = 7.644 \text{ olarak hesaplanır.}$$

Eğer I. tip hata olasılığı %1 olarak belirlenmiş ise, Z-dağılımında kontrol hipotezinin ret bölgesi, yapılan kontrol çift taraflı olduğu için, iki eşit parçaya ayrılmış olup her bir % 0.5'lik alan ±2.575 değerinden başlamaktadır. Hesaplanan test istatistiğinin değeri 7.644 olup 2.575 değerinden daha büyüktür ve kontrol hipotezinin ret bölgesine düşmektedir. Dolayısıyla, kontrol hipotezi reddedilerek İç Anadolu ve

Doğu Anadolu bölgeleri arasında bir üretim kooperatifine kayıt olma oranları arasındaki farkın istatistik olarak önemli olduğuna karar verilir.

ÖRNEK 2:

Bir doktor, bir hastalığın tedavisinde kullanılan iki farklı ilacın hastalığın iyileştirme oranları bakımından farklı olup olmadığını araştırmak istemektedir. Bu nedenle aynı hastalığa yakalanmış 260 hastanın 134'ünü A ilacı ile tedavi ederek 128 hastanın iyileştiğini ve geri kalan hastaları B ilacı ile tedavi ederek 120 hastanın iyileştiğini tespit etmiştir. Bu iki ilacın hastalığı iyileştirme oranları arasındaki farkın tesadüften ileri geldiği söylenebilir mi?

A ve B ilaçları ile tedavi edilen hastalarda iyileştirme oranları $p_A = \frac{128}{134} \cong 0.955$ ve $p_B = \frac{120}{126} \cong 0.952$olarak bulunur. Eğer ilaçlar arasında hastalığı iyileştirme oranı bakımından gözlenen fark tesadüften ileri geliyorsa, iki ilaç aynı populasyonu temsil eder ve populasyona ait π'nin, en iyi tahmini (5.12) numaralı eşitlikten; $p = \frac{128+120}{134+126} \cong 0.954$ olarak hesaplanır.

A ve B ilaçlarının söz konusu hastalığı iyileştirme oranları arasındaki farkın tesadüfi olup olmadığına karar vermede, hipotez kontrolünden elde edilen sonuçların güvenilir olmasında, söz konusu örnek oranlarının yaklaşık normal dağılım göstermeleri için örnek genişliğinin yeterli büyüklükte olup olmadığının kontrol edilmesi gerekir. A ve B ilaçları ile tedavi edilen gruplar için $n_A(1-\pi) = 134(1-0.954) = 6.164$ ve $n_B(1-\pi) = 126(1-0.954) = 5.796$ olarak bulunur. Her iki gruptaki gözlem sayısı da $n.(1-\pi) \geq 10$ şartını sağlamadığı için A ve B ilaçlarının söz konusu hastalığı iyileştirme oranları arasındaki farkın tesadüfi olup olmadığına ilişkin parametrik yöntemlerle hipotez kontrolü yapılamaz. Çünkü bu durumda elde edilecek sonuçlar güvenilir olmayacaktır.

ÖRNEK 3:

A bölgesinden tesadüfen seçilen 250 kişiden 200'nün yerel gazeteleri okuduğu belirlenmiştir. B bölgesinde ise 400 kişiden 300 kişinin yerel gazeteleri okuma eğiliminde olduğu saptanmıştır. A bölgesinde yerel gazeteleri okuma eğiliminin daha yüksek olduğu söylenebilir mi?

Yapılan çalışma A bölgesinde yerel gazete okuma eğiliminin B bölgesinden yüksek olup olmadığının araştırılmasıdır. A ve B bölgelerinde yerel gazete okuma oranları $p_A = \frac{200}{250} \cong 0.8$ ve $p_B = \frac{300}{400} \cong 0.75$dır. Eğer bölgeler arasında yerel gazete okuma oranı bakımından gözlenen fark tesadüften ileri geliyorsa, iki bölge yerel gazete okuma bakımından aynı populasyona aittir. Bu populasyona ait π'nin, en iyi tahmini (5.12) numaralı eşitlikten; $p = \frac{200+300}{250+400} \cong 0.769$ olarak hesaplanır.

Yapılan hipotez kontrolü sonucunda verilecek olan kararın güvenilir olması için üzerinde çalışılan örneklerin örnek genişliği, örnekten hesaplanan oranların yaklaşık normal dağılım göstermeleri için yeterli büyüklükte olmalıdır. A ve B grupları için $n_A\pi$=250(1-0.769)>10 ve $n_B\pi$=400(1-0.769)>10 olduğundan hipotez kontrolüne geçilebilir.

A bölgesinde yerel gazeteleri okuma eğiliminin B bölgesinden daha yüksek olup olmadığını araştırmak için uygulanan hipotez kontrolü Tablo 7.8'de verilmiştir.

Tablo 7.8. A bölgesinde yerel gazeteleri okuma eğiliminin B bölgesinden daha yüksek olup olmadığını araştırmak için uygulanan hipotez kontrolü

H_0: Bölgeler arasında yerel gazete okuma eğilimi bakımından fark yoktur. Kısaca, $\mu_{pA}-\mu_{pB}=0$ veya $\pi_A-\pi_B=0$'dır.

H_1: A bölgesinde yerel gazete okuma eğilimi B bölgesine nazaran daha yüksektir. Kısaca, ($\mu_{pA}>\mu_{pB}$) veya ($\pi_A>\pi_B$)'dır.

A ve B bölgelerinde yerel gazete okuma oranları;

$$p_A = \frac{200}{250} \cong 0.8 \text{ ve } p_B = \frac{300}{400} \cong 0.75$$

Yerel gazete okuma oranının, π'nin, en iyi tahmini (5.12) numaralı eşitlikten;

$$p = \frac{200+300}{250+400} \cong 0.769\text{'dir.}$$

Tablo 7.8 devam

Oranlar arasındaki farkın standart hatası (5.14) sayılı formülden;

$$\sigma_{(p_A-p_B)} = \sqrt{0.769(1-0.769)(\frac{1}{250}+\frac{1}{400})} = 0.034\text{'dir.}$$

(5.13) numaralı eşitlik kullanılarak Z-değeri (test istatistiği);

$$Z = \frac{(0.8-0.75)}{0.034} = 1.471 \text{ olarak hesaplanır.}$$

Eğer I. tip hata olasılığı %5 olarak belirlenmiş ise yapılan hipotez kontrolü tek taraflı olduğu için Z-dağılımında kontrol hipotezinin ret bölgesi olan %5'lik alan ±1.645 değerinden başlamaktadır. Hesaplanan test istatistiğinin değeri 1.471 olup 1.645 değerinden daha küçüktür. Dolayısıyla kontrol hipotezi ret edilemez. Yani, A bölgesinde yerel gazete okuma eğilimi B bölgesine nazaran daha yüksek değildir. Bölgeler arasında yerel gazete okuma eğilimi bakımından fark yoktur.

7.2.6. Oranlar Arası Farka ait Hipotez Kontrolünde Örnek Genişliği

İki grup arasında binomiyal dağılım gösteren bir özellik bakımından önemli bir fark olup olmadığı araştırılırken, grupların temsil ettiği populasyonlara ait oranlar arasındaki gerçek farkı ortaya çıkarmak için gerekli olan örnek genişliğini, belirlenen I. tip hata ve II. tip hata olasılığı ile tahmin etmede çeşitli yöntemler geliştirilmiştir. İki gruptaki örnek genişlikleri eşit ise gerekli olan en az örnek genişliği (7.2) numaralı eşitlik kullanılarak tahmin edilebilir.

$$n = \frac{\left[Z_\alpha \sqrt{2\bar{p}\,\bar{q}} + Z_\beta \sqrt{p_1 q_1 + p_2 q_2}\right]^2}{\delta^2} \qquad \ldots(7.2)$$

Eşitlikte $\bar{p} = \dfrac{p_1 + p_2}{2}$ ve $\bar{q} = 1 - \bar{p}$, $\delta = p_1 - p_2$, Z_α, belirlenen I. tip hata olasılığındaki Z-değeridir. Hipotez kontrolünün tek taraflı veya çift taraflı oluşuna bağlı olarak değişen Z-değeri kullanılır. Z_β, ise II. tip hata olasılığında tek taraflı Z-değeridir.

İki grubun örnek genişliklerinin eşit olması testin güvenirliliği ve belirlenen gücün gerçekleşmesi bakımından istenen durum olmasına karşın pratikte her zaman gruplardaki gözlem sayıları eşit olmayabilir. İki gruptaki gözlem sayıları arasında $r = n_2 / n_1$ (n_1 küçük örnek genişliğini belirtmektedir.) şeklinde bir oran belirlenir ise, birinci örneğin genişliği (7.3) numaralı eşitlik kullanılarak tahmin edilir.

$$n_1 = \frac{n}{4}\left[1 + \sqrt{1 + \frac{2(r+1)}{rn\delta}}\right]^2 \qquad \ldots(7.3)$$

(7.3) numaralı eşitlikte n (7.4) numaralı eşitlik kullanılarak hesaplanır.

$$n = \frac{\left[Z_\alpha \sqrt{(r+1)\bar{p}'\bar{q}'} + Z_\beta \sqrt{rp_1 q_1 + p_2 q_2}\right]^2}{r\delta^2} \qquad \ldots(7.4)$$

(7.4) numaralı eşitlikte $\bar{p}' = \dfrac{p_1 + rp_2}{r+1}$ ve $\bar{q}' = 1 - \bar{p}'$ şeklinde hesaplanır. Gerekli hesaplamalar yapıldıktan sonra ikinci grubun örnek genişliği (7.5) numaralı eşitlikten bulunur (Zar, 1999).

$$n_2 = r\,(n_1) \qquad \ldots(7.5)$$

ÖRNEK:

A bölgesinden alınan bir örnekte yerel gazeteleri okuma oranı 0.8 ve B bölgesinden alınan bir örnekte ise bu oran 0.65 olarak saptanmıştır. H_0: $p_A - p_B = 0$ hipotezini H_1: $p_B > p_A$ hipotezine karşı $\alpha = 0.01$ seviyesinde %90 güçle test etmek için örnek genişliği ne olmalıdır?

A ve B bölgelerinden eşit ve farklı sayıda bireyin seçilmesi durumunda örnek genişliklerinin tahmini Tablo 7.9'da verilmiştir.

Tablo 7.9. A ve B bölgesinden alınacak örnek genişliklerinin tahmini

a.Eşit örnek genişlikleri

$\delta = p_1 - p_2 = 0.80 - 0.65 = 0.15$

$$\bar{p} = \frac{p_1 + p_2}{2} = \frac{0.80 + 0.65}{2} = 0.725 \text{ ve } \bar{q} = 1 - 0.725 = 0.275$$

Hipotez kontrolü tek taraflı yapılacağı için α=0.01'de Z_α=2.33 dür.
Testin gücü %90 olarak belirlendiği için β=1-0.90=0.10'dur ve Z_β=1.28 olarak belirlendikten sonra (7.2) numaralı eşitlik kullanılarak örnek genişliği;

Tablo 7.9 devam.

$$n = \frac{\left[2.33\sqrt{2(0.725)(0.275)} + 1.28\sqrt{(0.80)(0.20) + (0.65)(0.35)}\right]^2}{(0.15)^2} = 228.6 \cong 229$$

olarak bulunur. A ve B bölgelerinden söz konusu kontrol hipotezini % 90 güçle test etmek için en az 229'ar birey seçilmelidir.

b. Farklı örnek genişlikleri

$\delta = p_1 - p_2 = 0.80 - 0.65 = 0.15$

$r = n_2 / n_1 = 2$

$$\bar{p}' = \frac{p_1 + rp_2}{r+1} = \frac{0.80 + 2x0.65}{2+1} = 0.7 \text{ ve } \bar{q}' = 1 - 0.7 = 0.3$$

Hipotez kontrolü tek taraflı yapılacağı için α=0.01'de Z_α=2.33 olup
Testin gücü %90 olarak belirlendiğinden β=1-0.90=0.10'dur ve Z_β=1.28 olarak bulunduktan sonra (7.3) numaralı eşitlik kullanılarak n_1'in hesaplanabilmesi için (7.4) numaralı eşitlik kullanılarak n'in tahmin edilmesi gerekir.

$$n = \frac{\left[2.33\sqrt{(2+1)(0.7)(0.3)} + 1.28\sqrt{2(0.8)(0.2) + (0.65)(0.35)}\right]^2}{2(0.15)^2} = 86.9 \cong 87$$

(7.3) numaralı eşitlikten n_1;

$$n_1 = \frac{87}{4}\left[1 + \sqrt{1 + \frac{2(2+1)}{(2)(87)(0.15)}}\right]^2 = 96.74 \cong 97$$

ve (7.5) numaralı eşitlikten;

n_2=2(97)=194 olarak bulunur. Söz konusu kontrol hipotezini, % 90 güçle test etmek için A bölgesinden en az 97 ve B bölgesinden en az 194 bireylik örnekler seçilmelidir.

7.2.7. Korelasyon Katsayısına ait Hipotez Kontrolü

Çalışmaların çoğu birden fazla özelliğe ait veri toplamaya yöneliktir. Sadece bir özelliğe ilişkin verinin toplandığı çalışma hemen hemen hiç yoktur veya çok nadirdir. Birden fazla özellik arasındaki ilişkiler daha önceden belirtildiği üzere korelasyon katsayısı veya regresyon katsayısı ile belirlenebilmektedir. Her hangi bir örnekten hesaplanmış X ve Y gibi iki özellik arasındaki ilişkinin doğrusallık derecesini belirten korelasyon katsayısının (r), bu örneğin alınmış olduğu populasyonda söz konusu özellikler arasındaki korelasyon katsayısını (ρ) ne ölçüde temsil ettiği hipotez testi ile belirlenebilir. Bunun için populasyona ait korelasyon katsayısı (ρ) ile örnekten hesaplanan korelasyon katsayısı (r) arasındaki farkın tesadüften ileri gelip gelmediğinin kontrol edilmesi, yani korelasyon katsayına (r) ait hipotez kontrolünün yapılması gerekir.

Hipotez kontrolü, aşağıdaki gibi hipotez takımının oluşturulmasıyla başlar.

H_0: Populasyon için bildirilen korelasyon katsayısı ile örnekten hesaplanan korelasyon katsayısı arasındaki fark tesadüften ileri gelmektedir. Bu fark sıfır olarak kabul edilebilir. Yani, $\mu_r=\rho$'dur.

Yapılan çalışmanın amacı doğrultusunda karşıt hipotez 3 farklı şekilde kurulabilir.

1. **H_1:** Populasyon ve örnekten hesaplanan korelasyon katsayıları arasındaki fark tesadüften ileri gelmemektedir. Bu fark sıfır kabul edilemez. Yani, $\mu_r\neq\rho$'dur.
2. **H_1:** Örnekten hesaplanan korelasyon katsayısı populasyona ait korelasyon katsayısından daha büyüktür. Yani, $\mu_r>\rho$'dur.
3. **H_1:** Örnekten hesaplanan korelasyon katsayısı populasyona ait korelasyon katsayısından daha küçüktür. Yani, $\mu_r<\rho$'dur.

Korelasyon katsayına ait hipotez kontrolünün yapılabilmesi için örnekten hesaplanan korelasyon katsayısının, korelasyon katsayısına ait örnekleme dağılımına dahil olma olasılığının hesaplanması gerekir.

Örneklerin alındığı populasyona ait korelasyon katsayısı, $\rho\neq0$ olduğu zaman bu populasyondan elde edilecek korelasyon katsayısına ait örnekleme dağılımının normal dağılım göstermesi için her bir örnekten hesaplanan korelasyon katsayısının (5.15) numaralı eşitlik kullanılarak Z_r-değerine dönüştürülmesi gerekir. Üzerinde çalışılan örnekten hesaplanan korelasyon katsayısına karşılık gelen Z_r-değerleri Tablo B'de verilmiştir. Korelasyon katsayısı Z_r-değerine dönüştürüldükten sonra, korelasyon katsayısına ait örnekleme dağılımına dahil olma olasılığının bulunması için (5.20) numaralı eşitlik kullanılarak Z-değeri hesaplanır.

ÖRNEK 1:

Yaş ile kan basıncı arasındaki korelasyon katsayısının 0.80 olduğu ileri sürülmüştür. Bunun doğruluğunu kontrol etmek amacıyla tamamen tesadüfen seçilen 125 bireyde adı geçen özellikler bakımından korelasyon katsayısı 0.86 olarak hesaplanmıştır. Örnekten hesaplanan korelasyon katsayısı ile söz konusu populasyona ait korelasyon katsayısı arasındaki farkın tesadüften ileri geldiği söylenebilir mi?

Bu çalışmada, yaş ile kan basıncı arasındaki korelasyon katsayısının bildirildiği gibi 0.80 olup olmadığının araştırılması amaçlanmıştır. Dolayısıyla korelasyon katsayısına ait hipotez kontrolü çift taraflı olarak yapılmalıdır. Yaş ile kan basıncı arasındaki korelasyon katsayısının bildirildiği gibi 0.80 olup olmadığının araştırılması için uygulanan hipotez kontrolü Tablo 7.10'da verilmiştir.

Tablo 7.10. Yaş ile kan basıncı arasındaki korelasyon katsayısının bildirildiği gibi 0.80 olup olmadığının araştırılmasına ilişkin hipotez kontrolü

H_0: Yaş ile kan basıncı arasındaki korelasyon katsayısının 0.80 olduğu söylenebilir. Kısaca, μ_r =0.80 veya ρ=0.80'dir.

H_1: Yaş ile kan basıncı arasındaki korelasyon katsayısının 0.80 olduğu söylenemez. Kısaca, $\mu_r \neq$0.80 veya $\rho \neq$0.80'dir.

Örnekten hesaplanan korelasyon katsayısına karşılık gelen Z_r değeri (5.17) numaralı eşitlikten;

$Z_r = \frac{1}{2}\log_e \frac{(1+0.86)}{(1-0.86)} = 1.2933$ olarak bulunur.

Z_r değerlerinin ortalaması (5.18) numaralı eşitlikten;

$\mu_{Z_r} = \frac{1}{2}\log_e \frac{(1+0.8)}{(1-0.8)} = 1.0986$ ve standart sapması da (5.19) numaralı eşitlikten $\sigma_{Z_r} = \sqrt{\frac{1}{(125-3)}} \cong 0.091$ olarak bulunur.

(5.20) numaralı eşitlikten örnekten hesaplanan korelasyon katsayısına karşılık gelen Z-değeri, yani test istatistiği;

$Z = \frac{1.2933-1.0986}{0.091} \cong 2.14$olarak hesaplanır.

Tablo 7.10. devam

Eğer I. tip hata olasılığı %5 olarak belirlenmiş ise yapılan hipotez kontrolü çift taraflı hipotez kontrolü olduğu için Z-dağılımında kontrol hipotezinin ret bölgesi, iki eşit parçaya ayrılmış olup, her bir % 2.5'luk alan ±1.96 değerinden başlamaktadır. Hesaplanan test istatistiğinin değeri 2.14 olup 1.96 değerinden daha büyüktür. Dolayısıyla, kontrol hipotezi ret edilerek yaş ile kan basıncı arasındaki korelasyon katsayısının bildirildiği gibi 0.80 olmadığı kararına varılır.

ÖRNEK 2:

Çavuş üzümü çeşidinde tane ağırlığı ile tanenin içerdiği früktoz miktarı arasında %70'lik doğrusal bir ilişkinin var olduğu bilinmektedir. Tesadüfen alınan 52 adet üzüm örneğinde adı geçen özellikler arasındaki doğrusal ilişki (korelasyon katsayısı) r=0.60 olarak bulunmuştur. Acaba bu iki özellik arasındaki korelasyon katsayısının hesaplanmış olduğu örnek, çavuş üzümü çeşidinden midir?

Bu çalışmanın amacı tesadüfen alınan 52 adet üzüm örneğinin tane ağırlığı ile tanenin içerdiği früktoz miktarı arasındaki korelasyon katsayısı bakımından Çavuş üzüm çeşidine ait olup olmadığının araştırılmasıdır. Bu amaçla uygulanan hipotez kontrolü Tablo 7.11'de verilmiştir.

Tablo 7.11. 52 adet üzüm örneğinin tane ağırlığı ile tanenin içerdiği früktoz miktarı arasındaki korelasyon katsayısı bakımından Çavuş üzüm çeşidine ait olup olmadığının araştırılmasına ilişkin hipotez kontrolü

H_0: Çalışılan üzüm örneğinin Çavuş üzüm çeşidine ait olduğu, yani tane ağırlığı ile früktoz miktarı arasındaki korelasyon katsayısının 0.70 olduğu söylenebilir. Kısaca, μ_r =0.70 veya ρ=0.70'dir.

H_1: Çalışılan üzüm örneğinin Çavuş üzüm çeşidine ait olduğu, yani tane ağırlığı ile früktoz miktarı arasındaki korelasyon katsayısının 0.70 olduğu söylenemez. Kısaca, $\mu_r \neq$0.70 veya ρ≠0.70'dir.

Örnekten hesaplanan korelasyon katsayısına karşılık gelen Zr-değeri (5.17) numaralı eşitlikten;

Tablo 7.11. devam

$$Z_r = \frac{1}{2}\log_e\frac{(1+0.6)}{(1-0.6)} = 0.6931 \text{olarak bulunur.}$$

Z_r-değerlerinin ortalaması (5.18) numaralı eşitlikten;

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

ve standart sapması (5.19) numaralı eşitlikten;

$$\sigma_{Z_r} = \sqrt{\frac{1}{(52-3)}} \cong 0.1429 \quad \text{olarak bulunur.}$$

(5.20) numaralı eşitlik kullanılarak, örnekten hesaplanan korelasyon katsayısına karşılık gelen Z-değeri, yani test istatistiği;

$$Z = \frac{0.6931-0.8673}{0.1429} \cong \text{-}1.219 \text{olarak hesaplanır.}$$

Eğer I. tip hata olasılığı %5 olarak belirlenmiş ise yapılan hipotez kontrolü çift taraflı hipotez kontrolü olduğundan Z-dağılımında kontrol hipotezinin ret bölgesi, iki eşit parçaya ayrılmış olup her bir % 2.5'luk alan ±1.96 değerinden başlamaktadır. Hesaplanan test istatistiğinin değeri -1.219 olup -1.96 değerinden daha büyüktür. Bu sebeple de kontrol hipotezi reddedilemez. Dolayısıyla tesadüfen alınan 52 adet üzüm örneğinin tane ağırlığı ile tanenin içerdiği früktoz miktarı arasındaki korelasyon katsayısı bakımından Çavuş üzüm çeşidine ait olduğu kararına varılır.

ÖRNEK 3:

Siyah Alaca ırkı ineklerde laktasyon süt verimi ile sütteki yağ miktarı arasındaki korelasyon katsayısının r= -0.70 olduğu bildirilmiştir. Tesadüfen seçilen 28 Siyah Alaca ırkı inekte söz konusu özellikler arasındaki korelasyon katsayısı r= -0.45 olarak bulunmuştur. Siyah Alaca ırkı ineklerde süt verimi ile sütteki yüzde yağ miktarı arasındaki ilişkinin derecesinin bildirilenden daha düşük olduğu söylenebilir mi?

Çalışmada, tesadüfen seçilen 28 adet Siyah Alaca ırkı inekte laktasyon süt verimi ile sütteki yağ miktarı arasında hesaplanan korelasyon katsayısı doğrultusunda, Siyah Alaca ırkı ineklerde söz konusu özellikler arasındaki ilişkinin bildirilenden daha düşük olup olmadığı araştırılmaktadır. Bu amaçla uygulanan hipotez kontrolü ve yapılan kontrol sonunda verilen karar Tablo 7.12'de verilmiştir.

Tablo 7.12. Siyah Alaca ırkı ineklerde süt verimi ile sütteki yüzde yağ miktarı arasındaki ilişkinin derecesinin bildirilenden daha düşük olup olmadığına ilişkin hipotez kontrolü

H_0: Siyah alaca ırkı ineklerde laktasyon süt verimi ile sütteki yüzde yağ miktarı arasındaki korelasyon katsayısı ile örnekten hesaplanan korelasyon katsayısı arasındaki fark tesadüften ileri gelmiştir. Siyah alaca ırkı ineklerde laktasyon süt verimi ile sütteki yüzde yağ miktarı arasındaki korelasyon katsayısının -0.70 olduğu söylenebilir. Kısaca, μ_r = -0.70 veya ρ = -0.70'dir.

H_1: Siyah alaca ırkı ineklerde laktasyon süt verimi ile sütteki yüzde yağ miktarı arasındaki korelasyon katsayısı ile örnekten hesaplanan korelasyon katsayısı arasındaki fark tesadüften ileri gelmemiştir. Siyah alaca ırkı ineklerde laktasyon süt verimi ile sütteki yüzde yağ miktarı arasındaki ilişki bildirilenden daha düşüktür. Yani, örnekten hesaplanan korelasyon katsayısı, ρ = -0.70'den daha düşük bir populasyona aittir. Kısaca, $\mu_r < -0.70$ veya $\rho < -0.70$'dir.

Örnekten hesaplanan korelasyon katsayısına karşılık gelen Z_r değeri (5.17) numaralı eşitlikten;

$$Z_r = \frac{1}{2}\log_e\frac{(1+0.45)}{(1-0.45)} = 0.4847$$ olarak bulunur.

Z_r-değerlerinin ortalaması (5.18) numaralı eşitlikten;

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

ve standart sapması (5.19) numaralı eşitlikten;

$$\sigma_{Z_r} = \sqrt{\frac{1}{(28-3)}} \cong 0.2 \text{ dir.}$$

(5.20) numaralı eşitlik kullanılarak, örnekten hesaplanan korelasyon katsayısına karşılık gelen Z-değeri, yani test istatistiği;

Tablo 7.12 devam.

$$Z = \frac{0.4847 - 0.8673}{0.2} \cong -1.913 \text{ olarak hesaplanır.}$$

Eğer I. tip hata olasılığı %5 olarak belirlenmiş ise yapılan hipotez kontrolü tek taraflı hipotez kontrolü olduğu için Z-dağılımında kontrol hipotezinin ret bölgesi olan %5'lik alan, ±1.645 değerinden başlamaktadır. Hesaplanan test istatistiğinin değeri -1.913 olup -1.645 değerinden daha küçüktür. Dolayısıyla kontrol hipotezi ret edilerek tesadüfen seçilen 28 adet Siyah Alaca ırkı inekten hesaplanan laktasyon süt verimi ile sütteki yağ miktarı arasında hesaplanan korelasyon katsayısına bakılarak Siyah Alaca ırkı ineklerde söz konusu özellikler arasındaki ilişkinin araştırmanın yapıldığı bölge itibariyle bildirilenden daha düşük olduğu kararına varılır.

7.2.8. Korelasyon Katsayısına ait Hipotez Kontrolünde Örnek Genişliği

Yapılan bir çalışmada testin gücü ve I. tip hata olasılığı belirlenmiş ise $H_0{:}\rho=0$ hipotezini gerçekten $\rho\neq0$ olduğu durumda reddetmek için gerekli örnek genişliği (7.6) numaralı eşitlik kullanılarak tahmin edilebilir.

$$n = \left[\frac{Z_\beta + Z_\alpha}{\xi}\right]^2 + 3 \qquad \ldots(7.6)$$

Eşitlikte ξ, belirlenen korelasyon katsayısına karşılık gelen (5.15) numaralı eşitlik kullanılarak veya Tablo B'den bakılarak bulunan Z_r değeridir. Z_α, yapılacak hipotez kontrolünün tek veya çift taraflı oluşuna bağlı olarak kullanılacak Z-dağılımı (Tablo A) değeridir. Z_β ise II. tip hata olasılığındaki tek taraflı Z dağılımı değeridir.

Populasyona ait korelasyon katsayısının sıfırdan farklı olduğu durumlarda da (7.6) numaralı eşitlik kullanılarak örnek genişliği tahmin edilebilir.

ÖRNEK:

Yapılan bir çalışmada, korelasyon katsayısına ait hipotez kontrolü yapıldığı zaman $H_0{:}\rho=0$ kontrol hipotezini $H_1{:}\rho=0.6$ hipotezine karşı $\alpha=0.05$ seviyesinde test etmek ve H_0 hipotezi ret edildiği zaman testin gücünün %90 olması için örnek genişliği en az ne olmalıdır?

Yapılacak hipotez kontrolü çift taraflı olduğu için α=0.05 için çift taraflı Z_α değeri 1.96'dır. Testin gücü %90 olarak belirlendiği için β=0.10'dur. β=0.10 için tek taraflı Z_β değeri 1.28'dir. Hipotezle belirlenen ρ=0.6 değerine karşılık gelen Z_r değeri (5.15) numaralı eşitlikten veya Tablo B'den 0.6931 olarak bulunur. (7.6) numaralı eşitlikten en az örnek genişliği;

$$n = \left[\frac{Z_\beta + Z_\alpha}{\xi}\right]^2 + 3 = \left[\frac{1.96 + 1.28}{0.6931}\right]^2 + 3 = 24.85 \cong 25$$ olarak tahmin edilir.

7.2.9. Korelasyon Katsayıları arasındaki Farka ait Hipotez Kontrolü

Üzerinde çalışılan iki özellik arasındaki korelasyon katsayısı iki ayrı bölgede, çeşitte veya iki farklı muamele uygulanmış gruplarda hesaplanabilir. Bu durumda iki örnekten hesaplanan korelasyon katsayıları arasındaki farkın tesadüfi olup olmadığının araştırılması gerekir. İki gruptan hesaplanan korelasyon katsayıları arasındaki farkın tesadüfi olup olmadığının kontrol edilmesi için korelasyon katsayıları arasındaki farka ait hipotez kontrolünün yapılması gerekir.

Hipotez kontrolü yapılırken ilk olarak hipotez takımı aşağıdaki gibi oluşturulur:

H_0: İki gruptan hesaplanan korelasyon katsayıları arasındaki fark tesadüften ileri gelmiştir ve bu fark sıfır kabul edilebilir. Yani, $\rho_1=\rho_2$'dir.

Yapılan çalışmanın amacı doğrultusunda karşıt hipotez 3 farklı şekilde kurulabilir.

1.H_1: İki gruptan hesaplanan korelasyon katsayıları arasındaki fark tesadüften ileri gelmemektedir ve sıfır kabul edilemez. Yani, $\rho_1\neq\rho_2$'dir.

2. H_1: Birinci grupta hesaplanan korelasyon katsayısı, ikinci gruptan hesaplanan korelasyon katsayısından daha büyüktür, yani birinci grupta söz konusu özellikler arasında daha yüksek bir ilişki vardır. Yani, $\rho_1>\rho_2$'dir.

3. H_1: Birinci grupta hesaplanan korelasyon katsayısı ikinci grupta hesaplanan korelasyon katsayısından daha küçüktür, yani birinci grupta söz konusu özellikler arasında daha düşük bir ilişki vardır. Yani, $\rho_1<\rho_2$'dir.

Test istatistiğinin hesaplanabilmesi için, n_A ve sonra n_B genişliğindeki gruplardan hesaplanan korelasyon katsayılarına karşılık gelen Z_r-değerleri hesaplanarak veya Tablo B'den söz konusu korelasyon katsayılarına karşılık gelen Z_r değerleri bulunarak korelasyon katsayıları arasındaki farka ait örnekleme dağılımının şeklinin normal dağılıma yaklaşması sağlanır. Daha sonra iki gruptan hesaplanan korelasyon katsayıları arasında gözlenen fark kadar ve daha fazla sapanların, korelasyon katsayıları arasındaki farka ait örnekleme dağılımına dahil olma olasılığı (5.20) numaralı eşitlik kullanılarak hesaplanarak, kontrol hipotezi kabul veya ret edilir.

ÖRNEK:

Amasya ve Starking elma çeşitleri arasında meyve ağırlığı ile meyvedeki şeker miktarı arasındaki doğrusal ilişki bakımından bir fark olup olmadığını araştırmak amacıyla; Amasya elma çeşidinden tesadüfen 84 adet elma alınmış ve korelasyon katsayısı r_A=0.40 olarak bulunmuştur. Tesadüfen alınan 124 adet Starking elma çeşidinde ise aynı özellikler arasındaki korelasyon katsayısı r_S=0.55 olarak bulunmuştur. Adı geçen özellikler arasındaki doğrusal ilişkinin, Starking elma çeşidinde daha yüksek olduğunu söyleyebilir misiniz?

Yürütülen çalışmanın amacı, Starking elma çeşidinde meyve ağırlığı ile meyvedeki şeker miktarı arasındaki doğrusal ilişkinin, Amasya elma çeşidinde söz konusu özellikler arasındaki ilişkiden daha yüksek olup olmadığının araştırılmasıdır. Bu amaçla uygulanan hipotez kontrolü ve varılan karar Tablo 7.14'de verilmiştir.

Tablo 7.14. Starking elma çeşidinde meyve ağırlığı ile meyvedeki şeker miktarı arasındaki doğrusal ilişkinin, Amasya elma çeşidinde söz konusu özellikler arasındaki ilişkiden daha yüksek olup olmadığını araştırmak üzere uygulanan hipotez kontrolü

H_0: Meyve ağırlığı ile meyvedeki şeker miktarı arasındaki doğrusal ilişki bakımından Amasya ve Starking elma çeşitleri arasında gözlenen fark tesadüften ileri gelmiştir. İki çeşit arasında, söz konusu özellikler arasındaki ilişki bakımından fark yoktur. Kısaca, $\mu_{rA}-\mu_{rS}=0$ ($\mu_{rA}=\mu_{rS}$) veya $\rho_A-\rho_S=0$ ($\rho_A=\rho_S$)'dır.

H_1: Meyve ağırlığı ile meyvedeki şeker miktarı arasındaki doğrusal ilişki bakımından Amasya ve Starking elma çeşitleri arasında gözlenen fark tesadüften ileri gelmemiştir. Starking elma çeşidinde söz konusu özellikler arasındaki ilişki daha yüksektir. Kısaca, $\mu_{rS}-\mu_{rA}>0$ ($\mu_{rS}>\mu_{rA}$) veya $\rho_S-\rho_A>0$ ($\rho_S>\rho_A$)'dır.

Örneklerden hesaplanan korelasyon katsayılarına karşılık gelen Z_r değerleri (5.17) numaralı eşitlikten;

$$Z_{rA}=\frac{1}{2}\log_e\frac{(1+0.40)}{(1-0.40)}=0.4236 \text{ ve } Z_{rS}=\frac{1}{2}\log_e\frac{(1+0.55)}{(1-0.55)}=0.6184 \text{ dür.}$$

(5.22) numaralı eşitlik kullanılarak, Amasya ve Starking elma çeşitlerinden hesaplanan korelasyon katsayıları arasındaki farka karşılık gelen Z-değeri, yani test istatistiği;

$$Z=\frac{Z_{rS}-Z_{rA}}{\sqrt{\frac{1}{(n_S-3)}+\frac{1}{(n_A-3)}}}=\frac{0.6184-0.4236}{\sqrt{\frac{1}{(124-3)}+\frac{1}{(84-3)}}}=\frac{0.1948}{0.1436}\cong 1.357$$

olarak hesaplanır.

Eğer I. tip hata olasılığı %1 olarak belirlenmiş ise yapılan hipotez kontrolü tek taraflı hipotez kontrolü olduğu için Z-dağılımında kontrol hipotezini ret bölgesi, %1'lik alan, ±2.33 değerinden başlamaktadır. Hesaplanan test istatistiğinin değeri 1.357 olup 2.33 değerinden daha küçüktür. Dolayısıyla kontrol hipotezi kabul edilerek, Starking elma çeşidinde meyve ağırlığı ile meyvedeki şeker miktarı arasındaki doğrusal ilişkinin, Amasya elma çeşidinde söz konusu özellikler arasındaki ilişkiden daha yüksek olduğu söylenemez.

7.2.10. Korelasyon Katsayıları Arası Farka ait Hipotez Kontrolünde Örnek Genişliği

Testin gücü, I. tip hata olasılığı ve Z_r değerlerine dönüştürülmüş iki korelasyon katsayısı arasındaki fark belirleniyorsa H_0 hipotezini reddetmek için gerekli olan örnek genişliği (7.7) numaralı eşitlik kullanılarak tahmin edilebilir.

$$n = 2\left[\frac{Z_\alpha + Z_\beta}{Z_{r1} - Z_{r2}}\right]^2 + 3 \qquad \ldots(7.7)$$

(7.7) numaralı eşitlikte Z_{r1} - Z_{r2}, belirlenen testin gücü ile bulunmak istenen farktır. Z_α, yapılacak hipotez kontrolünün tek veya çift taraflı oluşuna bağlı olarak kullanılacak Z-dağılımı (Tablo A) değeridir. Z_β ise II. tip hata olasılığındaki tek taraflı Z dağılımı değeridir.

Korelasyon katsayıları arasındaki farka ait hipotez kontrolü $n_1=n_2$ olduğu zaman en güçlüdür. Fakat bazı durumlarda n_1 belirlenmiştir ve değiştirilemez. Bu durumda n_2, (7.8) numaralı eşitlik kullanılarak tahmin edilir.

$$n_2 = \frac{n_1(n+3) - 6n}{2n_1 - n - 3} \qquad (7.8)$$

Eşitlikte n, (7.7) numaralı eşitlik kullanılarak tahmin edilen örnek genişliğidir. (7.8) numaralı eşitlik kullanılarak hesaplanan n_2 sıfırdan küçük çıkarsa n_1 artırılmalı veya testin gücü, önemlilik seviyesi veya ortaya konulmak istenen fark değiştirilmelidir.

ÖRNEK:

Bir fakültede kız ve erkek öğrenciler arasında vize ile final notları arasındaki korelasyon katsayısı bakımından 0.6 kadar farklılıkları, hipotez kontrolünü α=0.05 seviyesinde %90 güç ile ortaya koyabilmek için gerekli olan en az örnek genişliği ne olmalıdır?

Yapılacak hipotez kontrolü çift taraflı kontrol olduğu için α=0.05 için çift taraflı Z_α değeri 1.96'dır. Testin gücü %90 olarak belirlendiği için β=0. 10'dur. β=0.10 için tek taraflı Z_β değeri 1.28'dir.

Kız ve erkek öğrenciler arasında vize ile final notları arasındaki korelasyon katsayısı bakımından 0.6 kadar farklılıkları, hipotez kontrolünü α=0.05 seviyesinde %90 güç ile ortaya koyabilmek için gerekli en az örnek genişliği (7.7) numaralı eşitlik kullanılarak aşağıdaki gibi tahmin edilir.

$$n = 2\left[\frac{Z_{\alpha} + Z_{\beta}}{Z_{r1} - Z_{r2}}\right]^2 + 3 = 2\left[\frac{1.96 + 1.28}{0.6}\right]^2 + 3 = 61.32 \cong 61$$

Her bir grupta en az en az 61 öğrencinin olması gerekir. Bazı durumlarda iki gruptaki gözlem sayılarının eşit olması mümkün olmayabilir. Eğer birinci gruptaki gözlem sayısı 40 olarak belirlenmiş ise ikinci gruptaki gözlem sayısı (7.8) numaralı eşitlikten;

$n_2 = \dfrac{n_1(n+3) - 6n}{2n_1 - n - 3} = \dfrac{40(61+3) - 6(61)}{2(40) - 61 - 3}$ =137.13≅137 olarak tahmin edilir.

7.2.11. Varyasyon katsayısına ait hipotez kontrolü

Genellikle ihtiyaç duyulmamasına rağmen bazı durumlarda üzerinde çalışılan bir örneğin, varyasyon katsayısı $\dfrac{\sigma_x}{\mu_x}$ olan bir populasyonu temsil edip etmediğinin kontrol edilmesi gerekebilir. Bu durumda varyasyon katsayısına ait hipotez kontrolünün uygulanması gerekir. Hipotez kontrolü aşağıdaki şekilde hipotezlerin kurulması ile başlar:

H$_0$: Populasyona ait bildirilen varyasyon katsayısı ile örnekten hesaplanan varyasyon katsayısı arasındaki fark tesadüften ileri gelmektedir. Gözlenen fark sıfır kabul edilebilir, yani örnek söz konusu populasyonu temsil etmektedir.

Çalışmanın amacı doğrultusunda karşıt hipotez 3 farklı şekilde kurulabilir:

1. **H$_1$:** Populasyona ait bildirilen varyasyon katsayısı ile örnekten hesaplanan varyasyon katsayısı arasındaki fark tesadüften ileri gelmemektedir. Gözlenen fark sıfır kabul edilemez, yani örnek söz konusu populasyonu temsil etmemektedir.
2. **H$_1$:** Populasyona ait bildirilen varyasyon katsayısı ile örnekten hesaplanan varyasyon katsayısı arasındaki fark tesadüften ileri gelmemektedir. Populasyonun varyasyon katsayısı bildirilenden daha küçüktür.
3. **H$_1$:** Populasyona ait bildirilen varyasyon katsayısı ile örnekten hesaplanan varyasyon katsayısı arasındaki fark tesadüften ileri gelmemektedir. Populasyonun varyasyon katsayısı bildirilenden daha büyüktür.

Çift taraflı kontrol için, örneğin alındığı populasyon normal dağılım gösteriyor ve varyasyon katsayısı 0.67'den büyük değilse en az 10 bireylik örneğin söz konusu hipotez kontrolü için yeterli olduğu, tek taraflı hipotez kontrolünde ise eğer populasyona ait varyasyon katsayısı 0.33'den küçük ise örnek genişliğinin 10'dan büyük olmasının yeterli olduğu bildirilmiştir (Miller 1991).

Hipotez kontrolü için test istatistiği (7.9) numaralı eşitlik kullanılarak hesaplanır:

$$Z = \frac{\sqrt{(n-1)}[VK-(\sigma_x/\mu_x)]}{(\sigma_x/\mu_x)\sqrt{0.5+(\sigma_x/\mu_x)^2}} \qquad \ldots(7.9)$$

Eşitlikte, VK, örnekten hesaplanan varyasyon katsayısıdır. Hesaplanan test istatistiği Z-dağılımı değeri ile karşılaştırılarak karar verilir.

ÖRNEK:

Sürme derinliği 20 cm'ye ayarlanan traktörlerin sürme derinliklerinin, ortalaması μ_x = 20 cm, standart sapması da σ_x = 5 cm olan bir normal dağılım gösterdiği bildirilmiştir. Tesadüfen seçilen 30 parsel sürme derinliği 20 cm'ye ayarlanan traktörle sürülerek sürme derinlikleri ortalaması 25 cm ve standart sapması 8 cm olarak hesaplanmıştır. Sürme derinliği bakımından bu örnekte hesaplanan varyasyon katsayısı, traktör için varyasyon katsayısından farklı mıdır?

Sürme derinliği 20 cm'ye ayarlanan traktörlerin sürme derinliklerinin, ortalaması μ_x = 20 cm, standart sapması da σ_x = 5 cm olan normal dağılım gösterdiği bildirilmiştir. Sürme derinlikleri için (2.19) numaralı eşitlik kullanılarak varyasyon katsayısı hesaplandığı zaman sürme derinlikleri arasında (5/20).100=%25 (nispi olarak 0.25) oranında değişim olduğu saptanır.

Tesadüfen seçilen 30 parsel sürüldüğü zaman sürme derinliği ortalaması 25 cm ve standart sapması 8 cm olarak bulunduğu için örnekte sürme derinlikleri arasında gözlenen % değişim (2.19) numaralı eşitlikten (8/25).100=%32 (nispi olarak 0.32) olarak bulunur. Bu durumda araştırıcı sürme derinlikleri için bildirilen değişim ile örnekten hesaplanan değişim arasındaki farkın tesadüften ileri gelip gelmediğini kontrol etmek için hipotez kontrolünü Tablo 7.16'da verildiği gibi uygulamalıdır.

Tablo 7.16. Sürme derinlikleri arasındaki varyasyon katsayısı için uygulanan hipotez kontrolü

H_0: Sürme derinlikleri için bildirilen varyasyon katsayısı ile örnekten hesaplanan varyasyon katsayısı arasındaki fark tesadüften ileri gelmektedir. Gözlenen fark sıfır kabul edilebilir, yani örnek söz konusu populasyonu temsil etmektedir.

H_1: Sürme derinlikleri için bildirilen varyasyon katsayısı ile örnekten hesaplanan varyasyon katsayısı arasındaki fark tesadüften ileri gelmemektedir. Gözlenen fark sıfır kabul edilemez, yani örnek söz konusu populasyonu temsil etmemektedir.

Hipotezler kurulduktan sonra test istatistiği (7.9) numaralı eşitlik kullanılarak aşağıdaki şekilde hesaplanır.

$$Z = \frac{\sqrt{(n-1)}[VK-(\sigma_x/\mu_x)]}{(\sigma_x/\mu_x)\sqrt{0.5+(\sigma_x/\mu_x)^2}} = \frac{\sqrt{(30-1)}[0.32-(5/20)]}{(5/20)\sqrt{0.5+(5/20)^2}} = \frac{0.377}{0.1875} = 2.0101$$

olarak hesaplanır.

Eğer I. tip hata olasılığı %5 olarak belirlenmiş ise yapılan hipotez kontrolü çift taraflı olduğu için Z-dağılımında kontrol hipotezinin ret bölgesi olan her bir %2.5'luk alan, ±1.96 değerinden başlamaktadır. Hesaplanan test istatistiğinin değeri olan 2.0101, 1.96 değerinden daha büyük olduğundan kontrol hipotezi reddedilerek sürme derinlikleri için bildirilen varyasyon katsayısı ile örnekten hesaplanan varyasyon katsayısı arasındaki farkın tesadüften ileri geldiği söylenemez.

7.2.12. Varyasyon katsayıları arası farka ait hipotez kontrolü

Üzerinde çalışılan iki örnekten hesaplanan varyasyon katsayıları, örneklerdeki bireyler arasındaki değişimi % olarak ifade eder. Örnekler arasında varyasyon katsayısı bakımından fark olup olmadığını kontrol etmek için varyasyon katsayılarının hesaplanması gerekir. Varyasyon katsayıları hesaplandıktan sonra, örnekler arasında varyasyon katsayıları bakımından gözlenen farkın istatistik olarak önemli olup olmadığını kontrol etmek için varyasyon katsayıları arasındaki farka ait hipotez kontrolünün yapılması gerekir.

Hipotez kontrolüne hipotezler aşağıdaki şekilde kurularak başlanır.

H_0: Bireyler arasındaki değişim bakımından iki grup arasında gözlenen fark tesadüften ileri gelmektedir ve sıfır kabul edilebilir. Kısaca $VK_1=VK_2$

Çalışmanın amacı doğrultusunda karşıt hipotez 3 farklı şekilde kurulabilir:

1. **H_1:** Bireyler arasındaki değişim bakımından iki grup arasında gözlenen fark tesadüften ileri gelmemektedir ve sıfır kabul edilemez. Kısaca $CV_1 \neq CV_2$
2. **H_1:** Birinci grupta bireyler arasında gözlenen değişim, ikinci grupta gözlenen değişimden daha büyüktür. Kısaca $VK_1>VK_2$

3. **H_1:** Birinci grupta bireyler arasında gözlenen değişim, ikinci grupta gözlenen değişimden daha küçüktür. Kısaca $VK_1<VK_2$

Bu kontrol için farklı metotlar literatürde sunulmuştur. Bunlardan kolay uygulanabilir olanı Miller (1991) tarafından önerilmiştir. Miller (1991) tarafından önerilen yöntemde test istatistiği (7.10) numaralı eşitlikte verildiği şekilde hesaplanır.

$$Z = \frac{VK_1 - VK_2}{\sqrt{\left(\frac{VK_T^2}{(n_1-1)} + \frac{VK_T^2}{(n_2-1)}\right)(0.5+VK_T^2)}} \qquad \ldots(7.10)$$

Eşitlikte VK_1 ve VK_2 örneklerden hesaplanan varyasyon katsayılarıdır. VK_T, toplanmış varyasyon katsayısı olup (7.11) numaralı eşitlikte verildiği gibi hesaplanır ve populasyona ait varyasyon katsayının en iyi tahminidir.

$$VK_T = \frac{(n_1-1)VK_1 + (n_2-1)VK_2}{(n_1-1)+(n_2-1)} \qquad \ldots(7.11)$$

Hesaplanan test istatistiği Tablo A'da verilen standart normal dağılım tablosundan bulunacak tablo değeri ile karşılaştırılarak karar verilir.

ÖRNEK:

Tesadüfen seçilen 30 Siyah Alaca ırkı buzağının doğum ağırlığı ortalaması 45 kg ,doğum ağırlığına ait standart sapma 4 kg , 40 Simmental ırkı buzağının doğum ağırlığı ortalaması 32 kg ve doğum ağırlığına ait standart sapma da 2.5 kg olarak bulunmuştur. İki ırkın, buzağıları arasında gözlenen doğum ağırlığındaki değişimin birbirinden farklı olduğu söylenebilir mi?

Çalışmada araştırıcı, Siyah Alaca ve Simmental ırkları arasında buzağıların doğum ağırlığındaki değişim bakımından farklılığın önemli olup olmadığını araştırmak için varyasyon katsayıları arasındaki farka ilişkin hipotez kontrolünü Tablo 7.17'de verildiği şekilde uygulamalıdır.

Tablo 7.17. Siyah Alaca ve Simmental ırkları arasında doğum ağırlığındaki değişim bakımından farklılığın önemli olup olmadığını araştırmak için uygulanan hipotez kontrolü

H_0: Siyah Alaca ve Simmental ırkı buzağılar arasında doğum ağırlığına ilişkin varyasyon katsayılarının arasında görülen fark tesadüften ileri gelmektedir ve sıfır kabul edilebilir. Kısaca $VK_{SA}=VK_{SI}$

H_1: Siyah Alaca ve Simmental ırkı buzağılar arasında doğum ağırlığına ilişkin varyasyon katsayılarının arasında görülen fark tesadüften ileri gelmemektedir ve sıfır kabul edilemez. Kısaca $VK_{SA} \neq VK_{SI}$

$$n_{SA}=30 \quad \bar{X}_{SA}=45\,kg \quad S_{SA}=4\,kg$$
$$n_{SI}=40 \quad \bar{X}_{SI}=32\,kg \quad S_{SI}=2.5\,kg$$

olarak verildiği için ırklara ait varyasyon katsayıları ve (7.11) numaralı eşitlik kullanılarak toplanmış varyasyon katsayısı;

$$VK_{SA}=4/45=0.089$$
$$VK_{SI}=2.5/32=0.078$$

ve

$$VK_T=\frac{(30-1)0.089+(40-1)0.078}{(30-1)+(40-1)} \cong 0.0827$$

Tablo 7.17 devam.

olarak hesaplandıktan sonra test istatistiği (7.10) numaralı eşitlikten;

$$Z = \frac{(0.089-0.078)}{\sqrt{\left(\frac{(0.0827)^2}{(30-1)} + \frac{(0.0827)^2}{(40-1)}\right)(0.5+0.0827)}} \cong 0.71 \text{ olarak bulunur.}$$

Eğer I. tip hata olasılığı %5 olarak belirlenmiş ise yapılan hipotez kontrolü çift taraflı hipotez kontrolü olduğundan Z-dağılımında kontrol hipotezinin ret bölgesi olan her bir %2.5'luk alan ±1.96 değerinden başlamaktadır. Hesaplanan test istatistiğinin değeri 0.71 olup, 1.96 değerinden daha küçüktür. Dolayısıyla kontrol hipotezi kabul edilerek, Siyah Alaca ve Simmental ırkı buzağıların doğum ağırlığına ilişkin varyasyon katsayılarının arasında görülen farkın tesadüften ileri geldiği sonucuna varılır.

7.3. SORULAR

1. Belirli bir hastalıktan ölüm oranının %28 olduğu bilinmektedir. Yeni uygulanan bir tedavi yöntemi sonucunda 120 hastadan 22'si ölmüştür. Yeni uygulanan tedavinin ölüm oranını azaltıp azaltmadığını kontrol ediniz.
2. Sağlıklı yetişkinlerde HDL-kolesterol değerinin ortalaması 60 mg/dL, standart sapması ise 8 mg/dL olan normal dağılım gösterdiği bilinmektedir. Populasyondan rasgele seçilen 20 kişide HDL-kolesterol ortalaması 55 mg/dL olarak hesaplanmıştır. Bu 20 kişinin sağlıklı bireyler olup olmadığına karar veriniz.
3. Bir vitamin karmasındaki E vitaminine ait standart sapmanın 0.316 olduğu bilinmektedir. İki ayrı firma arasında drajelerdeki E vitamini miktarı bakımından bir farklılık olup olmadığını araştırmak üzere A firmasının farklı üretim partilerinden alınan 10 vitamin drajesinde E vitamini ortalaması 12 mg ve B firmasının farklı üretim partilerinden alınan 12 drajedeki E vitamini ortalaması 13,5 mg olarak bulunmuştur. İki firma arasında, drajelerdeki E vitamini bakımından farkın önemli olup olmadığına karar veriniz?

4. Arpalardaki protein oranının, ortalaması $\mu = 10.0$, standart sapması da $\sigma = 2.0$ olan, normal dağılım gösterdiği bilinmektedir. 36 parsele ekilen arpalara özel bir azotlu gübre uygulaması yapıldıktan sonra, bu parsellerden elde edilen arpalarda protein oranı ortalaması $\bar{x} = 11.5$ olarak bulunmuştur. Özel gübre arpalarda protein oranı ortalamasını artırmış mıdır?
5. Tarımsal üretimde ilaç kullanımına karşı olanların oranının %30 olduğu bilinmektedir. Son yıllardaki organik tarıma yönelmeler dolayısıyla bu oranın daha da arttığı görüşü ileri sürülmektedir. Yapılan bir kamuoyu araştırmasında rasgele seçilen 200 kişinin 80'ninin tarımsal üretimde ilaç kullanımına karşı olduğu belirlenmiştir. Buna göre, ilaç kullanımına karşı olanların sayısı artmış mıdır?
6. Populasyon ortalaması hakkında istatistik olarak bir fikir elde edilmek istendiğinde, Z-testi hangi durumlarda kullanılır? Açıklayınız.

7. Bir fakültede öğrencilerin spor faaliyetlerine katılma oranının %45 olduğu bilinmektedir. Sporun insan sağlığı için önemini anlatan bir kampanyadan sonra bu oranda bir artış olup olmadığını araştırmak üzere tamamen tesadüfen seçilen 90 öğrenciye uygulanan anket sonucunda 52 öğrencinin spor faaliyetlerine katıldığı saptanmıştır. Yürütülen kampanyanın spor faaliyetlerine katılma oranını artırıp artırmadığını kontrol ediniz.
8. 8 haftalık etlik piliçlerin ağırlıklarının, ortalaması μ=1850 g, standart sapması da σ= 250 g olan normal dağılım gösterdikleri bilinmektedir. Bir kümeste, özel bir yem ile beslenen 8 haftalık etlik piliçlerden rasgele alınan 25 tanesinde, ağırlık ortalaması $\overline{X}$= 1900 g olarak bulunmuştur. Özel yemin etlik piliçlerin ağırlık ortalamasını değiştirip değiştirmediğini kontrol ediniz.
9. Bir fakültedeki öğrencilerinin %30' unun sigara içtiği bilinmektedir. Sigaranın zararları konusunda öğrencilere belirli bir süre verilen eğitim çalışmasından sonra tesadüfen seçilen 300 öğrenciden 75 tanesinin sigara içtiği belirlenmiştir. Öğrencilere verilen eğitim çalışması, sigara içenlerin oranını azaltmış mıdır?
10. Aynı populasyondan alındığı ileri sürülen n_1=150 bireyde korelasyon katsayısı r_1=0.920 ve n_2=200 bireyde korelasyon katsayısı r_2=0.994 olarak hesaplanmıştır. Bu iki örneğin aynı populasyondan alınmış örnekler olduğu söylenebilir mi?
11. 28 adet Franka mısır çeşidinde bitki boyu ile bin tane ağırlığı arasındaki korelasyon katsayısı r=0.80 olarak bulunmuştur. 53 adet Ant-90 mısır çeşidinde ise aynı özellikler arasındaki korelasyon katsayısı r=0.50 olarak bulunmuştur. Adı geçen özellikler arasındaki ilişki bakımından, Franka ve Ant–90 mısır çeşitleri birbirinden farklı mıdır?

HİPOTEZ KONTROLLERİ 2: t-KONTROLLERİ

8.1. Giriş

Temel araştırmalarda üzerinde çalışılan populasyonun dağılımı ve parametreleri hakkında bilgi sahibi olunur. Araştırıcı üzerinde çalıştığı populasyonun parametreleri hakkında bilgi sahibi ise bir önceki bölümde açıklandığı gibi test istatistiği olarak Z-değeri ve test dağılımı olarak Z-dağılımı kullanarak gerekli hipotez kontrolü yapar.

Fakat yapılan çalışmaların çoğunda araştırıcı populasyonun parametreleri hakkında bilgi sahibi değildir. Dolayısıyla populasyona ait parametreleri, örnek yada örneklerden tahmin etmek durumundadır. Bu durumda yapılan hipotez kontrollerinde test istatistiği olarak t-değeri ve test dağılımı olarak da t-dağılımı kullanılır.

8.2. t Dağılımı (Student's t-distribution)

t-dağılımı, 1908 yılında Guinness Brewing şirketinde çalışan istatistikçi W. S. Gosset tarafından geliştirilmiştir. Guinness Brewing şirketinde çalışan Gosset, varyansı bilinmeyen populasyondan alınan örnekler için $\frac{\overline{X}-\mu_x}{S_{\overline{x}}}$ değerlerini hesaplamıştır. Örnek ortalaması ile populasyon ortalaması arasındaki farkı, örnekten tahmin edilen standart hataya böldüğü için hesaplanan değerleri **Z** yerine **t** olarak adlandırmış ve t değerlerinin dağılımını geliştirmiştir. Guinness Brewing şirketi çalışılanlarının kendi isimleri ile yayın yapmasına izin vermediğinden Gosset, dağılımını "Student's" takma adı ile yayınlamıştır. Bu sebeple t-dağılımı "Student's t-dağılımı" olarak da bilinir.

t-dağılımı, varyansı bilinmeyen bir populasyondan alınan n örnek genişliğindeki örneklerden her biri için (8.1) numaralı eşitlik kullanılarak hesaplanan t-değerlerinin dağılımıdır.

$$t = \frac{\overline{X} - \mu_x}{S_{\overline{x}}} \qquad \text{… (8.1)}$$

(8.1) numaralı eşitlikte $S_{\overline{x}}$, (5.2) numaralı eşitlik kullanılarak $\frac{S_x}{\sqrt{n}}$ şeklinde hesaplanan ortalamaya ait örnekleme dağılımının, örnekten hesaplanan standart sapmasıdır, kısaca ortalamanın standart hatası olarak ifade edilir. t-dağılımı serbestlik derecesine bağlı bir dağılım olup, t-dağılımının olasılık yoğunluk fonksiyonu (8.2) numaralı eşitlikte verildiği gibidir.

$$f(t) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi v}\ \Gamma(\frac{v}{2})}(1+\frac{t^2}{v})^{-(\frac{v+1}{2})} \qquad \text{… (8.2)}$$

eşitlikte v, serbestlik derecesidir.

-∞ ile +∞ arasında tüm değerleri içeren t-dağılımı simetrik bir dağılım olup ortalaması sıfırdır. Dağılım simetrik olduğu için |-t|=|t| şartı ile –t ile 0 arasında kalan alan ile 0 ile t arasında kalan alan birbirine eşittir.

t-dağılımı serbestlik derecesine bağlı bir dağılım olduğundan her serbestlik derecesi için farklı bir t dağılımı vardır. Serbestlik derecesi arttıkça t-dağılımı Z-dağılımına yaklaşır ve t-dağılımını oluşturan t değerleri arasındaki değişim, dolayısıyla t-dağılımının varyansı azalır. **(n)/(n-2)** şeklinde hesaplanan dağılımın varyansı, teorik olarak serbestlik derecesi sonsuz olduğu zaman Z- ve t-dağılımları üst üste çakışacağından 1'e eşit olur.

t-dağılımı serbestlik derecesine bağlı bir dağılım olduğu için sonsuz tane t-dağılımı vardır. Bu sebeple Tablo C'de görüldüğü gibi farklı serbestlik dereceli t-dağılımlarında belirli yüzde alanların başladığı t-değerleri verilerek t-dağılımı tabloları düzenlenir. Örneğin, Tablo C'den bakıldığı zaman 5 serbestlik dereceli t-dağılımında örnekten hesaplanan t-değerlerinin %2.5'unun 2.571'den büyük olduğu görülür. t-dağılımı simetrik bir dağılım olduğu için örnekten hesaplanan t-değerlerinin %2.5'u da -2.571'den daha küçüktür. Tablo C'den 10 serbestlik dereceli t-dağılımında %2.5'luk alanın ±2.228 değerinden başladığı görülür. 50 serbestlik dereceli t-dağılımında ise %2.5'lik alan 2.008 değerinden başlar. Serbestlik derecesi arttıkça t-dağılımını meydana getiren t-değerleri arasındaki değişim azaldığından, 5, 10 ve 50 serbestlik dereceli t-dağılımlarında %2,5'luk alanının başladığı t-değerleri giderek küçülmektedir.

Serbestlik derecesinin sonsuz olması durumunda %2,5'luk alan 1.96 değerinden %5'lik alan ise 1.645 değerinden başlamaktadır. Bu da serbestlik derecesi arttıkça t-dağılımının Z-dağılımına yaklaştığını ve sonsuz serbestlik derecesinde iki dağılımın çakıştığını göstermektedir.

8.3. Ortalamaya ait Hipotez Kontrolü

Normal dağılım gösteren bir populasyondan tamamen tesadüfen alındığı ileri sürülen bir örnekten hesaplanan ortalama ile populasyonun bilinen ortalaması arasındaki farkın tesadüften ileri gelip gelmediğine ortalamaya ait hipotez kontrolü yapılarak karar verilir. Üzerinde çalışılan populasyonun varyansı bilinmiyorsa test dağılımı olarak t-dağılımı kullanılır ve test istatistiği olarak da t-değeri hesaplanır.

ÖRNEK 1:

Sarıkuyruk balığı yetiştiriciliğinde kullanılan yemlerin yapısındaki protein oranının ortalama %22 olması gerektiği bildirilmiştir. Bu balık çeşidi için yem üreten bir yem fabrikasından tesadüfen seçilen 31 adet yem örneğinde protein oranı ortalamasının %20 ve standart sapmasının da %5 olduğu tespit edilmiştir. Acaba söz konusu fabrikada üretilen yemler Sarıkuyruk balığı yetiştiriciliği için uygun protein oranını içermekte midir?

Yapılan bu çalışmada araştırıcının amacı, söz konusu yem fabrikasında üretilen yemlerin protein içeriklerinin Sarıkuyruk balığı yetiştiriciliği için uygun olup olmadığına karar vermektir. Bu kararın verilebilmesi için hipotez kontrolü yapılması gerekir.

Burada araştırıcının yapması gereken hipotez kontrolü, ortalamaya ait hipotez kontrolüdür. Sarıkuyruk balık yetiştiriciliğinde kullanılan yemlerin yapısındaki ortalama protein miktarının %22 olması gerektiği bildirilmiş, fakat protein oranları arasındaki değişim için varyansın, yani populasyon varyansının ne olduğu verilmemiştir. Populasyon varyansının bilinmediği durumlarda hipotez kontrolünde kullanılması gereken test dağılımı, t-dağılımı ve hesaplanması gereken test istatistiği de t-değeridir.

Söz konusu yem fabrikasında üretilen yemlerin protein içeriklerinin Sarıkuyruk balığı yetiştiriciliği için uygun olup olmadığına karar vermek için uygulanacak hipotez kontrolünde, aşağıdaki gibi hipotez takımının oluşturulması gerekir.

H_0: Populasyon ile örnek ortalaması arasındaki fark tesadüften ileri gelmektedir. Gözlenen %2'lik fark istatistik olarak önemli değildir ve sıfır kabul edilebilir. Yani söz konusu fabrikada üretilen yemlerin Sarıkuyruk balığı yetiştiriciliği için uygun protein oranını içerdiği söylenebilir. Kısaca, μ_x=%22'dir.

H_1: Populasyon ile örnek ortalaması arasındaki fark tesadüften ileri gelmemektedir. Gözlenen %2'lik fark istatistik olarak önemlidir ve sıfır kabul edilemez. Yani söz konusu fabrikada üretilen yemlerin Sarıkuyruk balığı yetiştiriciliği için uygun protein oranını içerdiği söylenemez. Kısaca, $\mu_x \neq$%22'dir.

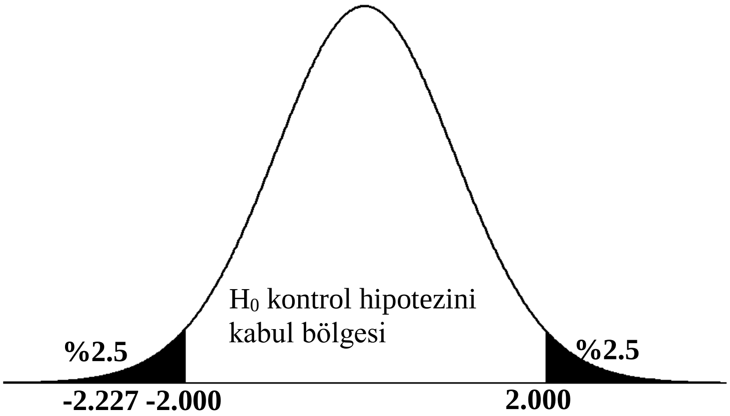
Hipotez kontrolünün çift taraflı yapılması gerekmektedir. Çünkü söz konusu fabrikada üretilen yemlerin Sarıkuyruk balık yetiştiriciliğine uygun olması için üretilen yemin protein oranının, önerilen protein oranından ne az ne de çok olmamalıdır. Yani araştırıcı, 31 yem örneğindeki ortalama protein oranının, populasyona ait ortalamadan küçük veya büyük oluşu ile değil sadece farklı olup olmadığı ile ilgilenmektedir.

Fabrikada üretilen yemlerin protein oranı, ortalaması %22 olan normal dağılım gösteriyor ise bu fabrikadan 31 yem örneği içeren çok sayıda örnekler alınsa ve ortalamaları hesaplansa, hesaplanan bu ortalamalar örnekten örneğe değişerek bir dağılım gösterir. Bu dağılıma, ortalamalara ait örnekleme dağılımı adı verilir. Bu dağılımın parametreleri, $\mu_{\bar{x}} = \mu_x$ = %22 ve standart sapması da (5.2) numaralı eşitlikten $S_{\bar{x}} = \frac{S_x}{\sqrt{n}} = \frac{5}{\sqrt{31}}$ =0.898 olarak örnekten tahmin edilir.

Üretilen yemlerin Sarıkuyruk balık yetiştiriciliği için uygun olup olmadığına karar vermek için yapılacak hipotez kontrolünde test dağılımı olarak t-dağılımı ve test istatistiği olarak da t-değeri kullanılır. Örnekten hesaplanan ortalamaya karşılık gelen t-değeri (8.1) numaralı eşitlikten;

$$t = \frac{\bar{X} - \mu_x}{S_{\bar{x}}} = \frac{(20-22)}{0.898} = -2.227$$ olarak hesaplanır.

Hesaplanan test istatistiği (n-1) = (31-1) = 30 serbestlik dereceli t-dağılımı gösterir. I. tip hata olasılığı, α=%5 olarak belirlenmiş ise çift taraflı hipotez kontrolü yapıldığından I. tip hata olasılığının yarısı (%2.5), ortalamadan büyük değerlerin bulunduğu, diğer yarısı (%2.5) ise ortalamadan küçük değerlerin bulunduğu tarafta alınır. Tablo C'den 30 serbestlik dereceli t-dağılımında %2.5'luk alan 2.00 değerinden başladığından test dağılımında, kontrol hipotezinin kabul ve ret bölgeleri Şekil 8.1'de görüldüğü gibi belirlenir.



Şekil 8.1. 35 serbestlik dereceli t-dağılımında çift taraflı hipotez kontrolünde ret ve kabul bölgeleri

Şekil 8.1'de siyah taralı alanlar H_0 hipotezinin ret bölgeleridir. Yapılan hipotez kontrolünde test istatistiği -2.227 olarak hesaplanmıştır. Hesaplanan bu test istatistiği H_0 hipotezinin ret bölgesine, yani taralı alana düştüğü için kontrol hipotezi ret edilir ve söz konusu fabrikada üretilen yemlerin Sarıkuyruk balık yetiştiriciliği için uygun olmadığı kararına varılır.

ÖRNEK 2:

Meyve suyu üretimi yapan bir fabrikada üretilen portakal sularında C vitamini ortalamasının 17 mg/L olduğu bildirilmiştir. Yeni bir üretim yöntemi kullanılarak üretilen portakal sularından tesadüfen 16 şişe alınmış ve C vitamini ortalaması 17.9 mg/L, standart sapması da 3 mg/L olarak bulunmuştur. Yeni üretim yönteminin portakal sularındaki C-vitamini miktarını artırdığı söylenebilir mi?

Meyve suyu fabrikasında yapılan araştırmanın amacı, önerilen yeni üretim yönteminin kullanılan yöntemden daha iyi olup olmadığının, yani üretilen portakal sularındaki C-vitamini miktarını artırıp artırmadığının araştırılmasıdır. Önerilen yeni yöntemin daha iyi olduğunun söylenebilmesi için yeni yöntem kullanılarak üretilen meyve sularındaki ortalama C-vitamini miktarının, halen kullanılan yöntem için bildirilen ortalamadan daha büyük olması gerekir. Bu sebeple tek taraflı hipotez kontrolü uygulanmalıdır.

Yeni üretim yönteminin daha iyi bir üretim yöntemi olup olmadığına karar vermek için uygulanan hipotez kontrolü Tablo 8.1'de verilmiştir.

Tablo 8.1. Yeni üretim yönteminin daha iyi bir üretim yöntemi olup olmadığına karar vermek için uygulanan hipotez kontrolü

H_0: Populasyon ile örnek ortalaması arasındaki fark tesadüften ileri gelmektedir. Gözlenen 0.9 mg/L'lik C-vitamini farkı istatistik olarak önemli değildir. Yeni üretim yönteminin daha iyi bir yöntem olduğu söylenemez. Kısaca, μ_x=17 mg/L'dir.
H_1: Populasyon ile örnek ortalaması arasındaki fark tesadüften ileri gelmemektedir. Gözlenen 0.9 mg/L'lik C-vitamini farkı istatistik

olarak önemlidir. Yeni üretim yönteminin daha iyi bir yöntem olduğu söylenebilir. Kısaca, $\mu_x > 17$ mg/L'dir.

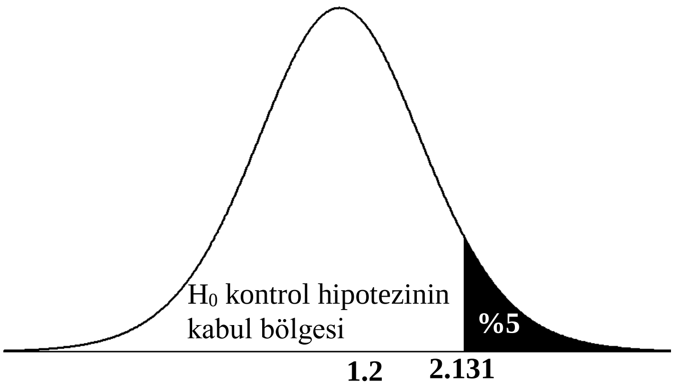
Meyve suyu fabrikasında çok sayıda 16 şişelik portakal suları yeni yöntem ile üretilse ve C-vitamini miktarı ortalamaları hesaplansa hesaplanan bu ortalamalar, $\mu_{\bar{x}} = \mu_x = 17$ mg/L ve standart sapması da (5.2) numaralı eşitlikten $S_{\bar{x}} = \frac{3}{\sqrt{16}} = 0.75$ mg/L olan normal dağılım gösterir.

Hipotez kontrolünde test dağılımı olarak t-dağılımı ve test istatistiği olarak da t-değeri kullanılır. Örnekten hesaplanan ortalamaya karşılık gelen t-değeri (8.1) numaralı eşitlikten;

Tablo 8.1 devam

$$t = \frac{\bar{X} - \mu_x}{S_{\bar{x}}} = \frac{(17.9 - 17)}{0.75} = 1.2 \text{ olarak hesaplanır.}$$

Hesaplanan test istatistiği (16-1)=15 serbestlik dereceli t-dağılımı gösterir. I. tip hata olasılığı, α=%5 olarak belirlenmiş ise tek taraflı hipotez kontrolü yapıldığında I. tip hata olasılığı, test dağılımının sağ tarafında, yani ortalamadan büyük t değerlerinin bulunduğu tarafta alınır. Tablo C'den 15 serbestlik dereceli t-dağılımında %5'lik alan 2.131 değerinden başladığından test dağılımında kontrol hipotezinin kabul ve ret bölgeleri Şekil 8.2'deki gibi belirlenir.



Şekil 8.2. 15 serbestlik dereceli t-dağılımında tek taraflı hipotez kontrolünde kontrol hipotezinin ret ve kabul bölgeleri

Yapılan hipotez kontrolünde test istatistiği 1.2 olarak hesaplandığından bu değer Şekil 8.2'de görüldüğü gibi kontrol hipotezinin kabul bölgesinde kalmaktadır. Bu sebeple H_0 hipotezi kabul edilir.

Meyve suyu fabrikasında yapılan araştırmada, Tablo 8.1'de uygulanan hipotez kontrolü sonunda yeni üretim metodu ile kullanılmakta olan üretim metodu arasındaki farkın tesadüfen ileri geldiğine karar verilmiştir. Bu durumda, yeni üretim yönteminin üretilen portakal sularındaki C-vitamini miktarını artırdığı söylenemez. Hipotez kontrolü sonunda verilen karar doğrultusunda fabrika yetkililerinin halen kullanılmakta olan üretim metodunu kullanmaya devam etmeleri daha akılcı olacaktır.

ÖRNEK 3:

Belirli bir rasyonla beslenen sakız ırkı koyunlarda laktasyon süt verimi ortalamasının 150 kg ($\mu = 150$ kg) olduğu bilinmektedir. Yeni bir rasyon, 30 adet sakız ırkı koyunda denenmiş ve laktasyon süt verimi ortalaması 165 kg, standart sapması da 35 kg olarak bulunmuştur. Kullanılan yeni rasyon, sakız ırkı koyunlarda laktasyon süt verimini artırmış mıdır?

Yeni rasyonun laktasyon süt verimini artırıp artırmadığına karar vermek için tek taraflı hipotez kontrolü Tablo 8.2'de görüldüğü gibi uygulanır.

Tablo 8.2. Yeni rasyonun laktasyon süt verimini artırıp artırmadığına karar vermek için yapılacak hipotez kontrolü

H_0: Populasyon ile örnek ortalaması arasında gözlenen 15 kg'lık fark tesadüften ileri gelmiştir ve istatistik olarak önemli değildir. Bu sebeple, hazırlanan yeni rasyonun sakız ırkı koyunlarda laktasyon süt verimi ortalamasını artırdığı söylenemez. Kısaca, μ_x=150kg'dır.

H_1: Populasyon ile örnek ortalaması arasında gözlenen 15 kg'lık fark tesadüften ileri gelmemiştir. Bu sebeple, hazırlanan yeni rasyonun sakız ırkı koyunlarda laktasyon süt verimi ortalamasını artırdığı söylenebilir. Kısaca, μ_x>150kg'dır.

30 bireylik örneklerden hesaplanacak laktasyon süt verimi ortalamaları, $\mu_{\bar{x}} = \mu_x = 150$ kg ve standart sapması da (5.2) numaralı eşitlikten $S_{\bar{x}} = \frac{35}{\sqrt{30}} \cong 6.39$ kg olan normal dağılım gösterir.

Hipotez kontrolünde test dağılımı olarak t-dağılımı ve test istatistiği olarak da t-değeri kullanılır. Örnekten hesaplanan ortalamaya karşılık gelen t-değeri (8.1) numaralı eşitlikten;

$$t = \frac{\bar{X} - \mu_x}{S_{\bar{x}}} = \frac{(165-150)}{6.39} = 2.347 \text{ olarak hesaplanır.}$$

Tablo 8.2 devam

Hesaplanan test istatistiği (30-1)=29 serbestlik dereceli t-dağılımı gösterir. I. tip hata olasılığı, α=%5 olarak belirlenmiş ise tek taraflı hipotez kontrolü yapıldığı için I. tip hata olasılığı, test dağılımının sağ tarafında, yani ortalamadan büyük t değerlerinin bulunduğu tarafta alınır. Tablo C'den 29 serbestlik dereceli t-dağılımında %5'lik alanın, yani kontrol hipotezinin ret edildiği kritik bölgenin 1.699 değerinden başladığı görülmektedir.

Yapılan hipotez kontrolünde test istatistiği 2.347 olarak hesaplandığından, bu değer kontrol hipotezinin ret edildiği kritik bölgenin başladığı değerden büyüktür. Bu sebeple, H_0 hipotezi ret edilerek, hazırlanan yeni rasyonun sakız ırkı koyunlarda laktasyon süt verimi ortalamasını artırdığı kararına varılır.

Uygulanan hipotez kontrolü sonunda araştırıcı kararını hem α=0.05 hem de α=0.01 seviyesinde verebilir. Hipotez kontrolünde α=0.05 seviyesinde kontrol yapılarak farkın istatistik olarak önemli olduğuna karar verilirse bu önemlilik tek yıldız, "*****" ve I. tip hata olasılığının da %5'ten az olduğu "**p<0.05**" şeklinde belirtilir.

Araştırıcı, %5 seviyesinde kararını verdikten sonra isterse %1 seviyesinde de kontrol yaparak karar verebilir. Tablo 8.2'de uygulanan hipotez kontrolünde α=0.01 seviyesinde karar vermek için 29 serbestlik dereceli t-dağılımında %1'lik alanın başladığı t değeri Tablo C'den 2.462 olarak bulunur. Hesaplanan test istatistiğinin değeri 2.347 , 2.462 değerinden küçük olduğundan H_0 hipotezinin kabul bölgesinde kalır. Hipotez kontrolü %1 seviyesinde yapıldığı zaman kontrol hipotezi kabul edilerek yeni rasyonun sakız ırkı koyunlarda laktasyon süt verimi ortalamasını artırmadığı kararına varılacaktır. Görüldüğü gibi %5 seviyesinde ret edilen kontrol hipotezi %1 seviyesinde kabul edilmiştir.

Yapılan bir hipotez kontrolünde, kontrol hipotezi %1 seviyesinde ret edilirse, farkın istatistik olarak önemli olduğu çift yıldız, "******" ile gösterilir ve I. tip hata olasılığının da %1'den az olduğu "**p<0.01**" şeklinde belirtilir.

8.4. Ortalamaya ait Hipotez Kontrolünde Örnek Genişliği

Eğer 8.1 numaralı bölümde açıklandığı gibi ortalamaya ait t-testi yapılmak isteniyorsa bu durumda araştırıcı, uygulanan hipotez kontrolü sonunda varılacak kararın güvenilirliği için belirlenen bir testin gücü ve fark ile örnek genişliğinin en az ne olması gerektiğini bilmek ister.

Populasyon varyansının (σ^2'nin) tahmini S^2 hesaplanmış ise örnek genişliği (n) tahmin edilebilir. Araştırıcı t-testini önceden belirlenen bir I. tip hata olasılığı (α) ve II. tip hata olasılığı (β) ile uygulamak ve populasyon ortalaması ile örnek ortalaması arasındaki farkın δ olarak belirlenmesi halinde; α yanılma seviyesinde t-testini 1-β güç ile populasyon ortalaması ile örnek ortalaması arasındaki farkı da δ olarak belirlemek için gerekli olan en az örnek genişliği (8.3) numaralı eşitlik kullanılarak hesaplanabilir.

$$n = \frac{S^2}{\delta^2}\left(t_{(\alpha,\ sd)} + t_{(\beta,\ sd)}\right)^2 \qquad ...(8.3)$$

(8.3) numaralı eşitlikte, sd, serbestlik derecesidir ve hesaplanacak olan örnek genişliğine bağlıdır. (8.3) numaralı eşitlik kullanılarak örnek genişliği tahmini yapılırken, örnek varyansı olan S^2, populasyon varyansının örnek genişliği yeteri kadar büyük (30 veya daha fazla birey içeren) bir örnekten hesaplanmış tahmini ise (8.3) numaralı eşitlik kullanılarak yapılacak örnek genişliği tahmini daha güvenilir olur.

Örnek genişliğinin tahmini Bölüm 8.3'de ÖRNEK 1 için Tablo 8.3'de verilmiştir.

Tablo 8.3. Bölüm 8.3'de ÖRNEK 1 için örnek genişliğinin tahmini

Araştırıcı 31 adet yem örneğinde protein oranı ortalamasını $\overline{X} = \%20$ ve standart sapmasını S_x=%5 olarak hesaplamıştır. Bu durumda protein oranına ait varyans, $S_x^2 = 25$'dir. Sarıkuyruk balığı yetiştiriciliği için uygun protein oranı %20 olarak bildirildiğine göre kontrol hipotezini ret etmek için gerekli örnek genişliği tahmin edilmek istenmektedir.

Burada yapılan hipotez kontrolü çift taraflı hipotez kontrolüdür. Eğer hipotez kontrolünün α=%5 seviyesinde çift taraflı olarak ve örnekten hesaplanan ortalama ile populasyon ortalaması arasındaki 3 birimlik farkı %95 olasılık ($1-\beta=0.95$) ile saptaması isteniyorsa örnek genişliği (8.3) numaralı eşitlik kullanılarak aşağıdaki şekilde tahmin edilir:

Tablo 8.3 devam.

Populasyon varyansı n=31 yem örneğinden tahmin edildiği için t-dağılımının sd=(31–1)=30'dur ve Tablo C'den $t_{(0.025,\ 30)}=2.042$ ve $t_{(0.05,\ 30)}=1.697$ olarak bulunur. Bu durumda;

$$n = \frac{25}{3^2}(2.042 + 1.697)^2 = 38.83 \cong 38 \text{ olarak hesaplanır.}$$

Tahmin edilen örnek genişliği kullanılarak örnek genişliği yine tahmin edilirse sd=38-1=37'dir ve $t_{(0.025,\ 37)}=2.026$ ve $t_{(0.05,\ 37)}=1.687$'dir. Bu durumda;

$$n = \frac{25}{3^2}(2.026+1.687)^2 = 38.29 \cong 38 \text{ olarak hesaplanır.}$$

Yapılan örnek genişliği tahminleri örnekten hesaplanan ortalama ile populasyon ortalaması arasındaki 3 birimlik farkı %95 olasılık ($1-\beta=0.95$) ile saptamak için en az 38 yem örneğinin gerektiğini göstermiştir.

8.5. Ortalamalar arası Farka ait Hipotez Kontrolü

8.5.1. Bağımsız İki Grubun Karşılaştırılması

Üzerinde çalışılan özelliğe etki bakımından iki muamele grubu ortalamaları arasındaki farkın tesadüfi olup olmadığının araştırılması gerekebilir. Çalışılan özelliğe etki bakımından iki muamele grubu arasındaki fark tesadüften ileri geliyorsa grupların, ortalaması μ_x olan populasyondan tesadüfen alınmış iki örnek olduğu ve grup ortalamaları arasındaki farkın sıfır kabul edilebileceği söylenir. Örneklerin tesadüfen alındığı populasyonun varyansı hakkında çoğu zaman bilgi yoktur ve üzerinde çalışılan örneklerden tahmin edilir. Varyansı bilinmeyen bir populasyondan iki tesadüf örneği alınarak her birinin varyansları hesaplansa, populasyon varyansının en iyi tahmini bu örnek varyanslarının tartılı ortalaması, yani toplanmış varyans olacaktır. Toplanmış varyansın güvenilir bir tahmin olabilmesi için örnek varyansları arasındaki fark, ancak tesadüften ileri gelecek kadar olmalıdır. Bir başka deyişle örnek varyanslarının homojen olması gerekir. Grup varyanslarının homojenlik kontrolü Bölüm 9'da açıklanacaktır.

Yapılan bir çalışmada dikkate alınan gruplardan populasyon varyansı tahmin edildikten sonra grup ortalamaları arasındaki farkın tesadüfi olup olmadığı, ortalamalar arası farka ait hipotez kontrolü yapılarak kontrol edilebilir. Populasyonun varyansı örneklerden tahmin edildiği için hipotez kontrolünde kullanılacak test dağılımı t-dağılımı ve hesaplanacak test istatistiği de t-değeridir. Grup ortalamaları arasındaki farka ait hipotez kontrolünde t-değeri (8.4) numaralı eşitlik kullanılarak hesaplanır.

$$t=\frac{(\mu_{\overline{A}}-\mu_{\overline{B}})-\mu_D}{S_D} \qquad \ldots(8.4)$$

(8.2) numaralı eşitlikte S_D (5.7) numaralı eşitlik kullanılarak

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

$$n_A=n_B=n \text{ ise } S_D=\sqrt{S_{\overline{A}}^2+S_{\overline{B}}^2}$$

şeklinde hesaplanır. (8.2) numaralı eşitlik kullanılarak hesaplanan t-değeri Serbestlik Derecesi=[(n_A-1)+(n_B-1)] olan t-dağılımı gösterir.

ÖRNEK 1:

Uykusuzluktan şikâyetçi 25 hastadan tesadüfen seçilen 10 tanesine A uyku ilacı, 15 tanesine de B uyku ilacı verilmiştir. Bu uygulamadan sonra söz konusu hasta gruplarının uyku süresi (saat) ortalamaları ve standart hataları sırası ile $\overline{A}\pm S_{\overline{A}}$ = 3.5 ± 0.70 ve $\overline{B}\pm S_{\overline{B}}$ = 4.5± 0.90 olarak bulunmuştur. Bu değerlere göre B ilacının uykusuzluğa karşı daha etkili olduğu söylenebilir mi?

Bu araştırmanın amacı, uykusuzluk şikayeti olan hastaların uyku süresine etki bakımından B ilacının A ilacından daha etkili olup olmadığının araştırılmasıdır. B ilacının daha etkili olduğunun söylenebilmesi için B ilacı verilen hastaların ortalama uyku sürelerinin, A ilacı verilen hastaların uyku sürelerinden daha uzun olması gerekir. Bu da yapılacak hipotez kontrolünün tek taraflı olduğunu gösterir. Bu duruma ilişkin hipotez takımı aşağıdaki gibi oluşturulur.

H₀: A ve B ilaçlarının ortalama uyku süreleri arasında gözlenen fark tesadüften ileri gelmektedir. B ilacının uykusuzluğa karşı daha etkili olduğu söylenemez. Kısaca $\mu_{\bar{A}} - \mu_{\bar{B}} = 0$ veya $\mu_{\bar{A}} = \mu_{\bar{B}}$'dir.

H₁: A ve B ilaçlarının ortalama uyku süreleri arasında gözlenen fark tesadüften ileri gelmemektedir. B ilacının uykusuzluğa karşı daha etkili olduğu söylenebilir. Kısaca $\mu_{\bar{B}} - \mu_{\bar{A}} > 0$ veya $\mu_{\bar{B}} > \mu_{\bar{A}}$'dir.

Kontrol hipotezi doğru ise hastaların ortalama uyku süresi bakımından A ve B ilaçları arasında gözlenen fark, ortalaması sıfır ve standart sapması (5.7) numaralı eşitlikte verildiği gibi olan normal dağılım gösterir.

(5.7) numaralı eşitlik kullanılarak ortalamalar arası farka ait standart hatanın hesaplanabilmesi için verilen bilgilerden yararlanarak grupların kareler toplamları hesaplanmalıdır.

$n_A = 10$ $\quad \bar{A} \pm S_{\bar{A}} = 3.5 \pm 0.70$

$n_B = 15$ $\quad \bar{B} \pm S_{\bar{B}} = 4.5 \pm 0.90$

Yapılan araştırmada A ve B gruplarının ortalaması ve ortalamanın standart hatası verilmiştir. Ortalamaların standart hatalarından kareler toplamları aşağıdaki şekilde hesaplanır. (5.2) numaralı eşitlikten standart hata;

$$S_{\bar{A}} = \frac{S_A}{\sqrt{n_A}} \text{ dır.}$$

(5.2) numaralı eşitlikten A grubunun varyansı;

$$(S_{\bar{A}})^2 = \left(\frac{S_A}{\sqrt{n_A}}\right)^2 = S_A^2 = S_{\bar{A}}^2.n \text{ şeklinde yazılabilir.}$$

(2.13) numaralı eşitlikten, varyansın $S_A^2 = \frac{\sum_{i=1}^{n_A}(A_i - \bar{A})^2}{(n_A - 1)} = \frac{\sum d_A^2}{(n_A - 1)}$ olduğu bilindiğine göre A grubunun kareler toplamı $\sum d_A^2 = S_{\bar{A}}^2.n_A.(n_A - 1)$ şeklinde, B grubunun kareler toplamı da $\sum d_B^2 = S_{\bar{B}}^2.n_B.(n_B - 1)$ şeklinde hesaplanır.

A ve B uyku ilacı alan hasta gruplarının uyku sürelerine ait kareler toplamları ve uyku süreleri ortalamaları arasındaki fark için t-değeri (8.4) numaralı eşitlikten ,

$$\sum d_A^2 = (0.7)^2.10.(10-1) = 44.1$$

$$\sum d_B^2 = (0.9)^2.15.(15-1) = 170.1 \text{ olarak bulunur.}$$

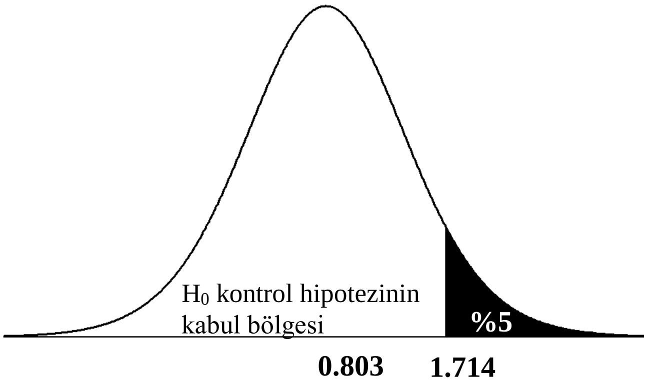
Bu değerler ile (5.7) numaralı eşitlikten ortalamalar arası farka ait standart hata;

$$S_D = \sqrt{\frac{44.1+170.1}{(10-1)+(15-1)}\frac{10+15}{(10)(15)}} \cong 1.246$$

ve (8.4) numaralı eşitlikten t-değeri;

$$t = \frac{4.5-3.5}{1.246} \cong 0.803 \text{ olarak bulunur.}$$

Uykusuzluk şikayeti olan hastaların uyku süresine etki bakımından B ilacının A ilacından daha etkili olup olmadığının araştırılması için uygulanan hipotez kontrolünde I. tip hata olasılığı önceden (α=0.05) %5 olarak belirlenmiş ise yapılan hipotez kontrolü tek taraflı olduğundan bu olasılığın tamamı ortalamadan büyük değerlerin bulunduğu tarafta alınır. Tablo C'den [(10-1)+(15-1)]=23 serbestlik dereceli t-dağılımında %5'lik alanın 1.714 değerinden başladığı bulunur. Böylece Şekil 8.3'de görüldüğü gibi kontrol hipotezinin ret bölgesinin 1.714 değerinden başladığı belirlenmiş olur.



Şekil 8.3. 23 serbestlik dereceli t-dağılımında tek taraflı hipotez kontrolünde kontrol hipotezinin ret ve kabul bölgeleri

Şekil 8.3'te görüldüğü gibi örnekten hesaplanan ortalamalar arası farka karşılık gelen t-değeri, yani test istatistiği olan 0.803 değeri, 1.714 değerinden küçük olup kontrol hipotezinin kabul bölgesine düşmektedir. Dolayısıyla, $t_{Hesaplanan} < t_{tablo}$ (0.803 < 1.714) olduğundan kontrol hipotezi kabul edilir. Yapılan hipotez kontrolü sonunda uykusuzluk şikayeti olan hastaların uyku süresine etki bakımından B ilacının A ilacından daha etkili olmadığı ve uyku süresine etki bakımından bu ilaçların birbirinden farksız olduğu kararına varılır.

ÖRNEK 2:

A ve B firmaları tarafından üretilen meyve sularındaki pH değerlerine ilişkin ortalama, standart sapma ve gözlem adedi aşağıdaki gibidir. Bu değerlere göre A ve B firmalarının üretmiş oldukları meyve sularında pH değerleri bakımından istatistik olarak önemli bir farkın olduğu söylenebilir mi?

Firmalar	Ortalama	Standart sapma	Gözlem sayısı
A	7.55	0.024	5
B	7.49	0.032	5

Meyve suyu fabrikalarında üretilen meyve sularındaki pH değerleri standartlar ile belirlenmiş ise fabrikalar arasında üretilen meyve sularındaki pH değeri bakımından fark tesadüften ileri gelmelidir. Bu araştırmada, A ve B firmalarının üretmiş oldukları meyve sularında pH değerleri ortalamaları arasındaki farkın istatistik olarak önemli olup olmadığının araştırılması amaçlanmıştır. Dolayısıyla bunun için ortalamalar arası farka ait çift taraflı hipotez kontrolü yapılması gerekir.

A ve B firmalarının ürettikleri meyve sularının ortalama pH değerleri arasındaki farkın önemli olup olmadığına karar vermek için (8.4) numaralı eşitlik kullanılarak t-değerinin hesaplanması gerekir. (8.4) numaralı eşitlikten t-değerinin hesaplanabilmesi için önce A ve B firmalarının pH değeri ortalamalarına ait standart hatalar bulunur, (5.7) numaralı eşitlikten ortalamalar arası farka ait standart hata ve (8.4) numaralı eşitlikten de t-değeri hesaplanarak hipotez kontrolü tamamlanır ve karar verilir.

A ve B firmalarının üretmiş oldukları meyve sularında ortalama pH değerleri arasındaki farkın istatistik olarak önemli olup olmadığının belirlenmesi için uygulanan hipotez kontrolü Tablo 8.4'de verilmiştir.

Tablo 8.4. A ve B firmalarının üretmiş oldukları meyve sularında ortalama pH değerleri arasındaki farkın istatistik olarak önemli olup olmadığının belirlenmesi için uygulanan hipotez kontrolü

H_0: A ve B firmalarının üretmiş oldukları meyve sularında ortalama pH değerleri arasındaki 0.06 birimlik fark tesadüften ileri gelmiştir. Bu fark sıfır kabul edilebilir ve istatistik olarak önemli değildir. İki firma arasında pH değerleri bakımından bir fark olduğu söylenemez. Kısaca $\mu_{\bar{A}} - \mu_{\bar{B}} = 0$ veya $\mu_{\bar{A}} = \mu_{\bar{B}}$'dir.

H_1: A ve B firmalarının üretmiş oldukları meyve sularında ortalama pH değerleri arasındaki 0.06 birimlik fark tesadüften ileri gelmemiştir. Bu fark sıfır kabul edilemez ve istatistik olarak önemlidir. İki firma arasında pH değerleri bakımından bir fark olduğu söylenebilir. Kısaca $\mu_{\bar{A}} - \mu_{\bar{B}} \neq 0$ veya $\mu_{\bar{A}} \neq \mu_{\bar{B}}$'dir.

n_A=5, S_A=0.024 ise (5.2) numaralı eşitlikten $S_{\bar{A}} = \dfrac{0.024}{\sqrt{5}} = 0.0107$ ve

n_B=5, S_B=0.032 ise (5.2) numaralı eşitlikten $S_{\bar{B}} = \dfrac{0.032}{\sqrt{5}} = 0.0143$ olarak

bulunur. (5.7) numaralı eşitlikten ortalamalar arası farka ait standart hata;

$$S_D = \sqrt{(0.0107)^2 + (0.0143)^2} = 0.0179$$

ve (8.4) numaralı eşitlikten de t-değeri;

$t = \dfrac{7.55 - 7.49}{0.0179} \cong 3.352$ olarak hesaplanır.

Hesaplanan test istatistiği [(5-1)+(5-1)]=8 serbestlik dereceli t-dağılımı gösterir. Araştırıcı I. tip hata olasılığını α=0.01 olarak belirlemiş ise çift taraflı hipotez kontrolü yapıldığı için I. tip hata olasılığının yarısı ortalamadan küçük ve diğer yarısı ise ortalamadan büyük t değerlerin bulunduğu tarafta alınır. Tablo C'den, 8 serbestlik dereceli t-dağılımında %0.5'lik alanının, yani kontrol hipotezinin ret bölgesinin 3.355 değerinden başladığı görülür. Hesaplanan test istatistiğinin değeri 3.352, kritik bölgenin başladığı t-değerinden (3.355) küçük olup H_0 hipotezinin kabul bölgesine düşmektedir. Bu sebeple de kontrol hipotezi kabul edilir ve firmalar arasında üretilen meyve sularındaki ortalama pH değeleri arasındaki farkın istatistik olarak önemli olmadığı kararına varılır.

Tablo 8.4'te verilen hipotez kontrolünde test istatistiğinin değeri t-dağılımı değerine çok yakın çıkmıştır. Bazı durumlarda uygulanan hipotez kontrolünde hesaplanan test istatistiğinin değeri ile test dağılımında kritik bölgenin başladığı değer aynı olabilir. Hesaplanan test istatistiğinin değeri, kritik alanın başladığı tablo değerine eşit veya büyükse kontrol hipotezinin ret edilmesi gerekmektedir.

ÖRNEK 3:

Keçi ve inek sütlerinde 100 mililitredeki kalsiyum (Ca) miktarları gr olarak aşağıdaki gibi bulunmuştur. Keçi sütündeki kalsiyum (Ca) miktarının inek sütündeki kalsiyum (Ca) miktarından daha fazla olduğu söylenebilir mi?

Keçi Sütü Ca (mL)	142	141	140	139	145
İnek Sütü Ca (mL)	120	122	127	130	

Bu çalışmada, keçi sütündeki kalsiyum miktarının inek sütündeki kalsiyum miktarından daha fazla olup olmadığı araştırılmaktadır. Keçi sütündeki kalsiyum miktarının inek sütündeki kalsiyum miktarından fazla olup olmadığına karar vermek için ortalamalar arası farka ait tek taraflı hipotez kontrolü yapılmalıdır.

Tablo 8.5. Keçi sütündeki kalsiyum miktarının inek sütündeki kalsiyum miktarından daha fazla olup olmadığını araştırmak üzere uygulanan hipotez kontrolü

H_0: Keçi ve inek sütleri arasında 100 mL'deki Ca miktarları bakımından gözlenen fark tesadüften ileri gelmektedir. Bu fark sıfır kabul edilebilir ve istatistik olarak önemli değildir. Keçi ve inek sütleri arasında Ca miktarı bakımından fark olduğu söylenemez. Kısaca $\mu_{\bar{K}} - \mu_{\bar{İ}} = 0$ veya $\mu_{\bar{K}} = \mu_{\bar{İ}}$'dir.

H_1: Keçi ve inek sütleri arasında 100 mL'deki Ca miktarları bakımından gözlenen fark tesadüften ileri gelmemektedir. Bu fark sıfır kabul edilemez ve istatistik olarak önemlidir. Keçi sütündeki kalsiyum miktarının inek sütündeki kalsiyum miktarından daha fazla olduğu söylenebilir. Kısaca $\mu_{\bar{K}} > \mu_{\bar{İ}}$'dir.

Tablo 8.5 devam.

Keçi ve inek sütlerinde tayin edilen Ca miktarlarına ait ortalamalar ve kareler toplamları hesaplandıktan sonra (5.7) numaralı eşitlikten ortalamalar arası farka ait standart hata aşağıdaki şekilde bulunur.

Keçi sütü Ca (ml)	İnek sütü Ca (ml)
142	120
141	122
140	127
139	130
145	

$\bar{K} = 141.4 \qquad \sum d_K^2 = 21.2$

$\bar{İ} = 124.75 \qquad \sum d_İ^2 = 62.75$

Ortalamalar arası farka ait standart hata (5.7) numaralı eşitlikten

$$S_D = \sqrt{\frac{21.2 + 62.75}{(5-1)+(4-1)} \cdot \frac{(5+4)}{(5)(4)}} = 2.323$$

ve test istatistiği (8.4) numaralı eşitlikten;

$$t = \frac{141.4\text{-}124.75}{2.323} \cong 7.168 \text{dir.}$$

Hesaplanan test istatistiği [(5-1)+(4-1)]=7 serbestlik dereceli t-dağılımı gösterir. Eğer araştırıcı I. tip hata olasılığını α=0.05 olarak belirlemiş ise tek taraflı hipotez kontrolü yapıldığı için I. tip hata ortalamadan büyük t değerlerinin bulunduğu tarafta alınır. Tablo C'den, 7 serbestlik dereceli t-dağılımında %5'lik alanının, yani kontrol hipotezinin ret bölgesinin 1.895 değerinden başladığı görülür. Hesaplanan test istatistiğinin değeri 7.168 olup kritik bölgenin başladığı t-değerinden büyüktür ve H_0 hipotezinin ret bölgesine düşmektedir. Bu sebeple kontrol hipotezi ret edilir ve keçi sütündeki kalsiyum miktarının inek sütündeki kalsiyum miktarından daha yüksek olduğu kararına varılır.

8.5.2. Bağımlı İki Grubun Karşılaştırılması

Bazı durumlarda, çalışılan iki gruptaki gözlemler birbirine bağımlı olabilir. Grupların birbirine bağımlı olması durumda her gruptaki gözlemler, aynı bireylerin farklı zaman veya koşullarda ölçülen değerleridir. İki grupta aynı bireyden ölçülen gözlemler birbirinin eşi niteliğindedir. Örneğin, herhangi bir hastalığın tedavisinde uygulanan tedavi yönteminin etkili olup olmadığının araştırılması için tedavi öncesi ve tedavi sonrasında aynı hastalardan gözlem yapılması, yöntemin etkinliğinin belirlenmesi için en doğru yoldur. Bu durumda iki grupta veri toplanan hastalar aynı hastalardır. Dolayısıyla gruplar birbirine bağımlıdır. Bir başka çalışmada ineklerde, sütteki yağ oranının yemlemeden önce ve yemlemeden sonra değişip değişmediği araştırılmak istenebilir. Bunun için aynı inekler hem yemlemeden önce hem de yemlemeden sonra sağılmalı ve sütlerindeki yağ oranı tayin edilmelidir. Bu durumda da yemlemeden önce ve yemlemeden sonra sağılan inekler aynı inekler olduğu için önce-sonra grupları birbirine bağımlıdır.

Birbirine bağımlı iki grubun karşılaştırılması için **Eş-yapma t-testi** uygulanır. Eş-yapma t-testinde, test dağılımı t-dağılımı ve test istatistiği t-değeri olup (8.5) numaralı eşitlik kullanılarak hesaplanır.

$$t = \frac{\overline{D} - \mu_{\overline{D}}}{S_{\overline{D}}} \qquad \ldots(8.5)$$

(8.5) numaralı eşitlikte $\overline{D}$, eşler arası farkların ortalaması, $S_{\overline{D}}$, eşler arası farkların ortalamasına ait standart hatadır. (5.2) numaralı eşitlik kullanılarak;

$$S_{\overline{D}} = \frac{S_D}{\sqrt{n}} = \frac{\sqrt{\frac{\sum d_D^2}{(n-1)}}}{\sqrt{n}}$$ olarak bulunur.

(8.5) numaralı eşitlik yardımıyla hesaplanan test istatistiği (n-1) serbestlik dereceli t-dağılımı gösterir.

ÖRNEK 1:

Toprak yıkama işleminin toprak pH'sı üzerine etkisini araştırmak üzere bir çalışma 5 parselde yürütülmüştür. Her bir parselde yıkamadan önce ve sonra pH değerleri ölçülmüş ve aşağıdaki gibi bulunmuştur.

Yıkama öncesi	3.05	3.45	3.95	4.05	5.00
Yıkama sonrası	7.05	5.65	5.85	6.45	7.00

Araştırıcı yıkama işleminin toprağın pH değerini artırıp artırmadığını araştırmak istemektedir. Yıkamadan önce ve sonra aynı parselin pH'sı ölçüldüğü için gruplardaki gözlemler aynı parselden elde edilmiştir ve iki grup birbirine bağımlıdır. Yıkama işleminin toprağın pH değerini artırıp artırmadığı **eş-yapma t-testi** kullanılarak kontrol edilmelidir.

Yıkama işlemenin toprağın pH değerini artırıp artırmadığını kontrol etmek için eş-yapma t-testi Tablo 8.6'da uygulanmıştır.

Tablo 8.6. Yıkama işlemenin toprağın pH değerini artırıp artırmadığını kontrol etmek için uygulanan eş-yapma t-testi

H_0: Yıkama işlemi öncesi ve sonrası ortalama pH değerleri arasındaki fark tesadüften ileri gelmiştir. Bu fark sıfır kabul edilebilir. Yıkama işlemi toprak pH'sını değiştirmemiştir. Kısaca $\mu_{\overline{D}} = 0$ 'dır.

H_1: Yıkama işlemi öncesi ve sonrası ortalama pH değerleri arasındaki fark tesadüften ileri gelme miştir. Bu fark sıfır kabul edilemez. Yıkama işlemi toprak pH'sını artırmıştır. Kısaca $\mu_{\overline{D}} > 0$ 'dır.

Uygulanacak hipotez kontrolünde karşıt hipotez $\mu_{\overline{D}} > 0$ şeklinde kurulmuştur. Yıkama işleminin toprak pH'sını artırıp artırmadığı araştırıldığı için yıkama sonrası pH değerlerinden yıkama öncesi pH değerleri çıkarılarak eşler arası farklar [D_i=(X_{2i}-(X_{1i})] bulunmuştur. (Bunun tersi de doğrudur. Yani, önce-sonra farkları ile de işlemler yapılabilir. Ancak bu durumda Fark ortalamalarının işaretinin değişeceği göz ardı edilmemelidir).

Yıkamadan önce(X_{1i})	Yıkamadan sonra (X_{2i})	D_i=(X_{2i}-(X_{1i})
3.05	7.05	4.0
3.45	5.65	2.2
3.95	5.85	1.9
4.05	6.45	2.4
5.00	7.00	2.0

$$\sum D_i = 12.5 \qquad \overline{D} = \frac{12.5}{5} = 2.5 \qquad \sum d_D^2 = 2.96$$

Yıkama öncesi ve sonrası pH değeri ölçülen parseller aynı parseller olduğu için test istatistiğinin (8.5) numaralı eşitlik kullanılarak hesaplanması gerekir.

Tablo 8.6 devam

(8.5) numaralı eşitlik kullanılarak test istatistiğinin hesaplanabilmesi için eşler arası farkların ortalamasına ait standart hata (5.2) numaralı eşitlikten;

$$S_{\overline{D}} = \frac{\sqrt{\dfrac{2.96}{(5-1)}}}{\sqrt{5}} = 0.385\,,$$

ve (8.5) numaralı eşitlikten test istatistiği;

$$t = \frac{2.5}{0.385} \cong 6.494 \text{ olarak bulunur.}$$

Hesaplanan test istatistiği, n-1 = (5-1) = 4 serbestlik dereceli t-dağılımı gösterir. Araştırıcı I. tip hata olasılığını α=0.05 olarak belirlemiş ise tek taraflı hipotez kontrolü yapıldığı için I. tip hata ortalamadan büyük t değerlerinin bulunduğu tarafta alınır. Tablo C'den, 4 serbestlik dereceli t-dağılımında %5'lik alanının, yani kontrol hipotezinin ret bölgesinin 2.132 değerinden başladığı görülür. Hesaplanan test istatistiğinin değeri 6.494 olup kritik bölgenin başladığı t-değerinden büyüktür ve H_0 hipotezinin ret bölgesine düşmektedir. Bu sebeple kontrol hipotezi ret edilir ve yıkama işlemenin toprağın pH değerini artırdığı kararına varılır.

ÖRNEK 2:

Herhangi bir dersten sınav stresinin öğrencilerin kanındaki adrenalin miktarı üzerine etkisini araştırmak için 7 öğrencide sınav öncesi ve sınavdan sonraki kandaki adrenalin miktarı (mg) aşağıdaki gibi bulunmuştur. Bu araştırmada sınavdan sonraki rahatlamanın kandaki adrenalin miktarını azaltıp azaltmadığının araştırılması amaçlanmıştır.

Sınav öncesi Adrenalin miktarı (mg)	40	45	50	52	54	53	55
Sınav sonrası Adrenalin miktarı (mg)	32	42	42	45	44	52	57

Sınav öncesi ve sınav sonrası kandaki adrenalin miktarı tespit edilen öğrenciler aynı öğrenciler olduğu için iki grup birbirine bağımlıdır ve tek taraflı eş-yapma t-testinin uygulanması gerekir.

Sınavdan sonraki rahatlamanın kandaki adrenalin miktarını azaltıp azaltmadığının araştırılması için eş-yapma t-testi Tablo 8.7'de uygulanmıştır.

Tablo 8.7. Sınavdan sonraki rahatlamanın kandaki adrenalin miktarını azaltıp azaltmadığının araştırılması için uygulanan eş-yapma t-testi

H_0: Sınav öncesi ve sınav sonrası gözlenen kandaki adrenalin miktarları arasındaki fark tesadüften ileri gelmiştir. Gözlenen farklar sıfır kabul edilebilir. Sınavdan sonraki rahatlamanın adrenalin miktarını azalttığı söylenemez. Kısaca $\mu_{\bar{D}} = 0$ 'dır.

H_1: Sınav öncesi ve sınav sonrası gözlenen kandaki adrenalin miktarları arasındaki fark tesadüften ileri gelmemiştir. Gözlenen farklar sıfır kabul edilemez. Sınavdan sonraki rahatlamanın adrenalin miktarını azalttığı söylenebilir. Kısaca $\mu_{\bar{D}} < 0$ 'dır.

Yapılan araştırmada sınavdan sonraki rahatlamanın adrenalin miktarını azaltıp azaltmadığı araştırıldığı için karşıt hipotez $\mu_{\bar{D}} < 0$ şeklinde kurulmuştur. Eğer sınavdan sonra adrenalin miktarı azalıyor ise sınav sonrası adrenalin miktarlarından sınav öncesi adrenalin miktarları çıkarıldığında elde edilen eşler arasındaki farkların sıfırdan küçük olması gerekir. Bu sebeple de sınav sonrası adrenalin miktarlarından sınav öncesi adrenalin miktarları çıkarılarak eşler arası farklar [$D_i=(X_{2i}-(X_{1i})$] aşağıdaki gibi bulunmuştur.

Öğrenciler	Sınavdan önce(X_{1i})	Sınavdan sonra (X_{2i})	$D_i=(X_{2i}-(X_{1i})$
1	40	32	-8
2	45	42	-3
3	50	42	-8
4	52	45	-7
5	54	55	1
6	53	52	-1
7	55	57	2

$$\sum D_i = -24 \qquad \bar{D} = \frac{-24}{7} \cong -3.429 \qquad \sum d_{\bar{D}}^2 = 109.71$$

Tablo 8.7 devam.

Sınavdan sonra rahatlamanın kandaki adrenalin miktarını azaltıp azaltmadığına karar vermek için test istatistiğinin (8.5) numaralı eşitlik kullanılarak hesaplanması gerekir.

(8.5) numaralı eşitlik kullanılarak test istatistiğinin hesaplanabilmesi için eşler arası farkların ortalamasına ait standart hata (5.2) numaralı eşitlikten;

$$S_{\bar{D}} = \frac{\sqrt{\dfrac{109.71}{(7-1)}}}{\sqrt{7}} = 1.616 \,,$$

ve (8.5) numaralı eşitlikten test istatistiği de;

$$t = \frac{-3.429}{1.616} \cong -2.122 \text{ olarak bulunur.}$$

Hesaplanan test istatistiği (7-1) =6 serbestlik dereceli t-dağılımı gösterir. Araştırıcı I. tip hata olasılığını α=0.01 olarak belirlemiş ise tek taraflı hipotez kontrolü yapıldığı için I. tip hata, ortalamadan küçük t değerlerinin bulunduğu tarafta alınır. Tablo C'den, 6 serbestlik dereceli t-dağılımında %1'lik alanının, yani kontrol hipotezinin ret bölgesinin -2.998 değerinden başladığı görülür. Hesaplanan test istatistiğinin değeri -2.122 olup kritik bölgenin başladığı t-değerinden büyüktür ve H_0 hipotezinin kabul bölgesine düşmektedir. Bu sebeple kontrol hipotezi kabul edilir ve sınavdan sonraki rahatlamanın kandaki adrenalin miktarında meydana getirdiği değişikliğin tesadüften ileri geldiğine karar verilir.

ÖRNEK 3:

Yemleme öncesi ve yemleme sonrası sütteki yağ oranları arasında gözlenen farkın istatistik olarak önemli olup olmadığını kontrol etmek için 8 ineğin yemleme öncesi ve sonrasında sütlerindeki (%) yağ miktarı aşağıdaki gibi tespit edilmiştir.

Yemleme öncesi Yağ miktarı (%)	3.35	3.32	3.45	3.40	3.60	3.41	3.70	3.52
Yemleme sonrası Yağ miktarı (%)	3.12	3.20	3.40	3.42	3.50	3.45	3.40	3.48

Yemleme öncesi ve yemleme sonrası sütteki yağ miktarı tespit edilen inekler aynı inekler olduğu için iki grup birbirine bağımlıdır. Yemlemenin sütteki yağ miktarını etkileyip etkilemediğini kontrol için çift taraflı eş-yapma t-testinin uygulanması gerekir.

Yemlemenin sütteki yağ miktarını etkileyip etkilemediğini kontrol için eş-yapma t-testi Tablo 8.8'de uygulanmıştır.

Tablo 8.8. Yemlemenin sütteki yağ miktarını etkileyip etkilemediğini kontrol için uygulanan eş-yapma t-testi

H_0: Yemleme öncesi ve sonrası sütteki yağ miktarları arasında gözlenen farklar tesadüften ileri gelmiştir. İstatistik olarak önemli değildir ve sıfır kabul edilebilir. Yemleme sütteki yağ miktarını değiştirmemiştir. Kısaca $\mu_{\overline{D}} = 0$ 'dır.

H_1: Yemleme öncesi ve sonrası sütteki yağ miktarları arasında gözlenen farklar tesadüften ileri gelmemiştir. İstatistik olarak önemlidir ve sıfır kabul edilemez. Yemleme sütteki yağ miktarını değiştirmiştir. Kısaca $\mu_{\overline{D}} \neq 0$ 'dır.

Uygulanacak hipotez kontrolü çift taraflı hipotez kontrolü olduğu için karşıt hipotez $\mu_{\overline{D}} \neq 0$ şeklinde kurulmuştur.

İnekler	Yemlemeden önce(X_{1i})	Yemlemeden sonra (X_{2i})	$D_i=(X_{1i}-(X_{2i})$
1	3.35	3.12	0.23
2	3.32	3.20	0.12
3	3.45	3.40	0.05
4	3.40	3.42	-0.02
5	3.60	3.50	0.10
6	3.41	3.45	-0.04
7	3.70	3.40	0.30
8	3.52	3.48	0.04

$$\sum D_i = 0.78 \qquad \overline{D} = \frac{0.78}{8} \cong 0.0975 \qquad \sum d_D^2 = 0.0974$$

Yemlemenin sütteki yağ miktarını değiştirip değiştirmediğine karar vermek için test istatistiği (8.5) numaralı eşitlik kullanılarak hesaplanır.

(8.5) numaralı eşitlik kullanılarak test istatistiğinin hesaplanabilmesi için eşler arası farkların ortalamasına ait standart hata (5.2) numaralı eşitlikten;

Tablo 8.8 devam

$$S_{\overline{D}} = \frac{\sqrt{\frac{0.0974}{(8-1)}}}{\sqrt{8}} = 0.0417 \ ,$$

ve (8.5) numaralı eşitlikten test istatistiği de;

$$t = \frac{0.0975}{0.0417} \cong 2.338 \text{ olarak bulunur.}$$

Hesaplanan test istatistiği (8-1) =7 serbestlik dereceli t-dağılımı gösterir. Araştırıcı I. tip hata olasılığını α=0.05 olarak belirlemiş ise çift taraflı hipotez kontrolü yapıldığı için Tablo C'den, 7 serbestlik dereceli t-dağılımında %2.5'luk alanının, yani kontrol hipotezinin ret bölgesinin 2.365 değerinden başladığı görülür. Hesaplanan test istatistiğinin değeri 2.338 olup kritik bölgenin başladığı t-değerinden küçüktür ve H_0 hipotezinin kabul bölgesine düşmektedir. Bu sebeple de kontrol hipotezi kabul edilir ve yemlemenin sütteki yağ miktarını etkilemediği kararına varılır.

8.6. Ortalamalar Arası Farka ait Hipotez Kontrolünde Örnek Genişliği

Araştırıcı, iki grup ortalaması arasındaki farkın tesadüften ileri gelip gelmediğine dair hipotez kontrolünden önce, hangi büyüklükteki farkların önemli olabilmesi için grupların örnek genişliklerinin en az kaç olması gerektiğini hesaplamak isteyebilir. Ortalamaları karşılaştırılacak grupların normal dağılım gösteren bir populasyondan alındığı varsayılarak örnek genişliği (8.6) numaralı eşitlik ile tahmin edilebilir.

$$n \geq \frac{2S_x^2}{\delta^2}(t_{(\alpha,sd)} + t_{(\beta,sd)})^2 \qquad \ldots(8.6)$$

Eşitlikte, δ, iki grubun temsil ettiği populasyon ortalamaları arasındaki farktır, I. tip hata olasılığı (α) testin tek taraflı mı yoksa çift taraflı mı yapılacağına bağlı olarak tek veya çift taraflı t-değeridir. β, II. tip hata olasılığıdır. S_x^2, populasyon varyansının en iyi tahmini, yani toplanmış varyanstır.

Ortalamalar arası farka ait hipotez kontrolü için gerekli olan en az örnek genişliği, populasyon ortalamaları arasında gözlenebilir en küçük farktan etkilenir. Çok küçük farklılıkların gözlenmesi durumunda kontrol hipotezinin ret edilmesi istenirse örnek genişliği büyür. Örnek içinde bireyler arasındaki değişimin büyük olması da örnek genişliğinin artmasına sebep olur. Grup ortalamaları arasındaki farkın önemli olup olmadığını kontrol etmek için uygulanacak hipotez kontrolündeki I. tip hata olasılığı küçüldükçe, iki grubun temsil ettiği populasyon ortalamaları arasındaki en küçük farkı saptamak için gerekli örnek genişliği artar. Ayrıca, testin gücünün artması da örnek genişliğinin artmasını, yani büyük örnekler ile çalışılmasını gerektirir.

İki grup ortalaması arasındaki fark karşılaştırılırken eğer n_1=n_2 ise testin gücü maksimumdur. Fakat iki gruptaki gözlem sayısının eşit olmaması sık rastlanan bir durumdur. Birinci grubun gözlem sayısının ne olacağı belirlendikten sonra örnek genişliği (8.6) numaralı eşitlik kullanılarak hesaplanır. Örnek genişliği (8.6) numaralı eşitlik kullanılarak hesaplandıktan sonra da ikinci grubun örnek genişliği (8.7) numaralı eşitlik kullanılarak tahmin edilir.

$$n_2 = \frac{n(n_1)}{2n_1 - n} \qquad \ldots(8.7)$$

Ortalamalar arası farka ait hipotez kontrolünde güvenilir sonuçlara ulaşılması için grupların örnek genişliklerinin tahmini Bölüm 8.5.1'de ÖRNEK 1 için Tablo 8.9'da verilmiştir.

Tablo 8.9. Bölüm 8.5.1'de ÖRNEK 1 için örnek genişliklerinin tahmini

Uykusuzluk şikayeti için kullanılan A ve B ilaçlarının uyku süreleri arasındaki farkın en fazla 2 saat duyarlılık ile %5 seviyesinde %90 güç ile saptanabilmesi için A ve B ilaçları ile kaçar hastanın tedavi edilmesi gerektiğinin tahmin edilmesi istenmektedir.

Bu amaçla yapılan bir çalışmada uykusuzluktan şikâyetçi 25 hastadan tesadüfen seçilen 10 tanesine A uyku ilacı, diğer 15 tanesine ise B uyku ilacı verilmiştir. Bu uygulamadan sonra söz konusu hasta gruplarının uyku süresi (saat) ortalamaları ve standart hataları sırası ile $\overline{A} \pm S_{\overline{A}}$ = 3.5 ± 0.70 ve $\overline{B} \pm S_{\overline{B}}$ = 4.5± 0.90 olarak bulunmuştur.

Tablo 8.9 devam

Bölüm 8.5.1 ÖRNEK 1'de açıklandığı gibi (5.2) numaralı eşitlikten A ve B gruplarının varyansları $S_A^2 = S_{\overline{A}}^2.n$ ve $S_B^2 = S_{\overline{B}}^2.n$

şeklinde yazılabilir. Buradan grupların varyansları $S_A^2 = (0.7)^2(10) = 4.9$ ve $S_B^2 = (0.9)^2(15) = 12.15$ olarak bulunur.

(8.6) numaralı eşitlikte S_x^2, populasyon varyansının en iyi tahmini, yani toplanmış varyanstır ve:

$$S_x^2 = \frac{(10-1)4.9+(15-1)2.15}{(10-1)+(15-1)} \cong 9.313 \text{ olarak hesaplanır.}$$

Yapılan araştırmada uyku süresi bakımından B ilacının daha iyi olup olmadığı araştırıldığı için hipotez kontrolü tek taraflı hipotez kontrolüdür. Hipotez kontrolünde, α=%5 seviyesinde tek taraflı olarak, B ilacının 2 saat kadar küçük farkla daha iyi olduğuna %90 olasılık (1-β=0.90) ile karar verilmesi isteniyorsa örnek genişliği (8.6) numaralı eşitlik kullanılarak aşağıdaki şekilde tahmin edilir:

$$n \geq \frac{2S_x^2}{\delta^2}(t_{(\alpha,sd)} + t_{(\beta,sd)})^2$$

Yapılan hipotez kontrolünde sd=(10-1)+(15-1)=23, $t_{(0.05,\ 23)}$=1.714 ve $t_{(0.10,\ 23)}$=1.319'dur. Bu durumda örnek genişliği;

$$n \geq \frac{2(9.313)}{(2)^2}(1.714+1.319)^2 \cong 42.84$$

yani yaklaşık 43 olarak hesaplanır.

Tahmin edilen örnek genişliği kullanılarak örnek genişliği yine tahmin edilirse sd=43-2=41 için Tablo C'de 41 serbestlik dereceli t-dağılımına ait değerler olmadığından bu serbestlik derecesine en yakın 50 serbestlik dereceli t-dağılımı değerleri kullanılmış ve $t_{(0.05,\ 41\cong 50)}$=1.676, $t_{(0.10,\ 41\cong 50)}$=1.299 için ,

$$n \geq \frac{2(9.313)}{(1.5)^2}(1.676+1.299)^2 \cong 41.21$$

yani yaklaşık 42 olarak hesaplanmıştır.

Yapılan araştırmada A ilacı verilen birinci gruptaki hasta sayısı 35 olarak belirlenmiş ise ikinci grupta olması gereken hasta sayısı (8.7) numaralı eşitlikten; $n_2 = \frac{n(n_1)}{2n_1 - n} = \frac{(42)(35)}{2(35)-42} = 52.5$ yani yaklaşık 53 olarak tahmin edilir. Bu durumda yapılan araştırma için 35 A ilacı ve 53 B ilacı uygulanacak toplam 88 hastaya ihtiyaç vardır.

8.7. Korelasyon Katsayısına ait Hipotez Kontrolü

Bir örnekte üzerinde çalışılan iki özellik arasındaki korelasyon katsayısı r olarak hesaplandığı zaman, gerçekten iki özellik arasında doğrusal bir ilişkinin olup olmadığı araştırılmak istenebilir. Örnekten hesaplanan korelasyon katsayısı populasyona ait korelasyon katsayısının (ρ) yani parametrenin bir tahminidir ve çoğu zaman populasyona ait korelasyon katsayısı bilinmez.

Populasyona ait korelasyon katsayısının bilinmemesi durumunda örneğin temsil ettiği populasyona ait korelasyon katsayısının sıfır olduğu kabul edilerek örnekten hesaplanan korelasyon katsayısının sıfırdan farklı olup olmadığına ilişkin hipotez kontrolü, t dağılımından yararlanılarak yapılır.

Populasyona ait korelasyon katsayısının sıfır olduğu kabul edilerek korelasyon katsayısına ait hipotez kontrolü uygulanırken kontrol ve karşıt hipotezler aşağıdaki şekilde kurulmalıdır.

H_0: Çalışılan iki özellik arasındaki korelasyon katsayısı tesadüften ileri gelmiştir. Sıfır kabul edilebilir ve istatistik olarak önemli değildir. Örnek, iki özellik arasındaki korelasyon katsayısı sıfır olan bir populasyonu temsil etmektedir, yani söz konusu iki özellik arasında doğrusal bir ilişki olduğu söylenemez. Kısaca $\rho_{xy}=0$ veya $\mu_r=0$'dır.

H_1: Çalışılan iki özellik arasındaki korelasyon katsayısı tesadüften ileri gelmemiştir. Sıfır kabul edilemez ve istatistik olarak önemlidir. Örnek, iki özellik arasındaki korelasyon katsayısı sıfır olan bir populasyonu temsil etmemektedir, yani söz konusu iki özellik arasında doğrusal bir ilişki olduğu söylenebilir. Kısaca $\rho_{xy}\neq 0$ veya $\mu_r\neq 0$'dır.

Kurulan kontrol hipotezi doğru olduğu zaman örneğin alındığı populasyonda söz konusu özellikler arasındaki korelasyon katsayısı, ρ=0 olacak ve bu populasyondan alınan örneklerden hesaplanacak korelasyon katsayılarının dağılımı normal dağılıma yaklaşacaktır. Üzerinde çalışılan örnekte X ve Y özellikleri arasındaki korelasyon katsayısının sıfır olup olmadığını kontrol etmek için t dağılımından yararlanılarak (8.8) numaralı eşitlik kullanılarak t-değeri hesaplanır.

$$t = \frac{r-\rho}{S_r} \qquad \ldots(8.8)$$

(8.8) numaralı eşitlikte S_r, korelasyon katsayısına ait örnekleme dağılımının örnekten tahmin edilen standart sapması, yani korelasyon katsayısının standart hatasıdır ve (5.16) numaralı eşitlik kullanılarak $S_r = \sqrt{\frac{(1-r^2)}{(n-2)}}$ şeklinde hesaplanır.

Örnekten hesaplanan korelasyon katsayısının istatistik olarak önemli olup olmadığını kontrol etmek için (8.8) numaralı eşitlik kullanılarak hesaplanan t-değeri (n-2) serbestlik dereceli t-dağılımı gösterir.

ÖRNEK 1:

32 adet tavuk yumurtasında yumurta çevresi ile yumurta sarısı ağırlığı arasındaki korelasyon katsayısı r=0.38 olarak bulunmuştur. Adı geçen özellikler arasında doğrusal bir ilişkinin olduğu söylenebilir mi?

Yumurta çevresi ve yumurta sarısı ağırlığı arasındaki korelasyon katsayısının tesadüfi olup olmadığı, yani söz konusu iki özellik arasında doğrusal bir ilişki olup olmadığı araştırılmak istenmektedir. Bu amaçla uygulanacak hipotez kontrolü, aşağıdaki şekilde hipotezlerin kurulması ile başlar.

H_0: Yumurta çevresi ve yumurta sarısı ağırlığı arasındaki korelasyon katsayısı sıfır kabul edilebilir, istatistik olarak önemli değildir. Söz konusu iki özellik arasında doğrusal bir ilişki olduğu söylenemez. Kısaca ρ_{xy}=0 veya μ_r=0'dır.

H_1: Yumurta çevresi ve yumurta sarısı ağırlığı arasındaki korelasyon katsayısı sıfır kabul edilemez, istatistik olarak önemlidir. Söz konusu iki özellik arasında doğrusal bir ilişki olduğu söylenebilir. Kısaca ρ_{xy}=0 veya μ_r=0'dır.

X ve Y özellikleri arasında hesaplanan korelasyon katsayısının sıfır olup olmadığını kontrol etmek için t-değeri olarak (8.8) numaralı eşitlik kullanılır. (8.8) numaralı eşitlik kullanılarak t-değerinin hesaplanabilmesi için ilk olarak (5.16) numaralı eşitlikten korelasyon katsayısının standart hatası;

$$S_r = \sqrt{\frac{(1-0.38^2)}{(32-2)}} \cong 0.169 \text{ bulunur.}$$

(8.8) numaralı eşitlikten ise test istatistiği;

$$t = \frac{r-\rho}{S_r} = \frac{0.38}{0.169} \cong 2.249 \text{ olarak hesaplanır.}$$

Hesaplanan test istatistiği (n-2) = (32-2) =30 serbestlik dereceli t-dağılımı gösterir. Araştırıcı I. tip hata olasılığını α=0.05 olarak belirlemiş ise çift taraflı hipotez kontrolü yapıldığı için Tablo C'den, 30 serbestlik dereceli t-dağılımında %2.5'luk alanının, yani kontrol hipotezinin ret bölgesinin 2.042 değerinden başladığı görülür. Hesaplanan test istatistiğinin değeri 2.249 olup kritik bölgenin başladığı t-değerinden büyüktür ve H_0 hipotezinin ret bölgesine düşer. Bu sebeple de kontrol hipotezi ret edilerek yumurta çevresi ve yumurta sarısı ağırlığı arasında istatistik olarak önemli bir doğrusal ilişki olduğu kararına varılır.

ÖRNEK 2:

Öğrencilerin ara sınavından almış oldukları notlar ile yarı yıl sonu sınavından almış oldukları notlar arasında doğrusal bir ilişkinin var olduğunu söyleyebilmek için, tesadüfen alınan 20 öğrenciden hesaplanan korelasyon katsayısının en az kaç olması gerektiği araştırılmaktadır.

Öğrencilerin ara sınav notları ile yarıyıl sonu sınav notları arasında doğrusal bir ilişkinin olduğunu söyleyebilmek için H_0: $\rho_{xy}=0$ kontrol hipotezinin reddedilmesi gerekir. Kontrol hipotezinin reddedilebilmesi için de (8.8) numaralı eşitlik kullanılarak hesaplanacak test istatistiğinin α seviyesinde (n-2) serbestlik dereceli t-dağılımı değerine eşit veya büyük olması gerekir. Bu örnekte korelasyon katsayısı için hesaplanacak test istatistiği (20-2)=18 serbestlik dereceli t-dağılımı gösterir. Araştırıcı I. tip hata olasılığını $\alpha=0.05$ olarak belirlemiş ise Tablo C'den, 18 serbestlik dereceli t-dağılımında %2.5'luk alanın başladığı t-değeri 2.101 olarak bulunur. Bu durumda (8.8) numaralı eşitlik 2.101 değerine eşitlenerek "r" çözülür.

$$2.101=\frac{r}{\sqrt{\frac{(1-r^2)}{(20-2)}}} \rightarrow (2.101)^2=\frac{r^2}{\frac{(1-r^2)}{18}} \rightarrow (2.101)^2=\frac{18r^2}{(1-r^2)} \rightarrow (2.101)^2(1-r^2)=18r^2$$

Yukarıdaki eşitlikten r çözüldüğünde, ara sınav notları ile yarıyıl sonu sınav notları arasında istatistik olarak önemli bir doğrusal bir ilişkinin olduğunu söyleyebilmek için 20 öğrenciden hesaplanan korelasyon katsayısının en küçük değerinin r=0.4438 olması gerektiği bulunur.

ÖRNEK 3:

X ve Y özelliklerinin 30 bireylik bir örnekten elde edilen verileri kullanılarak b_{yx} = -0.9, b_{xy} = -0.4 olarak hesaplanmıştır. Araştırıcı, regresyon katsayılarını hesapladıktan sonra X ve Y özellikleri arasında istatistik olarak önemli bir doğrusal bir ilişkinin olup olmadığını araştırmaya karar vermiştir.

Korelasyon ve regresyon katsayıları arasındaki ilişkiden yararlanılarak, (3.3) numaralı eşitlikte verilen korelasyon katsayısı, regresyon katsayıları cinsinden aşağıdaki şekilde yazılabilir:

$$r=\frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}}$$

$$r^2=\frac{(\sum d_x d_y)^2}{\sum d_x^2 \sum d_y^2}$$

Bu eşitlik; $r^2=\frac{\sum d_x d_y}{\sum d_x^2}\frac{\sum d_x d_y}{\sum d_y^2}$ şeklinde düzenlenebilir.

Yeni düzenlenen eşitlikte birinci terim b_{yx} ve ikinci terim b_{xy}'dir. Bu eşitliğin karekökü alınarak, korelasyon katsayısı $r=\sqrt{(b_{yx})(b_{xy})}$ şeklinde elde edilir.

X ve Y özelliklerinin 30 bireylik bir örnekten elde edilen verileri kullanılarak b_{yx} = -0.9, b_{xy} = -0.4 olarak hesaplanmış ise korelasyon katsayısı $r=\sqrt{(-0.9)(-0.4)}=-0.6$ olarak bulunur. (-0.9) ve (-0.4) çarpımlarının karekökü (0.6) olmasına karşın korelasyon katsayının işareti negatiftir. Çünkü hesaplanan regresyon katsayıları iki özellik arasında, bir özelliğin artarken diğer özelliğin azaldığı şeklinde ters bir ilişki olduğunu göstermekte dir. Bunun korelasyon katsayısındaki karşılığı da korelasyon katsayısının işaretinin negatif olmasıdır. Dolayısıyla, regresyon ve korelasyon katsayılarının işaretlerinin birbirinden farklı olması mümkün değildir.

X ve Y özellikleri arasında istatistik olarak önemli doğrusal bir ilişkinin olup olmadığını araştırmak üzere yapılan hipotez kontrolü Tablo 8.10'da uygulanmıştır.

Tablo 8.10. X ve Y özellikleri arasında istatistik olarak önemli doğrusal bir ilişkinin olup olmadığını araştırmak üzere uygulanan hipotez kontrolü

H_0: Örnek, iki özellik arasındaki korelasyon katsayısı sıfır olan bir populasyonu temsil etmektedir. Yani söz konusu iki özellik arasında doğrusal bir ilişki olduğu söylenemez. Kısaca $\rho_{xy}=0$ veya $\mu_r=0$'dır.

H_1: Örnek, iki özellik arasındaki korelasyon katsayısı sıfır olan bir populasyonu temsil etmemektedir. Yani söz konusu iki özellik arasında doğrusal bir ilişki olduğu söylenebilir. Kısaca $\rho_{xy}\neq 0$ veya $\mu_r\neq 0$'dır.

Korelasyon katsayısının standart hatası (5.16) numaralı eşitlikten;

$$S_r = \sqrt{\frac{(1-(-0.6)^2)}{(30-2)}} \cong 0.151,$$

X ve Y arasındaki korelasyon katsayısının sıfır olup olmadığını kontrol etmek için t-değeri de (8.8) numaralı eşitlikten;

Tablo 8.10 devam

$$t = \frac{-0.6}{0.151} \cong \text{-3.974 olarak bulunur.}$$

Hesaplanan test istatistiği (30-2) =28 serbestlik dereceli t-dağılımı gösterir. Araştırıcı I. tip hata olasılığını $\alpha=0.01$ olarak belirlemiş ise çift taraflı hipotez kontrolü yapıldığı için Tablo C'den, 28 serbestlik dereceli t-dağılımında %0.5'lik alanın, yani kontrol hipotezinin ret bölgesinin -2.763 değerinden başladığı görülür. Hesaplanan test istatistiğinin değeri -3.974, kritik bölgenin başladığı t-değerinden küçük olup H_0 hipotezinin ret bölgesine düşmektedir. Bu sebeple kontrol hipotezi ret edilerek X ve Y özellikleri arasında önemli bir doğrusal ilişki olduğu kararına varılır.

8.8. Regresyon Katsayısına Ait Hipotez Kontrolü

Yapılan bir çalışmada iki özellik üzerinde çalışılıyorsa, özelliklerden (değişkenlerden) biri diğerine bağlı olarak değişebilir dolayısıyla özelliklerden biri diğerinin fonksiyonu olarak dikkate alınabilir. Bu durumda araştırıcının bağımsız değişkenin bir birim artmasına karşılık bağımlı değişkende meydana gelen ortalama değişme miktarını bulabilmek için regresyon katsayısını hesaplaması gerekir. Regresyon katsayısı hesaplandığı zaman, bağımlı değişkende meydana gelen değişme miktarının istatistik olarak önemli olup olmadığı kontrol edilmek istendiğinde regresyon katsayısına ait hipotez kontrolünün yapılması gerekir.

Regresyon katsayısına ait hipotez kontrolü uygulanırken kontrol ve karşıt hipotezler aşağıdaki şekilde kurulmalıdır.

H_0: Üzerinde çalışılan iki özellik için hesaplanan regresyon katsayısı sıfır kabul edilebilir. Bağımsız değişkeninin bir birim artmasına karşılık bağımlı değişkende ortalama olarak meydana gelen değişme miktarı istatistik olarak önemli değildir. Kısaca, $\beta_{yx}=0$'dır.

H_1: Üzerinde çalışılan iki özellik için hesaplanan regresyon katsayısı sıfır kabul edilemez. Bağımsız değişkeninin bir birim artmasına karşılık bağımlı değişkende ortalama olarak meydana gelen değişme miktarı istatistik olarak önemlidir. Kısaca $\beta_{yx}\neq 0$'dır.

Test istatistiğinin hesaplanması için ilgilenilen örnekleme dağılımı regresyon katsayısına ait örnekleme dağılımıdır. Üzerinde çalışılan örneğin, $\beta_{yx}=0$ olan bir populasyondan tesadüfen alınmış bir örnek olup olmadığına karar vermek için populasyona ait β_{yx} bilinmediğinden kullanılması gereken t-değeri (8.9) numaralı eşitlikte verilmiştir.

$$t=\frac{b_{yx}-\beta_{yx}}{S_b} \qquad \ldots(8.9)$$

(8.9) numaralı eşitlikte, S_b, regresyon katsayısına ait standart hata olup (5.29) numaralı eşitlik kullanılarak $S_b=\sqrt{\frac{S_e^2}{\sum d_x^2}}$ şeklinde hesaplanır. (5.29) numaralı eşitlikte ise S_e^2, regresyondan sapma kareler ortalaması (regresyon varyansı) olup, (5.28) numaralı eşitlik kullanılarak hesaplanır.

Örnekten hesaplanan regresyon katsayısının istatistik olarak önemli olup olmadığını kontrol etmek için (8.9) numaralı eşitlik kullanılarak hesaplanan t-değeri (n-2) serbestlik dereceli t-dağılımı gösterir.

ÖRNEK 1:

10 bireylik bir örnekte X ve Y gibi iki özelliğe ilişkin varyanslar $S_X^2=S_Y^2=1.44$ ve bu iki özellik arasındaki korelasyon katsayısı da r = -0.5 olarak bulunduğuna göre Y'nin X'e göre regresyon katsayısının istatistik olarak önemli olduğu söylenebilir mi?

Yapılan çalışmada X ve Y özelliklerine ait standart sapmalar $\sqrt{S_X}=\sqrt{S_Y}=\sqrt{1.44}=1.2$ olarak bulunduktan sonra (8.10) numaralı eşitlik;

$r=b_{yx}\frac{S_x}{S_y} \Rightarrow b_{yx}=\frac{(r)S_y}{S_x}$ şeklinde düzenlenerek regresyon katsayısı;

$b_{yx}=\frac{(-0.5)\,(1.2)}{(1.2)}=-0.5$ olarak bulunur.

Korelasyon ve regresyon katsayısı arasında (8.10) ve (8.11) numaralı eşitliklerde verilen ilişkiler bilinmiyor ise regresyon katsayısı aşağıdaki şekilde de hesaplanabilir:

10 bireyden $S_X^2 = S_Y^2 = 1.44$ olarak hesaplanmış ise özellikler ait kareler toplamları $\sum d_x^2 = \sum d_y^2 = 12.96$ olarak hesaplanır. (3.3) numaralı eşitlikte verilen korelasyon katsayısı formülünden ise çarpımlar toplamı $\sum d_x d_y = -6.48$ olarak bulunur. (3.5) numaralı eşitlik kullanılarak regresyon katsayısı $b_{yx} = \dfrac{-6.48}{12.96} = -0.5$ olarak hesaplanır.

Regresyon katsayısı hesaplandıktan sonra X bağımsız değişkeninin bir birim artmasına karşılık Y bağımlı değişkeninde ortalama olarak gözlenecek değişimin tesadüfi olup olmadığını, yani Y'nin X'e olan regresyon katsayısının sıfırdan farklı olup olmadığını kontrol etmek için yapılan regresyon katsayına ait hipotez kontrolü Tablo 8.11'de uygulanmıştır.

Tablo 8.11. Y'nin X'e olan regresyon katsayısının sıfırdan farklı olup olmadığını kontrol etmek için yapılan regresyon katsayına ait hipotez kontrolü

H_0: Üzerinde çalışılan iki özellik için hesaplanan regresyon katsayısı sıfır kabul edilebilir. Bağımsız değişkenin bir birim artmasına karşılık bağımlı değişkende ortalama olarak meydana gelen değişme miktarı istatistik olarak önemli değildir. Kısaca, $\beta_{yx}=0$'dır.

H_1: Üzerinde çalışılan iki özellik için hesaplanan regresyon katsayısı sıfır kabul edilemez. Bağımsız değişkenin bir birim artmasına karşılık bağımlı değişkende ortalama olarak meydana gelen değişme miktarı istatistik olarak önemlidir. Kısaca $\beta_{yx}\neq 0$'dır.

(8.9) numaralı eşitlik kullanılarak t-değerinin hesaplanabilmesi için ilk olarak (5.28) numaralı eşitlikten regresyondan sapma kareler ortalamasının hesaplanması gerekir.

$$S_e^2 = \frac{\sum d_y^2 - b_{yx}\sum d_x d_y}{n-2} = \frac{12.96-(-0.5)(-6.48)}{10-2} = 1.215 \text{ olarak}$$

hesaplandıktan sonra regresyon katsayısının standart hatası (5.29) numaralı eşitlikten;

Tablo 8.11 devam

$$S_b = \sqrt{\frac{S_e^2}{\sum d_x^2}} = \sqrt{\frac{1.215}{12.96}} \cong 0.306$$ ve (8.9) numaralı eşitlikten t-değeri de;

$$t = \frac{-0.5}{0.306} \cong -1.634$$ olarak hesaplanır.

Hesaplanan test istatistiği (10-2) =8 serbestlik dereceli t-dağılımı gösterir. Araştırıcı I. tip hata olasılığını α=0.05 olarak belirlemiş ise çift taraflı hipotez kontrolü yapıldığı için Tablo C'den, 8 serbestlik dereceli t-dağılımında %2.5'luk alanın, yani kontrol hipotezinin ret bölgesinin -2.306 değerinden başladığı görülür. Hesaplanan test istatistiğinin değeri -1.634 olup kritik bölgenin başladığı t-değerinden büyüktür ve H_0 hipotezinin kabul bölgesine düşmektedir. Bu sebeple kontrol hipotezi kabul edilir. X bağımsız değişkeninin bir birim artmasına karşılık Y bağımlı değişkeninde ortalama olarak gözlenecek değişimin tesadüfî olduğu, yani Y'nin X'e olan regresyon katsayısının sıfırdan farklı olmadığı kararına varılır.

ÖRNEK 2:

Belirli bir yörede yetişen Titrek kavaklardan tesadüfen seçilen 8 tanesinin yaş ve boylarına ilişkin Tablo 8.12'deki ölçümler yapılmıştır. Bu çalışmada yaşın bir yıl artmasına karşılık titrek kavakların boylarında ortalama olarak gözlenen değişimin istatistik olarak önemli olup olmadığının araştırılması amaçlanmıştır.

Yaşın bir yıl artmasına karşılık titrek kavakların boylarında ortalama olarak gözlenen değişimin istatistik olarak önemli olup olmadığının araştırılabilmesi için Tablo 8.12'de verildiği gibi regresyon katsayısının hesaplanması gerekir.

Regresyon katsayısı hesaplandıktan sonra titrek kavaklarda yaşın bir yıl artmasına karşılık boylarında ortalama olarak gözlenen 0.3832 m'lik artışın istatistik olarak önemli olup olmadığının araştırılması için yapılan regresyon katsayına ait hipotez kontrolü Tablo 8.13'de uygulanmıştır.

Tablo 8.12. Titrek kavaklardan tesadüfen seçilen 8 tanesinin yaş ve boylarına ilişkin ölçümler ve boyun yaşa göre regresyon katsayısı

Yaş (yıl) (X)	Boy (m) (Y)	
23	19.5	$\sum d_x^2 = 40.875$
20	18.0	$\sum d_y^2 \cong 9.879$
19	17.8	$\sum d_x d_y \cong 15.663$
23	19.0	$b_{yx} = \frac{15.663}{40.875} \cong 0.3832$
18	17.5	Hesaplanan regresyon katsayısı titrek kavakların bir yıl yaşlanmasına karşılık boylarının ortalama olarak 0.3832 m arttığını göstermiştir.
16	16.0	
19	17.0	
21	16.5	

Tablo 8.13. Titrek kavaklarda yaşın bir yıl artmasına karşılık boylarında ortalama olarak gözlenen artışın tesadüfi olup olmadığının araştırılması için yapılan hipotez kontrolü

H_0: Titrek kavaklarda yaşın bir yıl artmasına karşılık boylarında ortalama olarak gözlenen 0.3832 m'lik artış tesadüften ileri gelmiştir ve istatistik olarak önemli değildir. Kısaca, β_{yx}=0'dır.

H_1: Titrek kavaklarda yaşın bir yıl artmasına karşılık boylarında ortalama olarak gözlenen 0.3832 m'lik artış tesadüften ileri gelmemiştir ve istatistik olarak önemlidir. Kısaca $\beta_{yx}\neq$0'dır.

(8.9) numaralı eşitlik kullanılarak t-değerinin hesaplanabilmesi için önce (5.28) numaralı eşitlikten regresyondan sapma kareler ortalamasının hesaplanması gerekir.

$$S_e^2 = \frac{\sum d_y^2 - b_{yx}\sum d_x d_y}{n-2} = \frac{9.879-(0.3832)(15.663)}{8-2} = 0.646 \text{ olarak}$$

hesaplandıktan sonra regresyon katsayısının standart hatası (5.29) numaralı eşitlikten;

Tablo 8.13. devam

$$S_b = \sqrt{\frac{S_e^2}{\sum d_x^2}} = \sqrt{\frac{0.646}{40.875}} \cong 0.126$$ ve (8.9) numaralı eşitlikten t-değeri de;

$$t = \frac{0.3832}{0.126} \cong 3.041 \text{ olarak hesaplanır.}$$

Hesaplanan test istatistiği (8-2) =6 serbestlik dereceli t-dağılımı gösterir. Araştırıcı I. tip hata olasılığını α=0.05 olarak belirlemiş ise çift taraflı hipotez kontrolü yapıldığı için Tablo C'den, 6 serbestlik dereceli t-dağılımında %2.5'luk alanın, yani kontrol hipotezinin ret bölgesinin 2.447 değerinden başladığı görülür. Hesaplanan test istatistiğinin değeri 3.041 olup kritik bölgenin başladığı t-değerinden büyüktür ve H_0 hipotezinin ret bölgesine düşmektedir. Bu sebeple kontrol hipotezi ret edilir.

Tablo 8.13'de uygulanan hipotez kontrolü sonunda hesaplanan boyun yaşa göre regresyon katsayısının tesadüfi olmadığına, bir başka deyişle, titrek kavakların bir yıl yaşlanmasına karşılık ortalama olarak boylarında gözlenen 0.3832 m'lik artışın istatistik olarak önemli olduğu kararına varılmıştır.

8.9. Regresyon Katsayıları Arasındaki Farka ait Hipotez Kontrolü

Üzerinde çalışılan iki özellik arasındaki regresyon katsayısı iki ayrı bölgede, çeşitte veya iki farklı muamele uygulanmış gruplarda hesaplanabilir. Bu durumda iki örnekten hesaplanan regresyon katsayıları arasındaki farklın, yani bağımsız değişkeninin bir birim artmasına karşılık bağımlı değişkende meydana gelen ortalama değişmeler arasındaki farklın tesadüfi olup olmadığının araştırılması gerekir. Bu kontrol regresyon katsayıları arasındaki farka ait hipotez kontrolü uygulanarak test edilebilir.

Regresyon katsayıları arasındaki farka ait hipotez kontrolü uygulanırken kontrol ve karşıt hipotezler aşağıdaki şekilde kurulmalıdır.

H_0: İki grup arasında, bağımsız değişkenin bir birim artmasına karşılık bağımlı değişkende meydana gelen değişme bakımından gözlenen fark sıfır kabul edilebilir. Bağımsız değişkenin bir birim artmasına karşılık bağımlı değişkende meydana gelen değişme bakımından örnekler aynı populasyonu temsil etmektedir. Kısaca, $\beta_{yxA} - \beta_{yxB} = 0$'dır.

H_1: İki grup arasında bağımsız değişkenin bir birim artmasına karşılık bağımlı değişkende meydana gelen değişme bakımından gözlenen fark sıfır kabul edilemez. Bağımsız değişkenin bir birim artmasına karşılık bağımlı değişkende meydana gelen değişme bakımından örnekler aynı populasyonu temsil etmemektedir. Kısaca, $\beta_{yxA} - \beta_{yxB} \neq 0$'dır.

Regresyon katsayıları arasındaki farka ait hipotez kontrolü, araştırmanın amacı doğrultusunda yukarıda ifade edildiği gibi çift taraflı veya tek taraflı uygulanabilir. Bu durumda hipotezler;

H_0: $\beta_{yxA} - \beta_{yxB} = 0$	veya	**H_0:** $\beta_{yxA} - \beta_{yxB} = 0$
H_1: $\beta_{yxA} < \beta_{yxB}$		**H_1:** $\beta_{yxA} > \beta_{yxB}$

şeklinde kurulur.

Burada araştırıcının ilgilendiği regresyon katsayıları arasındaki farka ait örnekleme dağılımıdır. İki regresyon katsayısı arasındaki farkın istatistik olarak önemli olup olmadığını, yani üzerinde çalışılan grupların regresyon katsayısı bakımından aynı populasyondan tesadüfen alınmış örnekler olup olmadıklarını kontrol için t-değeri (8.10) numaralı eşitlik kullanılarak hesaplanır.

$$t = \frac{b_{yx_A} - b_{yx_B}}{S_{(b_A - b_B)}} \qquad \ldots(8.10)$$

(8.10) numaralı eşitlikte $S_{(b_A - b_B)}$, regresyon katsayıları arasındaki farkın standart hatası olup (5.34) eşitlik kullanılarak;

$$S_{(b_A - b_B)} = \sqrt{S_e^2\left(\frac{1}{\sum d_{x_A}^2} + \frac{1}{\sum d_{x_B}^2}\right)}$$ şeklinde hesaplanır.

(5.34) numaralı eşitlikte S_e^2, n_A ve n_B örnek genişliğindeki örneklerden (5.28) numaralı eşitlik kullanılarak hesaplanan regresyondan sapma kareler ortalamalarının serbestlik dereceleri ile tartılı ortalamasıdır ve (5.35) numaralı eşitlik kullanılarak;

$$S_e^2 = \frac{(n_A - 2)S_{e_A}^2 + (n_B - 2)S_{e_B}^2}{(n_A - 2) + (n_B - 2)}$$

şeklinde hesaplanır.

Hesaplanan regresyon katsayıları arasındaki farkn önemli olup olmadığını kontrol için (8.10) numaralı eşitlik kullanılarak hesaplanan t-değeri [(n_A-2) +(n_B-2)] serbestlik dereceli t-dağılımı gösterir.

ÖRNEK 1:

İki ayrı ilden tesadüfen seçilen 10'ar bireyde ağırlık (kg) ve boy uzunlukları (cm) Tablo 8.14'de verildiği gibi ölçülmüştür. İki ayrı ilden elde edilen veriler kullanılarak boydaki bir cm artışa karşılık ağırlıkta gözlenen ortalama değişme, yani ağırlığın boya göre regresyon katsayısı ise yine Tablo 8.14'de verildiği gibi hesaplanmıştır.

Tablo 8.14. İki ayrı ilden tesadüfen seçilen 10'ar bireyde ağırlık (kg) ve boy uzunlukları (cm) ölçümleri ve ağırlığın boya göre regresyon katsayıları

İL 1		İL 2		
Ağırlık (Y)	Boy (X)	Ağırlık (Y)	Boy (X)	Birinci ilde:
56	165	57	166	$\sum d_{x1}^2 = 228.00$
75	176	76	178	$\sum d_{y1}^2 \cong 543.60$
70	175	60	169	$\sum d_{x1}d_{y1} \cong 343.00$
61	168	64	169	$b_{yx1} = 343/228 = 1.504$
61	167	63	170	İkinci ilde:
63	172	71	176	$\sum d_{x2}^2 = 214.00$
72	175	78	180	$\sum d_{y2}^2 \cong 492.00$
80	180	62	169	$\sum d_{x2}d_{y2} \cong 320$
76	179	75	177	$b_{yx2} = 1.495$
68	173	71	176	

Tablo 8.14'de hesaplanan regresyon katsayıları 1. ilde bireylerin boyunun 1 cm artmasına karşılık ağırlıklarının 1.504 kg arttığını, 2. ilde ise bireylerin boyunun 1 cm artmasına karşılık ağırlıklarının 1.495 kg arttığını göstermiştir. Araştırmada, boyun 1 cm artmasına karşılık ağırlıkta gözlenen ortalama artışın iki ilde de aynı olup olmadığı, yani ağırlıkta gözlenen artış bakımından iller arasındaki farkın önemli olup olmadığının test edilmesi amaçlanmıştır. Bunun için yapılan regresyon katsayıları arasındaki farka ait hipotez kontrolü Tablo 8.15'de verildiği şekilde uygulanır.

Tablo 8.15. Boyun 1 cm artmasına karşılık ağırlıkta gözlenen ortalama artışın iki ilde de aynı olup olmadığını kontrol için uygulanan hipotez kontrolü

H_0: Boyun 1 cm artmasına karşılık ağırlıkta gözlenen ortalama artış iki ilde de aynıdır, yani ağırlıkta gözlenen artış bakımından iller arasındaki fark tesadüften ileri gelmiştir. İstatistik olarak önemli değildir ve sıfır kabul edilebilir. Kısaca, β_{yx1} - β_{yx2} =0'dır.

H_1: Boyun 1 cm artmasına karşılık ağırlıkta gözlenen ortalama artış iki ilde de aynı değildir, yani ağırlıkta gözlenen artış bakımından iller arasındaki fark tesadüften ileri gelmemiştir. İstatistik olarak önemlidir ve sıfır kabul edilemez. Kısaca, $\beta_{yx1} \neq \beta_{yx2}$ =0'dır.

(8.10) numaralı eşitlik kullanılarak t-değerinin hesaplanabilmesi için ilk olarak $S_{(b_A - b_B)}$, yani regresyon katsayıları arasındaki farkın standart hatasının hesaplanması gerekir.

Her iki ile ait regresyondan sapma kareler ortalamaları (5.28) numaralı eşitlikten;

$$S^2_{e_1} = \frac{543.60 - 1.504(343)}{10-2} \cong 3.47$$

$$S^2_{e_2} = \frac{492.00 - 1.495(320)}{10-2} = 1.70 \text{ olarak bulunur.}$$

İller için hesaplanan regresyondan sapma kareler ortalamalarının serbestlik dereceleri ile tartılı ortalaması (5.35) numaralı eşitlik kullanılarak;

$$S^2_e = \frac{(10-2)(3.47) + (10-2)(1.70)}{(10-2)+(10-2)} = 2.585,$$

Regresyon katsayıları arasındaki farkın standart hatası (5.34) numaralı eşitlik kullanılarak;

$$S_{(b1-b2)} = \sqrt{(2.585)(\frac{1}{228.0} + \frac{1}{214.0}} = 0.153 \text{ve}$$

(8.10) numaralı eşitlikten de t-değeri;

$$t = \frac{1.504 - 1.495}{0.153} \cong 0.059 \text{ olarak hesaplanır.}$$

Hesaplanan test istatistiği [(10-2) +(10-2)]=16 serbestlik dereceli t-dağılımı gösterir. Araştırıcı I. tip hata olasılığını α=0.05 olarak belirlemiş ise çift taraflı hipotez kontrolü yapıldığı için Tablo C'den, 16 serbestlik dereceli t-dağılımında %2.5'luk alanın, yani kontrol hipotezinin ret bölgesinin 2.120 değerinden başladığı görülür. Hesaplanan test istatistiğinin değeri 0.059 olup kritik bölgenin başladığı t-değerinden çok küçüktür ve H_0 hipotezinin kabul bölgesine düşer. Bu sebeple kontrol hipotezi kabul edilir.

Tablo 8.15'de uygulanan hipotez kontrolü, boyun 1 cm artmasına karşılık ağırlıkta gözlenen ortalama artışın iki ilde de aynı olduğunu göstermiştir. Yapılan araştırmada, verilerin analizi sonucunda, ağırlıkta gözlenen artış bakımından iller arasındaki farkın tesadüften ileri geldiği kararına varılmıştır.

ÖRNEK 2:

2 ve 3 ay süt verilmiş kuzularda doğum ağırlığı (DA) ve 3. aydaki canlı ağırlıklar (CA) Tablo 8.16'da verildiği gibi ölçülmüştür. 2 ve 3 ay süt verilen kuzularda doğum ağırlığının bir kg artmasına karşılık 3. ay canlı ağırlıkta gözlenen ortalama değişme miktarları, yani CA'nın DA'na göre regresyon katsayıları ise Tablo 8.16'daki gibi hesaplanmıştır.

Tablo 8.16. 2 ve 3 ay süt verilen kuzularda doğum ağırlığı (DA) (kg) ve 3. ay canlı ağırlık (kg) ölçümleri ve 3. ay canlı ağırlığının doğum ağırlığına göre regresyon katsayıları

2 ay süt verilen		3 ay süt verilen		
DA (X)	CA (Y)	DA (X)	CA (Y)	2 ay süt verilenlerde:
3.2	16.1	4.2	19.0	$\sum d_{x1}^2 = 3.700$
4.8	19.1	5.1	19.8	$\sum d_{y1}^2 \cong 50.729$
4.9	19.6	4.3	19.6	$\sum d_{x1}d_{y1} \cong 11.320$
5.1	20.0	4.3	17.0	$b_{yx1} = \frac{11.320}{3.700} \cong 3.06$
5.0	18.0	4.6	18.6	3 ay süt verilenlerde:
3.4	11.6	3.8	16.0	$\sum d_{x2}^2 = 1.039$
4.4	16.6	4.7	19.3	$\sum d_{y2}^2 \cong 12.295$
		4.5	18.5	$\sum d_{x2}d_{y2} \cong 2.618$
				$b_{yx2} = \frac{2.618}{1.039} \cong 2.520$

Tablo 8.16'da hesaplanan regresyon katsayıları 2 ay süt verilen kuzularda doğum ağırlığının 1 kg artmasına karşılık 3. aydaki canlı ağırlığın ortalama 3.06 kg arttığını, 3 ay süt verilen kuzularda ise doğum ağırlığındaki 1 kg'lık artışa karşılık 3. aydaki ortalama canlı ağırlık artışının 2.520 kg olduğunu göstermiştir.

Bulunan bu sonuçlar doğrultusunda 2 ay süt verilen kuzularda doğum ağırlığındaki 1 kg'lık artışa karşılık 3. ay canlı ağırlıkta ortalama olarak gözlenen artışın, 3 ay süt verilen kuzulardan daha fazla olmasının istatistik olarak önemli olup olmadığının araştırılması hedeflenmiştir. Bunun için yapılan regresyon katsayıları arasındaki farka ait hipotez kontrolü Tablo 8.16'da verildiği şekilde uygulanır.

Tablo 8.16. 2 ay süt verilen kuzularda doğum ağırlığındaki 1 kg'lık artışa karşılık 3.ay canlı ağırlığında ortalama olarak gözlenen artışın 3 ay süt verilen kuzulara göre daha fazla olup olmadığının araştırılması için uygulanan hipotez kontrolü

H_0: 2 ay süt verilen kuzularda doğum ağırlığındaki 1 kg'lık artışa karşılık 3. ay canlı ağırlıkta ortalama olarak gözlenen artışın 3 ay süt verilen kuzulara göre daha fazla olduğu söylenemez. İki grup arasında 3. ay canlı ağırlığında gözlenen ortalama artış bakımından fark tesadüften ileri gelmiştir ve istatistik olarak önemli değildir. Kısaca, $\beta_{yx1} = \beta_{yx2}$'dir.

H_1: 2 ay süt verilen kuzularda doğum ağırlığındaki 1 kg'lık artışa karşılık 3. ay canlı ağırlıkta ortalama olarak gözlenen artışın 3 ay süt verilen kuzulara göre daha fazla olduğu söylenebilir. İki grup arasında 3. ay canlı ağırlığında gözlenen ortalama artış bakımından fark tesadüften ileri gelmemiştir ve istatistik olarak önemlidir. Kısaca, $\beta_{yx1} > \beta_{yx2}$'dir.

(8.10) numaralı eşitlik kullanılarak t-değerinin hesaplanabilmesi için ilk olarak $S_{(b_A - b_B)}$, yani regresyon katsayıları arasındaki farkın standart hatasının hesaplanması gerekir.

2 ay ve 3 ay süt verilen kuzulara ait regresyondan sapma kareler ortalamaları (5.28) numaralı eşitlikten;

$$S^2_{e_1} = \frac{50.729 - 3.06(11.32)}{7-2} \cong 3.218,$$

$$S^2_{e_2} = \frac{12.295 - 2.52(2.618)}{8-2} \cong 0.950,$$

Regresyondan sapma kareler ortalamalarının serbestlik dereceleri ile tartılı ortalaması (5.35) numaralı eşitlik kullanılarak;

$$S^2_e = \frac{(7-2)3.218 + (8-2)0.95}{(7-2)+(8-2)} \cong 1.981\ ,$$

Regresyon katsayıları arasındaki farkın standart hatası (5.34) numaralı eşitlik kullanılarak;

Tablo 8.16 devam

$$S_{(b1-b2)} = \sqrt{1.981(\frac{1}{3.7} + \frac{1}{1.039})} \cong 1.563 \, ,$$

ve (8.10) numaralı eşitlikten de;

$$t = \frac{3.06 - 2.52}{1.563} \cong 0.345$$

olarak hesaplanan t-değeri [(7-2) +(8-2)]=11 serbestlik dereceli t-dağılımı gösterir. Araştırıcı I. tip hata olasılığını α=0.01 olarak belirlemiş ise tek taraflı hipotez kontrolü yapıldığı için Tablo C'den, 11 serbestlik dereceli t-dağılımında %1'lik alanın, yani kontrol hipotezinin ret bölgesinin 2.718 değerinden başladığı görülür. Hesaplanan test istatistiğinin değeri 0.345 olup kritik bölgenin başladığı t-değerinden çok küçüktür ve H_0 hipotezinin kabul bölgesine düşmektedir. Bu sebeple kontrol hipotezi kabul edilir.

Tablo 8.16'da uygulanan hipotez kontrolü sonunda verilen karar doğrultusunda 2 ay süt verilen kuzularda doğum ağırlığındaki 1 kg'lık artışa karşılık 3.ay canlı ağırlığında ortalama olarak gözlenen artışın, 3 ay süt verilen kuzularda gözlenen artıştan daha fazla olduğu söylenemez. İki grup arasında 1 kg'lık artışa karşılık 3.ay canlı ağırlıkta ortalama olarak gözlenen artış bakımından farkın tesadüften ileri geldiği ve sıfır kabul edilebileceği kararına varılır.

8.3. SORULAR

1. t-dağılımı nasıl elde edilir açıklayınız?
2. t-dağılımının özellikleri nelerdir? Açıklayınız.
3. t-dağılımının varyansı neye göre değişir ve nasıl hesaplanır?
4. 10 serbestlik dereceli teorik t-dağılımı, deneysel olarak nasıl elde edilir? Açıklayınız.
5. Z-dağılımı ile t-dağılımı arasında ne gibi farklar vardır? Açıklayınız.
6. Uykusuzluktan şikâyetçi 20 hasta tesadüfen iki eşit gruba ayrılmış, bunlardan birine A uyku ilacı, diğerine de B uyku ilacı verilmiştir. Bu uygulamadan sonra söz konusu hasta gruplarının uyku süresi (saat) ortalamaları ve standart hataları sırası ile $\bar{A} \pm S_{\bar{A}}$ = 3.5±0.70, $\bar{B} \pm S_{\bar{B}}$ = 4.5±0.90 olarak bulunmuştur. Buna göre B ilacının uykusuzluğa karşı daha etkili olduğu söylenebilir mi?
7. R ve S gibi iki farklı ilacın sistolik kan basıncına etkisini araştırmak üzere seçilen 12 kişi 7 ve 5 kişilik iki gruba ayrılmış, bir gruba R, diğer gruba S ilacı verilmiştir. Sonuçlar aşağıdaki gibi bulunduğuna göre R ve S ilaçları arasında sistolik kan basıncına etki bakımından fark var mıdır?

R :	110	120	125	130	120	120	115
S :	125	130	125	130	135		

4. Tesadüfen seçilen 125 bireyde yaş ile kan basıncı arasında korelasyon katsayısı 0.81 olarak bulunmuştur. İki özellik arasında doğrusal bir ilişkinin olduğu söylenebilir mi?
5. Belirli bir ırka ait 18 adet fare, biri kontrol grubu (A), diğeri de deney grubu (B) olmak üzere eşit sayıda birey içeren rasgele iki gruba ayrılmış ve doğumdan 3 aylık oluncaya kadar deney grubu özel bir gıda rejimi ile beslenmiştir. Belirlenen süre sonunda grupların canlı ağırlık ortalamaları ve standart hataları sırasıyla $\overline{A} \pm S_{\overline{A}}$ = 45.6 ± 3.8 ve $\overline{B} \pm S_{\overline{B}}$ = 40.75 ± 2.1 olarak bulunmuştur. Uygulanan rejimin ortalama canlı ağırlığı değiştirip değiştirmediğini kontrol ediniz.
6. Birbirinden bağımsız iki grup (muamele; uygulama) ortalamasını karşılaştırmak için gerekli olan test istatistiği nasıl hesaplanır? Açıklayınız.
7. Birbirine bağımlı gözlemlerden elde edilmiş iki grubu karşılaştırmak için gerekli olan t-istatistiğinin Serbestlik Derecesinin 20 olabilmesi için örnekteki deney ünitesi adedi kaç olmalıdır? Nedenini belirterek açıklayınız.
8. Ortalaması ve varyansı bilinmeyen bir populasyondan rasgele alınan 25 bireylik bir örnekte, örnek ortalaması $\overline{X}$ = 12. ve standart sapması da S_x= 5. olarak hesaplanmıştır. Bu örnekte α = 0.05 yanılma olasılığı ile $H_0 : \mu = 10$ kontrol hipotezi, $H_1 : \mu < 10$ hipotezine karşı test edildiği zaman verilecek karar nedir? Gerekli kontrolleri yaparak açıklayınız.
9. 8 bireyin H ve P gibi iki özelliği arasında hesaplanan korelasyon katsayısı 0.6 olarak hesaplanmıştır. Bu iki özellik arasında doğrusal bir ilişki olmadığı söylenebilir mi? Gerekli kontrolleri yaparak karar veriniz.
10. 20 Adet mandalina ağacı tesadüfen A ve B gibi 2 eşit gruba bölünmüş, gruplardan biri kontrol grubu (A Grubu) olarak ayrılmış, diğer gruptaki (B Grubundaki) ağaçların gövdesine belirli bir eriyik sürülmüştür. Deneme sonunda kontrol grubunun ortalaması $\overline{A}$ = 30.0 kg/ağaç, standart sapması S_A=2.0, eriyik sürülen grubun ortalaması $\overline{B}$ = 32.0 kg/ağaç, standart sapması da S_B=3.0 olarak bulunmuştur. Eriyik sürme mandalina verimini etkilemiş midir?
11. Bir örnekten hesaplanan ortalamanın bilinen bir ortalamayla karşılaştırılması için gerekli olan t-istatistiği nasıl hesaplanır? Örnek vererek açıklayınız.
12. Eş-yapma t-testi hangi durumlarda kullanılmalıdır? Örnek vererek açıklayınız.
13. A firmasının ürettiği vitamin drajelerinden rasgele alınan 16 vitamin drajesinde B_1 vitamini miktarı ortalaması 20. mg ve standart sapması 0.2 mg, B firmasının ürettiği vitamin drajelerinden alınan 16 vitamin drajesinde B_1 vitamini miktarı ortalaması 20.5 mg ve standart sapması da 0.3 mg olarak bulunmuştur. Vitamin drajelerindeki B_1 vitamini miktarı bakımından A ve B firmaları arasındaki farklılığın tesadüfi olduğu söylenebilir mi?
16. Birbirine bağımlı gözlemlerden elde edilmiş iki ortalamayı karşılaştırmak için gerekli olan t-istatistiğinin Serbestlik Derecesinin 43 olabilmesi için örnekteki deney ünitesi sayısı kaç olmalıdır? Açıklayınız.
15. 5 sağlıklı bireyde sabah ve akşam ölçülen kan şekerleri miktarı aşağıdaki gibi bulunmuştur. Sabah ve akşam ölçülen kan şekerleri miktarı arasındaki farkın istatistik olarak önemli olduğunu söyleyebilir misiniz?

Sabah	Akşam
90	91
95	95
98	97
94	95
94	95

16. İki farklı karışım kullanılarak yapılan dondurmaların % yağ miktarlar aşağıdaki gibi bulunmuştur. Karışımlar arasında ortalama % yağ oranları bakımından farkın tesadüfî olup olmadığını kontrol ediniz.

Karışım1	5.7	4.5	6.2	6.3
Karışım2	6.3	5.7	5.9	6.4

17. Belirli bir hibrit mısırda koçan uzunlukları ortalamasının μ = 23 cm. olduğu bilinmektedir. Söz konusu hibrit mısırın ekili olduğu parsellere, yeni bir gübre karışımı uygulanmış ve 25 parselde ortalama koçan uzunluğu, $\overline{X}$ = 28. cm, standart sapma da S_x = 5. cm olarak bulunmuştur. Yeni gübre karışımının koçan uzunluğunu artırdığı söylenebilir mi? Kontrol ederek karar veriniz.
18. Yoğurtlarda maya gelişiminin zamana göre değişip değişmediğini araştırmak amacıyla, 20 adet yoğurt kâsesinin bulunduğu bir inkübasyon dolabından, mayalama işleminden 1 saat ve 7 saat sonra maya sayımı yapılmış ve elde edilen sayısal değerler kaydedilmiştir. 1 saat ve 7 saat sonraki maya sayımları arsasında farkların istatistik olarak önemli olup olmadığını test etmek için kullanılması gereken yöntem nedir? Açıklayınız.
19. Bir bölgedeki 205 adet parselde mercimek ekiminden önce ve mercimek ekiminden sonra topraktaki azot miktarları tespit edilmiştir. Bu parsellerdeki mercimek ekiminin topraktaki azot miktarını değiştirip değiştirmediği araştırılmak istense elde edilen sayısal değerler yardımı ile hangi hipotez testi yöntemi kullanılmalıdır? Açıklayınız.
20. 6 bireyin X ve Y gibi iki özelliği arasında hesaplanan korelasyon katsayısı 0.6 olarak hesaplanmıştır. Bu iki özellik arasında doğrusal bir ilişki olmadığı söylenebilir mi?
21. A ve B firmaları tarafından üretilen anti bakteriyel el temizleme jellerinde pH değerleri ölçülmüş ve aşağıdaki istatistikler hesaplanmıştır. A ve B firmalarına ait jellerdeki pH değerlerinin ortalamaları arasındaki farkın önemli olup olmadığına karar veriniz.

	n	Ortalama	Standart Hata
A-firması	10	5.51	0.03
B-firması	10	5.61	0.02

22. Astım hastalarına yeni bir tedavi uygulanmış ve tedavinin etkinliğini görmek amacıyla tedavi öncesinde ve tedavi sonrasında, saniyedeki baskılı ekspiratuar hacmi (FEV_1) değerleri ölçülmüştür. Buna göre yeni tedavi FEV_1 ölçümlerinin yükselmesini sağlamış mıdır?

Hasta	Tedavi Öncesi	Tedavi Sonrası
1	1.69	1.69
2	2.77	2.22
3	1.00	3.07
4	1.66	3.35
5	3.00	3.00
6	0.85	2.74

23. 10 bireyden elde edilen veriler kullanılarak X ve Y özellikleri arasındaki korelasyon katsayısı -0.899 ve regresyon denklemi $\hat{Y}=8.68-0.109X$ olarak hesaplanmıştır. X ve Y özellikleri arasında doğrusal ilişki bulunduğu söylenebilir mi? Kontrol ederek karar veriniz.
24. Belirli bir özellik bakımından A ve B gruplarında ortalama, varyans ve örnek genişliği aşağıdaki gibidir. A ve B gruplarının ortalamaları arasındaki fark tesadüften mi ileri gelmektedir?

$\overline{A}=1.5$	$S_A^2=2.5$	$n_A=25$
$\overline{B}=16.5$	$S_B^2=5.29$	$n_B=36$

25. 10 bireyde boy uzunlukları ile tüketilen süt miktarı arasındaki korelasyon katsayısı 0.56 olarak bulunmuştur. Buna göre bu 10 bireyin alındığı populasyonda, bu iki özellik bakımından doğrusal bir ilişki var mıdır?
26. 81 bireyde sağ ve sol kulak ile algılanabilen ses şiddetleri arasındaki farkların ortalaması 0.6 farkların standart sapması 0.27 olarak bulunmuştur. Sağ ve sol kulak ile algılanabilen ses şiddetleri arasındaki fark tesadüften mi ileri gelmektedir?
27. Türkiye'deki hastanelerinin mikrobiyoloji laboratuvarlarında yapılan testlerdeki hata oranının 0.10 olduğu bilinmektedir. Yeni açılan bir hastanenin mikrobiyoloji laboratuvarında yapılan ölçümlerden tesadüfi olarak seçilen 100 sonuçtan 13 tanesinin hatalı olduğu gözlemiştir. Yeni açılan hastanenin laboratuvar sonuçlarındaki hata oranını populasyondan daha fazla mıdır?
28. 5 sağlıklı bireyde sabah ve akşam ölçülen kan şekerleri miktarları aşağıdaki gibi bulunmuştur.

Birey	Sabah	Akşam
1	90	91
2	95	95
3	98	97
4	94	95
5	94	95

Sabah ve akşam ölçülen kan şekeri miktarları arasında istatistik olarak önemli bir farkın olduğunu söyleyebilir misiniz?

29. Bir diyet programının zayıflama üzerine etkisini araştırmak amacıyla tesadüfen seçilen 6 bireyin, başlangıç ve 2 aylık diyet programı sonrası ağırlıkları aşağıdaki gibi ölçülmüştür. Uygulanan bu diyet programının ağırlıkta önemli bir azalmaya sebep olduğu söylenebilir mi?

Başlangıç	72	70	65	80	75	70
Diyet Sonrası	70	68	62	70	74	69

30. 6 adet vitamin drajesinin üretimden hemen sonra ve depoda belirli bir süre bekletildikten sonraki pH değerleri aşağıdaki gibi tespit edilmiştir. Depolamadan sonra pH değerlerinde istatistik olarak önemli bir değişim olmuş mudur?

Başlangıç :	6.45	7.02	7.15	7.12	7.05	7.3
Depolamadan sonra:	5.04	5.08	6.02	5.80	5.75	6.0

31. 8 serbestlik dereceli teorik t-dağılımı, deneysel olarak nasıl elde edilir? Bu dağılımın ortalaması ve varyansı kaça eşittir? Bu dağılımı oluşturan t-değerleri %90 olasılıkla hangi değerler arasındadır?
32. Aynı yaş ve cinsiyetteki bebeklerden tesadüfen seçilen 5'ine A maması 7'sine B maması verilmiş ve bebeklerin 6. ayın sonundaki ağırlıkları (kg) aşağıdaki gibi bulunmuştur. Mamalar arasında 6. ayın sonundaki ağırlığı etkileme bakımından istatistik olarak önemli bir fark var mıdır?

A (kg)	8.0	9.5	9.2	8.6	8.7		
B (kg)	7.6	8.2	9.3	8.5	7.9	9.6	9.0

33. 90 adet bıldırcında yumurtadan çıkış ağırlığı ile 2. Hafta canlı ağırlığı arasındaki korelasyon katsayısı r=0.25 olarak bulunmuştur. Bıldırcınlarda adı geçen özellikler arasında hesaplanan korelasyon katsayısı istatistik olarak önemli midir?