ECZ158 BİYOİSTATİSTİK DERSİ

10. HAFTA DERS NOTLARI

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

Z DAĞILIMI VE Z-KONTROLLERİ

7.1. Z-Dağılımı

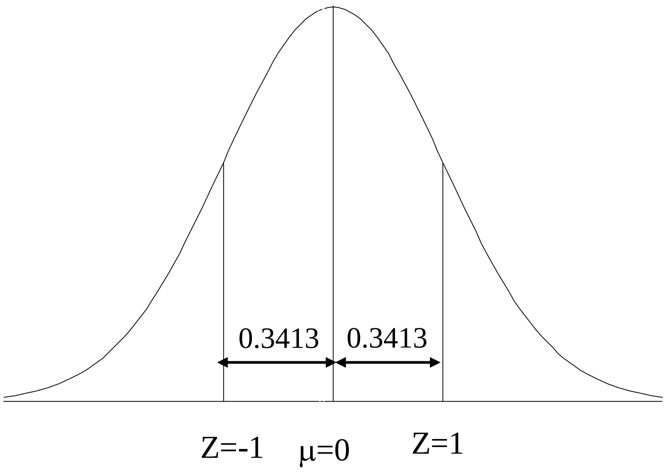
Bölüm 3.3.2'de belirtildiği gibi bütün normal dağılımlar, Z-dağılımına (standart normal dağılıma) dönüştürülebilir. Standart normal dağılımın ihtimal yoğunluk fonksiyonu;

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

şeklindedir. Bu dağılımın ortalaması, $\mu=0$ ve varyansı, $\sigma^2=1$'dir. Z-dağılımı Şekil 7.1'de verildiği gibi çan eğrisi şeklindedir. Dağılım ortalama etrafında simetriktir, yani;

$P(-Z_0<Z<0)=P(0<Z<Z_0)$'dır.

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



ŞEKİL 7.1. Standart normal dağılım

7.2. Z-Kontrolleri

Standart normal dağılımda değişkenin Z olduğu ve $(X-\mu_x)/\sigma_x$ olarak hesaplandığı Bölüm 3.3.2'de açıklanmıştı. Bundan yararlanarak populasyondan alınan bir gözlemin $(X_1<X<X_2)$ aralığında bulunması ihtimalinin hesaplanabileceği vurgulanmıştı. Aynı şekilde Bölüm 4.1'de verilen Örnek 6'da ortalaması, $\mu=60$ ve standart sapması, $\sigma=6.3$ olan normal dağılmış populasyondan çekilen ve 9 birey içeren (n=9) örneğin ortalamasının $\overline{X}=56.4$ veya bundan daha düşük değerlerin, ortalamaya ait örnekleme dağılımına dahil olma ihtimalinin, bu ortalamaya karşılık gelen Z-değeri ($Z=\frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{X}}}$) hesaplanarak bulunabileceği gösterilmişti. Bu yapılan aslında bir hipotez kontrolüdür.

Populasyon varyansının bilindiği durumlarda Z-dağılımı kullanılarak hipotez kontrolleri yapılır.

Bu bölümde çeşitli istatistikler için Z-kontrolleri, Bölüm 6'da açıklanan hipotez kontrollerinin adımları, karşıt hipotez, karşıt hipoteze göre tek ve çift taraflı kontroller ve I. tip hata gösterilerek açıklanacaktır.

7.2.1. Ortalamaya ait Hipotez Kontrolü

Ortalaması ve varyansı bilinen bir populasyondan tamamen tesadüfen alındığı ileri sürülen bir örnekten hesaplanan ortalama ile populasyon ortalaması arasındaki farkın tesadüften ileri gelip gelmediği, yani söz konusu örneğin ele alınan populasyondan rastgele çekilen örneklerden biri olup olmadığına karar verilir. Bu karar verildiğinde I. ve II. tip hatalar en çok α ve β kadardır.

Örneğin Bölüm VI'da bir ilaç fabrikasındaki bir yönetici üretilen vitamin-mineral karması drajelerindeki B_1 vitamininin ortalaması μ_x=25 mg ve standart sapması σ_x=1.2 mg olan bir normal dağılım gösterdiğini biliyor olsun. Bu yönetici gerçekten üretilen drajelerdeki ortalama B_1 vitamininin 25 mg olup olmadığını araştırmak üzere farklı partilerden 16 drajelik bir örnek alarak ortalama B_1 vitamini miktarını 24.5 mg olarak hesaplamış olsun. Burada yöneticinin yapması gereken gerçekten ortalama B_1 vitamini miktarının 25 mg olarak kabul edilip edilemeyeceğini kontrol etmektir.

Hipotez kontrolü yapılırken izlenecek adımlar sırası ile aşağıdaki gibidir:

1. ADIM: İlk olarak araştırıcı kontrol ve karşıt hipotezlerini oluşturmalıdır. Kurulan hipotezlerin bir anlamı olmalı ve kontrol edilebilmelidir. Kurulacak olan karşıt hipotez yapılacak hipotez kontrolünün tek taraflı veya çift taraflı olduğunu belirler.
Kontrol hipotezi:

H_0: Örnek ortalaması ile populasyon ortalaması arasındaki fark tesadüften ileri gelmektedir, söz konusu fark sıfır kabul edilebilir.
Kaşıt hipotez araştırıcının amacı doğrultusunda 3 farklı şekilde oluşturulabilir.

a. H_1: Populasyon ortalaması ile örnek ortalaması arasındaki fark tesadüften ileri gelmemektedir, söz konusu fark sıfır kabul edilemez. Başka bir deyişle, örnek söz konusu populasyondan tesadüfen alınmış bir örnek değildir.

Bu şekilde kurulacak karşıt hipotez yapılacak hipotez kontrolünün çift taraflı olduğunu gösterir. Çünkü araştırıcı örnekten hesapladığı ortalamanın söz konusu populasyondan çekilen örneklerden elde edilecek ortalamaya ait örnekleme dağılımına dahil olup olmadığı ile ilgilenmektedir.

b. H_1: Üzerinde çalışılan örnek, ortalaması söz konusu populasyonun ortalamasından daha küçük olan bir populasyondan tesadüfen alınmış bir örnektir.

Bu şekilde kurulacak karşıt hipotez yapılacak hipotez kontrolünün tek taraflı olduğunu gösterir (soldan test). Çünkü bu durumda araştırıcı örnekten hesaplanan istatistiğin ortalamaya ait örnekleme dağılımının ortalamasından küçük tarafa sapması ile ilgilenmektedir.

c. H_1: Üzerinde çalışılan örnek, ortalaması söz konusu populasyonun ortalamasından daha büyük olan bir populasyondan tesadüfen alınmış bir örnektir.

Bu karşıt hipotez de yapılacak hipotez kontrolünün tek taraflı olduğunu gösterir (sağdan test). Çünkü bu durumda da araştırıcı örnekten hesaplanan istatistiğin ortalamaya ait örnekleme dağılımının ortalamasından büyük tarafa sapması ile ilgilenmektedir.

2. ADIM: Örneğin çekildiği varsayılan populasyonun dağılımı belirlenir. Bu populasyon normal dağılıyor olabilir veya bilinen diğer dağılımlardan biri olabilir. Daha sonra üzerinde durulan örnekte ele alınan istatistiğin dağılımı belirlenir. Mesela ele alınan değişken normal dağılıyor olsun. Bu dağılımın ortalamasının μ_0, varyansının da σ^2 olduğu varsayılsın. Herhangi bir araştırma sonucu dokuz bireyden (n=9) elde edilen aritmetik ortalama $\overline{X}$ olsun. Belirlenen test hipotezi bu örneğin yukarıdaki populasyondan çekildiği şeklinde (H_0: $\mu_{\bar{x}} = \mu_0$) olabilir. Test hipotezinin geçerli olduğu durumda, ortalaması μ_0 ve varyansı, σ_x^2 olan normal dağılımdan çekilmiş örnek ortalamasına ait dağılımın ortalaması $\mu_{\bar{x}} = \mu_0$ ve standart sapması $\sigma_{\bar{x}} = \dfrac{\sigma_x}{\sqrt{9}}$ olan normal dağılım olacağı daha önce 4.1'de açıklanmıştı. Böylece örnek ortalamasının örnekleme dağılımı belirlenmiş olmaktadır.

3. ADIM: Bu adımda ise test birimi ve dağılımı belirlenir. Yani ele alınan istatistik test hipotezinin varlığı durumunda göstereceği birim ve dağılım belirlenir. Yukarıdaki örnekte ele alınan istatistik olan aritmetik ortalama ($\overline{X}$) standardize edildiğinde, yani test birimi olarak Z ele alındığında;

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

değeri hesaplandığında elde edilen sonuç Z'dir ve standart normal dağılım gösterir. Bu durumda test dağılımı standart normal dağılımdır.

4. ADIM: Bu adımda I. tip hata olasılığı (α) belirlenir. Karşıt hipotez de dikkate alınarak test dağılımı kontrol edilen hipotez için kabul ve ret bölgelerine ayrılır. Ret bölgesinin olasılığı (α) kadardır. Karşıt hipotez tek taraflı ise test dağılımı bir taraftan kabul ve ret bölgelerine ayrılır (Şekil 6.2). Karşıt hipotez çift taraflı ise test dağılımı iki ucundan kabul ve ret bölgelerine ayrılır. Her iki taraftaki ret bölgelerinin olasılığı $\frac{\alpha}{2}$'dir (Şekil 6.1).

5. ADIM: Veriler toplanarak test istatistiği hesaplanır. Bu test istatistiğinin bölgelerden hangisine dahil olduğuna bakılarak H_0'ın reddedilip edilemeyeceğine karar verilir.

ÖRNEK 1.

Bir ilaç fabrikasındaki bir yönetici üretilen vitamin-mineral karması drajelerdeki B_1 vitamininin ortalaması μ_0=25 mg ve standart sapması σ=1.2 mg olan bir normal dağılım gösterdiğini bilmektedir. Bu yönetici gerçekten üretilen drajelerdeki ortalama B_1 vitamininin 25 mg olup olmadığını araştırmak üzere farklı partilerden rastgele alınan 16 drajelik bir örnekte B_1 vitamini miktarının ortalamasını 24.5 mg olarak bulmuş olsun. Bu örneğin, ortalaması μ_0=25 mg ve standart sapması σ=1.2 mg olan normal dağılmış bir populasyondan çekilen örneklerin oluşturduğu dağılıma dahil olup olmadığına aşağıda adım adım açıklanan hipotez kontrolü yapılarak karar verilir.

1. ADIM: Yapılacak hipotez kontrolünün çift taraflı mı yoksa tek taraflı mı olacağı sorudan anlaşılır. Buradaki örnekte yöneticinin ilgilendiği örnekten hesaplanan ortalamanın 25 mg kabul edilip edilemeyeceğidir. Yani ortalamaya ait örnekleme dağılımına dahil olup olmadığıdır. Bu nedenle de çift taraflı hipotez kontrolü yapılacaktır.

H0: Örnek ortalaması ile populasyon ortalaması arasındaki fark tesadüften ileri gelmektedir, söz konusu fark sıfır kabul edilebilir. Yani $\mu_0 = 25\,\text{mg}$'dır.

H1: Populasyon ortalaması ile örnek ortalaması arasındaki fark tesadüften ileri gelmemektedir, söz konusu fark sıfır kabul edilemez. Başka bir deyişle, örnek söz konusu populasyondan tesadüfen alınmış bir örnek değildir ve söz konusu populasyonu temsil etmez. Yani, $\mu_0 \neq 25$ mg'dır.

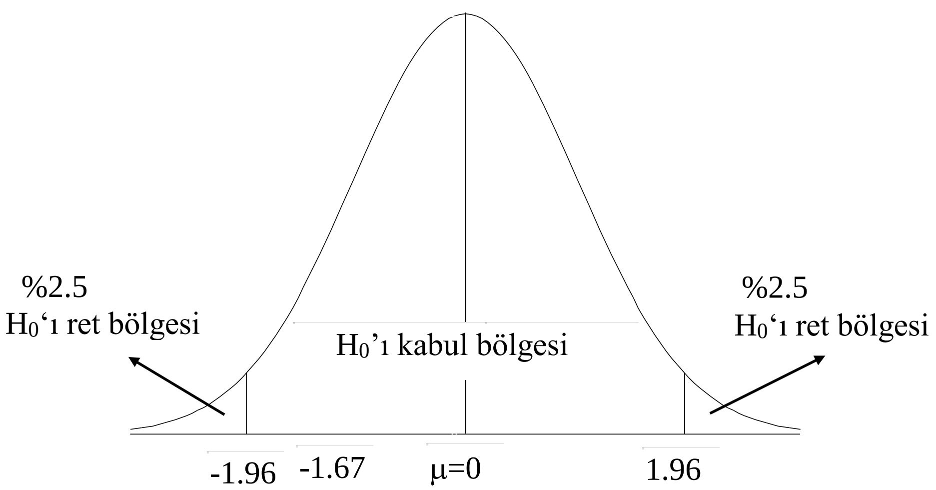
2. ADIM: Drajelerdeki B_1 vitamini miktarının normal dağıldığı bilinmektedir. Buna göre bu dağılımın ortalaması $\mu_0 = 25\,\text{mg}$ ve standart sapması σ=1.2 mg'dır. Bu dağılımdan çekilen 16'lık örneklerin ortalamasının dağılımı da normaldir ve parametreleri $\mu_{\bar{x}} = 25\,\text{mg}$ ve $\sigma_{\bar{x}} = \frac{1.2}{\sqrt{16}} = 0.3\,\text{mg}$'dır. Yani hipotez ile belirtilen dağılım ortalamaya ait örnekleme dağılımıdır.

3. ADIM: Örnekte ele alınan istatistik ($\bar{X}$) 2. adımdaki şartlar dikkate alınarak standardize edildiği zaman Z dağılımı gösterir.

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{X}}} = \frac{24.5 - 25.0}{0.3} \cong -1.67$$

Buna göre test dağılımı standart normal dağılımdır.

4. ADIM: Araştırıcı genellikle bu tip hipotez kontrollerinde yapıldığı gibi I. tip hata olasılığını, α=0.05 olarak kabul etmiş olsun. Karşıt hipoteze göre test iki taraflıdır. Buna göre test dağılımı Şekil 7.1'deki gibi kabul ve ret bölgelerine ayrılır.



ŞEKİL 7.1. Çift taraflı hipotez kontrolünde kontrol hipotezini ret ve kabul bölgeleri

5. ADIM: Hesaplanan Z=-1.67'lik test istatistiği kabul bölgesinde olduğundan, kontrol hipotezi reddedilemez. Ele alınan 16'lık örnekten hesaplanan $\overline{X}$=24.5 ve daha fazla sapan değerlerin test hipotezi ile belirlenen populasyona dahil olma olasılığı 0.05'ten daha fazladır. Bu durum kısaca P>0.05 olarak ifade edilir.

ÖRNEK 2.

Bir meyve suyu fabrikasında üretilen portakal sularında C vitamini miktarının μ_0=17 mg/lt ve σ_x^2=4.5 olan bir normal dağılım gösterdiği bilinmektedir. Yeni bir meyve suyu üretim tekniğinin portakal sularındaki C vitaminini artırdığını ileri sürülmektedir. Bu nedenle yeni teknikle üretilen meyve sularından, farklı partilerden tesadüfen alınan 25 portakal suyu şişesinde C vitamini incelenecektir. Yeni tekniğin portakal sularında C vitamini miktarını artırıp artırmadığını test etmek için Örnek 1'deki gibi aşağıdaki adımlar izlenir.

1. ADIM: Hipotezlerin kurulması:

H_0: Örnek ortalaması ile populasyon ortalaması arasındaki fark tesadüften ileri gelmektedir, söz konusu fark sıfır kabul edilebilir. Yani $\mu_0 = 17$ mg/lt 'dir.

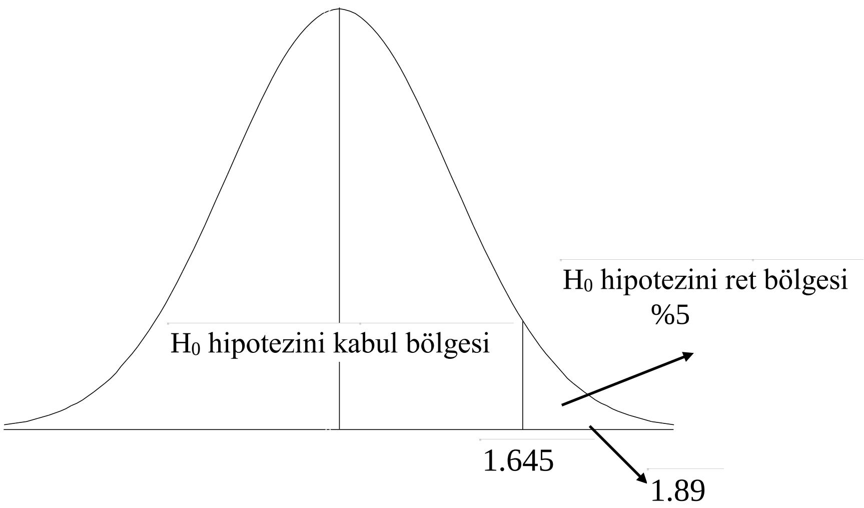
H_1: Populasyon ortalaması ile örnek ortalaması arasındaki fark tesadüften ileri gelmemektedir, söz konusu yeni teknik portakal sularındaki C vitamini miktarını artırmıştır. Yani, $\mu_0 > 17$ mg/lt 'dir.

Burada yapılacak hipotez kontrolünün tek taraflı bir hipotez kontrolüdür. Çünkü bu durumda da araştırıcı örnekten hesaplanan istatistiğin ortalamaya ait örnekleme dağılımının ortalamasından büyük tarafa sapması ile ilgilenmektedir.

2. ADIM: Daha önceki analiz sonuçlarından üretilen portakal suyundaki C vitamini miktarlarının normal dağıldığı bilinmektedir. Bu dağılımın parametreleri de problemde verilmiştir. Test hipotezine göre bu populasyondan alınan 25'lik örnek ortalamaları da ortalaması 17 mg/lt, standart sapması $\sigma_{\bar{x}} = \sqrt{\dfrac{4.5}{25}} \cong 0.424$ mg/lt olan bir normal dağılım gösterir.

3. ADIM olarak ele alınan $\overline{X}$ istatistiği standardize edildiğinde, yani $\dfrac{\overline{X} - \mu_{\bar{x}}}{\sigma_{\overline{X}}}$ hesaplandığında standart normal dağılan Z değeri elde edilir. Buna göre test birimi Z, test dağılımı da standart normal dağılımdır.

4. ADIM: Araştırıcı I. tip hata ihtimalini, α=0.05 olarak belirlemiş olsun. Yapılan hipotez kontrolü tek taraflı hipotez kontrolü olduğu için, Şekil 7.2'de gösterildiği gibi, bu ihtimali ortalamadan büyük değerlerin bulunduğu tarafta alır.



ŞEKİL 7.2. Tek taraflı hipotez kontrolünde kontrol hipotezini ret ve kabul bölgeleri

5. ADIM: Araştırıcı denemesini yürüterek 25 şişelik örneğinde C-vitamini ortalamasını 17.8 mg/lt olarak bulmuş olsun. Buna göre;

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{17.8 - 17.0}{0.424} \cong 1.887$$

olarak bulunur. Bu Z-değerine göre örnekten hesaplanan ortalama ve daha büyüklerin örnekleme dağılımına dahil olma ihtimali 0.5-P(0<Z<1.887)'dir. Tablo A'dan, Z değerlerinin 0 ile yaklaşık 1.89 arasında olma ihtimali 0.4706 olarak bulunur. Ve bu durumda örnekten hesaplanan ve daha büyük ortalamaların örnekleme dağılımına dahil olma ihtimali, 0.5-P(0<Z<1.89)= 0.5-0.4706=0.0294'tür. Bu ihtimal I. tip hata ihtimalinden (%5'ten) daha küçük olduğu için kontrol hipotezi ret edilerek karşıt hipotez kabul edilir, yani yeni üretim tekniğinin portakal sularındaki C vitamini miktarını artırdığı söylenebilir.

Hangi hipotezin kabul edileceğine şu şekilde de karar verilebilir: Tablo A'da 0 ile 1.64 değeri arasında kalan alan %44.95, 0 ile 1.65 arasında kalan alan ise %45.05'dir. Burada bulunması gereken Z-değeri ise %45.0'lik alanı ayıran Z-değeridir. Bunun için interpolasyon yapılır. %44.95 ile %45.05'in tam orta noktası (yani yarısı) %45.0'e karşılık gelmektedir. Aynı şekilde 1.64 ile 1.65'in de orta noktası bulunur ise elde edilecek Z-değeri 1.645'tir. Bu değer Z-dağılımında %5'lik alanın başladığı noktadır. Hesaplanan test istatistiğinin (Z'nin) değeri ise 1.887'dir. Bu değer Şekil 7.2'de de görüldüğü gibi kontrol hipotezini ret bölgesine düşmektedir. Yani kontrol hipotezi reddedilir ve karşıt hipotez kabul edilir. Bu durumda verilen karar; yeni üretim tekniğinin portakal sularındaki C vitamini miktarını artırdığıdır. Bu durum P<0.05 olarak ifade edilir.

Eğer araştırıcı hipotez kontrolü yaparken I. tip hata ihtimalinin α=0.01 olmasına karar vermiş ise bu durumda Tablo A'da 0 ile hangi Z-değerinin arasında kalan alanının %49 olduğunu bulması gerekir. Tablo A'ya bakıldığı zaman 0 ile 2.33 arasında kalan alanın %49.01 olduğu görülür. Araştırıcı 2.33'ü %1'lik alanı ayıran Z-değeri olarak kullanabilir veya interpolasyon yaparak tam %1'lik alanı ayıran Z-değerini hesaplayabilir.

Bu örnekte eğer I. tip hata ihtimali %1 olarak kararlaştırılmış olsa idi, 5.adımda örnekten hesaplanan ortalamanın örnekleme dağılımına dahil olma ihtimali %2.94 olarak hesaplanmıştı. Bu ihtimale göre kontrol hipotezi kabul edilecekti yani yeni üretim tekniğinin C-vitaminini artırmadığı kararı verilecekti. Ayrıca %1'lik alanının başladığı Z-değeri 2.33 olduğundan 1.89 değeri kabul bölgesine düşmektedir. Bu da H_0 hipotezinin reddedilmemesini gerektirir.

ÖRNEK 3:

Bir fakültede öğrencilerin biyoistatistik dersinden aldıkları notların ortalaması 60 ve standart sapması 8 olan normal bir dağılım gösterdiği bilinmektedir. Fakülteye ders başlamadan 30 dakika önce gelen öğrencilerin daha ilgili ve başarılı öğrenciler olduğu öne sürülmektedir. Bu sebeple, ders başlamadan 30 dakika önce gelen öğrencilerden tesadüfen seçilen 16 öğrenci içeren bir örnekte not ortalaması incelenmiştir. Erken gelen öğrencilerin başarılı öğrenciler olup olmadığı aşağıdaki şekilde incelenir:

1.ADIM: Hipotezlerin kurulması:

H_0: Örnekten hesaplanan not ortalaması ile fakülte öğrencilerinin not ortalaması arasındaki fark tesadüften ileri gelmektedir, söz konusu fark sıfır kabul edilebilir. Başka bir deyişle örneği oluşturan öğrenciler söz konusu fakültenin öğrencileridir. Yani $\mu_0 = 60$ 'tır.

H_1: Örnekten hesaplanan not ortalaması ile fakülte öğrencilerinin not ortalaması arasındaki fark tesadüften ileri gelmemektedir, söz konusu fark sıfır kabul edilemez. Başka bir deyişle, erken gelen öğrencilerin biyoistatistik dersi not ortalaması daha yüksektir. Yani $\mu_0 > 60$ 'tır.

2. ADIM: Örneğin çekildiği varsayılan populasyonun normal dağıldığı ve parametreleri problemde verilmiştir. Buradan rastgele alınan 16 öğrenci içeren örneklerin ortalamaları da normal dağılır ve parametreleri aşağıdaki gibidir:

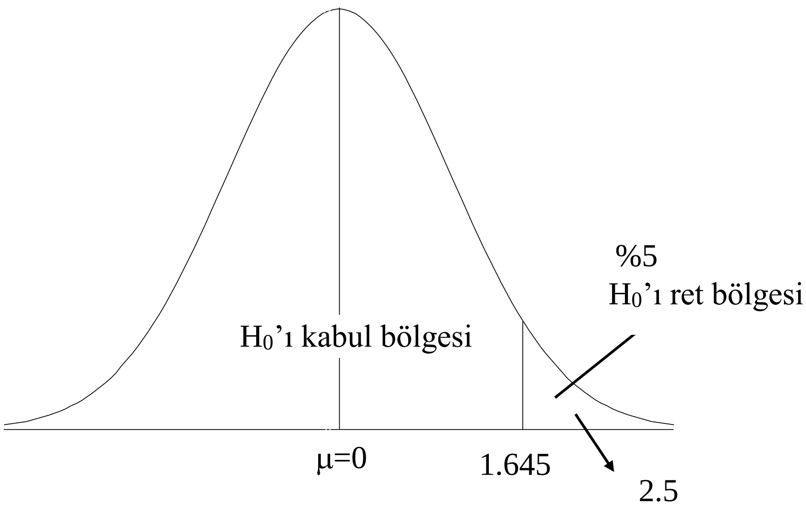
$$\mu_{\overline{X}} = 60$$

$$\sigma_{\overline{X}} = \frac{8}{\sqrt{16}} = 2$$

3. ADIM: Örnekten hesaplanan ortalamaya karşılık gelen Z-istatistiği;

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}}$$

4. ADIM: Araştırıcı I. tip hata ihtimalini, α=0.05 olarak belirlemiş olsun. Yapılan hipotez kontrolü tek taraflı kontrol olduğu için, Şekil 7.3'de gösterildiği gibi, bu ihtimal ortalamadan büyük değerlerin bulunduğu tarafta alınır.



ŞEKİL 7.3. Tek taraflı hipotez kontrolünde kontrol hipotezini ret ve kabul bölgeleri

5. ADIM: Erken gelen öğrencilerden oluşturulan 16 öğrencilik örnekte biyoistatistik not ortalaması 65 olarak hesaplanmış olsun. Buna göre;

$$Z = \frac{65 - 60}{2} = 2.5$$

olarak bulunur. Bulunan bu Z-değerine göre örnekten hesaplanan ortalama ve daha fazla sapan değerlerin örnekleme dağılımına dahil olma ihtimali 0.5-P(0<Z<2.5)'dir. Tablo A'dan, P(0<Z<2.5) = 0.4938 olarak bulunur. Ve 0.5-P(0<Z<2.5)= 0.5-0.4938=0.0062'dir. Bu ihtimal, I. tip hata ihtimalinden (%5'ten) daha küçük olduğu için kontrol hipotezi reddedilir, yani örnek ortalaması ile populasyon ortalaması arasındaki 5 notluk fark sıfır olarak kabul edilemez. Derslere erken gelen öğrenciler daha çalışkan olduklarından not ortalamaları yüksektir. Örneğe giren 16 öğrenci çalışkan olanları kapsamaktadır.

7.2.2. Ortalamalar arası Farka ait Hipotez Kontrolü

Ortalamalar arası farkların hipotez kontrolünde izlenecek yol daha önce etraflı olarak adım adım açıklanan genel hipotez kontrolünde olduğu gibidir. Bu durum aşağıda bir örnek ile açıklanmıştır.

ÖRNEK 1:

Vitamin ve mineraller üreten bir ilaç fabrikasında uzun yıllar yapılan analizler sonucuna göre drajelerdeki B_1 vitamini içeriğine ait varyansın 4 olduğu bilinmektedir. Aynı miktar B_1 vitamini içeren, A ve B tipi kaplama şekli uygulanan ambalajlardan bir yıl sonra örnekler alınıp ortalamaları bulunmuştur. A ve B tipi kaplama uygulanan drajelerdeki B_1 vitamini miktarlarının bir yıl sonra farklı olup olmadığı kontrol edilmek isteniyor olsun. Yapılacak hipotez kontrolünde izlenecek yol aşağıdaki gibidir:

1. H0: A ve B tipi kaplama uygulanan drajelerdeki bir yıl sonundaki B_1 vitamini miktarları farklı değildir. Yani kısaca

$$\mu_{\bar{x}A} = \mu_{\bar{x}B} = \mu \text{ veya}$$
$$\mu_{\bar{x}A} - \mu_{\bar{x}B} = 0$$

$\mu_{\bar{x}A} - \mu_{\bar{x}B} = \mu_D$ şeklinde gösterilirse test hipotezi $\mu_D = 0$ şeklinde de ifade edilebilir

. Karşıt hipotez ise;

H1: $\mu_{\bar{x}A} - \mu_{\bar{x}B} \neq 0$; veya kısaca
$\mu_D \neq 0$

şeklinde olacaktır. Çünkü araştırıcı iki kaplama tipinin birbirinden farklı olup olmadığını araştırmaktadır. Daha denemeyi kurarken birinin diğerinden üstün olabileceği şeklinde bir iddiası yoktur.

2. Drajelerdeki B_1 vitamini miktarlarının normal dağılım gösterdiği ve varyansının 4 olduğu bilinmektedir. Kaplama şekline göre zamanla varyansın değişmediği varsayılmaktadır.

Buna göre test hipotezi geçerli olduğunda örnek ortalamaları arası farklar, ortalaması sıfır (0) ve standart sapması 4.3 numaralı eşitlikte verildiği gibi,

$$\sigma_{(\bar{X}_A - \bar{X}_B)} = \sigma_D = \sqrt{\sigma^2 \frac{n_A + n_B}{n_A n_B}}$$ olan bir normal dağılım gösterir.

3. Ele alınan problemdeki istatistik, yani ortalamalar arası fark (D= $\bar{X}_A - \bar{X}_B$) standardize edildiğinde standart normal dağılım gösterir. Böylece test birimi Z, test dağılımı da standart normal dağılım olduğu kendiliğinden ortaya çıkmaktadır.

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_{\bar{X}_A} - \mu_{\bar{X}_B})}{\sigma_{(\bar{X}_A - \bar{X}_B)}}$$ veya daha kısa olarak

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - \mu_D}{\sigma_D}$$ 'dir.

4. Araştırıcı I. tip hata ihtimalini α=0.05 olarak belirlemiş olsun. Yapılan hipotez kontrolü çift taraflı olduğu için bu ihtimalin yarısı (%2.5'i) ortalamadan küçük değerlerin bulunduğu tarafta diğer yarısını da ortalamadan büyük değerlerin bulunduğu tarafta alır. Z-dağılımında %2.5'lik alanın başladığı değer ise 1.96'dır.

5. A ve B kaplama şekilleri uygulanan drajeler bir yıl süre ile aynı şartlarda muhafaza edilmiştir. Bir yıl sonra A tipi kaplama uygulanan 25 drajede yapılan analizlerde $\bar{X}_A = 22.0$ mg, B tipi kaplama uygulanan 36 drajede ortalama $\bar{X}_B = 22.8$ mg bulunmuştur. Buna göre;

$$Z = \frac{22-22.8}{\sqrt{4(\frac{1}{25}+\frac{1}{36})}} = -1.5361 \cong -1.54 \text{ olarak bulunur.}$$

Örnekten hesaplanan -1.54'lük test değeri kabul bölgesinde bulunmaktadır, yani kontrol hipotezi reddedilemez. Daha kısa olarak $|Z|<Z_\alpha$ veya $|-1.54|<1.96$ olduğundan kontrol hipotezi reddedilemez. Ele alınan ortalamalar arası fark olan 0.8 mg'lık farkın hipotezle belirlenen populasyona dahil olma olasılığı α'dan (0.05'ten) büyüktür ($P>0.05$).

7.2.3. Oranlara ait Hipotez Kontrolü

Araştırıcının üzerinde durduğu özellik binom dağılımı gösteren bir özellik olabilir, örneğin bozuk ve sağlam, başarılı ve başarısız, iyileşenler ve iyileşemeyenler gibi olaylar söz konusudur. Bu durumda üzerinde durulan olayın oluş ihtimali Bölüm 3.1'de açıklandığı şekilde binom dağılımının parametresi π ile tanımlanır. Ve araştırıcının amacı hesaplanan bir olasılık (p) ile ilgili hipotez kontrolü yapmak olabilir. Oranlara ait hipotez kontrolü yapılırken yerine getirilmesi gereken varsayımlar ve hipotez kontrolünde izlenecek adımlar aşağıdaki gibidir:

Hipotez kontrolü yapılırken yerine getirilmesi gereken varsayımlar:

- **Örnekler, binom dağılımı için gerekli şartları yerine getirir.**
- **Örneklerden hesaplanan p'ler yaklaşık normal dağılım göstermelidir.**

Hipotez kontrolü yapılırken izlenecek adımlar sırası ile aşağıdaki gibidir:

1. ADIM: İlk olarak araştırıcı kontrol ve karşıt hipotezlerini oluşturmalıdır.

Kontrol hipotezi:

H_0: Örnekten hesaplanan olasılık (p) ile populasyona ait olasılık arasındaki fark tesadüften ileri gelmektedir. Yani $\mu_p=\pi$'dir.

Kaşıt hipotez araştırıcının amacı doğrultusunda 3 farklı şekilde oluşturulabilir.

a. H_1: Örnekten hesaplanan olasılık (p) ile populasyona ait olasılık arasındaki fark tesadüften ileri gelmemektedir. Yani $\mu_p\neq\pi$'dir.

b. H_1: Örnekten hesaplanan olasılık (p), populasyona ait olasılıktan (π) daha büyük olasılığa sahip bir populasyonu temsil etmektedir. Yani $\mu_p>\pi$'dir.

c. H_1: Örnekten hesaplanan olasılık (p), populasyona ait olasılıktan (π) daha küçük olasılığa sahip bir populasyonu temsil etmektedir. Yani $\mu_p<\pi$'dir.

2. ADIM: Ele alınan oran ortalaması π olan bir binom dağılımı gösterir. Bu oranların dağılımının varyansı $\sigma_p^2 = \frac{\pi(1-\pi)}{n}$'dir. Daha önce belirtildiği gibi n yeteri kadar büyükse p'ler normal dağılır.

3. ADIM: Kontrol hipotezi ile belirtilen dağılım oranlara ait örnekleme dağılımıdır. Bu nedenle söz konusu örnekten hesaplanan oranın, hipotezle belirtilen örnekleme dağılımına dahil olma ihtimalini hesaplamak için Z-değeri Bölüm 4.3'te de açıklandığı şekilde aşağıdaki gibi hesaplanacaktır:

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

Burada, σ_p 4.8 numaralı eşitlikte verildiği gibi;

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \text{ 'dir.}$$

4. ADIM: Bölüm 7.2.1 ve 7.2.2'de açıklandığı şekilde araştırıcı I. tip hata ihtimalini, yani α'yı belirlemelidir.

5. ADIM: Araştırma yapılıp veriler toplanarak test istatistiği hesaplandıktan ve örnekten hesaplanan oranın söz konusu örnekleme dağılımına dahil olma ihtimali bulunduktan sonra Bölüm 7.2.1 ve 7.2.2'de açıklandığı şekilde karar verilir.

ÖRNEK 1:

Herhangi bir hastalığın tedavisinde kullanılan bir ilacın hastalığı iyileştirme oranının %75 olduğu bilinmektedir. Bir doktor söz konusu hastalığın tedavisinde kullanılan ilacın iyileştirme oranını araştırmak üzere yaptığı bir araştırmada 150 hastadan 120'sinin iyileştiğini tespit ediyor. Kullanılan ilacın hastalığı iyileştirme oranının %75 olduğu söylenebilir mi?

Hipotezlerin kurulması:

H_0: Örnekten hesaplanan iyileşme olasılığı ile populasyona ait olasılık (π=0.75) arasındaki fark tesadüften ileri gelmektedir. Yani μ_p=0.75 'dir.

H_1: Örnekten hesaplanan olasılık ile populasyona ait olasılık (π=0.75) arasındaki fark tesadüften ileri gelmemektedir. Yani $\mu_p \neq 0.75$ 'dir.

Söz konusu örnekten hesaplanan oranın, hipotezle belirtilen oranlara örnekleme dağılımına dahil olma ihtimalini hesaplamak için Z-değerini hesaplamak için önce p ve σ_p 'nin bulunması gerekir.

$$p = \frac{120}{150} = 0.80$$

$$\sigma_p = \sqrt{\frac{0.75(1-0.75)}{150}} = 0.0354$$

Araştırıcı kararını verirken I. tip hata ihtimalini, α=0.05, %5 olarak belirlemiş olsun. Karşıt hipotez çift taraflı kontrolüüüüü gerektirdiği için %5'in yarısı ortalamadan büyük değerlerin bulunduğu tarafta, diğer yarısı ise ortalamadan küçük değerlerin bulunduğu tarafta alınır.

Oranlara ait hipotez kontrolü yapılırken, örnekten hesaplanan oranların normal dağılım varsayımını yerine getirmiş olması gerekir. Bölüm 4.3'de açıklandığı gibi, eğer π>1/2 ise oranlara ait örnekleme dağılımının normal dağılım göstermesi için $n(1-\pi) \geq 10$

olmalıdır. Bu örnekte söz konusu populasyona ait π=0.75 olup, 0.5'ten büyüktür. Bu populasyondan çekilen örneklerin genişliği en az, n(1-0.75)=10 yani n=40 olmalıdır. Yapılan araştırmada üzerinde çalışılan örnek 150 birey içerdiği için bu şart yerine getirilmiştir. Yani π=0.75 olan populasyondan tesadüfen alınan ve 150 birey içeren örneklerden elde edilecek oranlara ait örnekleme dağılımı, normal dağılım gösterir. Bu oranlara ait Z kontrolünün yapılması için gerekli varsayımın yerine getirilmiş olduğunu gösterir. Bu durumda yukarıda hesaplanan p ve σ_p değerleri kullanılarak, Z-değeri;

$$Z = \frac{(0.80 - 0.75)}{0.0354} = 1.4124 \text{ olarak bulunur.}$$

Hesaplanan Z-değeri, %2.5'lik alanın başladığı 1.96 değerinden küçük olduğu için kontrol hipotezi reddedilemez. Diğer bir deyişle hesaplanan test değeri olan Z kabul bölgesindedir. Yani ele alınan örneğin hipotezle belirtilen örnekleme dağılımına dahil olma olasılığı 0.05'ten büyüktür (P>0.05).

ÖRNEK 2:

Bir fakültede öğrencilerin spor faaliyetlerine katılma oranının %45 olduğu bilinmektedir. Sporun insan sağlığı için önemini anlatan bir kampanyandan sonra bu oranda bir artış olup olmadığını araştırmak üzere 90 öğrenciye uygulanan anket sonucunda 52 öğrencinin spor faaliyetlerine katıldığı anlaşılmıştır. Acaba söz konusu fakültede öğrencilerin spor faaliyetlerine olan ilgisinin arttığı söylenebilir mi?

H_0: Söz konusu fakültede öğrencilerin spor faaliyetlerine katılma oranı 0.45 olarak kabul edilebilir. Gözlenen fark tesadüften ileri gelmektedir ve sıfır kabul edilebilir. Yani μ_p=0.45 'tir.

H_1: Söz konusu fakültede öğrencilerin spor faaliyetlerine katılma oranı artmıştır. Yani μ_p>0.45 'tir.

Bölüm 4.3'de açıklandığı gibi, eğer $\pi \leq 1/2$ ise oranlara ait örnekleme dağılımının normal dağılım göstermesi için $n\pi \geq 10$ olmalıdır. Bu örnekte söz konusu populasyona ait π=0.45 olup, 0.5'ten küçüktür. Bu populasyondan çekilen örneklerin genişliği en az, n(0.45)=10 yani $n \cong 23$ olmalıdır. Yapılan araştırmada üzerinde çalışılan örnek 90 öğrenci içerdiği için bu şart yerine getirilmiştir. Yani π=0.45 olan populasyondan tesadüfen alınan ve 90 birey içeren örneklerden elde edilecek oranlara ait örnekleme dağılımı, normal dağılım gösterir. Bu oranlara ait Z kontrolünün yapılması için gerekli varsayımın yerine getirilmiş olduğunu gösterir. Ve Z kontrolü aşağıda açıklandığı şekilde yapılır.

Söz konusu örnekten hesaplanan oranın, hipotezle belirtilen oranlara örnekleme dağılımına dahil olma ihtimalini hesaplamak için;

$$p = \frac{52}{90} = 0.578$$

$$\sigma_p = \sqrt{\frac{0.45(1-0.45)}{90}} = 0.0524$$

değerleri bulunur. Ve bu değerler kullanılarak, Z-değeri;

$$Z = \frac{(0.578 - 0.45)}{0.0524} \cong 2.443 \text{ olarak hesaplanır.}$$

Araştırıcı kararını verirken I. tip hata ihtimalini, $\alpha=0.01$, %1 olarak belirlemiş olsun. Karşıt hipotez tek taraflı kontrolü gerektirdiği için %1'in ortalamadan büyük değerlerin bulunduğu tarafta alınır.

Standart normal dağılımda, Tablo A'ya bakıldığı zaman %1'lik alanın yaklaşık 2.33 değerinde başladığı görülür. Burada hesaplanan Z-değeri yaklaşık 2.44 olup, bu değer %1'lik H_0 ret bölgesine düşmektedir. Bu da H_0 hipotezinin reddedilmesini gerektirir. Yani söz konusu fakültedeki öğrencilerin spor faaliyetlerine katılma olasılığı artmıştır. Bu durum araştırma sonuçlarında kısaca $P<0.01$ olarak gösterilir.

7.2.4. Oranlar arası Farka ait Hipotez Kontrolü

Yapılan bir araştırmanın amacı üzerinde durulan ve binom dağılımı gösteren bir özellik bakımından iki farklı muameleyi, bölgeyi veya metodu karşılaştırmak olabilir. Örneğin araştırıcının amacı, herhangi bir hastalığı tedavi için kullanılan iki farklı ilacı, hastalığı iyileştirme oranları bakımından karşılaştırmak olabilir. Bu durumda yapılması gereken oranlar arası farka ait hipotez kontrolüdür.

Hipotez kontrolü yapılırken yerine getirilmesi gereken varsayımlar:

- **Örnekler, binom dağılmış populasyondan bağımsız ve tamamen tesadüfen alınmış örnekler olmalıdır.**
- **Örneklerdeki birey sayısı, örneklerden hesaplanan oranların normal dağılım göstermesini sağlayacak büyüklükte olmalıdır. Yani örneklerden hesaplanan oranlara ait örnekleme dağılımı normal dağılım göstermelidir.**

İlk olarak araştırıcı kontrol ve karşıt hipotezlerini oluşturmalıdır.

Kontrol hipotezi:

H_0: İki oran arasındaki fark tesadüften ileri gelmektedir. Bu fark sıfır kabul edilebilir. Yani $\mu_{p1}=\mu_{p2}$ veya $\mu_{p1}-\mu_{p2}=0$ $(\pi_1 - \pi_2 = 0)$'dır.

Kaşıt hipotez araştırıcının amacı doğrultusunda 3 farklı şekilde oluşturulabilir.

a. H_1: İki oran arasındaki fark tesadüften ileri gelmemektedir. Bu fark sıfır kabul edilemez. Yani, $\mu_{p1}\neq\mu_{p2}$ veya $\mu_{p1}-\mu_{p2}\neq0$ $(\pi_1 - \pi_2 \neq 0)$'dır.

b. H_1: Birinci örneğin ait olduğu populasyondaki oran ikinci örneğin ait olduğu populasyondaki orandan daha büyüktür. Yani, $\mu_{p1}>\mu_{p2}$'dir.

c. H_1: Birinci örneğin ait olduğu populasyondaki oran ikinci örneğin ait olduğu populasyondaki orandan daha küçüktür. Yani, $\mu_{p1}<\mu_{p2}$'dir.

Kontrol hipotezi doğru olduğu zaman $\pi_1= \pi_2=\pi$ olacaktır. Bu durumda π'nin en iyi tahmini;

$$p = \frac{x_1 + x_2}{n_1 + n_2}$$ olacaktır.

Örneklerden hesaplanan oranların arası farkın, oranlar arası farka ait örnekleme dağılımına dahil olma ihtimalini hesaplamak için Z değeri;

$$Z = \frac{p_1 - p_2}{\sigma_{(p1-p2)}}$$ şeklinde hesaplanır.

Yukarıda verilen eşitlikte $\sigma_{(p1-p2)}$ ise aşağıda verilen eşitlik kullanılarak hesaplanır:

$$\sigma_{(p1-p2)} = \sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})}$$

Bölüm 7.2.1 ve 7.2.2'de açıklandığı şekilde araştırıcı I. tip hata ihtimalini, yani α'yı belirlemelidir.

Test istatistiği hesaplandıktan ve oranlar arası farkın söz konusu örnekleme dağılımına dahil olma ihtimali bulunduktan sonra yine Bölüm 7.2.1 ve 7.2.2'de açıklandığı şekilde karar verilir.

ÖRNEK 1:

Bir doktor, bir hastalığın tedavisinde kullanılan iki farklı ilacın hastalığın iyileşmesine etki bakımından farklı olup olmadığını araştırmak istemektedir. Bu nedenle aynı hastalığa yakalanmış 60 hastanın 34'ünü A ilacı ile tedavi ederek 28 hastanın iyileştiğini ve geri kalan hastaları B ilacı ile tedavi ederek 20 hastanın iyileştiğini tespit ediyor. Bu iki ilacın hastalığın iyileştirme oranları bakımından arasındaki farklılığın tesadüften ileri geldiği söylenebilir mi?

H_0: Hastalığı iyileştirme yüzdesi bakımından iki ilaç arasındaki fark tesadüften ileri gelmektedir. Bu fark sıfır kabul edilebilir. Yani $\mu_{pA}=\mu_{pB}$'dir.

H_1: Hastalığı iyileştirme yüzdesi bakımından iki ilaç arasındaki fark tesadüften ileri gelmemektedir. Bu fark sıfır kabul edilemez. Yani $\mu_{pA}\neq\mu_{pB}$'dir.

Kontrol hipotezi doğru olduğu zaman $\pi_1= \pi_2=\pi$ olacağı için π'nin en iyi tahmini;

$$p = \frac{28 + 20}{34 + 26} = 0.8$$

A ve B ilaçları için hastalığı iyileştirme oranları ise;

$$p_A = \frac{28}{34} = 0.8235$$

$$p_B = \frac{20}{26} = 0.7692$$

Oranlar arası farka ait örnekleme dağılımının standart sapması ise:

$$\sigma_{(p_A - p_B)} = \sqrt{0.8(1-0.8)(\frac{1}{34}+\frac{1}{26})} = 0.1042$$

Oranlar arası farkın, söz konusu örnekleme dağılımına dahil olma ihtimalini hesaplamak için Z değeri;

$$Z = \frac{0.8235 - 0.7692}{0.1042} = 0.5211 \text{ olarak bulunur.}$$

Araştırıcı I. tip hata ihtimalini, α=0.05 olarak belirlemiş olsun. Karşıt hipotez ile yapılacak hipotezin çift taraflı bir kontrol olduğu belirlendiğine göre I. tip hata ihtimali ortalamanın iki tarafında yarı yarıya alınır.

Z dağılımında %2.5'lik kontrol hipotezini ret bölgesi 1.96 değerinde başlamaktadır (Tablo A). Burada hesaplanan Z değeri 0.5211 olup 1.96 değerinden küçüktür yani H_0 hipotezini kabul bölgesine düşmektedir. Bu durumda kontrol hipotezi reddedilemez, yani hastalığı iyileştirme yüzdesi bakımından iki ilaç arasında fark tesadüften ileri gelmektedir. Söz konusu hastalığın tedavisinde iki ilaçtan biri rahatlıkla kullanılabilir.

7.2.5. Korelasyon Katsayısına ait Hipotez Kontrolü

Yapılan bir araştırmada üzerinde çalışılan örnekten iki özelliğe ait veri toplanabilir ve araştırıcı örneğin tamamen tesadüfen alındığı populasyona ait korelasyon katsayısı ile örnekten hesaplanan korelasyon katsayısı arasındaki farkın tesadüfi olup olmadığını kontrol etmek isteyebilir. Bu durumda yapılması gereken korelasyon katsayısına ait hipotez kontrolüdür.

Hipotez kontrolü yapılırken araştırıcı ilk olarak kontrol ve karşıt hipotezlerini oluşturmalıdır.

Kontrol hipotezi:

H_0: Örnekten hesaplanan korelasyon katsayısı ile populasyona ait korelasyon katsayısı arasındaki fark tesadüften ileri gelmektedir ve bu fark sıfır kabul edilebilir. Yani, $\mu_r=\rho$'dur.

Kaşıt hipotez araştırıcının amacı doğrultusunda 3 farklı şekilde oluşturulabilir.

a. H_1: Örnekten hesaplanan korelasyon katsayısı ile populasyona ait korelasyon katsayısı arasındaki fark tesadüften ileri gelmemektedir ve bu fark sıfır kabul edilemez. Yani, $\mu_r\neq\rho$'dur.

b. H_1: Örnekten hesaplanan korelasyon katsayısı populasyona ait korelasyon katsayısından daha büyüktür. Yani, $\mu_r>\rho$'dur.

c. H_1: Örnekten hesaplanan korelasyon katsayısı populasyona ait korelasyon katsayısından daha küçüktür. Yani, $\mu_r<\rho$'dur.

Burada yapılması gereken örneklerden hesaplanan korelasyon katsayısının, korelasyon katsayısına ait örnekleme dağılımına dahil olma ihtimalinin hesaplanmasıdır. Populasyona ait korelasyon katsayısının sıfırdan farklı olması durumunda, örnekten hesaplanan korelasyon katsayısı için 5.9 numaralı eşitlik kullanılarak hesaplanacak Z_r değerlerinin yaklaşık normal dağılım gösterdiği Bölüm 5.1'de etraflı olarak açıklanmıştı. Örnekten hesaplanacak korelasyon katsayısına karşılık gelen Z_r-değerleri Tablo B'de verilmiştir. Bu Z_r değerlerinin ortalaması 5.10 numaralı eşitlikte verildiği gibi;

$$\mu_{Z_r} = \frac{1}{2}\log_e \frac{1+\rho}{1-\rho}$$

ve standart sapması 5.11 numaralı eşitlikte verildiği gibi;

$$\sigma_{Z_r} = \sqrt{\frac{1}{(n-3)}}$$ 'tür.

Örneklerden hesaplanan korelasyon katsayısının, korelasyon katsayısına ait örnekleme dağılımına dahil olma ihtimalinin hesaplanması için test istatistiği (Z-değeri);

$$Z = \frac{Z_r - \mu_{Z_r}}{\sigma_{Z_r}}$$ eşitliği kullanılarak hesaplanır.

Yukarıda açıklanan yaklaşım örnek genişliğinin 50'den büyük olduğu durumlarda (ve hatta örnek genişliğinin 25'den büyük olduğu durumlarda da) kullanılabilir. Çünkü Z_r değerleri populasyon korelasyon katsayısı ne olursa olsun yaklaşık normal dağılım gösterecektir. Fakat örnek genişliğinin daha küçük, yapılan hipotez kontrolünün ve hesaplanan güven sınırlarının kritik olduğu durumlarda, yapılan hipotez kontrolünün ve bulunan güven sınırlarının güvenilirliği bakımından Hotelling bir başka transformasyon geliştirmiştir. Hottelling'in yaklaşımında;

$$Z^* = Z_r - \frac{3Z_r + r}{4(n-1)}$$ olarak verilir. Bu değerlerin standart sapması;

$$\sigma_{Z^*} = \frac{1}{\sqrt{(n-1)}}$$ 'dir. Bu durumda test istatistiği;

$$Z = \frac{Z^* - \mu_{Z^*}}{\sigma_{Z^*}} = (Z^* - \mu_{Z^*})(\sqrt{n-1})$$ 'dir. Eşitlikte;

$$\mu_{Z^*} = \mu_{Z_r} - \frac{(3\mu_{Z_r} + \rho)}{4n}$$ 'dir.

Daha önceki bölümlerde açıklandığı şekilde araştırıcı I. tip hata ihtimalini, yani α'yı belirlemelidir.

Test istatistiği hesaplandıktan ve örnekten hesaplanan korelasyon katsayısı ve daha fazla sapanların söz konusu örnekleme dağılımına dahil olma ihtimali bulunduktan sonra yine daha önceki bölümlerde açıklandığı şekilde karar verilir.

ÖRNEK 1:

Yaş ile kan basıncı arasındaki korelasyon katsayısının (ρ'nun) 0.92 olduğu ileri sürülmüştür. Bunun doğruluğunu kontrol etmek üzere tamamen tesadüfen seçilen 7 bireyde yaş ve kan basıncı ölçümleri aşağıdaki gibi tespit edilmiştir.

Yaş (X)	Kan basıncı (Y)
35	117
40	121
72	160
56	142
64	158
31	111
73	166

Yaş ve kan basıncı arasındaki korelasyon katsayısını hesaplayarak, örnekten hesaplanan korelasyon katsayısı ile populasyona ait korelasyon katsayısı arasındaki farkın tesadüfi olduğu söylenebilir mi?

7 bireyden elde edilen veriler kullanılarak, Bölüm 5.1'de açıklandığı şekilde işlemler yapılacak olursa korelasyon katsayısı 0.994 olarak bulunur. Örnekten hesaplanan korelasyon katsayısı ile populasyon katsayısı arasındaki farkın tesadüfi olup olmadığı aşağıda açıklandığı şekilde hipotez kontrolü yapılarak kontrol edilir.

H_0: Örnekten hesaplanan korelasyon katsayısı ile populasyona ait korelasyon katsayısı (ρ=0.92) arasındaki fark tesadüften ileri gelmektedir ve bu fark sıfır kabul edilebilir. Yani, $\mu_r=0.92$'dir..

H_1: Örnekten hesaplanan korelasyon katsayısı ile populasyona ait korelasyon katsayısı (ρ=0.92) arasındaki fark tesadüften ileri gelmemektedir ve bu fark sıfır kabul edilemez. Yani, $\mu_r \neq 0.92$'dir.

Örnekte korelasyon katsayısı 0.994 olarak bulunmuştur. Tablo B'de r=0.99'a karşılık gelen Z-değeri $Z_r=2.6467$ olarak bulunur. Veya 5.9 numaralı eşitlik kullanılarak;

$$Z_r = \frac{1}{2}\log_e\frac{(1+0.994)}{(1-0.994)} = 2.9031 \text{ olarak hesaplanır.}$$

Aradaki fark tablodan 0.99'un karşılığının kullanılmasından ileri gelmiştir. Günümüzde cepte taşınan hesap makinalarında bile logaritma ve antilogaritma alma imkanı olduğundan Z_r'nin hesapla bulunması daha doğrudur.

Bu Z-değerlerine ait ortalama;

$$\mu_{Z_r} = \frac{1}{2}\log_e \frac{1+0.92}{1-0.92} = 1.589$$

ve standart sapma;

$$\sigma_{Z_r} = \sqrt{\frac{1}{(7-3)}} = 0.5\text{'tir.}$$

Örneklerden hesaplanan korelasyon katsayısının, korelasyon katsayısına ait örnekleme dağılımına dahil olma ihtimalinin hesaplanması için test istatistiği (Z-değeri);

$$Z = \frac{2.9031-1.589}{0.5} = 2.6282 \text{ olarak bulunur.}$$

Araştırıcı I. tip hata ihtimalini, yani α'yı %5 olarak belirlemiş olsun. Karşıt hipotez, çift taraflı hipotez kontrolünü gerektirmektedir. Bu nedenle, yanılma ihtimali ortalamanın iki tarafında eşit miktarda alınır.

Hesaplanan test istatistiğinin değeri 2.6282 olup, Z dağılımında %2.5'lik alanın başladığı 1.96 değerinden büyüktür, yani ret bölgesine düşmektedir. Bu durumda kontrol hipotezi ret edilir, yani söz konusu örnek korelasyon katsayısı 0.92 olan populasyondan alınmamıştır. Yani yaş ile kan basıncı arasındaki korelasyon katsayısının bu örneğin alındığı populasyonda 0.92 olduğu söylenemez.

Bu örnek için Hotelling yaklaşımı kullanılarak hipotez kontrolü ise aşağıdaki gibidir:

$$Z^* = 2.9031 - \frac{3(2.9031) + 0.994}{4(7-1)} = 2.499$$

$$\sigma_{z^*} = \frac{1}{\sqrt{(7-1)}} = 0.408$$

$$\mu_{z^*} = 1.589 - \frac{3(1.589) + 0.92}{4(7)} = 1.386$$

ve Z istatistiği;

$$Z = \frac{2.499 - 1.386}{0.408} = 2.728 \text{ olarak bulunur.}$$

Bulunan bu değer de kontrol hipotezinin reddedilmesini gerektirir.

7.2.6. Korelasyon Katsayıları arasındaki Farka ait Hipotez Kontrolü

İki örnekten hesaplanan korelasyon katsayıları arasındaki farkın tesadüften ileri gelip gelmediği kontrol edilmek istenebilir. Bu durumda korelasyon katsayıları arasındaki farka ait hipotez kontrolü yapılması gerekir.

Hipotez kontrolü yapılırken izlenecek adımlar sırası ile aşağıdaki gibidir:

İlk olarak araştırıcı kontrol ve karşıt hipotezlerini oluşturmalıdır.

Kontrol hipotezi:

H_0: Korelasyon katsayıları arasındaki fark tesadüften ileri gelmektedir ve bu fark sıfır kabul edilebilir. Yani, $\rho_1=\rho_2$'dir.

Kaşıt hipotez araştırıcının amacı doğrultusunda 3 farklı şekilde oluşturulabilir.

1. H_1: Korelasyon katsayıları arasındaki fark tesadüften ileri gelmemektedir ve bu fark sıfır kabul edilemez. Yani, $\rho_1\neq\rho_2$'dir.

2. H_1: Birinci örnekten hesaplanan korelasyon katsayısı ikinci örneğe ait korelasyon katsayısından daha büyüktür. Yani, $\rho_1>\rho_2$'dir.

3. H_1: Birinci örnekten hesaplanan korelasyon katsayısı ikinci örneğe ait korelasyon katsayısından daha küçüktür. Yani, $\rho_1<\rho_2$'dir.

Örneklerden hesaplanan korelasyon katsayıları arası fark kadar ve daha fazla sapanların, korelasyon katsayıları arasındaki farka ait örnekleme dağılımına dahil olma ihtimalinin hesaplanması için yine korelasyon katsayıları için Z_r değerleri hesaplanır veya Tablo B'den söz konusu korelasyon katsayılarına karşılık gelen Z_r değerleri bulunur.

Örneklerden hesaplanan korelasyon katsayıları arası fark kadar ve daha fazla sapanların, korelasyon katsayıları arası farka ait örnekleme dağılımına dahil olma ihtimalinin hesaplanması için test istatistiği (Z-değeri);

$$Z=\frac{Z_{r1}-Z_{r2}}{\sqrt{\frac{1}{(n_1-3)}+\frac{1}{(n_2-3)}}}$$ şeklinde hesaplanır.

Daha önceki bölümlerde açıklandığı şekilde araştırıcı I. tip hata ihtimalini, yani α'yı belirlemelidir.

Test istatistiği hesaplandıktan ve örnekten hesaplanan korelasyon katsayıları arasındaki fark kadar ve daha fazla sapanların söz konusu örnekleme dağılımına dahil olma ihtimali bulunduktan sonra yine daha önceki bölümlerde açıklandığı şekilde karar verilir.

ÖRNEK 1:

Bir fakültede yapılan bir araştırmada, fakültenin birinci sınıf öğrencilerinden tamamen tesadüfen seçilen 50 kız öğrencide boy ile kilo arasındaki korelasyon katsayısı 0.60 ve yine tamamen tesadüfen seçilen 40 erkek öğrencide boy ile kilo arasındaki korelasyon katsayısı 0.45 olarak bulunmuştur. Boy ile kilo arasındaki korelasyon katsayıları bakımından iki cinsiyet arasındaki farklılığın önemli olup olmadığını kontrol ediniz.

Hipotezlerin oluşturulması;

H_0: İki cinsiyet için hesaplanan korelasyon katsayıları arasındaki fark tesadüften ileri gelmektedir ve bu fark sıfır kabul edilebilir. Yani, $\rho_1=\rho_2$'dir veya $\rho_1-\rho_2=0$'dır.

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

Örneklerden hesaplanan korelasyon katsayıları için Z_r değerleri hesaplanır (veya Tablo B'den söz konusu korelasyon katsayılarına karşılık gelen Z değerleri bulunur.).

$$Z_{r1} = 1.1513 \log \frac{(1+0.60)}{(1-0.60)} = 0.6931$$

$$Z_{r2} = 1.1513 \log \frac{(1+0.45)}{(1-0.45)} = 0.4847$$

olarak bulunur.

Örneklerden hesaplanan korelasyon katsayıları arası farkın, korelasyon katsayıları arası farka ait örnekleme dağılımına dahil olma ihtimalini hesaplamak için Z-değeri;

$$Z = \frac{0.6931 - 0.4847}{\sqrt{\frac{1}{(50-3)} + \frac{1}{(40-3)}}} = 0.948 \text{ olarak hesaplanır.}$$

Araştırıcı I. tip hata ihtimalini, yani α'yı %5 olarak belirlemiş olsun. Karşıt hipotez, çift taraflı hipotez kontrolünü gerektirmektedir. Bu nedenle, yanılma ihtimali ortalamanın iki tarafında eşit miktarda alınır.

Hesaplanan test istatistiğinin değeri 0.948 olup, Z dağılımında %2.5'lik alanın başladığı 1.96 değerinden çok küçüktür, yani kontrol hipotezini kabul bölgesine düşmektedir. Bu durumda kontrol hipotezi reddedilemez, yani iki cinsiyet için hesaplanan korelasyon katsayıları arasındaki fark tesadüften ileri gelmektedir. Yani boy ile kilo arasındaki ilişki cinsiyetlere göre değişmemektedir.

Burada dikkat edilmesi gereken nokta korelasyon katsayıları örnekten hesaplanmış olmasına karşın Z-kontrolü yapılmıştır. Çünkü örnek genişlikleri Z-dağılımını kullanmak için yeteri büyüklüktedir. Eğer korelasyon katsayılarının hesaplandığı örnek genişlikleri küçük ise bu durumda t-kontrolü yapılmalıdır. Bu, 8. bölümde açıklanacaktır.