

ÖNEMLİLİK TESTLERİNDE HİPOTEZLER, HATA TIPLERİ VE ÖRNEK BÜYÜKLÜĞÜ İLİŞKİLERİ ÜZERİNE BİR DEĞERLENDİRME

(Derleme)

M. Emin TEKİN¹

A short review on hypothesis test types and their characteristics, and error types in analytical statistics (A review)

SUMMARY

In this short review, hypothesis test types and their characteristics, and error types in analytical statistics are explained. Also, the possible ways for reducing these error types are reviewed. As a result, for reducing errors, it was recommended that large sample should be analyzed. But if this is impossible, then the sample should be blocked and grouped in according to effective factors.

KEY WORDS : Statistic, hypothesis, error types, sample size.

ÖZET

Bu makalede, analitik istatistiklerde kurulan hipotezlerin çeşit ve özellikleri ile hipotez testlerinde yapılması muhtemel olan hata tipleri ve bu hataları azaltma yolları üzerinde kısaca durulmuştur. Sonuç olarak; hatayı azaltmak için, uygun ve doğru hipotez kurulması; örnek hacminin mümkün olduğunca yüksek tutulması; sınırlı örnek hacmi ile çalışmak zorunlu ise örneğin etkili faktörlere göre gruplandırılması, bloklara ayrılması gerektiği vurgulanmıştır.

ANAHTAR KELİMELELER: İstatistik, hipotezler, hata tipleri, örnek büyüklüğü.

GİRİŞ

Hipotezler

İstatistikte çoğu zaman gruplar arası farkın önemli olup olmadığı araştırılır. Yani önemlilik testleri yapılır. Bu işlemler yapılırken bir hipotez test edilir. Bir önyargı veya ön kabul demek olan başlangıç hipotezini test etmek için, kabul edilmediği takdirde bunun yerine geçecek olan hipotez de belirlenir. Dolayısıyla bir hipotez testinde iki tip hipotez vardır;

- 1- Sıfır hipotezi,
- 2- Alternatif hipotez.

Sıfır hipotezi (H_0 ile gösterilir) farksızlık hipotezi veya eşitlik hipotezidir; iki grup ortalaması arasında bir fark olmadığı veya iki grup ortalamasının birbirine eşit olduğunun kabulüdür. Sıfır hipotezi her çalışmada aynıdır, tektir. Alternatif hipotez (H_1 ile gösterilir) ise, sıfır hipotezinin kabul edilmediği durumda otomatik olarak onun yerine geçecek olan hipotezdir. Ancak

alternatif hipotez tek değildir; iki türdür, araştırma biçimine göre ikisinden biri tercih edilir. Bunlar;

- 1- İki yönlü hipotez,
- 2- Tek yönlü hipotez.

İki yönlü hipotez, sıfır hipotezinin zıttı olarak iki grup ortalaması arasında fark vardır veya iki grup ortalaması birbirine eşit değildir şeklinde kurulur. Tek yönlü hipotez ise, gruplardan birinin ortalamasının diğerinin ortalamasına göre küçük veya büyük olduğu kabulüne dayanır.

H_0 kabul edilirse H_1 reddedilir; H_0 reddedilirse H_1 kabul edilir.

$H_0 : \mu_1 = \mu_2$ veya $\mu_1 - \mu_2 = 0$ (Sıfır hipotezi)

$H_1 : \mu_1 \neq \mu_2$ veya $\mu_1 - \mu_2 \neq 0$ (İki yönlü alternatif hipotez)

$H_1 : \mu_1 < \mu_2$ (Tek yönlü alternatif hipotez)

$H_1 : \mu_1 > \mu_2$ (Tek yönlü alternatif hipotez)

Bu hipotezler z ve t testleri içindir. Çoklu karşılaştırma testlerinde de hipotezler böyledir. F

testinde sıfır hipotezi ve alternatif hipotez biraz değişiktir.

F testinde hipotezler:

H_0 : Hata varyansı ($\sigma^2_{iç}$) = gruplar arası varyans (σ^2_{ara}) veya $\sigma^2_{ara}/\sigma^2_{iç}=1$

H_1 : $\sigma^2_{iç} \neq \sigma^2_{ara}$

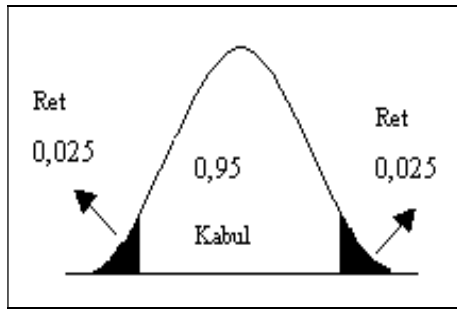
$\sigma^2_{iç} < \sigma^2_{ara}$; en az bir grup diğerlerinden farklıdır.

Regresyon analizinde hipotezler:

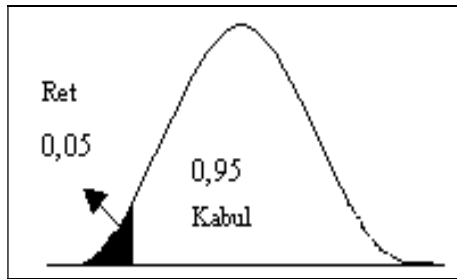
H_0 : Regresyon katsayısı = 0

H_1 : Regresyon katsayısı $\neq 0$

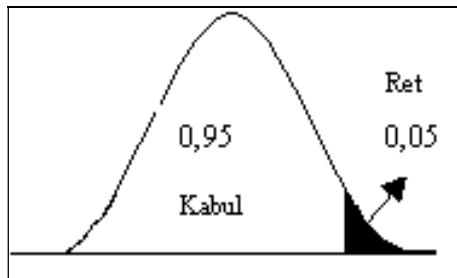
Normal dağılıma uyan verilerle ilgili yapılan testlerde alternatif hipotezin tipine göre ret ve kabul bölgeleri Şekil 1'de verilmiştir. Şekil 1'den anlaşılacağı gibi, iki yönlü ve tek yönlü hipotezler arasındaki fark, tek yönlü hipotezde ret bölgesinin iki yönlünün iki katı olmasıdır. Yani tek yönlü hipotez ile sıfır hipotezini reddetme şansı iki misli fazladır.



(a)



(b)



(c)

Şekil 1. Alternatif Hipotez Tipine Göre, $\alpha = 0,05$ Olduğunda Ret (0,05) ve Kabul (0,95) Bölgeleri; a: İki Yönlü Hipotez; b: Tek Yönlü Hipotez ($\mu_1 > \mu_2$); c: Tek Yönlü Hipotez ($\mu_1 < \mu_2$).

Hata Tipleri

Hipotez testleri sonunda hipotezlerden biri ret diğeri kabul edilir. Verilen bu karar her zaman doğru olmayabilir. Doğru karar verilmişse sorun yoktur. Ancak yanlış karar verilmişse bir hata yapılmış demektir. Dolayısıyla istatistikte hipotez test edilirken hata ihtimalleri ve tipleri belirlenmiş, mümkün olduğunca bunların denetim altına alınması öngörülmüştür. Hipotez testlerinde iki tip hata ihtimali vardır. Bunlara I. Tip hata ve II. Tip hata denir. I. Tip hata ihtimali önemlilik düzeyi veya yanlışlık ihtimali olarak bilinir ve alfa (α) ile gösterilir. Bu nedenle I. Tip hataya alfa hatası da denir. II. Tip hata beta (β) ile gösterilir ve beta hatası olarak bilinir. Doğru bir sıfır hipotezinin (gerçekte iki ortalama birbirine eşitken) yanlışlıkla ret edilmesi ile alfa hatası yapılmış olur. Yani gruplar arası fark olmadığı halde fark var denmişse yapılan hata alfa hatasıdır. Yanlış bir sıfır hipotezinin (gerçekte iki ortalama birbirine eşit değilken) yanlışlıkla kabul edilmesi ile beta hatası yapılmış olur. Yani gruplar arası fark olduğu halde fark yok denmişse yapılan hata beta hatasıdır. Bu durum aşağıdaki tabloda formüle edilmiştir.

Tablo1. Hipotez Testlerinde Hatalar

Gerçek durum	H_0 ret	H_0 kabul
H_0 doğru	I. Tip hata	Doğru karar
H_0 yanlış	Doğru karar	II. Tip hata

Zihinde kolay kalması bakımından Tablo 1'deki hata ihtimallerini bir cümle ile şöyle formüle edebiliriz:

Doğruyu ret I. tip hata, yanlış kabul II. tip hatadır

Bazı kaynaklarda I. Tip hata için "yanlış negatif", II. Tip hata için "yanlış pozitif" ifadeleri kullanılmaktadır.

Birinci Tip hatadan kaçınmak için düşük α seviyesi, 2. Tip hatadan kaçınmak için düşük β seviyesi kullanılmalıdır. Ancak alfa hatasını denetlemek kolay iken beta hatasını denetlemek zordur. α küçüldükçe β büyür.

Beta hatası dolaylı olarak azaltılabilir. Bunları maddeler halinde şöyle sıralayabiliriz:

1- Örnek hacmini büyütme; imkan varsa en kestirme ve kesin yol budur.

2- Tekerrür sayısını artırarak; materyali artırma imkanı yoksa aynı materyalde aynı muameleyi fazla tekrarlar yapmakla da beta hatasından kaçınılabilir.

3- Materyal yetersiz ise mevcut materyalde birörnekliliği bozan faktöre göre materyali bloklara ayırmak. Tesadüf blokları veya latin karesi deneme düzeni bu amaçla geliştirilmiştir. Eğer eldeki materyal bloklara ayırmaya müsaitse beta hatasını azaltmanın önemli bir yolu da budur. Ancak materyali bloklara ayırmak uygun değilse fayda yerine zarar gelir; beta hata ihtimali artar.

4- Mümkünse muameleyi bağımsız gruplarda değil de bağımlı gruplarda uygulamak. Yani deneme düzenini, bağımsız t testi veya bağımsız gruplarda varyans analizi yerine *paired t* testine veya tekrarlı ölçümlerde varyans analizine uygun biçimde düzenlemek.

5- Alternatif hipotezi doğru kurmak. Alternatif hipotezi doğru kurmaktan kasıt, eğer eldeki problem büyük müdür veya küçük müdür? gibi ise iki yönlü yerine tek yönlü hipotez testi yapmaktır. Çünkü iki yönlü hipotez ile reddedilemeyen sıfır hipotezi tek yönlü test ile reddedilebilir. Böylece, aynı güven düzeyinde olduğu halde beta hatası yapılmamış olur.

İstatistik analizlerde beta hatası yapmayı önlemenin yolları bunlarla sınırlı değildir. İstatistikte birçok test yöntemi vardır ve her testin kendine özgü hatayı azaltma yolları vardır. Burada bu detaylara girilmemiştir.

Yukarıdaki maddelerden her biri için bir örnek göstermek, konuyu detaylıca ortaya koymak mümkündür. Fakat yazının kapsamı buna müsait değildir; amaç da o değildir zaten. Ancak burada alternatif hipotez ve örnek büyüklüğü maddeleri ile ilgili bir örnek gösterebiliriz.

Örnek 1: t testi için bağımsız iki örnek

Koyunlarda kültür ırkı ile yerli ırk arasında yapılan bir melezleme çalışmasında saf yerli ırk kuzuların besi performansı ile F₁ melez kuzuların besi performansını karşılaştırmak isteyelim. Besi performansının ölçüsü olarak günlük canlı ağırlık artışını kriter kabul edelim. Burada amaç, besi performansı bakımından iki grup arasında fark olup olmadığını bulmaktır. Ancak F₁ melezlerinin saf yerli ırktan üstün olması (ortalamasının yüksek olması) beklenir. Bu nedenle, sorumuzu İki grup arasında bir fark var mı/yok mu? şeklinde sorabileceğimiz gibi (iki yönlü hipotez); F₁'lerin saflardan yüksek olup olmadığı şeklinde de sorabiliriz (tek yönlü hipotez) Bu hipotezleri test etmek için araştırma yaptığımızı ve her genotip grubundan belli sayıda kuzuda besi süresince günlük canlı ağırlık artışı verilerini topladığımızı düşünelim.

Her grupta N =15 olduğunda sonuçlar Tablo 2'deki gibi alınmış olsun (veriler türetilerek elde edilmiştir, gerçek değildir).

Tablodan görüldüğü gibi, aynı veri sayısı ile 0.05 yanılma olasılığında iki yönlü hipotez ile iki grup arasında fark yok derken (bir beta hatası yapma ihtimali varken), tek yönlü hipotezde iki grup arasında fark var sonucuna varılmıştır.

Tablo 2. Her Genotipten 15 Baş Kuzunun Günlük Canlı Ağırlık Artışı Ortalamaları

Genotipler	N	Ortalama	Stand. sapma	Stand. hata
F ₁ melezi	15	274,5	25,9	6,7
Yerli ırk	15	257,3	24,7	6,4

Hipotez tipi	T değ.	P değ.	Sonuç
İki yönlü	1,86	0,073	P>0,05; H ₀ kabul; iki grup arasında fark yok
Tek yönlü	1,86	0,036	P<0,05; H ₀ ret; iki grup arasında fark var

Her grupta N = 20 olduğunda sonuçlar Tablo 3'deki gibi alınmış olsun.

Tablo 3. Her Genotipten 20 Baş Kuzunun Günlük Canlı Ağırlık Artışı Ortalamaları

Genotipler	N	Ort.	St. sapma	St. hata
F ₁ melezi	20	275,2	20,0	4,5
Yerli ırk	20	255,2	20,4	4,6

Hipotez tipi	T değeri	P değeri	Sonuç
İki yönlü	3,13	0,0033	P<0,01; H ₀ ret; iki grup arasında fark var
Tek yönlü	3,13	0,0017	P<0,01; H ₀ ret; iki grup arasında fark var

Tablo 3' te görüldüğü gibi, bu sefer n sayısının artmasıyla her iki hipotezle de sıfır hipotezi reddedilebilmiştir. Burada bir başka konu iki grup arasındaki fark miktarıdır. Her iki tabloda da gruplar arası fark yaklaşık 20 g kadardır. Gruplar arası fark ikisinde de aynı olduğu halde, n sayısı artmakla bu farkın önemli hale geldiği açıkça görülmektedir.

Etki Büyüklüğü

Etki büyüklüğü, gruplar arası farkın veya ilişkinin gücünü gösterir, etki büyükse bunu örnekte tespit etmek kolayken, etki küçükse örnekte tespit etmek zordur. Uygun etki büyüklüğünü belirlemek için, ya bir pilot çalışma yapılır veya önceki çalışma bulgularından yararlanılır veya uygulamada anlamlı olan en küçük etki büyüklüğü seçilir. Genel olarak, İki grubun ortalamasını karşılaştırmada kullanılan t testinde beklenen küçük etki, standart sapmanın %20'si, orta etki %50'si, büyük etki %80' i olarak kabul edilmektedir (Akgül 1997). Örneğin ortalama 350, standart sapma 20 birim ve iki grup arasındaki fark 4 birim ise (20 x 0,2) küçük etki; aradaki fark 10 birim ise (20 x 0,5) orta etki; fark 16 birim ise (20 x 0,8) büyük etki var demektir. Etki büyüklüğü örneğin varyansı ile yakından ilgilidir.

Testin Gücü

Testin gücü, sıfır hipotezini (gruplar arası fark yok kabulünü) reddetme düzeyidir. Yani II. Tip hatadan kaçınmaktır. Araştırmalarda kullanılabilir makul örnek hacmine paralel olarak, testin gücünün %80 olması yeterli görülmektedir. Güç ne kadar büyüksün o kadar örnek hacmi gereklidir. Örnek büyüklüğünü arttırmak zaman ve para masrafı gerektireceğinden, küçük bir beta hatası için gerekli minimum örnek büyüklüğü oluşturmaya çalışılır.

Önemlilik Seviyesi, Hata Tipi, Testin Gücü, Örnek Büyüklüğü ve Etki Büyüklüğü İlişkileri

Önemlilik seviyesi, doğru sıfır hipotezinin reddedilmesi ihtimali (I. Tip hata) olup α ile gösterilir ve genellikle 0,05 taban olarak alınır. Çoğu zaman I. tip hatanın 0,05 ihtimalle olması, II. Tip hatanın ise 0,20 ihtimalle olması istenir. Birinci tip hatanın zıttı güven düzeyidir; hata az ise güven artar; $\alpha = 0,05$ olursa güven düzeyi 0,95 (%95) iken $\alpha = 0,01$ olursa güven düzeyi 0,99 (%99) olur. İkinci tip hatanın zıttı ise testin gücüdür hata az ise testin gücü artar; $\beta = 0,20$ olursa testin gücü 0,80 iken $\beta = 0,15$ olursa testin gücü 0,85 olur. Bunun tersi de doğrudur; testin gücü yüksekse hata az olur.

Birinci tip hatayı azaltmak (veya kontrol etmek) kolay iken ikinci tip hatayı denetlemek zordur. Ancak örnek hacmini mümkün olduğunca artırmak suretiyle ikinci tip hata azaltılabilir; denek sayısı ne kadar çok olursa β o kadar küçüldür. Böylece testin gücü 1- β kadar büyür.

α ile β belirlendikten sonra örnek sayısını bulmak için, gruplar arasında ne kadar bir fark bulunmak istendiği (etki büyüklüğü) de tespit edilmelidir. Bunun için; farkın pratik önemi de göz önünde tutulur. Örneğin pratik önemi olan asgari farkın önemli çıkması istenir ve bunu sağlayacak örnek hacmi belirlenir.

İki grup arasındaki farkın büyüklüğü (etki büyüklüğü) azaldıkça, bunun önemli bulunması için daha büyük örnek genişliğine ihtiyaç vardır. Ayrıca, önemlilik seviyesi ne kadar küçük ise (yani alfa hatası küçük olsun istenirse) o kadar büyük örnek genişliği belirlenmelidir.

Ki kare, regresyon ve korrelasyon analizleri, örnek büyüklüğünden çok etkilenirler. Örneğin aynı derecedeki bir korrelasyon katsayısı büyük bir örnekte hesaplanmışsa önemli, küçük örnekte hesaplanmışsa önemsizdir. Ki kare analizlerinde de örneğin iki yüzde arası %5 gibi bir fark, büyük örnek hacminde önemli küçük örnek hacminde önemsizdir.

İdeal olarak, yanlış pozitif ve yanlış negatif sonuçların ortadan kaldırılması için, yani hiç hata yapmamak için, α ve β sıfır olmalıdır. Ancak, bunların küçültülmesi daha büyük örnek hacmi gerektirdiği için uygulamada nerede ise başarmak mümkün değildir.

Örnek Büyüklüğü Belirleme

- Sıfır ve alternatif hipotez tanımlanır.
- Değişken tipine ve grup sayısına göre uygun istatistik test seçilir.
- Önceki bulgulara göre en küçük etki büyüklüğü (ortalamalar arası fark) belirlenir.
- Önceki bulgulara/bilgilere göre, incelenen değişkene ait standart sapma belirlenir.
- Uygun α ile β seviyeleri belirlenir.

Analitik İstatistiklerde (z ve t Testleri, Tek ve İki Yönlü Varyans Analizi) Örnek Büyüklüğü Belirleme

Değişik istatistik kaynaklarında bu amaçla geliştirilmiş formüller bulunmaktadır (2, 6, 8, 12). Ancak bugün bazı paket programlarda bu işlem kolayca yapılabilmektedir.

Minitab paket programında *Stat* menüsündeki *Power and Sample Size* seçeneği yardımıyla örnek büyüklüğü ve testin gücü hesaplamaları kolayca yapılabilmektedir.

Populasyon standart sapması önceden bilinmek koşuluyla, bu menü aracılığı ile;

1- Belirlenen anlamlı en küçük fark için değişik örnek büyüklüklerinde testin gücü (*Calculate power for each sample size*).

2- Belirlenen bir örnek büyüklüğünde değişik fark değerleri için testin gücü (*Calculate power for each difference*)

3- Belirlenen anlamlı en küçük fark için değişik test gücü değerleri için örnek büyüklüğü (*Calculate sample size for each power value*) hesaplamaları yapılabilmektedir.

Yapılan hesaplamalar, belirli bir alfa yanılma düzeyi ve alternatif hipotezin tipine göre, isteğe bağlı biçimde değiştirilebilmektedir.

Standart sapma, ya önceden bilinmelidir veya pilot çalışmalar ile tahmin edilmelidir veya ilgili başka araştırma sonuçlarından alınabilir.

Örnek 2: Bağımsız İki Örnek t testi için örnek hacmi belirleyelim

Örnek 1'de araştırılan konu ile ilgili olarak yapılan bir ekonomik değerlendirme ile, eğer iki grup arasındaki ağırlık artışı farkı 25 g dan az olmazsa bunun ekonomik yönden anlamlı olduğu kabul edilmiş olsun. Yerli ırkta daha önce yapılan besi çalışmalarında standart sapma 20 g olarak hesaplanmış olduğunu varsayalım,

Gruplar arasındaki 25 g'lık farkı, alfa = 0,05 hata ile, testin gücünün %80 olmasını isteyerek, tek yönlü hipoteze göre önemli bulabilmek için (sıfır hipotezini reddetmek için, veya beta hatası yapmamak için) gerekli olan en az örnek hacmini belirleyelim.

Bu amaçla Minitab programının ilgili menüsünü kullandığımızda, her bir örnek/grup için (yerli ırk ve F₁ melez grupları) hesaplanan örnek büyüklüğü; iki yönlü hipotez testi yapılacaksa 12, tek yönlü hipotez testi yapılacaksa 9 dur.

Eğer testin gücü %85 alınırsa yani beta hatası biraz küçük tutulursa hesaplanan değerler sırasıyla 13 ve 10 dur.

Eğer alfa = 0.01 seçilirse ve testin gücü %80 olsun istenirse gerekli n sayısı; iki yönlü test için 19, tek yönlü test için 16 dir.

Görüldüğü gibi, belirli bir farkın önemli çıkmasını istediğimizde, gerek alfa hatasını küçültürsek, gerek beta hatasını küçültürsek, gerekli olan örnek büyüklüğü o oranda artmaktadır. Aynı sonucu almak için iki yönlü hipotez testine göre tek yönlü hipotez testi için daha az sayıda materyal gerekli olduğu da ortaya çıkmaktadır.

Sonuç olarak, istatistik analizlerde doğru karar verebilmek veya hatayı en az yapmak için;

- Alternatif hipotezin doğru kurulması,
- Alfa yanılma olasılığının 0.05 ten, beta hatasının 0.20 den büyük olmaması
- Hatayı minimum yapmak için örnek büyüklüğünün imkan ölçüsünde geniş tutulması
- Yeterli materyal yoksa tekerrüre başvurulması
- Materyalin kendi içinde bloklara ayrılması
- Muamele etkisinin mümkünse bağımlı gruplarda araştırılması gibi önlemler önerilebilir.

KAYNAKLAR

- Akgül A (1997) Tıbbi Araştırmalarda İstatistiksel Analiz Teknikleri, SPSS Uygulamaları. Yüksek Öğretim Kurulu Matbaası Ankara.
- Düzgüneş O, Kesici T, Gürbüz F (1983) İstatistik Metotları - 1. Ank. Ün. Zir. Fak. Yay. 861, Ders kitabı 229, Ankara.
- Düzgüneş O, Kesici T, Kavuncu O, Gürbüz F (1987) Araştırma ve Deneme Metotları (İstatistik Metotları - II) Ank. Ün. Zir. Fak. Yay. 1021, Ders kitabı:295, Ankara.
- Kutsal A, Alpan C, Arpacık R (1990) İstatistik Uygulamalar, Ankara Ün. Veteriner Fakültesi Zootečni Bölümü. Ankara.
- Minitab for Windows Release 11.12, (1996) Minitab Inc. 814-238-3280.
- Özdamar K (1985) Biyoistatistik, Bilim ve Teknik Yayınevi, I. Baskı, İstanbul.
- Özdamar K (1997) Paket Programlar ve İstatistiksel Veri Analizi - I. Anadolu Üniversitesi yay. No: 1001. Fen Fakültesi Yay. No: 11, Eskişehir.
- Petrie A, Watson P (1999) Statistics for Veterinary and Animal Science, Blackwell Science Ltd.
- Ryan BF, Joiner BL, Ryan Jr TA (1985) Minitab, Handbook. Second Edition, Revised printing, PWS-KENT Publishing Company, Boston.
- Snedecor GW, Chocran WG (1980) Statistical Methods. Seventh edition, The IOWA State Press, Ames Iowa, U.S.A.
- SPSS for Windows. Release 10,0.1. Standart version (1999) SPSS inc.
- Sümbüloğlu K, Sümbüloğlu V (1987) Biyoistatistik. 6. Baskı. Özdemir Yayıncılık, Ankara.