

İSTATİSTİKSEL KARAR VERME SÜRECİNDE GÜÇLÜ TEST SEÇİMİNİN ÖNEMİ VE GENEL İLKELERİ

Prof. Dr. Neclâ ÇÖMLEKÇİ*

Hipotez testi istatistiksel karar vermede çok önemli bir rol oynar; hipotez testi karar sürecinin özel bir türüdür ve mantıksal dayanağı oldukça karmaşıktır. Sosyal bilimlerde «hipotez» terimi farklı anlamlarda kullanılarak bazen dolaylı olarak bir önerme için kullanılmakta, bazen de fiilen istatistiksel olarak test edilebilen bir önerme türü için benimsenmektedir. Bu makalede, istatistiksel olarak testi mümkün önerme için kullanılan hipotezin testi, istatistiksel karar vermede karşılaşılabilecek hata türleri ve bunların dengelenmesi ile güçlü test seçimindeki genel ilkelerin açıklanmasına yer verilmiştir.

1. Hipotez Testinin Niteliği

Ana kütle için ifade edilen hipotezler, «istatistik» veya «örneklemdeğer» olarak isimlendirilen ve örneklem değerlerinin bir fonksiyonu olan tesadüfi değişkenler yardımıyla test edilirler. Bilindiği gibi «sıfır hipotezi» olarak nitelendirilen hipotez bir alternatif hipoteze karşı test edilir.

Bir istatistik hipotezi, tesadüfi bir değişkenin sıklık fonksiyonu için yapılan bir varsayım olarak tanımlanır ve hipotezin testi, sıfır hipotezinin kabulü veya reddinin kararlaştırılması için yapılan işlemler bütünüdür.

(*) Fen-Edebiyat Fakültesi İstatistik Bölümü Başkanı

İstatistik hipotezlerinin çoğu bir sıklık dağılımının parametreleri için yapılan varsayımlardır. Eğer sıfır hipotezi bir sıklık dağılımının bütün parametrelerinin değerlerini belirtiyorsa yalın hipotez olarak isimlendirilir; aksi durumda sıfır hipotezi birleşik hipotez olarak nitelendirilir.

2. Kararlardaki Hatalar

Sıfır hipotezi aslında doğru olduğu halde gözlemlenen örneklemdeğeri önceden belirlenen kritik bölgede yer aldığı için reddedilirse, hatalı bir karar verilmiş olur; bu tip hataya birinci tip hata denilir. Diğer taraftan karşıt hipotez aslında doğru olduğu halde örneklem değerinin kritik değerden mutlak olarak daha büyük olmaması sebebiyle sıfır hipotezi kabul edilecektir; bu da hatalı bir karardır, bu hata türü ikinci tip hata olarak isimlendirilir.

Bilindiği gibi birinci tip hatanın büyüklüğü, sıfır hipotezinin doğru olması durumunda bu hipotezin reddedilmesi olasılığıdır. İkinci tip hatanın büyüklüğü, hatalı olan sıfır hipotezinin kabul edilmesi, veya başka bir anlatımla doğru olan alternatif hipotezin reddedilmesi olasılığıdır. Bu iki hata türü büyüklükleri gözönünde tutularak, hipotezler için güvenilir testler belirlemede uyulacak yalın bir ilke sözkonusudur: hepsi aynı büyüklükte birinci tip hataya sahip testler arasından, ikinci tip hatası en küçük olanı seçilir. Güvenilir test için gözönünde tutulacak bir diğer ilke, iki hata türü büyüklükleri toplamının minimize edilmesidir. Ancak sözü edilen ilk ilkenin çok daha geçerli olduğunu belirtmeliyiz, çünkü her istatistikçi çalışmasının başlangıcında razı olacağı birinci tür hatanın büyüklüğünü belirler. Dolayısıyla örneklemindeki birim sayısı değişmeyecekse araştırmacı testini, ikinci tür hata büyüklüğünü minimize edecek şekilde oluşturacaktır. Ancak örneklemin büyüklüğü değişmedikçe birinci tür hatanın küçültülmesiyle ikinci tür hatanın artacağı unutulmamalıdır; bu durumda birinci tür hata serbestçe ve istenildiği kadar küçük tutulamayacaktır.

Teorik olarak, yanlış bir sıfır hipotezinin doğru diye kabul edilmesi olasılığı olan β 'yi sıfırlamak da mümkündür; deneme veya örnekleme sonucuna bakmaksızın sıfır hipotezini daima reddederek bunu sağlamak mümkündür. Karar verme ilkesi olarak bunun benimsenebilmesi için sıfır hipotezinin reddinin istenmesi gerekir çünkü β 'nin sıfır olması, α 'nın 1 olmasıdır. Böyle bir durumda doğru olan sıfır hipotezi her durumda reddedilecektir. Böyle bir karar ilkesinin temelsizliği, yanlışlığı açıktır.

Örnekleme veya denemeyle elde edilen sonuçlar karar ilkesi bakımından iki grupta toplanabilir. Sıfır hipotezinin reddini gerektiren sonuçların kritik bölgede, diğerlerinin ise kabul bölgesinde yer aldıkları ifade edilir.

İstatistiksel çalışmalarda iki hata türü büyüklükleri arasında yeterli bir denge oluşuncaya kadar birinci hata türünün ayarlanması gerekebilir. Uygulamada, özellikle sosyal bilim dallarında istatistikçilerce benimsenen birinci tür (α) hata büyüklüğü % 5'dir.

2.1. Güç Fonksiyonu

Test edilen sıfır hipotezinin sadece bir tek alternatifi varsa, hipotezler $H_0 : \Theta = 2$ ve $H_1 : \Theta = 1$ şeklindeyse, uygun testin seçiminin açıklanması oldukça kolaydır. Böyle bir durumda ikinci tür (β) hatanın büyüklüğünü belirlemek mümkündür. Ancak araştırmacı çoğu zaman test edeceği parametrenin teorik veya görgül nedenlerle belirleyebildiği halde, H_0 hipotezinin yanlış olması durumunda sözkonusu parametrenin alacağı belirli değeri nadiren bilebilir.

β hatasının büyüklüğü, anlaşılacağı gibi gözönünde tutulan Θ parametresinin alternatif değerine bağlı olacaktır. Seçilen testin uygulanabilecek bir değerine göre ne derecede iyi olduğunu belirleyebilmek için Θ parametresinin bir tek değeri için değil, alabileceği tüm değerler için ikinci tür hataları karşılaştırmak gerekir. Bu hatanın büyüklüğü β (Θ) olarak gösterilecektir. β (Θ), parametrenin gerçek değerinin Θ olması durumunda örneklem değerinin sıfır hipotezinin kabul bölgesinde yer alması olasılığıdır. Ancak sıfır hipotezinin red bölgesi genellikle gözönünde bulundurulduğu için $\{1-\beta$ (Θ) $\}$ değerinin hesaplanması alışkanlığı vardır. $\{1-\beta$ (Θ) $\}$ gerçek parametre değerinin Θ olması durumunda gözlemlenen örneklem değerinin kritik bölgede bulunması olasılığıdır; $\{1-\beta$ (Θ) $\}$ fonksiyonu «güç fonksiyonu» olarak isimlendirilir. Bir testin $P(\Theta)$ güç fonksiyonu, gözönünde bulundurulan parametrenin gerçek değerinin Θ olması durumunda örneklem değerinin testin kritik bölgesinde yer alması olasılığını veren parametrenin fonksiyonudur.

Ana kütle parametresi için birden fazla alternatif değer var olması durumunda karşılaştırma yapıp en uygun testi belirlemede yararlanılan güç fonksiyonu ayrıca parametre değerinin bir fonksiyonu olan doğru karar vermede belirli bir testin etkinliğinin be-

lirlenmesinde de faydalıdır. Güç fonksiyonunu inceleyerek araştırmacı ana kütle parametresinin olabilecek çeşitli alternatif değerlerini ortaya çıkarma olasılıklarını, buna bağlı olarak da teste dayanarak vereceği herhangi bir karardaki güveni belirleyebilir.

2.2. Testin Gücü

Bir testin gücü, alternatif hipotezin doğru olması durumunda sonuçların sıfır hipotezinin red bölgesinde yer alması olasılığıdır. Başka bir anlatımla testin gücü, doğru olan alternatif hipotezinin kabul edilmesi olasılığıdır; bu durumda testin gücü $(1-\beta)$ 'ya eşit olacaktır. Aynı α değeri için belirli bir testin gücü diğer testlerin güçlerinden daha büyükse, sözkonusu test en güçlü test olarak benimsenir.

3. İki Tip Hatanın Dengelenmesi

α ve β hatalarını olabildiğince dengelemek için, gözönünde tutulduğunda faydalı olabilecek bazı ilkeler geliştirmek mümkündür; eğer sıfır hipotezinde ifade edilen olguya güven duyuluyorsa α hatası mümkün olduğunca küçük tutulmağa çalışılmalıdır. Diğer taraftan α ve β hatalarının büyüklüklerini belirlemede her iki hata türü için sözkonusu olacak görelî maliyet etkili olur. α tipi hata yapmanın maliyeti yüksek, β tipi hata yapmanın maliyeti düşükse, β hatası büyüyeceği halde α hatasının mümkün olduğunca küçük tutulmasına çalışılır. Ancak bilimsel araştırmaların birçoğunda sözü edilen hususların belirli bir güvenilirlikle belirlenmesi zordur. Bununla beraber α tipi hatalar genellikle yeterince önemli olduklarından, çoğu zaman % 5 veya % 1 gibi küçük değerlerin benimsenmesi sözkonusudur.

Hipotez testinde örneklem büyüklüğü çok önemli bir rol oynar; örneklem çok büyük ise, sıfır hipotezini reddetmek çok tehlikeli olabilir. Aynı şekilde örneklem çok küçük ise sıfır hipotezini kabul etmek de çok tehlikelidir. Bilindiği gibi örneklem hacmine bağlı olarak aynı olayda standart hata, dolayısıyla kritik değer değişir.

4. Sıfır Hipoteziyle İlgili Kimi Düşünceler

4.1. Sıfır Hipotezinin Kabulüyle İlgili Bir Uyarı

Sıfır hipotezini kabul etmek hata ise ve bu hipotezi red etmek için yeterli dayanağımız yoksa ne olacak? Böyle bir durumda ya-

pılabilecek tek şey bir süre için karar vermekten vazgeçmek, veya daha açık bir ifadeyle sıfır hipotezini redetmemektir. Ancak bu şekilde ikinci tip hata yapma (yanlış olan sıfır hipotezinin kabulü) riskinden kaçınılabılır. Sıfır hipotezindekinden apayrı bir değerde olan parametrenin alternatif hipotezde yer alması (basit alternatif hipotez) durumunda β hatası riskine katlanmak mümkün olmakla beraber, birleşik bir alternatif hipotezin formüle edilmesiyle sözkonusu risk % 95'e yükselebilmektedir. Bu bakımdan, sözü edilen durumda sıfır hipotezi reddedilmeyerek geçici bir süre için kararın ertelenmesi yoluna gidilir veya örneklem istatistiğinin istatistiksel olarak anlamlı olmadığı sonucuna varılır; ancak sıfır hipotezi doğrudan doğruya kesinlikle benimsenmez. Sıfır hipotezinin ne kabul, ne de reddine karar verilemediği durumlarda bir diğer alternatif «olasılık değeri» hesaplamaktır. Kesin karar için belirlenen sözkonusu ölçüt de yeterli olmazsa başkaca tanıt elde edilmesine çalışılır, yani yeni gözlemler yapılır.

4.2. Olasılık Değerinin Hesabı

Çeşitli nedenlerle hipotez testi dikkatle kullanılması gereken bir tekniktir. Aslında güven aralığı belirlemek, bir hipotez testine daima tercih edilmelidir. Güven aralığı gözlemlenen örneklem sonucunu çok açık seçik betimlemekte, buna karşın hipotez testi örneklem istatistiğinin istatistiksel olarak anlamlı olup olmadığını göstermektedir.

Diğer taraftan hipotez testinde özellikle örneklem küçük olduğunda güvensiz bir sıfır hipotezinin kabulünde gerçek problemler sözkonusudur. Örneklem küçük olduğunda testin olasılık değeri (prob-value) hesaplanmalıdır. Gözlemlenen istatistiği T , sıfır hipotezinde yer alan parametreyi Θ , standart hatayı σ_T simgesiyile gösterirsek, olasılık değeri aşağıdaki gibi belirlenir:

$$\text{olasılık değeri} = P \left(Z \geq \frac{T - \Theta}{\sigma_T} \right)$$

Anlaşılacağı gibi olasılık değerinin hesaplanması, α 'nın hesaplanmasına çok benzer. Gözlemlenen T aşırı bir değer ise, olasılık değeri çok küçük olacaktır. Bu bakımdan olasılık değeri sıfır hipotezinin güvenilirliğini ölçer. Eğer olasılık değeri α değerinden küçükse sıfır hipotezi reddedilir, çünkü sıfır hipotezinin güvenilir-

liđi α 'nın altına dūřmūřtūr. Olasılık deđeri anlařılacađı gibi sıfır hipotezinin reddedilebileceđi α 'nın mūmkūn en kūçük deđeridir.

4.3. Sıfır Hipotezinin Seđimi

Genelde yalın sıfır hipotezi birleřik bir alternatif hipoteze karřı test edilir ($H_0: \Theta = a$; $H_1: \Theta > a$ veya $H_1: \Theta < a$ veya $H_1: \Theta \neq a$). Ancak uygulamada her iki hipotezin de birleřik olduđu durumlarla karřılařılabilir. Bōyle bir durumda birleřik sıfır hipotezi eřitli yalın hipotezlerden oluřur; bu olgu ařađıda ōrneklenmiřtir:

$$\begin{aligned} H_0 : \pi_1 = \pi_2 = a \\ \pi_1 = \pi_2 = b \\ \vdots \\ \vdots \\ \pi_1 = \pi_2 = x \quad 0 \leq x \leq 1 \end{aligned}$$

Yukarıda genel bir biimde ifade edilen birleřik sıfır hipotezinin karřıtı daha da karmařıktır:

$$\begin{aligned} H_1 : \pi_1 = a \text{ ve } \pi_2 = b \\ \pi_1 = b \text{ ve } \pi_2 = a \\ \pi_1 = c \text{ ve } \pi_2 = d \\ \vdots \\ \vdots \\ \pi_1 = x \text{ ve } \pi_2 = y \quad 0 \leq x \leq 1 \\ \quad \quad \quad 0 \leq y \leq 1 \\ \quad \quad \quad x > y \end{aligned}$$

Bōyle bir durumda yeni bir δ ana kūtle parametresi tanımlamak en uygun davranıř olur:

$$\delta = \pi_1 - \pi_2$$

Yeni tanımlamadan sonra hipotezler ařađıdaki gibi formūle edilir:

$$H_0 : \delta = 0 \quad H_0 : \delta \neq 0$$

veya

$$H_1 : \delta > 0 \quad H_1 : \delta \neq 0 \{ \delta > 0 \text{ veya } \delta < 0 \}$$

Bazen sıfır hipotezinin deđil, alternatifinin formūle edilmesi istenir; sıfır hipotezi reddedilerek alternatif hipotezin kanıtlanma-

sı sağlanır. Böyle bir durumda sıfır hipotezinin kanıtlanması söz konusu değildir ve alternatif hipotez de ancak sıfır hipotezinin çürütülmesiyle kanıtlanmış olur. Bunun sonucu olarak, kanıtlanması istenen önerme alternatif hipotez, reddi umulan varsayım da sıfır hipotezi olarak formüle edilir.

5. Güçlü Test Belirlemenin Genel İlkeleri

Benimsenecek bir testin güçlü olması için gözönünde bulundurulacak bazı kriterler olmakla beraber, bazı basit genellemeler geliştirmek hiç de kolay değildir. Bununla birlikte, uygulanacak bir hipotez testinin uygulanabilir diğer testlere göre daha güçlü olabilmesi için benimsenebilecek bazı genel ilkeler vardır.

Herşeyden önce belirli bir testin güç etkinliği, oluşturulan örneklem büyüklüğüne bağlıdır: test büyük örneklem için çok etkin olabilmeye karşın küçük örneklem için çok daha az etkin olacaktır.

Diğer taraftan parametrik testler parametrik olmayanlara göre daha güçlüdür, dolayısıyla imkân olduğu takdirde parametrik testler tercih edilmelidir. Ancak uygulanacak bir test, test edilmesi mümkün olmayan bazı varsayımlara dayandırılıyorsa gözardı edilmesi gerekir. Böyle durumlarda sözü edilen varsayımları gerektirmeyen testler daima tercih edilmeli, daha az güçlü olmalarına rağmen parametrik olmayanlar benimsenmelidir.

Parametrenin değeri sıfır hipotezinde ileri sürülen değerden fazla farklı değilse testin gücü düşük, aksi durumda yüksek olacaktır. Başka bir anlatımla β riski, gerçek değer ile sıfır hipotezinde ileri sürülen değer birbirine yakın olması durumunda oldukça büyük olacaktır.

Ayrıca tek ve çift taraflı testlerin de testin gücünü etkilediği bilinmektedir: tek taraflı test benimsenip örnekleme dağılımının sağ ucu kritik bölge olarak tanımlanmışsa, parametrenin gerçek değeri de sıfır hipotezinde ileri sürülenden daha büyük ise, tek taraflı testin gücü, çift taraflı bir teste göre daha büyük olacaktır. Fakat ana kütle parametresinin gerçek değerinin sıfır hipotezinde ileri sürülen değerden daha küçük olması durumunda sıfır hipotezini red etme imkânımız hiç olmayacağı için tek taraflı testin gücü çok düşük olacaktır. Doğal olarak tek taraflı bir testte kritik bölge örnekleme dağılımının sol ucunda belirlenmiş, ancak ana kütle

parametresinin gerek deęeri sıfır hipotezinde ileri srlenden ok daha byk ise, testin gc dřk olacaktır.

zetleyecek olursak, alternatif hipotezde belirtilen deęerler kritik blgenin ynnde ise aynı anlam dzeyi iin tek taraflı testlerin gc, ift taraflılara gre daha fazladır. Buna karřın ana kt-le parametresinin gerek deęeri sıfır hipotezinde belirtilenin tersi ynnde ise, ift taraflı testin gc tek taraflı testinkine gre daha byk olur.

KAYNAKLAR

- 1 — Blalock, H.M., Social Statistics, Mcgraw-Hill MLADINSKA KNJIGA Ljubljana, 1960.
- 2 — Bryant, E.C., Statistical Analysis, McGraw-Hill Tokyo 1966.
- 3 — Hoel, P.G., Jntroduction to Mathematical Statistics, John Willey and Sons, Inc., New York 1966.
- 4 — Kurtuluř, K., Pazarlama Aratırmaları, İstanbul 1981.
- 5 — Mendenhall, W., Introduction to Linear Models and the Design and Analysis of Experiments, Wadsworth Publishing Company, Inc., Belmont 1968.
- 6 — Wonnacott, R.J., Wonnacott, T.H., Introductory Statistics, John Willey and Sons, Inc., New York 1969.