



## DÜŞÜNCE YAZISI-OPINION PAPER

### Denence Testi ve $H_0$ Denencesinin Reddedilememesinin Dayanılmaz Ağırlığı

Adnan Erkuş, *Emekli Profesör*, [adnanerkuspsi@gmail.com](mailto:adnanerkuspsi@gmail.com)

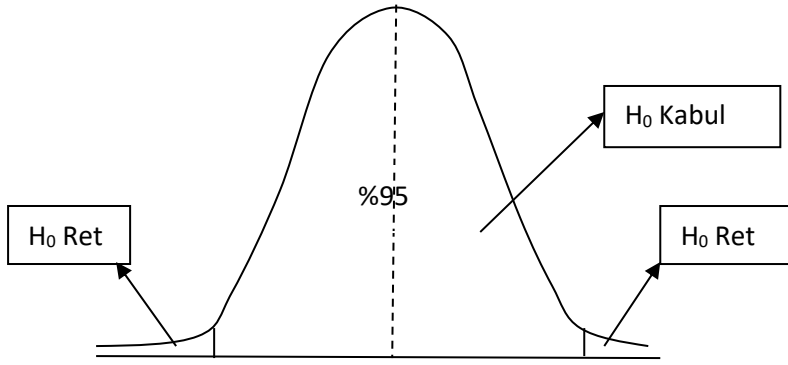
Aşağıdaki yazı, özellikle belirli bir zaman sınırlaması içinde bitirilmesi gereken yüksek lisans ve doktora tez çalışmaları sırasında, araştırma denencelerinin (hypothesis) “doğrulanmaması” üzerine danışmanlarından veya jüri üyelerinden “zılgıt yiyen” genç arkadaşlarımla “gayrimeşru” savunmalarına katkı sunmak ve eğer ulaşılabılırsa söz konusu danışman ve jüri üyelerinin bir nebze olsun farkındalıklarını artırmaya (haddimiz olmadan) hizmet etmek amacıyla kaleme alınmıştır. Elbette sözümüz, görgül (empirical) çalışma yapan genç arkadaşlara...

Her araştırmacı, araştırmasına bir iddia/önerme ile başlar, bu iddianın karşılığı istatistiksel olarak  $H_1$ ; tersi ise  $H_0$  denencesidir. Örneğin;

Yokluk Denencesi ( $H_0$ ): *Paranoyaklar ile normal bireylerin kuşku ölçeğindeki ortalamaları birbirinden farklı değildir.*

Araştırma (Karşıt) Denencesi ( $H_1$ ): *Paranoyaklar ile normal bireylerin kuşku ölçeğindeki ortalamaları birbirinden farklıdır.*

Bunlar “çift uçlu” denencelerdir; yani fark ne yönde (eksi veya artı) çıkarsa çıksın araştırma denencesi doğrulanmış olur ve çift uçlu denenceler, farkın yönünün kestirilemediği yeni bir konu üzerinde yapılan çalışmalarda daha çok tercih edilir. Denence, “büyüktür”-“küçüktür” gibi yön bildirseydi, “tek uçlu” denence şeklinde olur ve bu da daha çok farkın yönünün önceden kestirildiği ve kontrolün sağlandığı deneysel çalışmalarda tercih edilir. İki arasındaki ayırım; beklentideki kesinlik/muğlaklığın yanında, farkın yönü (işareti) ile I. Tip hata yapma olasılığının ( $\alpha$ ; bilgisayar çıktısında “p”) büyüklüğünde ortaya çıkar. I. Tip hata yapma olasılığı, çoğu çalışmada maksimum %5 (0,05) olarak alınır. Çift uçlu denencede bu hata yüzdesi, dağılımın eksi ve artı yönünde ikiye bölünüp dağıtılır ve her yön için 0,025’e düşer; böyle olunca dağılımın X eksenini boyunca, aşılması gereken kuramsal/beklenen değer *büyür*. Bu, çift uçlu denenceye ödenen bir bedeldir. Oysa, tek uçlu denencede fark sadece belirli bir yönde beklediği için beklenen değer (0,05 dağılımının bir ucunda olacağı için) *küçülür* ve yokluk denencesini reddetmek *kolaylaşır*. Bu da, ne beklediğini bilen ve kontrollü çalışan tek uçlu denence sahibinin ödülüdür.



**Şekil:** İki uçlu denencede  $H_0$  ret ve kabul bölgeleri

Toplanan görgül veriler üzerinde uygun istatistiksel analizler yapılarak öncelikle bir “gözlenen istatistiksel değer” elde edilir. Örneğin, yukarıdaki örneğimizde kuşku ölçeğindeki puanların aritmetik ortalamasının ve standart sapmasının hesaplanması *uygunsa*, iki grup olduğu için (elbette veriler normal dağılıyorsa ve varyanslar eşitse) bu denencenin istatistiksel olarak test edilmesi “*bağımsız gruplar için t-testi*” ile test edilebilir. Formüldeki değerler bulunup yerlerine konduktan sonra bulunan t-değeri, *gözlenen değer*dir. Bundan sonra bu *gözlenen değer*, t-kuramsal değer dağılımında belirli bir I. Tip hata yapma olasılığında ve örneklemimizdeki birey sayısından hareketle bulunan serbestlik derecesindeki *beklenen değer* ile karşılaştırılır. Eğer, Gözlenen Değer > Beklenen Değer olursa, yokluk denencesi reddedilir, araştırma denencesi kabul edilmiş (doğrulanmış) olur. Bu karşılaştırmayı bir bilgisayar programıyla yapıyorsanız, siz beklenen değeri görmezsiniz, sadece hata olasılığı (p) verilir; farkın (ya da ilişkinin) istatistiksel anlamlılığı (significance) için bu değer de 0,05’ten küçük olması istenir. Elbette, tek uçlu denence kurulmuşsa, bulunan gözlenen değer in işareti de dikkate alınır: “Büyüktür” deyip, eksi bir değer çıkmışsa, gözlenen değer, mutlak değer içinde beklenen değeri aşmış olsa da araştırma denencesi doğrulanmamış olur.

Haydi, K. R. Popper’ın (1961) çok bilinen “kuğu” örneğini verelim: “Tüm kuğular beyazdır” ( $H_1$ ) araştırmacının test etmek istediği denence olsun; bunun karşıtı (aslında tersi olacak ya neyse) ise “Tüm kuğular beyaz değildir” ( $H_0$ ) denencesidir. Hayal edin, siz bu çalışmayı yapacak bir tez adayısınız ve bunun için tonlarca alanyazın taradınız, öneri yazdınız ve bu önerinizi danışman ve jüri önünde kabul gördü, sonra çıktınız veri topladınız, analizler yaptınız, vee... Tüm kuğuların beyaz olmadığı (lanet olası *bir tek* siyah kuğu, her şeyi bozuyor!) ortaya çıktı, ne yaparsınız; siz değil belki ama danışmanınız veya jüri üyelerinden bazıları (ülkemizde olsalar) ne der? “Böyle araştırma mı olur, ne yap yap bize tüm kuğuların beyaz olduğunu kanıtla!” Neyse uzatmayalım, bazılarınız olasıdır ki buna benzer bir olay yaşadınız (kendim yaşamadım ama çalışma alanımdan dolayı çok tanık oldum) veya bundan sonra bazılarınızın buna benzer bir durumla karşılaşma olasılığı yüksek ne de olsa...

Bu genç arkadaşımıza bir nebze olsun yardım edelim de “aforoz” edilmesini önlemek için en azından kaynak olarak bu yazıyı gösterebilsin...

1. Yukarıdaki denencelerden herhangi birinin (yokluk veya karşıt) kabul ya da reddedilmesinin, “iyi”-“kötü” gibi *insani değer yargıları* ile herhangi bir ilişkisi bulunmamaktadır. Çünkü, ret ya da kabul kararı, tamamıyla görgül olarak toplanan verilere ve bu verilere uygun yapılmış olan istatistiksel sonuçlara bağlıdır. Elbette, her araştırmacının “gönlünden geçen” araştırmaya başlarken öne sürdüğü iddianın ( $H_1$ ) kabul edilmesidir; ancak, araştırmaya başlarken araştırmacının oluşturduğu iddia sadece ve sadece bir *bilişsel* üründür (*fikir düzeyinde*), ‘gerçek’ değildir; zaten veri toplayıp araştırma yapmanın esprisi de buradadır. Aksi halde, hiç araştırma yapmaz, sadece “kendi fikirlerimizi” öne sürer dururduk (son moda olan ‘bazı’larına ‘bazı’ şeyleri hatırlatıyor olsa gerek)... “Denence test etme” adı üstünde, bilişsel/kurgusal iddiamızın *verilere dayanarak* “test” edilmesidir. Bir *denence testinde (bazı istisnalar hariç) hemen daima  $H_0$  denencesi test edilir; çünkü örneklem üzerinde çalışılmaktadır ve evren dağılımı bilinirken örneklem dağılımı örnekleme değişir, bilinmez.* Denenceleri test etmede aşağıdaki olası durumlar ve hatalar söz konusudur:

**Tablo 1:** Denence Test Etmede Olası Durumlar

		Gerçek Durum	
		$H_0$ doğru	$H_0$ yanlış
Verilen Karar	$H_0$ Ret	I. Tip Hata yapma olasılığı ( $\alpha$ )	Doğru karar
	$H_0$ Kabul	Doğru karar	II. Tip Hata yapma olasılığı ( $\beta$ )

2. Üstelik iddiamızın kabulü (yani  $H_0$  ret,  $H_1$  kabul), bulunan sonucun “mutlak/değiştirilemez gerçek” olduğu anlamına da gelmez. Çünkü (eğer tüm evren üzerinde çalışılmamışsa), yukarıda da değinildiği gibi, örneklemden elde edilen verilerin analiziyle elde edilen *gözlenen istatistiksel değer, kuramsal/beklenen evren değeriyle belirli bir olasılık içerisinde (I. tip hata yapma olasılığı) ve serbestlik derecesinde karşılaştırılarak* ret ya da kabul kararı verilir. Bu karar da (normal dağılım için)  $-\infty$  ile  $+\infty$  integral aralığında (yani kapalı/kesin sınırları olmayan) bir boyutta test edilir. Bu bakımdan, bilimsel çalışmaların sonuçları asla “kesin” değildirler, daima bir hata olasılığı taşırlar. K. R. Popper’ın öne sürdüğü gibi, “bilim yanlışlanabilirlik” üzerine kuruludur. “Tüm kuğular beyazdır” denencesi “siyah bir kuğu” bulununcaya kadar geçerlidir. Bulduğunuz sonuç isterse, milyonda bir hata olasılığıyla bile doğru olsa, bu *çürütülünceye kadar geçerlidir*. Üstelik bir tek çalışma bile bunu sağlamaya yeter. Günlük yaşamdaki yalanlar ile bilimsel ‘yalanlar’ arasındaki fark burada yatar ve bu da bilimde asla yalan/intihal yapılamayacağıının garantisini oluşturur. Bilimsel çalışmayı dogmalardan ayıran en önemli özellik yanlışlanabilirliktir, aksi halde her öne sürülen iddia bir dogma niteliği taşırdı.
3. Denence testinde, madem ki değer yargılarına yer yok, öyleyse *test sonucunda çıkan her sonuç bir bilgi taşır.  $H_1$ ’in kabulü nasıl ki geçici bir doğruluk bilgisi taşırsa, reddi de bir doğruluk bilgisi taşır*. Bilim tarihi, belirli olasılıklarla defalarca doğrulanan iddiaların, gerçekte doğru olmadığına ilişkin çalışmalarla doludur. “Yumurta-kolesterol

ilişkinini hatırlamanız yeterli. Aksi halde, “saçı uzun aklı kısa”, “siyahların hayvanlara yakın bir ırk olduğu” gibi arkaik ‘önergelerin’ halâ geçerli olması gerekirdi.  $H_1$ ’in kabul,  $H_0$ ’ın reddedilmesi elbette, araştırmacıyı duygusal anlamda mutlu eder, ama bu mutluluk çok uzun sürmeyebilir.  $H_0$ ’ın kabul edilmesi de elbette duygusal anlamda hayal kırıklığı yaratır, ama dünyanın sonu değildir; çünkü o da bir bilgidir; en azından iddianın inşa edildiği var olan destek bilgilerinden kuşku duyulmaya başlanmasına yol açar ve yeni arayışları beraberinde getirir ve yeni çalışmalarla belki de var olan destek bilgilerinin tümünden reddine ve yeni bilgiler üretilmesine neden olabilir.

Peki  $H_0$  kabul çıkarsa (yani öngörülerimiz doğrulanmazsa) bu ne anlama gelir veya neden kaynaklanabilir?

- a) Araştırma kurgusu ve problemi uygun olmayabilir; zaten fark çıkması olanaklı olmayan bir karşılaştırma yapıyor olabilirsiniz, yani bağımlı değişken (kuşku) veya bağımsız (paranoya/normal) uygun olmayabilir, daha uygun bir başka bağımlı/bağımsız değişkenle çalışılması gerekiyordur. Bağımlı değişken, bağımsız değişkenlerdeki değişimlere bağlı olarak değişiklik göstermesi beklenen değişkendir. Ya ilgilendiğiniz özelliğin ayırt edici özelliği başka bir şeyse? Paranoyaklar ile normallerin boy uzunluklarının farklı olup olmadığını incelemek ne kadar uygun olur? Elbette, araştırma probleminin veya kurgusunun uygun olması sizden çok, danışmanınızın problemidir...
- b) Her denence testi, görgül verilere dayanır ve bu veriler bir *örneklemdeki deneklerden* elde edilir; bu bakımdan, örneklemin nasıl belirlendiği (birim sayısı, seçilme yönteminin seçilenlerin ilgili özellikleri açısından uygun olup olmadığı) çok önemli hale gelir. Belirlenen uygun sayıda birim var olabilir, ama deneklerden beklenen nitelikler uygun olmayabilir: Paranoya tanısı konmuş olan aslında paranoyak değil (veya paranoya açısından değişik düzeylerde), normal kabul edilenler içinde de gizli paranoyaklar olabilir. Bu durum, araştırma örnekleminin belirlenmesinin (örnekleme birim seçilmesinin) ne kadar önemli olduğunu gösterir.

Her denence testi, istatistiksel dağılımlar ve testler gereği, örneklemdeki birey sayısına oldukça duyarlıdır. Bu nedenle serbestlik derecesi, belirli bir hata olasılığındaki beklenen değeri saptayabilmek için en önemli parametredir. Nedeni de basitçe şudur: Örneklemdeki birey sayısı ( $n$ ), evrendeki birey sayısına ( $N$ ) yaklaştıkça, gözlenen değer de evrenin beklenen değerine, hata da sıfıra yaklaşır. Bu bakımdan,  $n$  arttıkça ( $H_0$  gerçekten yanlışsa)  $H_0$  denencesini reddetmek de kolaylaşır. Küçük  $n$ ’lerle yapılan çalışma sonuçları bu nedenle daima kuşku ile karşılanır: 10-15 kişilik bir paranoyak grubunun tüm paranoyakların özelliklerini taşıması beklenebilir mi, üstelik paranoid bozukluk açısından paranoyaklar arasında derece farklılıkları (ağır/hafif gibi) varsa? Hele hele bu çalışmaya dayanarak tüm paranoyaklar için *tanı ölçütleri çıkarılacaksa* çok çok büyük örneklem üzerinde çalışılması gerekir. Öte yandan, istatistiksel testler serbestlik derecelerine duyarlı olduğu için, 2,02 gibi bir gözlenen  $t$  değeri 15-20 birey için anlamlı bir fark çıkarmazken, 100 kişi için anlamlı bir fark ortaya çıkarır. Küçük örneklem büyüklüklerinden elde edilen veriyle (veri toplama vb sağlıklı olsa da) denencenin doğrulanmaması şaşırtıcı değildir. Bu durumda, a) büyük örneklemde çalışmanın tekrar edilmesi tavsiye edilir (veya yapılır); özellikle bazı ölçümlerde fark

çıkma yakın değerler elde edilmesi bunu destekler; b) SPSS gibi paket programların çoğu çift uçlu p değerleri verir, oysa tek uçlu denencede ilgili p değeri anlamlı olabilir (X ekseninde tablo değeri küçüleceğinden) veya c) denence testindeki gibi örneklem büyüklüğünden etkilenmeyen “etki büyüklüğü” indeksi hesaplanabilir. Ha, küçük örneklerde fark çıkmaz mı, elbette çıkar. Bunun için, deney ve kontrol gruplarının çok iyi seçildiği (bağımlı değişken açısından her grubun kendi içinde varyansın minimum olduğu), ancak bağımsız değişken açısından farkın beklendiği kontrollü deneysel çalışmalarda bu sağlanabilir.

Dolayısıyla, örneklem seçimi ve büyüklüğü yokluk denencesinin kabulünde kullanılacaklar arasındadır.

- c) Veri toplama aracı hatalı (güvenilir ve geçerli olmayan) ölçümler yapıyor olabilir. Bu nedenle “kuşku puanları” da kuşkulu duruma düşer ve salt bu nedenle gruplar arasında fark çıkmamış olabilir.
- d) Veri toplayan, veri toplama standartlarına uymamış olabilir: Bir beyin dalgası ölçümünde başlangıç ve sonlandırma kurallarına uymama, elektrodları uygun yerlere uygun biçimde yerleştirmeme gibi.
- e) Veriler uygun toplanmış olabilir, ama verileri analize uygun duruma getirmek için başvuru ve güvenilir olduğu sanılan yöntem/teknik/indeks vb gerçekte doğru olmayabilir. Bunun doğruluğunun test edilmesi başlıca başka bir temel çalışma gerektirir. Ancak yukarıdaki koşullar (sayı, seçim, veri toplama vb) doğru ve yeterli ise, bu yöntem/teknik/indeks vb den kuşulanmak için yeterince kanıt ortaya çıkar.
- f) Yukarıdakilerin hepsi uygun olabilir, ama en son başvuru denence testindeki istatistiksel teknikler uygun olmayabilir: Örneğin veriler kesiklidir (oran vb hesaplamaya uygundur), ama parametrik (t-test gibi) test yapılmıştır veya tersi...

Sonuç olarak şunlar söylenebilir:

- Eğer fark bekleniyorken çıkmamışsa, çalışmanızda öncelikle yukarıda ele alınan noktalarda ölçümlerinize hata karışıp karışmadığını irdelemeniz gerekir. Eğer olanağınız varsa bu hata kaynaklarını ve eksikliklerini giderip çalışmanızı tekrar etmeniz en doğrusudur. Eğer olanağınız yoksa, ki çoğu tez çalışmasında yoktur, farkına vardığınız bu eksiklikler ve hatalar konusunda bu çalışmayı okuyup başka çalışmalar yapacak olanları uyarmalısınız. Ancak fark çıkmadı diye verilerle vb oynayarak zorla denence doğrulatma yoluna asla başvurmamalısınız! Asıl yanlış olan budur!
- Eğer, yukarıda ele alınan eksik ve hatalar yoksa, (araştırma kurgusu ve probleminin yanlış olması dışında) hiç endişelenmeyin, doğru olduğu sanılanların doğru olmayacağına ilişkin bir bilgi üretmiş olabilirsiniz. Çünkü, denencelerinizin doğrulanmaması da önemli bir bilgidir. Herkesin doğru olarak bildiği bir bilginin yanlış olduğunun gösterilmesi beklentilerinizin doğrulanmasından daha da önemlidir: “Saçı uzun, ama akli da uzun!”

Popper, K. R. (1961). *The logic of scientific discovery*. New York: Science Editions.