

İSTATİSTİKSEL MANİDARLIK - NE? (*)

Thoams C. O'BRIEN ve
Bernard J. SHAPIRO

Çeviren : Doç. Dr. İlhan AKHUN

Eğitimde yapılan arařtırmaların bir çoğunun amacı iki (veya daha fazla) eğitim politikasından hangisinin daha etkili olduğunu saptamaktır. Bu konu üzerinde Johnson ve Brownell tarafından yazılan mükemmel makaleler son zamanlarda yayınlanmıştır.¹ Bu makalenin amacı böyle arařtırmanın önemli olan fakat kısmen ihmâl edilen yönüne ışık tutmaktır.

Eğitim politikalarının istatistiksel olarak yapılan karşılařtırmalarının çoğunda - ve bu konuda yapılan diğeri birçok istatistiksel karşılařtırmalarda - arařtırmacı istatistiksel manidarlık kavramını kullanır. Aritmetik ortalamalar ve diğeri istatistikler arasındaki farklar, bunların şans ile ortaya çıkmasının olasılığı, arařtırmacının saptadığı sınırdan daha az ise, bunların tesadüfen ortaya çıkmasının olasılığı 100 kezde 5 kez veya daha azdır. Arařtırmacı bu .05 sınırını önceden benimsemiş ise ve elde ettiği sonuçların yalnızca tesadüfi faktörlere göre olmasının olasılığını .05 veya daha az olarak bulur ise, o zaman yalnızca tesadüfi faktörlerin etkisinin olduğu hipotezini reddeder, bazı sistematik etkinin örneğinin, belli bir eğitim politikasının etkisi gibi iş başında olduğunu söyler ve bunu söylerken 100 kezde 5 kez yanılmış olmayı göze alır.

Mesleki literatürde, birçok önemli arařtırmanın istatistiksel bakımdan manidar sonuçları rapor ettiği görülecektir. Bu-

(*) Thomas C. O'Brien and Bernard J. Shapiro. «Statistical Significance-What?», *The Mathematics Teacher*, 1968, 61, 673-676.

1. Donovan A. Johnson, «A Pattern for Research in the Mathematics Classroom», *The Mathematics Teacher*, LIX (May 1966), 418-25; and W. A. Brownell, «The Evaluation of Learning Under Dissimilar Systems of Instruction», *The Arithmetic Teacher*, XIII (April 1966), 267-74.

nun kendisi ilginin nedeni olmayabilir. Bununla beraber, bu durumun editör ve araştırmacıların istatistiksel manidarlığı başarı ve önem ile tanımlama ve aynı zamanda istatistiksel bakımdan manidar olmayan sonuçları hemen reddetme biçimindeki bir eğilimi yansıttığını söyleyebiliriz. Bu makalenin çabası istatistiksel manidarlığın çoğu kez daha sonra yapılacak bir araştırmaya temel olmaktan çok kendisinin bir amaç olduğunu ortaya koymaktır.

Somut örnekler bakımından istatistiksel manidarlığı incelemek amacıyla, dört madeni paranın atılması ile ilgili iki dizi denemeyi düşünelim. Bu denemelerin her birinde elde edilen turaların oranı sabittir: Birinde .52 ve diğerinde .70. Açıkcası, böylesine mükemmel oranların pratikte meydana gelmesi oldukça enderdir. Fakat bunlar bundan sonraki tartışmaya büyük katkıda bulunacaktır.

Bütün bu durumlardaki hipotez turaların toplam atış sayısına göre beklenen oranının .5 veya başka bir anlatımla, her denemede kullanılan madeni paranın yansız olduğu konusundaki deneme hipotezidir (H_0). Yüzde 5 manidarlık düzeyi seçilmiştir. Bu, H_0 'e göre, .5'den her iki yönde olan sapmaların gözlenen sapmalara eşit veya daha büyük olmasının olasılığı 100 kezde 5 kez veya daha az ise, H_0 'ın reddedilmesi ve madeni paranın yanlı olduğu sonucuna varılması demektir. .5'den olan sapmaların deneysel sonuçların her birinde elde edilene eşit veya daha büyük olmasının olasılığı (P) Edward² tarafından verilen normal eğriye yaklaşan binom dağılımı ile saptanır. Deneysel veriler Tablo: 1 ve 2'de gösterilmiştir.

Tablo : 1. Elde Edilen Turaların Oranının .52 Olduğu Madeni Para Atma Denemeleri

Deneme	Atış Sayısı	Beklenen	Elde Edilen	P	H_0 Reddet
		Turaların Oranı	Turaların Oranı		
1a	25	.50	.52	.84	Hayır
1b	250	.50	.52	.52	Hayır
1c	2500	.50	.52	.04	Evet
1d	25000	.50	.52	.0001	Evet

2 A. L. Edwards, *Experimental Design In Psychological Research* (Rev. ed.; New York: Holt, Rinehart and Winston, 1969), s. 49.

Tablo : 2. Elde Edilen Turaların Oranının .70 Olduğu Madeni Para Atma Denemeleri

Deneme	Atış Sayısı	Beklenen Turaların Oranı	Elde Edilen Turaların Oranı	P	H ₀ Reddet
2a	10	.50	.70	.20	Hayır
2b	20	.50	.70	.07	Hayır
2c	40	.50	.70	.01	Evet
2d	80	.50	.70	.001	Evet

Bu tablolarda temel olarak dört tür bulgu vardır :

1. Küçük farklar vardır ve bunlar istatistiksel bakımdan manidar değildir (deneme 1a ve 1b).

2. Küçük farklar vardır, fakat bunlar istatistiksel bakımdan manidardır (deneme 1c ve 1d).

3. Büyük farklar vardır, fakat bunlar istatistiksel bakımdan manidar değildir (deneme 2a ve 2c).

4. Büyük farklar vardır ve bunlar istatistiksel bakımdan manidardır (deneme 2c ve 2d).

Bir an için, yalnızca gözlenen farkların istatistiksel bakımdan manidar olduğu yani null hipotezin reddedildiği sonuçlar (2 ve 4) üzerinde duralım. Diğer şeyler eşit olduğunda, bu veriler, yalnız gözlem sayısının (bu durumda, atış sayısı) yeteri kadar büyük olduğunda, hem küçük (Tablo 1.) ve hem büyük (Tablo 2.) farklar için istatistiksel manidarlığı (en azından belli bir evren için) gösterir.³

Bununla beraber, veriler istatistiksel bakımdan manidar olan bütün farkların eşit büyüklükte olmadığını da belirtir. Paralardan bir tanesi çok az yanlı olarak düşünülürken, diğeri kesinlikle daha çok yanlı olarak görülecektir. Bu, istatistiksel manidarlığın yalnızca beklenmedik bir olayın ortaya çıktığını gös-

3. Bu olması gereken şeydir. İstatistiksel manidarlık testlerinde bir örneklemden giderek bu örneklemin alındığı evren hakkında yapılan anlam çıkarmalar ile ilgileniriz. Evrenin elemanları hakkında ne kadar çok bilgiye sahip olunursa, şans ve hatanın rolü o kadar azalır.

termesi demektir. Olayın derecesi ve önemi hakkında hiç bir şey belirlenmez. Birçok araştırmacı çoğu kez istatistiksel manidarlık ve pratik bakımdan önemlilik arasındaki bu fark üzerinde yeterince durmaz.

Gözlenen farkların miktarının değerlendirilebileceği çeşitli yaklaşımlar vardır. Araştırmacı şu bakımlardan konuşabilir : (1) standart sapmalar (örneğin, grup A'nın ortalaması grup B'nin ortalamasından X standart sapmaları kadar büyüktür), (2) W^2 , belli bir deneme faktörü⁴ tarafından açıklanan varyans yüzdesi veya (3) Cohen⁵ tarafından açıklanan bir veya daha çok çeşitli göstergeler. En azından, araştırmacı bu farkları bunların pratik önemi bakımından her zaman açıklayabilir. Örneğin, farklar normlaştırılmış testler için sınıf puanları (örneğin, grup A'nın ortalaması grup B'den üç ay daha yüksektir) veya hatta farkın temsil ettiği test maddelerinin sayısı bakımından değerlendirilebilir. Bunların veya diğer yaklaşımların hangisinin en uygun olduğu eldeki belli bir probleme bağlı olacaktır.

Şimdi istatistiksel bakımdan manidar olmayan sonuçların oluşturduğu, yani H_0 reddedilmediği, denemeleri düşünelim. Bir kimsenin ilk isteği başlangıçta araştırmanın yapılmasına neden olan karşıt hipotezi yeniden formüle etmesi veya terk etmesidir. Çünkü araştırmacı istatistiksel manidarlık olmaksızın istatistiksel anlam çıkarmaya göre evren hakkında bir genelleme yapamaz.⁶

Bununla beraber, bir kimsenin hemen karar vermemesi gerekir. İstatistiksel manidarlığa ek olarak, araştırmacı deneme hipotezini doğru olarak reddetmenin (örneğin, H_0 gerçekte yanlış iken onu reddetmek) olasılığı olan güç kavramı ile de ilgilenmelidir.

-
4. W. L. Hays, *Statistics for Psychologist* (New York : Holt, Rinehart and Winston, 1963). Bl. X, XII, XIII.
 5. J. Cohen, «The Statistical Power of Abnormal-Social Psychological Research,» *Journal of Abnormal and Social Psychology*. LXV, (July, 1962).
 6. İstatistiksel manidarlık olmaksızın evren hakkında anlam çıkarmalar yapılamamasına karşın, araştırmacı farkın değerini mevcut durumu olduğu gibi yargısal olarak her zaman değerlendirebilir. Eşler arasındaki 50 ZB'lik bir fark istatistiksel manidarlık için her zaman test edilemez, fakat muhakkak ki bunun pratik öneminin olması olasıdır.

Bir an için, bir biyolojistin mikropları araştırmak için bir mikroskop kullandığını düşünün. Mikroskop ne kadar güçlü olursa, daha küçük olan mikropları daha iyi görebilir.

Manidar olmayan farkların bulunması durumunda araştırmacının çarçabuk karar vermeden önce onun kullandığı mikroskopun tipi, örneğin testin gücü hakkında bilgisinin olması daha iyi olur. Araştırmacı mikroskopun (örneğin, gücünün) aranan mikropları (farkı) ayırt etme gücünün yalnızca yüzde 20 olduğunu saptarsa, denemesini tekrarlamayı düşünebilir ve daha büyük N kullanarak gücünü artırabilir. Bazı araştırmacıların inanmış olarak görüldüğü gibi önemli bir şeyi bulamamanın ötesinde, araştırmacı çok önemli bir şeyi ortaya çıkarmanın eşliğinde olabilir. Bu yalnızca onun mikroskopunun yeteri kadar güçlü olmamasıdır. Bu durumda araştırmacının yapacağı şey en azından bulgularını rapor etmek - mikroskopunun zayıf olduğunu olgusunu da ekleyerek - ve diğerlerinin araştırmayı kendisinin bıraktığı yerden devam etmesini sağlamaktır. Beş parametre - örneklem büyüklüğü, N; manidarlık düzeyi; belli bir farkın ortaya çıkartılmasının olasılığı (güç); farkların büyüklüğü; ve σ , evren standart sapması - bunlardan dördünün saptanması beşinciye belirleyecek biçimde birbiri ile ilişkilidir. Gerçek bir N ile çalışarak, belli bir farkı ve σ 'nın değerini elde etmekle ve bir manidarlık düzeyini önceden kararlaştırmakla, araştırmacı testinin gücünü basit hesaplamalarla saptayabilir. Çeşitli istatistiksel testler için gücün hesaplanmasına ilişkin yöntemler Walker ve Lev⁷, Guilford⁸, Hays⁹, Dixon ve Massey¹⁰, Feldt¹¹ ve Edwards¹² tarafından tartışılmıştır.

7. H. Walker and J. Lev, *Statistical Inference* (New York: Holt, Rinehard and Winston, 1953), Bl. II, IV.
8. J. P. Guilford, *Fundamental Statistics In Psychology and Education* (4 th ed.; New York: McGraw-Hill, 1951), s. 210 - 14.
9. Hays, a.g.e., s. 269 - 80.
10. W. Dixon and F. Massey, *Introduction to Statistical Analysis*, (New York: McGraw-Hill, 1951), s. 207 - 20.
11. L. S. Feldt and N. W. Mahmoud, «Power Function Charts for Specification of Sample Size In Analysis of Variance,» *Psychometrika*, XXIII (1958), 203 - 10.
12. Edwards, a.g.e. s. 94 - 100.

Buraya kadar yalnızca halihazırda tamamlanmış olan deneylerin sonuçlarını ele aldık. Bununla beraber, sadece **post hoc** yorumlara dayanmak yerine, araştırmacının olası sonuçları başlangıçta planlaması hem olası ve hem de tercih edilebilir olması not edilmelidir. Deneme - öncesi planlamada araştırmacı manidarlık düzeyini önceden saptayabilir. (Son araştırmaların çoğunda genellikle önceden saptanan tek değer), arzu edilen güç derecesi, ortaya çıkarmak istediği bir fark; ve çoğu kez, önceki deneyimine göre evren standart sapmasının en azından yaklaşık değerini hesaplayabilir. Bu durumda sorun yine bu koşulları gerçekleştirmek için gerekli olan örneklem büyüklüğünü (N) saptamak için yalnızca bir hesaplama yapmaktır.

En iyisinden, N önceden seçilen manidarlık düzeyi, güç, farkların büyüklüğü ve evren standart sapmasının yaklaşık değeri ile saptanmalıdır. En kötüsünden, araştırmacı en azından yukarıda önerilen **post hoc** yaklaşımları kullanmalıdır. Her iki durumda, elde edilen farkların büyüklüğünün açıklanmasına gereken önem verilmelidir.