

İstatistiksel Anlamlılık (/Statistical Significance) ve İstatistiksel Güç (/Statistical Power)

Prof.Dr. Ömer Cevdet Bilgin
Atatürk Üniversitesi

Tüm popülasyona ait bir nüfus sayımı istatistiksel çıkarımı gereksiz kılar, çünkü herhangi bir fark veya ilişki, ne kadar küçük olursa olsun, “hakikidir” ve mevcuttur. Mamafih, bir nüfus sayımı nadiren yürütülür. Bundan ötürü, araştırmacı ilgilendiği bir popülasyondan çektiği bir örnekten o popülasyon için çıkarımlar yapmak zorundadır.

İstatistiksel güç aslında yanlış olan bir sıfır hipotezini örnek verinin sağladığı kanıta dayanarak reddetme olasılığıdır ve şunların bir fonksiyonudur:

1. Araştırmacı tarafından I. Tip hata için tespit edilen istatistiksel anlamlılık düzeyi (α)
2. Analizde kullanılan örnek hacmi
3. İncelenen etki miktarı

İstatistiksel çıkarımların interpretasyonu, nüfus sayımı yerine örnek kullanmaktan dolayı meydana gelen istatistiksel hatanın (örnekleme hatası) kabul edilebilir bir düzeyini spesifikleştirmeyi gerektirir. En müşterek yaklaşım I. Tip hatanın düzeyini spesifikleştirmektir ki aynı zamanda α olarak bilinir. I. Tip hata aslında doğru olan bir sıfır hipotezini yanlışlıkla reddetme olasılığıdır. Bu hata işlendiğinde, aslında mevcut olmayan bir istatistiksel anlamlılık gösterilmiş olur ki buna aynı zamanda “yanlış pozitif” denir. Bir α düzeyi spesifikleştirmek suretiyle, araştırmacı hata için izin verdiği sınırları tespit eder ve aslında mevcut olmayan anlamlılığın mevcut olduğu neticesine varma olasılığını belirtir.

I. Tip hata düzeyini spesifikleştirdiği zaman, araştırmacı aynı zamanda II. Tip hata veya β adı verilen ilişkili bir hatayı da belirler. II. Tip hata aslında yanlış olan sıfır hipotezini reddedememe olasılığıdır. Daha ilginç bir olasılık ise $1 - \beta$ dır ki istatistiksel çıkarım testinin gücü olarak adlandırılır. Güç, aslında yanlış olan sıfır hipotezini haklıca reddetme olasılığıdır. Böylece, güç, istatistiksel anlamlılığın, eğer mevcut ise, gösterilmesi olasılığıdır. Herhangi bir istatistiksel anlamlılık testine ait hipotetik durumlarda farklı hata olasılıklarının birbirleriyle ilişkisi aşağıdaki üzere gösterilir:

		Popülasyondaki hakikat	
		H_0 doğru	H_0 yanlış
Örneğe dayalı karar	H_0 reddedilir	α (I. Tip hata)	$1 - \beta$ (güç)
	H_0 reddedilmez	$1 - \alpha$	β (II. Tip hata)

Her ne kadar α nın spesifikleştirilmesi kabul edilebilir istatistiksel anlamlılık düzeyini tesis etse dahi, eğer aslında mevcut ise, anlamlılığı gösterme olasılığını güç düzeyi dikte eder. Öyleyse, niçin hem α hem de β için kabul edilebilir düzeyler tespit etmiyoruz? Çünkü I. Tip hata ve II. Tip hata ters ilişkilidir ve I. Tip hata küçüldükçe, II. Tip hata büyür. Bu sebeple, I. Tip hatayı

azaltmak, istatistiksel testin gücünü azaltır. Böylece, araştırmacı tespit edeceği α düzeyi ve bundan mütevellit güç arasında bir denge kurmalıdır.

Güç tek başına α nın bir fonksiyonu değildir. Onunla beraber, fiilen şu üç faktör tarafından belirlenir:

1. *Etki miktarı.* İstatistiksel anlamlılığın gösterilmesi olasılığı yalnızca istatistiksel mülahazalara dayanmaz, aynı zamanda ilgilenilen etkinin popülasyondaki asıl büyüklüğüne dayanır (mesela iki grup aritmetik-ortalama arasındaki fark veya değişkenler arasındaki korelasyon). Bekleneceği üzere, daha büyük bir etkinin örnek veride bulunabilirliği daha küçük bir etkinin bulunabilirliğine göre daha yüksektir ve bu da istatistiksel testin gücüne yansır. Herhangi bir istatistiksel testin gücünü değerlendirmek için, araştırmacı ilk önce incelenen etkiyi anlamalıdır. Etki miktarı karşılaştırmaları kolaylaştırmak için standartlaştırılmış terimler ile tanımlanır. Aritmetik-ortalama farklar standart sapmalar cinsinden ifade edilir, öyle ki mesela 0.5 lik bir etki miktarı bize aritmetik-ortalama farkın standart sapmanın yarısı olduğunu belirtir. Korelasyonlar için, etki miktarı değişkenler arasındaki korelasyonun kendisidir çünkü korelasyon katsayısı zaten standartlaştırılmış bir terimdir.
2. *Anlamlılık düzeyi.* Yukarıda ifade edildiği üzere, I. Tip hata veya α testin gücü azaldıkça azalır. Bundan ötürü, araştırmacı aslında anlamlı olmayan bir etkiyi yanlışlıkla anlamlı bulma şansını ne kadar azaltırsa, bir etkiyi haklıca doğru bulma olasılığını da o kadar azaltmış olur. Müstereken kullanılan α düzeyleri 0.05 ve 0.01 dir. Mamafih, araştırmacı bu kararı verirken, tespit edeceği α düzeyinin testin gücüne etkisini mulaka göz önünde bulundurmalıdır.
3. *Örnek hacmi.* Verilen herhangi bir α düzeyinde, daha geniş örnek hacimleri istatistiksel testin gücünü daha artırır. Bu halde, potansiyel bir problem çok fazla güç kesbetmektir. “Çok fazla güç” ile kastedilen şudur ki örnek hacmi artırıldıkça, daha küçük ve daha küçük etkiler istatistikselce anlamlı bulunacaktır, ta ki çok geniş örnek hacimleriyle neredeyse herhangi bir etki anlamlı bulunana kadar. Araştırmacı tespit edeceği örnek hacminin yapacağı istatistiksel testi şu iki yönde etkileyebileceğinin farkında olmalıdır: ya duyarsızlaştırır (dar örnek hacimleriyle) ya da aşırı duyarlılaştırır (çok geniş örnek hacimleriyle).

Anlamlılık düzeyi, örnek hacmi, etki miktarı ve güç arasındaki ilişki hayli komplikedir ve ortalıkta kılavuzluk edecek pek çok kaynak ta mevcuttur. Araştırmacı, arzu ettiği güç düzeylerine ulaşmak için, her üç faktörü – anlamlılık düzeyi, örnek hacmi ve etki miktarı – eş zamanlıca mülahaza etmelidir.