

İSTATİSTİK SONUÇLARININ YORUMU: P DEĞERİ VE GÜVEN ARALIĞI NEDİR?

INTERPRETATION OF STATISTICAL RESULTS: WHAT IS P VALUE AND CONFIDENCE INTERVAL?

Seval Kul

Gaziantep Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Gaziantep, Türkiye

e-posta: sevalkul@gantep.edu.tr

DOI:10.5152/pb.2014.003

© Telif hakkı 2014 Türk Toraks Derneği

Özet

İstatistiksel anlamlılık bir kararın önemli olma derecesidir. Tıpta klinik ve istatistiksel anlamlılık birbirinden farklılık gösterir. Bir bulgunun klinisyen açısından önemli olması için hipotez testi sonucunda istatistiksel anlamlılığın bulunmasının yani bulgunun klinik anlamlılığı da tartışılır. Bir bulgunun istatistiksel olarak anlamlılığının derecesi p değeri ile gösterilirken klinik anlamlılık için genellikle etki büyüklüğü veya güven aralığı tahminlerine dayalı tartışılır. Bu çalışmada p değeri ve güven aralığının ne anlama geldiği ve nasıl kullanıldığından bahsedilmiştir.

Abstract

Statistical significance is defined as an importance level of a decision. Clinical and statistical significance differ in medicine. For clinicians, importance of a finding depends on statistical significance which is found after the hypothesis testing. Level of statistical significance is shown by p value; however clinical significance is discussed by effect size or confidence interval. In this study, what p and confidence interval mean and how to interpret those are mentioned.

Key words: Significance; p value, confidence interval

Anahtar kelimeler: Anlamlılık, p değeri; güven aralığı

Giriş

Bir bilimsel çalışmanın yürütülmesinde ilk adım test edilecek hipotez takımının oluşturulmasıdır (1,2). Anneleri sigara içen 0-6 yaş çocuklarda astım gözlenme oranının anneleri sigara içmeyen 0-6 yaş çocuklara göre anlamlı düzeyde daha yüksek olabileceği düşüncesi ile bir araştırma planlandığını varsayalım.

Böyle bir çalışmada oluşturulacak hipotez takımı şöyledir;

H_0 (yokluk hipotezi): Annesi sigara içen ve içmeyen 0-6 yaş çocuklarda astım gözlenme oranı eşittir.

H_1 (alternatif hipotez): Annesi sigara içen 0-6 yaş çocuklarda astım gözlenme oranı annesi sigara içmeyenlere göre daha yüksektir.

Böyle bir hipotez testinde, örneklemi karşılaştırarak, örneklem arasındaki fark tesadüften mi ileri geliyor yoksa gerçek farklılık mıdır? Kararlarından birini vermeye çalışırız. Bu hipotezin test edilmesi için çalışmamızda iki bağımsız grup olması gerektiği açıktır. Bu gruplar hem yeterli büyüklükte hem de kendi popülasyonunu en iyi şekilde temsil etme yeteneğinde olmalıdır (1). Grupları kendi popülasyonuna benzer oluşturmak için doğru örnekleme tekniğini kullanmak gerekir. Yeterli sayıda kişi ile çalışmak için ise çalışmanın başında güç analizi yapmak gerekir. Çalışmanın mümkün olduğunca yanlılıklardan arındırılmış bir şekilde düzenlemesi ve yürütülmesi gereklidir (3). Sonunda doğru istatistiksel testler kullanarak hipotez testinin sonuçlandırılması şarttır (3).

Peki çalışmanın bu aşamalarından herhangi bir yerinde bilerek ya da bilmeyerek hatalar yapılırsa karışımıza ne gibi durumlar çıkabilir? Bir hipotez testi sonucunda karşılaşılabilecek olası tüm durumlar Tablo 1’de belirtilmiştir. Hipotez testlerinde karşılaşılabileceğimiz en riskli durum, gerçekte iki grup arasında anlamlı farklılık olmadığı halde hipotez test sonucunda anlamlı bir fark bulmaktır. Buna hata istatistikte I. Tip hata denir ve düzeyi A ile gösterilir (3).

I. Tip hata bir çalışmada yapılan her karşılaştırmada gerçekleşebileceği için bu riskin katlanılabilecek maksimum düzeyi çalışmanın başında araştırmacı tarafından belirlenerek I. Tip hata kontrol altına alınır. Tıptaki çalışmalarda I. Tip hata düzeyi genellikle 0,05 düzeyinde tutulur, böylelikle kararlar %95 güvenle verilmiş olur. Bu hatanın düzeyini %1 e düşürmek mümkündür. Bu durumda %99 güvenle çalışılmış olur ama bu durumda da gerekli minimum örnek genişliği artar (2).

P değeri

P (Probability; Olasılık) değeri istatistiksel anlamlılığın (statistical significance) varlığının ve varsa da var olan farklılığın kanıtının düzeyinin belirlenmesi amacı ile kullanılan bir değerdir (4). Her istatistiksel testin sonucunda kullanılan test istatistiğine ait bir P değeri hesaplanır. Bu P değeri “ilgili hipotez testi sonucunda anlamlı fark vardır” denileceği durumda hatalı karar verme olasılığının ne olduğunu gösterir. Her ne kadar p değerinin 0,05 den küçük olması tıp literatüründe “istatistiksel olarak anlamlı” kabul edilse de araştırmacının isteğine göre 0,01 olarak ta kabul edilebilir. P değeri ne kadar küçük olursa H_0 hipotezini reddetmek için elimizdeki kanıt o kadar yüksek olur. P değerinin nasıl yorumlanmasındaki genel yaklaşım Tablo 2’de verilmiştir (3).

P değeri, genellikle paket programlarda “Sig.” kısaltması ile ifade edilir. Burada “Sig.” “significance; anlamlılık” kelimesinin kısaltılmışı halidir.

Sonuçları raporlarken dikkat edilmesi gereken bir başka nokta p değerini $\leq 0,05$ ya da $> 0,05$ vermek yerine tam değerini vermek en doğrusudur. Sonuçları raporlarken genellikle p değeri virgülden sonra 3 hane kadar belirtilir.

Güven aralığı

P değeri, bir çalışmanın klinik anlamlılığı (clinical significance) hakkında bilgi veremez (3,4). Buna ek olarak büyük örnekleme yapılmış bir çalışmadan elde edilmiş küçük bir p değeri, klinik olarak hiçbir anlam ifade etmiyor olabilir (5). Başlangıçta verdiğimiz örneğe dönecek olursak; ilgili çalışmanın yapıldığını ve sonuçta annesi sigara içen çocuklarla içmeyen çocuklar

Tablo 1. Bir hipotez testinde karşılaşılabilecek olası durumlar

		Gerçek Durum	
Araştırma Sonucunda Varılan Karar	H_0 Doğru	H_0 Doğru Doğru Karar 1-A=Güven aralığı	H_1 Doğru II. Tip Hata B
	H_1 Doğru	I. Tip Hata A	Doğru Karar 1-A=Testin Gücü

Tablo 2. p değerinin yorumlanması

P değeri	Yorumu
$0.01 \leq p < 0.05$	İstatistiksel anlamlılık
$0.001 \leq p < 0.01$	Yüksek düzeyde istatistiksel anlamlılık
$p < 0.001$	Çok yüksek istatistiksel anlamlılık
$0.05 \leq p < 0.10$	Anlamlılık eğilimi(sınırdan anlamlılık)
$p > 0.10$	Fark tesadüften ileri gelmiştir (istatistiksel olarak anlamlı farklılık saptanmamıştır)

arasında astım gözlenme oranı bakımından istatistiksel olarak anlamlı bir farklılık olduğuna karar verildiğini varsayalım. Bu aşamada araştırmacının yapması gereken bu farkın ne kadar olduğuna bakmaktır. Astım olma oranlarında %3’lük bir fark olduğu gözlenmiş olsun. Oluşan bu %3’lük fark etki büyüklüğü olarak adlandırılır ve tek bir denemeden elde edilmiş bir sonuçtur. Halbuki bu deneme yüzlerce farklı örneklemden tekrarlanabilir ve her seferinde de farklı oranlar elde edilebilir. Bu denemelerin defalarca tekrarlandığı durumda etki büyüklüğünün hangi aralıkta olabileceği güven aralığı ile tahmin edilir. I. Tip hatanın %5’de tutulduğu bir hipotez testinde %95 güven aralığı, I. Tip hatanın %1’de tutulduğu durumda ise %99 güven aralığı hesaplanmış olur.

Bizim örneğimizde iki grup astım oranları arasındaki farka ait elde edilen %95 güven aralığının %2,5-%4 olduğunu varsayalım. Bu durumda aynı çalışma defalarca tekrarlandığında oranlar arasındaki fark %95 olasılıkla %2,5-%4 arasında yer alır yorumu yapılır.

Klinik Anlamlılık

Tıpta klinik ve istatistiksel anlamlılık birbirinden farklılık gösterir. İstatistiksel anlamlılık hastaların oluşturduğu örneklemden yola çıkarak yine aynı hastalara ait popülasyon hakkında kararlar almak ya da tahminler yapmaktır (3,6). İstatistiksel anlamlılık test edildikten sonra anlamlı bulunan bir bulgunun klinikte kullanılabilişliğini, yararlılığına karar vermek amacı ile klinik

anlamlılık tartışılır. Bir bulgunun klinik olarak anlamlı olması için öncelikle istatistiksel olarak anlamlı olması gerekir (6). Fakat istatistiksel olarak anlamlı her bulgu klinik olarak anlamlı olmayabilir. Araştırma sonucunda bulunan fark çalışmanın başında belirlenen etki büyüklüğünden büyükse bulgunun klinik olarak anlamlı yani yararlı ya da kullanılabilir olduğu söylenebilir (6).

Kaynaklar

1. Sümbüloğlu K, Sümbüloğlu V. Sağlık Bilimlerinde Araştırma Yöntemleri, 6. Basım. Ankara: Hatipoğlu Yayınevi.
2. Kul S. Klinik Araştırmalarda Örnek Genişliği Belirleme. Plevra Bülteni 2011;2:129-32. [\[CrossRef\]](#)
3. Rosner B. Fundamentals of Biostatistics, 7th edition. Şehir: Brooks/Cole. 2010.
4. Dawson B, Trapp RG. Basic & Clinical Biostatistics (LANGE Basic Science) Şehir: McGraw-Hill Medical. 2004.
5. Leung WC. Balancing statistical and clinical significance in evaluating treatment effects. Postgrad Med J. 2001;77:201-4. [\[CrossRef\]](#)
6. Sainani KL. Clinical versus statistical significance.PM R 2012;4:442-5. [\[CrossRef\]](#)