

Uygun Olmayan İstatistiksel Analiz Kullanımı ve İstatistiksel Sonuçların Hatalı Yorumlanması

Use of Inappropriate Statistical Analysis and Misinterpretation of Statistical Results: Letter to the Editor

Ekim PEKÜNLÜ^a

^aEge Üniversitesi Beden Eğitimi ve Spor Yüksekokulu, İzmir

Geliş Tarihi/Received: 08.10.2015
Kabul Tarihi/Accepted: 27.10.2015

Yazışma Adresi/Correspondence:
Ekim PEKÜNLÜ
Ege Üniversitesi Beden Eğitimi ve Spor Yüksekokulu, İzmir,
TÜRKİYE/TURKEY
ekimpekunlu@gmail.com

Anahtar Kelimeler:

Karma tasarımlı varyans analizi; etkileşim; kontrollü çalışma; değişim farkı

Key Words:

Mixed design analysis of variance; interaction; controlled study; difference in change score

Türkiye Klinikleri J Sports Sci
2016;8(1):35-6

doi: 10.5336/sportsci.2015-48236

Copyright © 2016 by Türkiye Klinikleri

Aras ve Akalan'ın Türkiye Klinikleri Spor Bilimleri Dergisi 2015;7(1):16-22'de yayımlanan "Sekiz haftalık spor kaya tırmanışı antrenmanının günlük fiziksel aktivite düzeyi ve enerji harcama-sına etkisi" başlıklı çalışmasındaki uygun olmayan istatistiksel analiz kullanımı ve istatistiksel analiz sonuçlarının hatalı yorumlanması (tartışılması) ile ilgili bazı görüşlerimi paylaşmak istiyorum.

Araştırma tasarımlarında kontrol grubu kullanılmasının temel amacı, incelenen bağımlı değişken(ler)de zamana bağlı ortaya çıkabilecek olası değişim düzeyini belirlemektir. Bu şekilde "uygulama grubuna herhangi bir uygulama yapılmayı bu grupta da ilgili bağımlı değişkende bu düzeyde bir değişim gözlenecekti." varsayımından yola çıkılarak, araştırmada kullanılan uygulamanın gerçek etkisinin belirlenmesi amaçlanmaktadır. Aras ve Akalan'ın araştırmasının hipotezlerini test etmek için gerçekleştirilmesi uygun olan istatistiksel analizleri şu şekilde sıralayabiliriz:

1. Parametrik test varsayımlarını karşılayan bağımlı değişkenler için 2x2 [Grup (kontrol, deney) x Zaman (ön test, son test)] Karma Tasarımlı Varyans Analizi (Mixed Design Analysis of Variance [ANOVA]): "İncelenen bağımlı değişkenlerde zamana bağlı/uygulamaya bağlı ortaya çıkan değişim düzeyleri gruplar arasında benzer bir yapı mı gösteriyor?" sorusuna yanıt vermektedir. Araştırmacıların tartışma bölümünde kullandıkları "Bununla birlikte UK'da da yüzde 84,0±6,93'ten 81,67±8,67'ye azalma görüldü ... daha fazla etkilendiğini göstermektedir." gibi yorumlar bu istatistiksel analiz sonuçlarına göre yapılabilmektedir.

2. Eşleştirilmemiş (bağımsız örneklem) Student t-testi (parametrik test varsayımlarının karşılanmadığı durumlarda Mann-Whitney U test): "Deney grubunda uygulamaya bağlı ortaya çıkan değişim (Δ), kontrol grubunda zamana bağlı ortaya çıkan değişimden (Δ) istatistiksel olarak farklı mı?" sorusuna yanıt vermektedir [Öncesinde ilgili bağımlı değişkenlere yönelik olarak her katılımcı için ön test ve son test değerleri arasındaki farklar (Δ) hesaplanmalıdır. Araştırmacıların bağımlı değişkenlere ait ön test değerle-

rinde (tanımlayıcı istatistiklerde olduğu gibi) gruplar arasında istatistiksel olarak anlamlı bir fark olup olmadığını rapor etmemeleri araştırmanın bir diğer eksikliği olarak düşünülebilir. Grupların ön testlerindeki olası istatistiksel farklılık durumları araştırmanın sonuçlarında yanlışlık oluşturmuş olabilir].

Ayrıca araştırmacılar, bulgulara yönelik tablolarda hangi bağımlı değişken(ler)in parametrik olmayan testlerle analiz edildiğini belirtmemişlerdir. Tabloların bu bilgiyi okuyuculara iletebilecek şekilde düzenlenmesi ve ilgili değişkenlere ait “ortalama±standart sapma” yerine (Wilcoxon İşaretli Sıralar Testi “ortalama” karşılaştırmamaktadır).¹ “medyan ve belirli yüzdelik değerlerin (örneğin; 25-75 persentil)” rapor edilmesi daha uygun bir yaklaşım olabilir.

İstatistiksel analizlerde rapor edilen “p” değerlerinin gerçekleştirilmiş olan ilgili istatistiksel analizin “koşullu olasılığını” rapor ettiği unutulmamalıdır (“H0” hipotezinin doğru olduğu varsayıldığında). Bir hipotez testi “p>0,05” sonucunu rapor ediyorsa bu, incelenen bağımlı değişken(ler)de ortaya çıkan sayısal farkın istatistiksel olarak anlamlı olmadığını belirtmektedir. Araştırmacılar tablolarda her iki gruptaki bütün bağımlı değişkenler için “p” değerini “0,05”ten büyük rapor etmiş olmalarına rağmen “tartışma” bölümünde kullanmış oldukları “*Bu bilgi, sekiz haftalık SKT antrenmanlarından sonra deney grubunda enerji harcamalarında görülen artışı doğrular niteliktedir ve her ne kadar istatistiksel olarak anlam taşımaya da, sekiz*

haftalık SKT'nin günlük enerji harcamasını artırdığının göstergesidir.” gibiden ifadeler istatistiksel açıdan doğru değildir. Araştırmacıların ilgili değişkenle ilgili buldukları p>0,05 değeri, olası bir tekrar (replikasyon) çalışmasında “sekiz haftalık SKT'nin %95 olasılıkla günlük enerji harcamasını istatistiksel olarak anlamlı “olmayacak” şekilde artırabileceği gibi istatistiksel olarak anlamlı “olmayacak” şekilde “azaltabileceğini” de göstermektedir”.²⁻⁴

Araştırmacıların tartışmalarını, “literatürdeki diğer fiziksel egzersiz çalışmalarının aksine (orijinal metin kaynak no: 8, 31, 32, 33) sekiz haftalık SKT'nin günlük enerji harcamasını neden “artırmadığı” konusu üzerine yoğunlaştırmaları gerekirdi. Bunun yerine araştırmacılar -hatalı olarak-, araştırmalarında bulmuş oldukları “istatistiksel olarak anlamlı *olmayan* artışı” literatürdeki “istatistiksel olarak *anlamlı* artışlarla” tutarlı olan sonuçlanmış gibi tartışmışlardır. Bu tartışma şeklinin bilimsel açıdan “doğru” olduğunu söylemek olanaklı değildir.^{2,4}

Sayın Aras ve Akalan'ın eleştirilerimi çalışmalarına bir katkı ve gelecek araştırmalarında dikkate alabilecekleri konu başlıkları olarak değerlendirmelerini ümit ediyorum. Ayrıca, Türkiye Klinikleri Spor Bilimleri Dergisi hakemlerinin makale değerlendirme sürecinde araştırmacılara bu konularda uygun geri-bildirim zamanında sağlama konusunda daha hassas olmaları gerektiğini düşünüyorum.

KAYNAKLAR

1. Field AP. Non-parametric tests. Discovering Statistics Using IBM SPSS Statistics (and Sex and Drugs and Rock 'n' Roll). 3rded. London: SAGE Publications; 2009. p.539-83.
2. Ercan İ, Yazıcı B, Yang Y, Özkaya G, Cangür Ş, Ediz B, et al. Misusage of statistics in medical research. Eur J Gen Med 2007;4(3):128-34.
3. Kline RB. Fundamental concepts. Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research. 1sted. Washington, DC: American Psychological Association; 2004. p.19-60.
4. Pekünlü E. Misinterpretation of statistical (in)significance in scientific researches. MÜS-BED 2013;3(2):114.