

Standart Sapma mı Yoksa Standart Hata mı?

Hanefi Özbek*, Sıddık Keskin**

Özet:

Amaç: Standart sapma ile standart hatanın farklarını ortaya koymak ve makalelerde aritmetik ortalama değerinden sonra hangisinin yazılmasının daha doğru olduğu konusunu tartışmak.

Sonuç: Çalışma gruplarına ait veriler, sadece ilgili olduğu grubun özelliğini/özelliklerini gösteriyorsa “ortalama \pm standart sapma” tercih edilmelidir. Verileri birbiri ile karşılaştırarak, gruplar arasında fark olup olmadığı öğrenilmek isteniyorsa; “ortalama \pm standart hata” kullanılması daha uygun olacaktır.

Anahtar kelimeler: Standart sapma, standart hata.

Bilimsel araştırmalarda genel amaçlardan birisinin: üzerinde durulan özelliğin (değişkenin) diğer özelliklerle olan ilişkisini ve bu özelliğe etkili olan faktör veya faktörleri belirleyebilmek, böylece ilgili özelliğin değişimine istenilen şekilde yön verebilmek olduğu söylenebilir. Bu amaca yönelik olarak yapılacak araştırmalarda; genelde maliyet ve zaman sınırlaması, tüm popülasyonun çalışılmasını hemen hemen olanaksız kılar. Bu nedenle üzerinde durulan özellik bakımından, popülasyonu temsil edebilen sınırlı sayıda birey içeren örneklerle çalışılır (1).

Bilindiği gibi, ilgili özellik bakımından popülasyonu karakterize eden değerlere *parametre*; örneği karakterize eden değerlere de *istatistik* denir. Popülasyona ait değerler (parametreler) genel olarak Grek (Yunan) alfabesinin harfleri ile (“ μ ” ve “ σ ” gibi), örneğe ait değerler (istatistik) ise Latin alfabesinin harfleri ile (“x” ve “s” gibi) (2).

İstatistik biliminin amacı, sayılarla ilgili bilgilerin özetlenmesi ve veri analizi olarak kabul edilebilir. Bu durumda istatistik bilimi “tanımlayıcı (descriptive) istatistik” ve “çıkarımsal (inferential) istatistik” olarak iki alana ayrılabilir. Tanımlayıcı istatistikler, verilerin özetlenmesi ve sunumu ile ilgili bilgileri içerir, popülasyon parametreleri ile ilgili olarak doğrudan tahmin yapmada kullanılamaz. Bu bağlamda tanımlayıcı istatistikler; merkezi eğilim ölçülerini (ortalamalar) ve değişim ölçülerini (dağılım veya varyasyon) içerir.

Ayrıca ilişki (korelasyon) katsayıları da tanımlayıcı istatistiğin alanına sokulabilir. Çıkarımsal istatistikler ise genel olarak örnekten elde edilen bilgileri kullanmak suretiyle popülasyon parametreleri hakkında tahmin veya yorum yapmak amacıyla kullanılır. Yapılacak tahminin veya yorumun doğruluk derecesinin yüksek olması, seçilecek bireylerin (örneklem) popülasyonu iyi temsil etmesi ile mümkündür. Bu durum, bireylerin rasgele (randomize) olarak seçilmesi, üzerinde durulan özellik bakımından bireylerin homojen olması ve birey sayısının (örneklem büyüklüğünün) yeterli olması ile sağlanabilir. En yaygın kullanılan merkezi eğilim ölçüleri aritmetik ortalama (mean), tepe değeri (mode) ve ortanca değeri (median), en yaygın olarak kullanılan değişim ölçüleri ise standart sapma (standard deviation), standart hata (standard error of the mean) ve varyasyon katsayısıdır (coefficient of variation) (3,4,5).

Bir veri grubunu tanımlamak için; bir yerleşim ölçüsü ve bir değişim ölçüsü kullanılır. Yerleşim ölçüleri (merkezi dağılım ölçüleri) verilerin toplanma eğilimi gösterdiği değeri yani verilerin almak istediği değeri gösterirken, değişim ölçüleri yerleşim ölçüsü etrafındaki yayılışı yani dağılımı gösterir. En yaygın olarak kullanılan merkezi eğilim ölçüsü aritmetik ortalama olup, kısaca ortalama olarak belirtilir. Aritmetik ortalamanın yanında verilecek olan değişim ölçüsü ise bazı durumlarda standart sapma, bazı durumlarda da standart hatadır (6).

Makale yazarken karşılaşılan, ama üzerinde pek fazla durulmayan konulardan biri de yukarıdaki sorudur. Yani “verilerin ortalama değerinden sonra, standart sapma değeri mi yoksa standart hata değeri mi yazılmalı?”. Bu soruya iki türlü cevap verilebilir; “farketmez, ikisi de olur” ya da “farkeder, çünkü

*Yüzüncü Yıl Üniversitesi, Tıp Fakültesi Farmakoloji AD, Van.

**Yüzüncü Yıl Üniversitesi, Tıp Fakültesi Biyoistatistik AD, Van.

Yazışma Adresi: Dr. Hanefi Özbek

Yüzüncü Yıl Üniversitesi

Tıp Fakültesi Farmakoloji AD.

65300 Van.

ikisinin de ifade ettiği kavramlar birbirinden farklıdır”.

Öncelikle her iki terimin tanımı yapılarak konu açıklığa kavuşturulmaya çalışılabilir:

Standart sapma (standard deviation, SD): Bir çalışma grubundaki herbir verinin ortalamaya göre ne kadar uzaklıkta olduğunu, bir diğer deyişle dağılımın ne yaygınlıkta olduğunu gösteren bir ölçüdür. Başka bir şekilde tanımlanacak olursa; belirli bir popülasyonda incelenen özelliğin (veya özellikle ilgili değerlerin ya da ölçümlerin) ne genişlikteki bir aralıkta (dar veya geniş) dağıldığının göstergesi varyans ve onun bir türevi olan standart sapmadır. (Varyansın karekökü standart sapmayı verir). Standart sapma büyüdükçe dağılım yaygınlaşır (7). Örneğin; boy ortalaması “165 cm” olan 50 kişilik bir denek grubunun standart sapması “4.0” olsun. Bu grubun elemanlarının büyük bir kısmı (istatistiğe göre % 68’i) boy yönünden “165±4, yani 161-169 cm arasındadır” anlamına gelmektedir. Burada standart sapmanın iki misli alınır, bu durumda “grubun elemanlarının % 95’i” boy yönünden “165±8, yani 157-173 cm arasındadır” anlamına gelir. Kısaca söylemek gerekirse; ortalama ± standart sapma, grubu oluşturan deneklerin belli bir özellik yönünden ortalamaya göre sağa-sola ne kadar saptığını yaklaşık bir ifade ile ortaya koymaktadır. Standart sapma, popülasyonu oluşturan bireyler arasındaki farklılığın bir ölçüsüdür.

Standart hata (standard error of mean, SEM): Aynı popülasyondan seçilecek, aynı büyüklükteki örneklemelerin ortalamalarının yayılmasını gösteren ölçüt, *ortalamanın standart hatası* (standard error of mean)’dır. Standart sapma değerinin denek sayısının kareköküne bölünmesi ile elde edilen değerdir (7). Ortalamanın standart hatası, ortalamanın dağılımındaki varyasyonu (değişimi) gösterir, örneklem sayısının artması ile küçülür. Standart hatanın küçük olması popülasyon parametresine ait yapılacak olan tahminler açısından ve daha dar güven aralığı sınırlar bulma açısından önemlidir. Ekstrem bir örnek verilecek olursa; örneğin bir çalışma grubunun ortalamasına ait standart hata sıfır olarak bulunsun. Bu sıfır değeri, çalışma grubuna ait ortalama değerinin, grubu oluşturan örneklerde değişmediğini ve popülasyon parametresini tahmin bakımından en iyi tahmini yaptığını gösterir. Yani standart hata ne kadar küçükse, popülasyon ait tahminlerimiz de o kadar isabetli olmaktadır.

Standart hata kavramının iyi anlaşılması gerektiğinden bir örnek vermek uygun olacaktır. Herbirinde 100’er kişi (denek) bulunan iki ayrı grubumuz (1. grup ve 2. grup) olsun. Buradaki her bir denegün hemoglobinin (Hb) değeri ölçülsün ve kaydedilsin. Birinci ve ikinci grubun Hb değerlerinin

ortalaması sırayla 13.0 ve 15.7, standart sapması ise sırayla 9.0 ve 4.0 olsun. Grupların standart hatalarını bulmak için standart sapma değerini denek sayısının kareköküne bölünmesi gerekir (denek sayısı=100 olduğuna göre, bunun karekökü=10 olacaktır). Sonuç olarak standart hata değerleri sırasıyla “9/10=0.9” ve “4/10=0.4” olarak elde edilecektir. Buna göre grupların ortalama değerlerinin yanında standart hata değerleri de verilecek olursak; “13.0 ± 0.9, yani 12.1-13.9” ve “15.7 ± 0.4, yani 15.3-16.1” değerleri elde edilir. Bu değerler her iki grup ortalamasının, % 68 olasılıkla, 12.1-13.9 ve 15.3-16.1 olarak yayıldığını göstermektedir. Standart hata değerinin 2 katı alınır, bu kez % 95 olasılıkla grupların yayılımı bulunmuş olur (6).

Bilimsel çalışmaların amacı, örneklem değerlerinden yola çıkarak evren (popülasyon) değerlerinin kestirilmesidir. Bu amaçla ya “güven aralığı ve sınırları” ya da “hipotez testleri” kullanılır. Burada “güven aralığı ve sınırları” tercih edilmiş olsun. Buna göre her iki grubun güven aralığı sınırları saptanıp, bu sınırların birbiri ile kesişip kesişmediğine bakılmalıdır. Eğer kesişme yok ise (yani her iki grubun güven aralığı değerleri aynı sayı doğrusu üzerine koulduğunda, birbiri üzerine binmiyor ise); her iki grubun ortalaması birbirinden farklıdır, dolayısı ile birinci grubun Hb değerlerinin ortalaması (13.0) ikinci grubun Hb değerlerinin ortalamasından (15.7) istatistiksel olarak düşüktür denilebilir. Eğer her iki grubun güven aralığı değerleri birbiri ile kesişiyor ise (yani her iki grubun güven aralığı değerleri aynı sayı doğrusu üzerine koulduğunda, birbiri üzerine biniyor ise); bu durumda grupların Hb değerlerinin ortalamaları birbirinden farklı değildir, dolayısı ile gruplar arasında fark yoktur denilebilir.

Çalışma gruplarının ortalama değerlerinin yanında “± standart sapma” veya “± standart hata” değerleri üç şekilde verilebilir; yani standart sapma veya standart hata değerinin 1, 2 veya 3 katı ortalamanın yanına yazılır. Daha açık bir ifade ile:

Ortalama ± 1 (standart sapma),

Ortalama ± 2 (standart sapma),

Ortalama ± 3 (standart sapma),

veya

Ortalama ± 1 (standart hata)

Ortalama ± 2 (standart hata)

Ortalama ± 3 (standart hata)

Ortalamadan sonra gelen standart sapma veya standart hatanın 1, 2 veya 3 katı şu anlama gelmektedir:

Ortalama ± 1 (standart sapma): Hb değerlerinin % 68’inin “ortalama ± 1 (standart sapma)” kadar ortalamanın sağına-soluna yayıldığını gösterir.

Ortalama \pm 2(standart sapma): *Hb değerlerinin % 95'inin "ortalama \pm 2 (standart sapma)" kadar ortalamanın sağına-soluna yayıldığını gösterir.*

Ortalama \pm 3(standart sapma): *Hb değerlerinin % 99'unun "ortalama \pm 3 (standart sapma)" kadar ortalamanın sağına-soluna yayıldığını gösterir.*

Aynı durumu standart hata için yapılacak olursa:

Ortalama \pm 1(standart hata): *Hb ortalama değerinin % 68 olasılıkla " \pm 1 (standart hata)" kadar sağa-sola yayıldığını gösterir. (Yani, bu çalışma grubu, aynı popülasyondan 100 kez tekrar tekrar oluşturulacak olursa, bunların 68'inin ortalaması, bu sınırlar arasında kalacaktır).*

Ortalama \pm 2(standart hata): *Hb ortalama değerinin % 95 olasılıkla " \pm 2 (standart hata)" kadar sağa-sola yayıldığını gösterir.*

Ortalama \pm 3(standart hata): *Hb ortalama değerinin % 99 olasılıkla " \pm 3 (standart hata)" kadar sağa-sola yayıldığını gösterir.*

Yukarıdaki aralıklardan pratik olarak 2. ve 3.'sü (% 95 ve % 99) kullanılır. Grupların herbiri tek tek ele alınırken (örneğin; birinci grubun yaş ortalaması, boy ortalaması gibi) "*ortalama \pm standart sapma*" değerleri kullanılır. Böylece, sadece o grubun, yaş veya boy yönünden nasıl bir dağılım gösterdiği anlaşılabilir olur. Ancak gruplar birbiri ile karşılaştırılırken "*ortalama \pm standart sapma*" değil "*ortalama \pm standart hata*" değerleri kullanılır.

Yukarıdaki bilgilere göre iki grubun ortalama \pm standart hata değerleri karşılaştırılacak olursa:

1. grup:

ortalama= 13.0
standart hata= 0.9

2. grup:

ortalama= 15.7
standart hata= 0.4
Her iki grubun 2 ve 3 kat standart hataya göre güven aralığı sınırları tespit edilsin:

1. grup:

13 \pm 2(0.9); yani 11.2-14.8 (Hb için güven sınırları % 95 olasılıkla 11.2-14.8 değerleri arasında olacaktır).
13 \pm 3(0.9); yani 10.3-15.7 (Hb için güven sınırları % 99 olasılıkla 10.3-15.7 değerleri arasında olacaktır).

2. grup:

15.7 \pm 2(0.4); yani 14.9-16.5 (Hb için güven sınırları % 95 olasılıkla 14.9-16.5 değerleri arasında olacaktır).
15.7 \pm 3(0.4); yani 14.5-16.9 (Hb için güven sınırları % 99 olasılıkla 14.5-16.9 değerleri arasında olacaktır).

1. grup ile 2. grubun güven aralığı değerleri birbiri ile karşılaştırılacak olursa:

11.2-14.8 ve 14.9-16.5: Her ikisi de birbirini aynı sayı doğrusu üzerinde *kesmiyor*.

10.3-15.7 ve 14.5-16.9: Her ikisi de birbirini aynı sayı doğrusu üzerinde *kesiyor*.

Yukarıdaki ilk karşılaştırmada, her iki grubun güven aralıkları birbiri ile kesişmediği için "yani ortalaması küçük olan grubun (birinci grup) üst güven aralığı sınırı (14.8), ortalaması büyük olan grubun (ikinci grup) alt güven aralığı sınırından (14.9) küçük olduğu için" grupların Hb değerleri birbirinden istatistiksel olarak farklıdır denilebilir. İkinci karşılaştırmada ise güven aralıkları birbiri kestiği için (15.7 ve 14.5) grupların Hb değerleri birbirinden istatistiksel olarak farklı değildir denilebilir. Birinci karşılaştırmada % 95 olasılıkla (yani % 5 hata ile) ikinci karşılaştırmada ise % 99 olasılıkla (yani % 1 hata ile) güven aralıkları belirlenmiş oldu.

İstatistikte "p" değeri, "probability" yani "olasılık" teriminin kısaltılmış halidir. % 5 hatayı kabul ederek karşılaştırma yapıldığında; $p < 0.05$ (yani p değeri % 5'ten küçük ise) "*gruplar arasında istatistiksel olarak anlamlı fark vardır*" sonucu çıkarılır. Eğer % 1 hata oranı kabul edilerek karşılaştırma yapılıyor ise; $p < 0.01$ (yani p değeri % 1'den küçük ise) "*gruplar arasında istatistiksel olarak anlamlı fark vardır*" sonucu çıkarılır (6). Bu durumda birinci karşılaştırma kabul edilirse; "*her iki grubun Hb değerleri arasındaki fark anlamlıdır ($p < 0.05$)*" denilebilir. Eğer ikinci karşılaştırmada (yani "Ortalama \pm 3 (standart hata)") hesaplanan güven aralığı sınırları birbiri ile kesişmeseydi, bu durumda "*her iki grubun Hb değerleri arasındaki fark anlamlıdır ($p < 0.01$)*" denilmesi daha doğru olacaktı. Ama kesiştikleri için gruplar arasındaki karşılaştırmaların sonucunun $p < 0.01$ şeklinde ifade edilmesi yanlış olacaktır.

Yukarıdaki açıklamalardan da görüleceği gibi, grupların ortalamaları birbiri ile karşılaştırılırken *standart sapmadan* değil *standart hatadan* yararlanılmaktadır. Öyle ise grupların ortalama değerleri tabloda verilirken "*ortalama \pm standart sapma*" değil "*ortalama \pm standart hata*" şeklinde göstermek elbette ki daha doğru olacaktır.

Sonuç olarak; çalışma gruplarına ait veriler, sadece ilgili olduğu grubun özelliğini/özelliklerini (boy, kilo, yaş gibi) göstermek amacıyla verilmiş ise; bu verileri "*ortalama \pm standart sapma*" şeklinde vermek daha doğru olacaktır. Amaç, bu verileri birbiri ile karşılaştırarak, gruplar arasında bir fark olup olmadığını öğrenmek ise; bu durumda verileri "*ortalama \pm standart hata*" şeklinde göstermek daha doğru olacaktır. Örnekteki birey sayısı verildiğinde, standart hata ile standart sapma arasındaki ilişkiden yararlanmak suretiyle; standart

hata dan standart sapmayı veya standart sapmadan standart hatayı hesaplamak oldukça kolay olacaktır. Dolayısı ile “*standart hatanın ortalama ile birlikte verilmesi tercih edilmekle birlikte*”, standart sapma veya standart hatadan birinin verilmesi, aradaki ilişkiyi bilen biri için yeterli olacaktır.

Standard Deviation or Standard Error?

Abstract:

Aim: The aim is to explain the difference between standard deviation and standard error of the mean and to discuss which one should be used with the arithmetic mean when writing an academic paper.

Result: When data of the experimental groups describes group parameters “mean \pm standard deviation” should be used. “Mean \pm standard error of the mean” should be used when comparing data from two or more groups to find a difference between the groups.

Key words: Standart deviation, Standard error of the mean.

Kaynaklar

1. Heperkan Y. Tıpta İstatistik Yöntem ve Uygulamaları, 1981, Ankara Üniversitesi Tıp Fakültesi Yayını, s: 2-5.
2. Kesici T, Kocabaş Z. Biyoistatistik. Ankara Üniversitesi Basımevi, 1998, s : 1-2.
3. Esin A, Ekni M, Gamgam H. Sağlık Bilimlerinde İstatistik. Gazi Üniversitesi Basın-Yayın Yüksek Okulu Matbaası, 1991, Ankara, s: 9-10.
4. Şenocak M. Biyoistatistik. İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Yayını, 1998, İstanbul, s: 18-19.
5. Lemeshow S, Hosmer DW, Klar J, Wanga SK. Sağlık Araştırmalarında Örneklem Büyüklüğünün Yeterliliği. Çeviren: S. Oğuz Kayaalp. Hacettepe-TAŞ Kitabevi, Ankara, 2000.
6. Hayran M, Özdemir O. Bilgisayar İstatistik ve Tıp. Hekimler Yayın Birliği, Ankara, 1995, s: 248-250.
7. Sümbüloğlu, K, Sümbüloğlu V. Biyoistatistik. 10. Baskı, Hatiboğlu Yayınları No: 10, Ankara, 2002.