

Korelasyon ve İnteraksiyon (Etkileşim)

Sıddık Keskin¹

¹ Biyoistatistik Anabilim Dalı, Temel Tıp Bilimleri, Bölümü / Van Yüzüncü Yıl Üniversitesi Tıp Fakültesi, Türkiye

skeskin@yyu.edu.tr

Özet – Bilimsel çalışmalarda ele alınan özellikler veya değişkenler arasındaki ilişkiler, farklı yapıda ve farklı düzeyde de olabilir. Bu ilişkiler, ilişkinin yapısına göre genel olarak, doğrusal ilişkiler ve doğrusal olmayan ilişkiler olarak sınıflandırılabilir, aralarında ilişki bulunmak istenilen değişken veya özellik sayısına göre de basit ve çoklu ilişkiler olarak sınıflandırılabilir. Değişkenler (özellikler) arasındaki ilişkinin yönünü ve derecesini (düzeyini) belirlemek üzere farklı ölçütler veya istatistikler geliştirilmiştir. Bu ölçütler, genel olarak korelasyon kavramı ile ifade edilir. Korelasyon veya korelasyon katsayısı (r) iki özelliğin (veya sürekli değişkenin) birlikte değişiminin ölçüsüdür. Diğer bir ifade ile korelasyon katsayısı, iki özellik veya değişken arasındaki ilişkinin yönünü ve derecesini (yüzde) belirtir. Kısmi korelasyon katsayısı ise iki değişken arasında, bu iki değişkene birlikte ettiği varsayılan kovaryet değişken(ler)in etkisi giderildikten sonra hesaplanmış olan korelasyon katsayısıdır. Buna karşılık interaksiyon (etkileşim) kavramı ise kısaca; iki değişkenin, üçüncü bir değişken veya bağımlı değişken üzerine olan etkilerinin, bu değişkenlerin farklı düzeylerinde veya değerlerinde aynı olup olmadığını veya değişiklik gösterip göstermediğini ifade eden kavramdır. Diğer bir ifade ile iki değişkenin üçüncü bir değişken üzerine birlikte olan etkilerinin, yalnız başına olan etkilerinden farklılık gösterip göstermediğini ifade eder. Bu çalışmada korelasyon ve interaksiyon kavram genel özellikleri ve farklı yönleriyle açıklanmaya çalışılmış ve sonuçta bu iki kavramın birbirinden farkı olduğu dikkat çekilmiştir.

Anahtar Kelimeler –: Faktöriyel deneme, etkileşim, esas etki, kısmi korelasyon, faktör seviyesi

I. GİRİŞ

Bilimsel çalışmalara konu olan veya ilgilenilen değişkenler arasındaki ilişkiler, farklı yapıda olabileceği gibi farklı düzeyde de olabilir. Bu ilişkiler, ilişkinin yapısına göre genel olarak, doğrusal ilişkiler ve doğrusal olmayan ilişkiler olarak sınıflandırılabilir, aralarında ilişki bulunmak istenilen değişken veya özellik sayısına göre de basit ve çoklu ilişkiler olarak sınıflandırılabilir. Doğrusal ilişkilerin hesaplanması ve yorumlanması, doğrusal olmayan ilişkilere göre kısmen kolay olduğundan, uygulamalarda daha çok doğrusal ilişkilerin yönü ve derecesi (miktarı, düzeyi) belirlenmeye çalışılır.

Genellikle 'X' ve 'Y' ile gösterilen iki özellik veya değişken arasındaki ilişki, basit ilişki olarak ifade edilirken, ikiden fazla özellik veya değişken arasındaki ilişkiler çoklu ilişki olarak ifade edilir.

Değişkenler (özellikler) arasındaki ilişkinin yönünü ve derecesini (düzeyini) belirlemek üzere farklı ölçütler veya istatistikler geliştirilmiştir. Bu istatistiklerin kullanılabilmesi ve geçerli-güvenilir sonuçların elde edilebilmesi için veri setleri ve değişken tiplerine ilişkin bazı varsayımların (ön şartların) sağlanması gerekmektedir. Diğer bir ifade ile farklı değişken tiplerine ve varsayımlara göre kullanılabilen farklı ilişki ölçüleri bulunmaktadır. Örneğin, iki sürekli değişken arasındaki ilişkinin yönünü ve derecesini belirlemede, varsayımların sağlanması durumunda, Pearson korelasyon katsayısı yaygın olarak kullanılırken, (varsayımların) sağlanmaması durumunda Spearman korelasyon katsayısı kullanılmaktadır. Benzer şekilde, iki kategorik değişken arasındaki ilişki için (bağımsızlık testi) Ki-kare testi kullanılmaktadır.

İnteraksiyon (etkileşim) kavramı ise kısaca iki değişkenin, üçüncü bir değişken veya bağımlı değişken üzerine olan etkilerinin, bu değişkenlerin farklı düzeylerinde veya değerlerinde aynı olup olmadığını veya değişiklik gösterip göstermediğini ifade eder. Diğer bir ifade ile iki değişkenin üçüncü bir değişken üzerine birlikte olan etkilerinin, yalnız başına olan etkilerinden farklılık gösterip göstermediğini ifade eder. Böylece, iki değişken arasındaki interaksiyonun önemli (veya anlamlı) olması durumu; iki değişkenden birisinin, üçüncü değişkene (bağımlı değişkene) olan etkisinin, diğer değişkenin farklı düzeylerinde (seviyelerinde veya değerlerinde) değişiklik (farklılık) gösterdiğini ifade eder.

İnteraksiyon (etkileşim) kavramı, uygulamada çoğu kez korelasyon yerine, yanlış olarak kullanılmaktadır. Diğer bir ifade ile istatistikte yaygın olarak kullanılan interaksiyon ve korelasyon kavramları birbirinden farklı olmasına rağmen, yanlışlıkla birbiri yerine kullanılmaktadır. Bu nedenle çalışmada, bu iki kavram genel özellikleri ve farkı yönleriyle açıklanmaya çalışılmıştır.

II. MATERYAL VE YÖNTEM

Değişkenlerden ikisinin de sürekli değişken olması durumunda korelasyon katsayısı yaygın olarak kullanılır. Ancak birinin kategorik, diğerinin sürekli olması durumunda ilişkiyi ölçmede yaygın kullanılan bir ölçü bulunmamaktadır. Bu yapıdaki iki değişken arasındaki ilişki için; bağımlı değişkenin sürekli olması durumunda, Varyans analizi kullanılabılırken, kategorik olması durumunda; Lojistik regresyon analizinden yararlanılabilir. Kategorik değişken, ikili (binary veya dikotom) yapıda ise Point-biserial korelasyon da kullanılabilir [1].

A. Korelasyon

Korelasyon veya korelasyon katsayısı (r) iki özelliğin (veya sürekli değişkenin) birlikte değişiminin ölçüsüdür. Diğer bir ifade ile korelasyon katsayısı, iki özellik veya değişken arasındaki ilişkinin yönünü ve derecesini (yüzde) belirtir. Değişkenlerin sırasının önemi yoktur. Örneğin, Bel çevresi (B, cm) ile Ağırlık (A, kg) arasındaki korelasyon (r_{BA}) ile Ağırlık (kg) ile Bel çevresi (cm) arasındaki korelasyon (r_{AB}) aynıdır. Yaygın olarak kullanılan Pearson korelasyon katsayısı da dahil olmak üzere; birçok korelasyon katsayısı, -1 ile +1 aralığında değer alır. İki özellik

arasında; -1 veya +1 olarak bulunan korelasyon katsayısı, adı geçen özellikler arasındaki doğrusal ilişkinin %100 olduğunu ifade eder ki, bu durum Biyolojik olaylarda hemen hemen hiç mümkün değildir. Korelasyon katsayısının negatif işaretli olması, özellikler arasında negatif yönlü ilişki olduğunu, diğer bir ifadeyle özelliklerden biri artarken diğerinin azaldığını, pozitif olması ise özelliklerden bir artarken diğerinin de arttığını belirtir. Farklı değişken tiplerine göre farklı korelasyon katsayıları geliştirilmiş olmasına rağmen, yorumları genel olarak birbirine benzerdir.

Korelasyon katsayısı, ele alınan değişkenlerin ya da özelliklerin durumuna göre sebep-sonuç ilişkisi olarak ifade edilebilir. Ancak her zaman sebep-sonuç ilişkisi olmayabilir. Diğer yandan, korelasyon katsayısı, özellikler arasındaki ilişkinin doğrusallık kısmını ölçer. Bu nedenle, hesaplanan korelasyon katsayısının 0 olarak ($r = 0$) bulunmuş olması, özellikler arasında hiçbir ilişki olmadığı anlamına gelmez. Adı geçen özellikler arasında doğrusal ilişkinin haricinde muhtelif (parabolik, logaritmik, üssel, sinüzoidal vb) ilişkiler olabilir.

Kısmi (partial) korelasyon katsayısı: İki sürekli değişken (X ve Y) arasındaki ilişkiye üçüncü değişken (Z) etki ediyor olabilir. Bu durumda, üçüncü değişkenin (veya değişkenlerin) etkisi istatistik olarak düzeltildikten (veya giderildikten) sonra, bu iki değişken arasında hesaplanan korelasyon katsayısına kısmi korelasyon katsayısı denir. Diğer bir ifadeyle kısmi korelasyon katsayısı, iki değişken arasında, bu iki değişkene birlikte etki ettiği varsayılan kovaryet değişken(ler)in etkisi giderildikten sonra hesaplanmış olan korelasyon katsayısıdır.

B. İnteraksiyon (Etkileşim)

İnteraksiyon veya etkileşim, çoğunlukla varyans analizi modellerinde; bağımlı (cevap) değişken olarak alınan sürekli bir değişken üzerine etkisi olabileceği düşünülen veya bu değişkenle ilişkili olduğu varsayılan iki kategorik değişkenden (faktörden veya etkenden) birinin seviyelerindeki; bağımlı değişkenin ortalamaları arasındaki farkın, diğer faktörün farklı seviyelerinde değişip değişmediğini ifade eden kavramdır. Buna göre bağımlı değişkenin ortalaması bakımından; modele dahil edilen faktörlerden (bağımsız değişkenlerden) herhangi birisinin seviyeleri arasındaki fark, diğer faktörün farklı seviyelerinde sabit kalmayıp değişiyorsa ve bu değişim de istatistik olarak

anlamli (önemli) bulunmuşsa, bu iki kategorik deęişken (faktör) arasında interaksyon (etkileşim) vardır. Buna karşılık, faktörlerden herhangi birisinin seviyelerindeki bağımlı deęişken ortalamaları arası fark, dięer faktörün (farklı) seviyelerinde deęişmeyip sabit kalıyorsa bu iki kategorik deęişken arasında interaksyon (etkileşim) yoktur.

Bağımlı deęişken olarak alınan deęişkenin ortalamasının, bağımsız deęişken olarak alınan herhangi bir kategorik deęişkenin farklı seviyelerinde (düzeylelerinde) deęişip deęişmediğini belirlemek üzere Tek Yönlü Varyans Analizi (One-way ANOVA veya Basit Varyans Analizi) modelinden yararlanılır. Dięer bir ifade ile ilgilenilen herhangi bir sürekli deęişken ile kategorik deęişken arasında ilişkiyi belirlemek üzere Tek Yönlü Varyans Analizi kullanılır. Örneğin hipertansiyonun (Yüksek kan basıncı) tedavisinde kullanılan; A (Alfa Blokerler), B (Beta Blokerler), C (ACE İnhibitörleri) ve D (Diüretikler) grubu ilaçların, Sistolik Kan basıncı (Y, mmHg) ile olan ilişkisi belirlemek üzere Tek Yönlü Varyans Analizi modelinden yararlanılır.

Benzer şekilde, Sistolik kan basıncı (SKB) ortalaması bakımından erkek ve kadınlar arasında fark olup olmadığı, dięer bir ifade ile SKB ile cinsiyet arasında ilişki olup olmadığı da Tek Yönlü Varyans Analizi ile belirlenebilir. Bu iki faktör veya kategorik deęişken (İlaç grubu ve Cinsiyet) ile SKB ilişkisi; ayrı ayrı yapılacak olan iki adet Tek Yönlü Varyans analizi ile belirlenebileceği gibi bu iki kategorik deęişken birlikte modele dahil edilerek, Faktöriyel Varyans Analizi ile de belirlenebilir.

İki ya da daha fazla kategorik deęişken ile bağımlı deęişken arasındaki ilişkileri belirlemek üzere; her faktör için ayrı ayrı yapılacak olan Tek Yönlü Varyans Analizleri yerine, bu kategorik deęişkenlerin birlikte modele dahil edildiği Faktöriyel (Two-way ANOVA) Varyans Analizi modellerinin tercih edilmesi gerekir. Böylece Faktöriyel Varyans Analizi ile I. tip hatanın kontrolü sağlanmış ve faktörlerin esas etkilerinin yanı sıra interaksyon (etkileşim) etkisi de belirlenmiş olacaktır. Zira kategorik deęişkenlerin ayrı ayrı modele dahil edilmesi ile I. Tip hata artacağı gibi faktörler arasındaki olası interaksyon (etkileşim) etkisi de belirlenemeyecektir.

Faktöriyel Varyans Analizi modelinde; SKB için ilaç çeşitlerinin ortalamaları arasındaki fark, her iki cinsiyette deęişiyorsa (veya farklı ise) ve bu deęişim (farklılık) de istatistik olarak anlamlı (önemli) ise

'ilaç çeşidi x cinsiyet' interaksyonu (etkileşimi) vardır.

Benzer şekilde, erkek ve kadın hastaların SKB ortalamaları arasındaki fark, ilaç çeşitlerinde deęişiyorsa (veya farklı ise) ve bu deęişim (farklılık) de istatistik olarak anlamlı (önemli) ise 'ilaç çeşidi x cinsiyet' interaksyonu (etkileşimi) vardır. Buna karşılık, SKB değeri bakımından erkek ve kadınların ortalaması arasındaki fark, ilaç çeşitlerinde deęişmeyip sabit kalıyorsa (aynı ise) veya ilaç çeşitleri arasındaki farklılık, her iki cinsiyette de deęişmeyip sabit kalıyorsa (aynı ise) 'ilaç çeşidi x cinsiyet' interaksyonu yoktur veya istatistik olarak önemli deęildir

İnteraksyon ve ilişki (etkileşim): Bireylerdeki Kolesterol düzeyinin (Y) Bağımlı deęişken (sürekli), Bel çevresinin (X_1) Bağımsız deęişken (sürekli) ve Vücut kütle indeksinin (X_2) de 'Normal' ve 'Yüksek' olmak üzere iki seviyeli Kategorik bağımsız deęişken olduğu varsayılınsın.

I. Durum: İlişki var, interaksyon yok: Y deęişkeni ile olan ilişki göz ardı edildiğinde; X_2 deęişkeninin yüksek düzeyinde (veya seviyesinde, $X_2 = 1$) X_1 deęişkeninin ortalamasının yüksek, normal (düşük) düzeyinde (veya seviyesinde, $X_2 = 0$) ise düşük olduğu ve bunlar arasında interaksyon (etkileşim) olmadığı varsayılınsın.

Bağımsız kategorik deęişken olan Vücut kütle indeksinin (X_2) normal ve yüksek düzeylerinde ayrı ayrı olmak üzere; bağımlı deęişken olan Kolesterol düzeyi (Y) ile bağımsız sürekli deęişken olan Bel çevresi (X_1) arasında regresyon hesaplanıp, regresyon doğruları çizildiğinde, bu doğrular birbirine paralel olacaktır. Zira Vücut kütle indeksinin her iki kategorisinde (Normal ve Yüksek) Kolesterol düzeyi ile Bel çevresi arasındaki ilişki aynıdır ve Bel çevresi ile Vücut kütle indeksi arasında bir ilişki vardır. Böyle bir durumda, Vücut kütle indeksi, kontrol deęişkeni olarak modele dahil edilmediğinde veya araştırmada böyle bir amaç bulunmadığında, regresyon doğruları birleştirilerek tek bir regresyon doğrusu çizilebilir.

II. Durum: İlişki yok, interaksyon var: İkinci durumda X_1 ve X_2 deęişkenleri arasında ilişki olmadığı dięer bir ifade ile bunların ilişkisiz olduğu varsayılınsın. Bu durumda, X_1 'in (Bel çevresi) ortalaması, X_2 'nin (VKİ) her iki kategorisi için de aynıdır. Ancak X_1 'in Y'ye olan etkisi, X_2 'nin iki kategorisinde (Normal, Yüksek) farklılık gösteriyorsa, bu durumda; X_1 ve X_2 deęişkenleri

arasındaki interaksiyon (etkileşim) etkisi olduğundan söz edilir. Diğer bir ifade ile X_1 ve X_2 değişkenleri arasındaki interaksiyon vardır.

III. Durum: İlişki var, interaksiyon var: Üçüncü durumda, X_1 ve X_2 değişkenleri arasında ilişki ve interaksiyon bulunmaktadır. Bu durumda; X_1 'in ortalaması X_2 'nin yüksek düzeyinde ($X_2 = 0$) düşük, normal düzeyinde ise yüksek, ($X_2 = 1$) olsun. Bağımsız kategorik değişken olan Vücut kütle indeksinin (X_2) normal ve yüksek düzeylerinde ayrı ayrı olmak üzere; Bağımlı değişken olan Kolesterol düzeyi (Y) ile Bağımsız değişken olan Bel çevresi (X_1) arasında regresyon hesaplanıp, regresyon doğruları çizildiğinde, bu doğruların eğimleri farklı olacaktır. Bu durumda, X_1 ve X_2 değişkenleri arasındaki interaksiyon vardır.

III. TARIŞMA VE SONUÇ

İnteraksiyon (etkileşim) kavramı, korelasyon kavramına göre biraz daha karmaşıktır. Korelasyon, iki özellik veya değişken arasındaki ilişkinin yönünü ve derecesini (yüzde) belirtir.

İnteraksiyon kavramı, modeldeki faktörlerden birinin seviyelerindeki bağımlı değişkenin ortalamaları arasındaki farkın, diğer faktörün seviyesinden seviyesine değişip değişmediği ile ilgili kavramdır. İnteraksiyon (etkileşim) ayrıca, ilgilenilen sürekli değişken veya cevap değişkeni bakımından, iki değişkenin üçüncü bir değişken üzerine birlikte olan etkisinin, yalnız başına olan etkilerinden farklılık gösterip göstermediğini ifade eder. Diğer bir ifade ile iki kategorik değişkenin (veya faktörün), üçüncü bir bağımlı değişken veya cevap değişkeni üzerine olan etkilerinin, bu değişkenlerin farklı düzeylerinde veya değerlerinde aynı olup olmadığını veya değişiklik gösterip göstermediğini ifade eder. Bu tanımı ile Kısmi korelasyon katsayısına benzerlik gösterdiği düşünülse de, Kısmi korelasyon katsayısından farklı olduğu aşıkardır. Zira kısmi korelasyon katsayısı, iki değişken arasındaki ilişkiye etkili olduğu düşünülen üçüncü değişkenin etkisi, istatistik olarak giderildikten sonraki hesaplanan korelasyon katsayısıdır. Oysaki interaksiyon, iki bağımsız veya açıklayıcı değişkenin, bağımlı veya cevap değişkeni üzerine ayrı ayrı veya birlikte olan etkilerinin farklılık gösterip göstermediğini ifade eder.

Modelde yer alan faktörlerin ayrı ayrı eklemeli etkileri toplamı, birlikte modele alındığındaki toplama eşit olmuyorsa bu durum, eklemeli etkilerin yanında çarpımsal etkilerin de olduğuna işaret eder.

Böylece interaksiyon, bir çeşit eklemeli olmayan etkidir [2].

Bağımlı ve bağımsız değişkenlerin sürekli değişken olduğu regresyon modellerinde, interaksiyon etkisi, eklemeli etkiden olan sapmanın ölçüsü olarak çarpımsal etki ile ifade edilebilir [3].

Böylece regresyon modellerine ana (esas) etkilerle birlikte interaksiyon etkisinin eklenmesi, daha doğru tahminlerin yapılmasını sağlayabilir [4].

Sonuç olarak, interaksiyon (etkileşim) ve korelasyon kavramlarının birbirinden farklı olduğu ve çalışmalarda, bu kavramlara dikkat edilmesi gerektiği önerilmektedir.

KAYNAKLAR

- [1] Anonymous, (2023) The difference between interaction and association. By Karen Grace-Martin. <https://www.theanalysisfactor.com/interaction-association/> Erişim Tarihi: 22.12.2023,
- [2] A. Berrington de González, and D. R. Cox. "Interpretation of interaction: A review." *The Annals of Applied Statistics* 1(2), 371–385 DOI: 10.1214/07-AOAS124. 2007.
- [3] R. P. Duncan and B. J. Kefford, "Interactions in statistical models: three things to know". *Methods in Ecology and Evolution*, 12(12), 2287-2297. 2021.
- [4] J. Li, "Comparison of Correlation, Partial Correlation, and Conditional Mutual Information for Interaction Effects Screening in Generalized Linear Models. Graduate Theses and Dissertations Retrieved from" <https://scholarworks.uark.edu/etd/2860>. 2018.